



**HAL**  
open science

# Approches bayésiennes pour le débruitage des images dans le domaine des transformées multi-échelles parcimonieuses orientées et non orientées

Larbi Boubchir

► **To cite this version:**

Larbi Boubchir. Approches bayésiennes pour le débruitage des images dans le domaine des transformées multi-échelles parcimonieuses orientées et non orientées. Traitement des images [eess.IV]. Université de Caen, 2007. Français. NNT: . tel-00161573

**HAL Id: tel-00161573**

**<https://theses.hal.science/tel-00161573>**

Submitted on 11 Jul 2007

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

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



UNIVERSITÉ de CAEN/BASSE-NORMANDIE

U.F.R de Sciences

ÉCOLE DOCTORALE SIMEM

THÈSE

présentée par

Larbi BOUBCHIR

et soutenue

le 04 juillet 2007

en vue de l'obtention du

DOCTORAT de l'UNIVERSITÉ de CAEN

Spécialité : Traitement du Signal et des Images

(Arrêté du 07 août 2006)

**Approches bayésiennes  
pour le débruitage des images dans le domaine  
des transformées multi-échelles parcimonieuses  
orientées et non orientées**

---

MEMBRES du JURY

---

<i>Rapporteurs</i>	Laure BLANC-FERAUD	DR CNRS, I3S-INRIA
	Jean-Luc STARCK	Chercheur CEA&HDR, CEA-Saclay
<i>Examineurs</i>	Jian-Feng YAO	Professeur, Université de Rennes 1
	François KAUFFMANN	Maître de Conférence, Université de Caen
<i>Directeur</i>	Marinette REVENU	Professeur, ENSICAEN
<i>Co-directeur</i>	Jalal M. FADILI	Maître de Conférence, ENSICAEN



*À celui à qui je dois tout  
À la mémoire de mon cher frère Mourad  
À mes très chers parents, qui m'ont permis de devenir  
ce que je suis aujourd'hui  
À mes très chers frères Mohammed-Amine, Abdelhafid et Fethallah  
À ma bien aimée soeur Chahrazed  
À toute ma famille  
À mes amis*



## Remerciements

---

Cette thèse a été effectuée au sein de l'équipe Image du centre de recherche GREYC, dirigé par le professeur Régis CARIN, que je remercie pour m'avoir accueilli.

Je remercie sincèrement Laure BLANC-FERAUD (directrice de recherche au CNRS, I3S/INRIA Nice Sophia-Antipolis) et Jean-Luc STARCK (chercheur, CEA Saclay) d'avoir accepté la lourde tâche de rapporteurs malgré le temps très restreint que je leur ai laissé.

J'aimerais également remercier le professeur Jian-Feng YAO (laboratoire IRMAR, université de Rennes 1) et François KAUFFMANN (laboratoire de mathématiques, université de Caen) pour avoir accepté de juger ces quelques années de travail en participant au jury de ma thèse.

Je remercie encore une fois tous les membres de jury pour l'intérêt qu'ils ont bien voulu porter à mon travail.

Je remercie également ma directrice de thèse le professeur Marinette REVENU pour sa confiance et le soutien qu'elle m'a apporté durant ces années de thèse et qui m'ont aidé à accomplir ce travail.

Je tiens à témoigner publiquement à Jalal FADILI, mon directeur scientifique de recherche, toute la reconnaissance que je lui dois. Ce dernier a suscité, développé, puis accompagné mes premiers pas dans le domaine de la recherche avec une grande patience et avec une pédagogie extraordinaire. Le soutien moral et intellectuel de mon encadrant Jalal FADILI, fut essentiel. Non seulement il m'a fourni une aide indispensable à l'avancement de mes travaux de recherche, mais aussi, quand j'étais dans certaine situation délicate, il a su m'aider avec ses conseils à reprendre le chemin. Chercheur reconnu et source inépuisable de nouvelles idées, il restera mon mentor.

Aux membres de l'équipe Image qui m'ont accueilli comme doctorant puis comme ATER, pour leur aide, les discussions que nous avons eues et leurs précieux conseils. Je leur exprime ma plus grande gratitude.

Mes remerciements s'adressent aussi à Daniel CARRE, Luc BRUN, Régis CLOUARD, Sébastien FOUREY avec qui j'ai partagé mon goût de l'enseignement. Ils m'ont été d'une aide précieuse.

Je n'omettrai pas de remercier David TSCHUMPERLE, chercheur CNRS, avec le quel j'ai partagé mon espace de travail durant mes dernières années de thèse pour ses discussions, ses conseils et pour l'ambiance sympathique.

Je suis aussi redevable à Luigi LANCIERI, chercheur à France Télécom R&D, pour m'avoir encadrer pendant mon stage de DEA, aussi bien, pour ses conseils et ses qualités humaines. Également, je remercie tous les enseignants qui ont contribué à ma formation durant mes années d'études.

Je voudrai également remercier tous les doctorants et particulièrement Sébastien, Arnaud, Jonathan, Jean-Hugues, François, Nicolas, Haz-Edine, Guillaume et François-Xavier pour l'ambiance sympathique. Je vous souhaite bon courage!

Je voudrai aussi exprimer mes reconnaissance à mes amis(es) sans exception. En particulier, Abdelmadjid, Belkacem, Sabih, Abdelmalek, Boubakeur, Malik, Ahmed, Imad, Djilali, Nazim et Mohammed qui ont constitué ma deuxième famille pendant ces années de thèse.

Je garde le meilleur pour la fin, ma famille qui a supporté toutes les difficultés morales et matérielles pour me soutenir tout au long de mes études supérieures. A mes très chers parents, que Dieu les garde et les bénisse. Tous les mots ne suffisent pas pour exprimer le respect et l'amour que j'avoue pour eux. A mes très chers frères Mohammed Amine, Abdelhafid et Fethallah. A ma bien aimée soeur Chahrazed, je lui souhaite tout le bonheur du monde. Ils savent combien ils comptent pour moi...

Enfin, je remercie tous ceux qui ont contribué à l'élaboration de ce travail de près ou de loin et qui méritent d'y trouver leur nom. J'espère qu'ils me pardonneront.

A tous un sincère et chaleureux merci.

L'auteur  
Larbi Boubchir

# Table des matières

Table des figures	xi
Liste des tableaux	xv
Glossaire	xvii
Introduction Générale	1
Avant-propos	3
Partie I Représentations parcimonieuses multi-échelles pour le débruitage	7
1 État de l'art : débruitage multi-échelle des images	9
1.1 Régression non-paramétrique . . . . .	10
1.2 Transformées multi-échelles et débruitage . . . . .	12
1.2.1 Formulation générale du problème . . . . .	12
1.2.2 Débruitage classique . . . . .	13
1.2.3 Débruitage bayésien . . . . .	14
1.2.4 Estimation du niveau de bruit . . . . .	14
1.3 Débruitage dans le domaine des transformées multi-échelles non-orientées . .	15
1.3.1 Approches classiques sans <i>a priori</i> . . . . .	15
1.3.1.1 Méthodes terme à terme . . . . .	15
1.3.1.2 Méthodes classiques par blocs . . . . .	20
1.3.2 Approches bayésiennes . . . . .	21
1.3.2.1 Méthodes bayésiennes univariées . . . . .	21
1.3.2.2 Méthodes bayésiennes multivariées . . . . .	24
1.4 Débruitage dans le domaine des transformées multi-échelles orientées . . . .	26

1.4.1	Approches classiques sans <i>a priori</i> . . . . .	26
1.4.2	Approches bayésiennes . . . . .	26
1.5	Conclusion . . . . .	28
<b>2</b>	<b>Modélisation statistique des images</b>	<b>31</b>
2.1	Pourquoi une modélisation statistique? . . . . .	31
2.2	Axiomes et propriétés relatives aux images naturelles . . . . .	32
2.2.1	Invariance par changement d'échelle . . . . .	32
2.2.2	Persistance à travers les échelles . . . . .	32
2.2.3	Dépendance intra-échelle . . . . .	32
2.3	Modèles statistiques des images . . . . .	34
2.3.1	Modèles génératifs parcimonieux . . . . .	34
2.3.1.1	Représentation parcimonieuse . . . . .	34
2.3.1.2	Images compressibles . . . . .	35
2.3.2	Autres modèles . . . . .	35
2.3.2.1	Modèles de superposition . . . . .	35
2.3.2.2	Les champs aléatoires de Markov . . . . .	36
2.4	Lois dans les représentations parcimonieuses . . . . .	36
2.4.1	Loi marginale . . . . .	36
2.4.2	Loi conditionnelle/jointe . . . . .	38
2.5	Conclusion . . . . .	39
<b>Partie II</b>	<b>Statistiques univariées pour la restauration</b>	<b>41</b>
<b>3</b>	<b>Modélisation des statistiques marginales</b>	<b>43</b>
3.1	Quel <i>a priori</i> ? . . . . .	44
3.2	Modèle de mélange d'échelles de gaussiennes : cadre général . . . . .	44
3.2.1	Définition . . . . .	44
3.2.2	Propriétés . . . . .	44
3.3	L' <i>a priori</i> SMG et l'espace de Besov . . . . .	47
3.4	Estimation des hyperparamètres . . . . .	49
3.5	Cas 1 : <i>a priori</i> $\alpha$ -stable . . . . .	49
3.5.1	Définition . . . . .	49
3.5.2	Propriétés . . . . .	50
3.5.2.1	Variables aléatoires $\alpha$ -stables symétriques . . . . .	50
3.5.2.2	Quelques opérateurs arithmétiques . . . . .	50
3.5.2.3	Comportement asymptotique des queues de distribution . . . . .	51

---

3.5.2.4	PDFs $\alpha$ -stable connues . . . . .	51
3.5.3	Approximation analytique de la PDF $\alpha$ -stable . . . . .	52
3.5.4	Du mélange d'échelles au mélange fini . . . . .	54
3.5.4.1	Modèle de mélange de gaussiennes . . . . .	54
3.5.4.2	Approximation analytique des PDFs $S_{\alpha S}$ . . . . .	55
3.5.5	Sélection de modèles . . . . .	56
3.5.5.1	Le critère MDL (" <i>minimum description length</i> ") . . . . .	57
3.5.5.2	Simulations Monte-Carlo et divergence KL . . . . .	57
3.5.6	Estimation des hyperparamètres . . . . .	59
3.5.6.1	Méthodes des quantiles . . . . .	59
3.5.6.2	Méthodes basées sur la fonction caractéristique . . . . .	60
3.5.6.3	Maximum de vraisemblance . . . . .	62
3.5.6.4	Méthodes des moments fractionnaires . . . . .	62
3.5.6.5	Méthodes basées sur le comportement asymptotique des queues . . . . .	62
3.5.7	Performance des différents estimateurs . . . . .	62
3.6	Cas 2 : <i>a priori</i> des Formes K de Bessel . . . . .	65
3.6.1	Définition . . . . .	65
3.6.2	Propriétés . . . . .	65
3.6.2.1	Quelques propriétés fondamentales . . . . .	65
3.6.2.2	Moments d'ordre $p$ . . . . .	67
3.6.3	L' <i>a priori</i> BKF et l'espace de Besov . . . . .	68
3.6.4	Estimation des hyperparamètres . . . . .	68
3.6.4.1	Méthode des cumulants . . . . .	68
3.6.4.2	Algorithme EM . . . . .	69
3.6.5	Performance des différents estimateurs . . . . .	72
3.7	Qualité de l' <i>a priori</i> : application à la modélisation des statistiques marginales . . . . .	73
3.7.1	Objectifs de l'expérience . . . . .	73
3.7.2	Protocole expérimental . . . . .	73
3.7.3	Discussion des résultats . . . . .	73
<b>4</b>	<b>Débruitage bayésien avec <i>a priori</i> univarié</b> . . . . .	<b>77</b>
4.1	Rappels sur l' <i>a priori</i> . . . . .	77
4.2	Estimation bayésienne : cas univarié . . . . .	78
4.3	Estimateur de l' <i>espérance conditionnelle a posteriori</i> (ECP) . . . . .	79
4.3.1	Estimateur ECP avec l' <i>a priori</i> $\alpha$ -stable . . . . .	79
4.3.1.1	La PDF marginale des coefficients bruités . . . . .	79
4.3.1.2	Estimation des hyperparamètres . . . . .	79

4.3.1.3	Débruiteur bayésien terme-à-terme . . . . .	80
4.3.2	Estimateur ECP avec l' <i>a priori</i> BKF . . . . .	82
4.3.2.1	La PDF marginale des coefficients d'ondelettes . . . . .	82
4.3.2.2	Estimation des hyperparamètres . . . . .	85
4.3.2.3	Débruiteur bayésien terme-à-terme . . . . .	87
4.4	Estimateur <i>maximum a posteriori</i> (MAP) . . . . .	91
4.4.1	Estimateur MAP avec l' <i>a priori</i> BKF . . . . .	91
4.5	Expérimentation et résultats . . . . .	93
4.5.1	Objectifs de l'expérience . . . . .	93
4.5.2	Protocole expérimental . . . . .	94
4.5.3	Discussion des résultats . . . . .	94
4.6	Conclusion . . . . .	101

### Partie III Statistiques multivariées pour la restauration 103

<b>5</b>	<b>Modélisation des statistiques jointes</b> . . . . .	<b>105</b>
5.1	Étude et analyse statistique . . . . .	106
5.1.1	Curvelets et notations . . . . .	106
5.1.2	Étude statistique . . . . .	106
5.1.2.1	Distribution jointe . . . . .	106
5.1.3	Étude quantitative . . . . .	108
5.1.3.1	Information mutuelle . . . . .	108
5.2	Modèle SMG multivarié : cadre général . . . . .	109
5.2.1	Définition . . . . .	109
5.2.2	Propriétés . . . . .	109
5.2.3	Moments absolus d'ordre $p$ . . . . .	110
5.2.4	Cas particuliers du SMG multivarié . . . . .	110
5.3	Estimation des hyperparamètres . . . . .	110
5.4	Cas 1 : modèle BKF multivarié . . . . .	111
5.4.1	Définition . . . . .	111
5.4.2	Propriétés . . . . .	111
5.4.3	Estimation des hyperparamètres . . . . .	113
5.4.3.1	Méthode des moments . . . . .	113
5.4.3.2	Algorithme EM . . . . .	113
5.5	Qualité de l' <i>a priori</i> : application à la modélisation des statistiques jointes . . . . .	115
5.5.1	Objectifs de l'expérience . . . . .	115
5.5.2	Protocole expérimental . . . . .	115

---

5.5.3	Discussion des résultats . . . . .	115
5.6	Conclusion . . . . .	116
<b>6</b>	<b>Débruitage bayésien avec a priori multivarié</b>	<b>121</b>
6.1	Rappels sur l'estimation bayésienne : cas multivarié . . . . .	121
6.2	Estimateur ECP avec l' <i>a priori</i> MBKF . . . . .	122
6.2.1	La PDF jointe des coefficients bruités . . . . .	122
6.2.2	Estimation des hyperparamètres . . . . .	123
6.2.3	Débruiteur bayésien ECP . . . . .	124
6.3	Expérimentation et résultats . . . . .	125
6.3.1	Objectifs de l'expérience . . . . .	125
6.3.2	Protocole expérimental . . . . .	125
6.3.3	Discussion des résultats . . . . .	125
6.4	Conclusion . . . . .	126
	<b>Conclusion et perspectives</b>	<b>133</b>
	<b>Conclusion et perspectives</b>	<b>135</b>
	<b>Annexes</b>	<b>137</b>
<b>A</b>	<b>Éléments de la théorie de l'estimation</b>	<b>139</b>
A.1	Le maximum de vraisemblance . . . . .	139
A.1.1	Définition générale . . . . .	139
A.1.2	Quelques propriétés . . . . .	140
A.2	L'algorithme EM . . . . .	140
A.2.1	Définition générale . . . . .	140
A.2.2	Propriétés . . . . .	141
A.2.3	Autres variantes de l'algorithme EM . . . . .	141
A.3	Paradigme bayésien en images . . . . .	141
A.3.1	Lois a priori et a posteriori . . . . .	142
A.3.2	Estimation bayésienne . . . . .	142
A.3.2.1	Maximum a posteriori (MAP) . . . . .	143
A.3.2.2	Espérance conditionnelle a posteriori (ECP) . . . . .	143
A.3.2.3	Médiane conditionnelle a posteriori (MCP) . . . . .	143

---

<b>B</b>	<b>Estimation de McCulloch</b>	<b>145</b>
B.1	Tables . . . . .	145
<b>C</b>	<b>Modèle multivarié gaussien généralisé anisotrope (AMGGD)</b>	<b>147</b>
C.1	Modèle multivarié analytique . . . . .	147
C.2	Estimation des hyperparamètres . . . . .	148
<b>D</b>	<b>Transformée de Curvelet 2<sup>ème</sup> génération</b>	<b>151</b>
D.1	Transformée multi-échelle orientée discrète : FDCT . . . . .	151
D.2	Propriétés . . . . .	153
<b>E</b>	<b>Publications &amp; diffusion des résultats</b>	<b>155</b>
	<b>Bibliographie</b>	<b>157</b>

# Table des figures

1	Exemple de débruitage d'images. . . . .	4
1.1	Application de la transformée d'ondelettes séparable 2D sur une image. . . . .	13
1.2	Estimateurs par seuillage d'ondelettes. . . . .	16
1.3	Autres variantes des estimateurs par seuillage d'ondelettes. . . . .	17
1.4	Exemple de représentation par blocs sans recouvrement. . . . .	21
2.1	Sous-bandes de détails diagonaux de la DWT (ondelette Daubechies-8) appliquée sur l'image de Lena (de taille [512x512]) aux échelles : 1, 2 et 3, montrant la persistance des détails à travers les échelles. . . . .	33
2.2	Sous-bandes de la DWT (ondelette Daubechies-8) de l'image de Lena à l'échelle 1 : détails verticaux ( <i>HL</i> ), diagonaux ( <i>HH</i> ), horizontaux ( <i>LH</i> ), illustrant la dépendance entre les sous-bandes d'une même échelle. . . . .	33
2.3	Distributions marginales des coefficients de l'image de Lena avec : DWT (a), FDCT (b) et UDWT (c). Ces lois marginales sont clairement non gaussiennes, caractérisées pas des densités symétriques centrées en 0 avec des queues relevées (leptokurtique). Les kurtosis ont des valeurs significativement plus élevées que la valeur gaussienne de 3. . . . .	37
2.4	Exemple de PDFs (observées et ajustées par des modèles) jointe et conditionnelle de deux coefficients d'ondelettes à deux échelles différentes (le père est à une échelle plus grossière par rapport au fils). Les coefficients d'ondelettes sont ceux de l'image de Lena. L'intensité sur chaque figure reflète l'amplitude de la PDF. . . . .	38
3.1	(a) Densités des lois stables (PDFs) de paramètres $\alpha \in \{2, 1.5, 1, 0.5\}$ , $\beta = 0$ , $\sigma = 1$ et $\mu = 0$ . (b) Évolution de la queue de distribution avec le paramètre $\alpha$ . . . . .	53
3.2	(a) L'effet du paramètre d'asymétrie $\beta$ , avec $\alpha = 1.5$ , $\sigma = 1$ et $\mu = 0$ . (b) L'effet du paramètre d'échelle $\sigma$ , avec $\alpha = 1$ , $\beta = 0$ et $\mu = 0$ . . . . .	53
3.3	Comparaison sur une échelle log-log de la PDF exacte, calculée à partir de l'Eq.3.33 (trait plein), et la PDF approchée par un mélange de 8 gaussiennes (- - -) pour différentes valeurs du paramètre $\alpha$ . La divergence de "Kullback-Leibler" (KL) entre les deux PDFs est notée au dessus de chaque graphe. . . . .	56
3.4	Évolution du critère MDL en fonction du nombre de gaussiennes $N$ pour différentes valeurs du paramètre $\alpha$ . . . . .	58
3.5	Évolution de la divergence de KL, calculée entre la PDF exacte et la PDF approchée par un mélange de gaussiennes, en fonction du nombre de gaussiennes pour différentes valeurs du paramètre $\alpha$ . . . . .	58
3.6	Estimation du paramètre $\alpha$ par la méthode de McCulloch pour $\sigma = 0.1$ 3.2 6.9 10 . . . . .	63

3.7	Estimation du paramètre $\alpha$ par la méthode de Fama-Roll pour $\sigma = 0.1$	3.2	6.9	10	63
3.8	Estimation du paramètre $\alpha$ par la méthode $M_1$ pour $\sigma = 0.1$	3.2	6.9	10	64
3.9	Estimation du paramètre $\alpha$ par la méthode par régression pour $\sigma = 0.1$	3.2	6.9	10	64
3.10	(a) PDFs des lois BKF de paramètres $\alpha \in \{0.5, 0.75, 1, 2\}$ avec $c = 10$ . (b) Évolution de la queue de distribution avec le paramètre de forme $\alpha$ .				66
3.11	L'effet du paramètre d'échelle $c$ avec $\alpha = 0.8$ .				66
3.12	Comparaison entre l'estimation des hyperparamètres $\alpha$ et $c$ par la méthode des cumulants et l'algorithme EM. Les estimés de $\alpha$ et $c$ par l'algorithme EM sont plus proches des valeurs initiales comparés aux estimés par la méthode des cumulants.				72
3.13	Comparaison sur une échelle log-log entre la PDF marginale observée et les PDFs estimées des coefficients de détail d'ondelettes pour trois d'images test, à deux échelles et trois orientations $HH$ , $HL$ et $LH$ . L'histogramme observé (-●-) a été ajusté par trois modèles d' <i>a priori</i> : BKF (trait plein), $\alpha$ -stable mixture (trait pointillé), $\alpha$ -stable (alternance de points et de traits) et GGD (points).				75
3.14	Évolution de la divergence KL, calculée entre la PDF observée et les PDFs estimées par l' <i>a priori</i> BKF (avec la méthode des cumulants et l'algorithme EM), $\alpha$ -stable [Achim et al., 2001], $\alpha$ -stable mixture et GGD. La divergence KL a été moyennée pour chaque orientation à chaque échelle sur une base de 100 images.				76
4.1	Comparaison entre l'histogramme marginal observé (trait en pointillé) et la PDF $\alpha$ -stable estimée (trait plein) des coefficients de détail d'ondelettes pour trois d'images tests corrompues par un bruit blanc gaussien (SNR= 15dB). Seulement, deux niveaux de décomposition sont montrés pour chacune des trois images test : Barbara, Boat et House. Les trois colonnes correspondent aux orientations $HH$ , $HL$ et $LH$ .				81
4.2	Évolution de la fonction $s(d)$ de l'estimateur bayésien ECP $\alpha$ -stable. (a) En fonction du rapport $\frac{\sigma}{\sigma_\epsilon}$ . (b) En fonction du paramètre $\alpha$ .				83
4.3	Comparaison entre l'histogramme marginal observé (trait en pointillé) et la PDF BKF estimée (trait plein) des coefficients de détail d'ondelettes pour trois d'images tests corrompues par un bruit blanc gaussien (SNR= 15dB). Seulement, deux niveaux de décomposition sont montrés pour les trois images test : Barbara, Boat et House. Les trois colonnes correspondent aux orientations $HH$ , $HL$ et $LH$ .				88
4.4	Évolution de la fonction $s(d)$ de l'estimateur bayésien ECP BKF. (a) En fonction du rapport $\frac{\sqrt{c}}{\sigma_\epsilon}$ . (b) En fonction du paramètre $\alpha$ .				90
4.5	Évolution de la fonction $s(d)$ de l'estimateur bayésien MAP BKF en fonction de $d$ pour différentes valeurs de SNR (en faisant varier $c$ ).				93
4.6	Comparaison des différents débruiteurs sur l'image test "Lena". Cette image a été corrompue par un bruit additif gaussien de 15dB en entrée.				96
4.7	Comparaison des différents débruiteurs sur l'image test "Barbara" zoomée sur une région texturée du pantalon. Cette image a été corrompue par un bruit additif gaussien de 15dB en entrée. Le zoom montre que les débruiteurs que nous proposons réalise un bon compromis entre le rejet du bruit et la conservation des détails fins de l'image.				97
4.8	Autre comparaison des différents débruiteurs sur l'image test "Mandrill". Cette image a été corrompue par un bruit additif gaussien de 15dB en entrée.				98
4.9	Évaluation des performances de débruiteurs BKF de type ECP et MAP basés sur la méthode de cumulants et l'algorithme EM pour l'estimation des hyperparamètres.				99

4.10	Comparaison de différents estimateurs de débruitage pour les trois images test (Lena, Barbara et Mandrill), pour un SNR à l'entrée de 5 à 20dB. . . . .	100
4.11	Comparaison de différents estimateurs de débruitage pour une base de 100 images. . . . .	101
5.1	Transformée de curvelets discrète de 2 <sup>ème</sup> génération (FDCT) de l'image test Bateau. Trois niveaux d'échelles et six orientations sont montrés (j : indique l'échelle, o : indique l'orientation). . . . .	107
5.2	Dépendances inter- et intra- individuelles en position, échelle et orientation entre les coefficients de curvelet. . . . .	107
5.3	Distributions conjointes (en log) des coefficients de curvelet pour l'image test Bateau dans le cas bivarié, $\mathcal{P}(X, .)$ , avec leurs parents $PX$ (a), voisins $VX$ (b) et cousins $CX$ (c). . . . .	108
5.4	Exemple de comparaison sur une échelle log – log entre la PDF jointe observée et les PDFs estimées des coefficients de détail d'ondelettes pour l'image Barbara dans le cas : bivarié ( $d = 2$ ) et trivarié ( $d = 3$ ). La distribution jointe observée (en rouge) a été ajustée par trois modèles d' <i>a priori</i> : MBKF (en bleu), AMGGD (en vert) et Jeffrey (en bleu clair) [Portilla et al., 2003]. Les trois colonnes correspondent aux dépendances inter- et intra-échelles entre coefficients. . . . .	117
5.5	Exemple de comparaison sur une échelle log – log entre la PDF jointe observée et les PDFs estimées des coefficients de détail de curvelet pour l'image Barbara dans le cas bivarié ( $d = 2$ ). La distribution jointe observée (en rouge) a été ajustée par trois modèles d' <i>a priori</i> : MBKF (en bleu), AMGGD (en vert) et Jeffrey (en bleu clair) [Portilla et al., 2003]. Les trois colonnes correspondent aux dépendances inter- et intra-échelles entre coefficients. . . . .	118
5.6	Évaluation de la divergence de KL, calculée entre la PDF observée et les PDF estimées par les <i>a priori</i> multivariés : MBKF, AMGGD et Jeffrey [Portilla et al., 2003]. La divergence KL a été moyennée pour chaque orientation et à chaque échelle sur une base de 100 images. Trois orientations ont été considérées pour le UDWT. Pour la FDCT et pour mieux synthétiser les résultats, nous avons aussi moyenné les valeurs KL sur toutes les orientations faisant partie d'un même quadrant : Est, Ouest, Nord, Sud. . . . .	119
6.1	Comparaison des différents débruiteurs sur l'image test de Lena (peu texturée). . . . .	128
6.2	Comparaison des différents débruiteurs sur l'image test de Lena zoomée sur une région texturée du chapeau. Le zoom montre que la qualité visuelle des débruiteurs avec la FDCT est meilleure à ceux avec l'UDWT. . . . .	129
6.3	Comparaison des différents débruiteurs sur l'image test de Barbara (texturée). . . . .	130
6.4	Comparaison des différents débruiteurs sur l'image test de Barbara zoomée sur une région texturée du pantalon. . . . .	131
D.1	Comparaison de l'approximation non-linéaire des ondelettes et des curvelets. . . . .	152
D.2	Décomposition pyramidale orientée par la FDCT : (a) Partition spectrale de la FDCT. (b) FDCT de l'image de Lena. Chaque rectangle représente une portion angulaire. . . . .	152
D.3	Une curvelet. . . . .	152



# Liste des tableaux

3.1	Choix de $K$ en fonction de $\alpha$ . . . . .	61
3.2	Protocole expérimental . . . . .	73
3.3	La moyenne (et l'écart-type) de la divergence KL entre la PDF marginale observée et les PDFs estimées par l' <i>a priori</i> : BKF, $\alpha$ -stable mixture, $\alpha$ -stable et GGD, à trois échelles et trois orientations sur une base de 100 images. . . . .	76
4.1	Protocole expérimental . . . . .	94
5.1	L'information mutuelle moyennée sur une base d'images (100 images) [url : base d'images, ], calculée pour deux échelles successives $J$ et $J - 1$ où $J$ est l'échelle la plus grossière. . . . .	109
5.2	Protocole expérimental . . . . .	115
6.1	Protocole expérimental . . . . .	125
6.2	Comparaison en temps de calcul de différents débruiteurs sur l'image test de barbara de taille $512 \times 512$ (Fig.6.3), pour un PSNR=16 dB en entrée ( $\sigma_\epsilon = 40$ ). Tous les débruiteurs sauf le filtrage combiné ont été codés en Matlab avec des routines accélérées écrites en C, et ont été testés sur une station Linux équipée d'un processeur Intel(R) Xeon(TM) à 2.8 GHz. Le filtrage combiné était implémenté en C++. Nous avons obtenu son executable par [Starck et al., 2003] compilé uniquement pour Solaris que nous avons fait tourner sur une station Sun Blade 2000/1000 avec un processeur UltraSparc-III+ à 1.8 Ghz. . . . .	127
6.3	Les moyennes et les écart-type (sur 10 simulations) du PSNR données par les divers débruiteurs. Pour chaque image test, nous montrons en gras les deux meilleurs PSNR correspondant à chaque valeur de $\sigma$ . . . . .	132
B.1	Estimé du paramètre $\alpha$ en fonction de $\nu_\alpha$ et $\nu_\beta$ . . . . .	145
B.2	Estimé du paramètre $\beta$ en fonction de $\nu_\alpha$ et $\nu_\beta$ . . . . .	146
B.3	Les différentes valeurs de la fonction $\vartheta_3(\alpha, \beta)$ . . . . .	146



# Glossaire

## Abréviations

---

i.i.d	indépendant et identiquement distribué
p.s.	presque sûrement
AMGGD	Modèle multivarié gaussien généralisé anisotrope
BKF	Formes K de Bessel (de l'anglais Bessel K Forms)
DLP	en anglais Diagonal Linear Projection (projection diagonale linéaire)
DLS	en anglais Diagonal Linear Shrinker (contraction diagonale)
DWT	Transformée d'ondelettes discrète décimée
ECP	Espérance Conditionnelle a Posteriori
EQM	Erreur Quadratique Moyenne
EM	Espérance-Maximisation
FAC	Fonction d'auto-corrélation
FDCT	Transformée de curvelet discrète rapide
FDR	en anglais False Discovery Rate
FFT	Transformée de Fourier Rapide
GGD	Distribution gaussienne généralisée
IDWT	Transformée d'ondelettes discrète décimée inverse
IFFT	Transformée de Fourier Inverse
IM	Information mutuelle
KL	Divergence de Kullback-Leibler
LDCT	Transformée en cosinus discrète locale
MAD	L'estimateur robuste de Donoho & Johnstone pour estimer le niveau de bruit
MDL	en anglais Minimum Description Length
Median	Médiane
MBKF	Modèle des Formes K de Bessel multivarié
MRF	Champ aléatoire de Markov (de l'anglais Markov Random Field)
MSMG	Modèle de mélange d'échelles de gaussiennes multivarié
MP	Médiane a Posteriori
MV	Maximum de vraisemblance
PDF	Fonction de densité de probabilité
PSNR	Rapport pic du signal sur risque (de l'anglais Peak Signal-to-Noise Ratio)
RB	Règle de Bayes
SNR	Rapport signal sur bruit (de l'anglais Signal-to-Noise Ratio)
SMG	Modèle de mélange d'échelles de gaussiennes
SURE	Estimateur du risque quadratique sans biais de Stein

UDWT	Transformée d'ondelettes discrète non-décimée
SaS	Distribution $\alpha$ -stable symétrique
VA	Variable aléatoire (les symboles en majuscule pour représenter les VAs et en minuscule pour leurs réalisations)
VVA	Vecteur de variables aléatoires (les symboles gras en majuscule pour représenter les VVAs et en minuscule pour leurs réalisations)

## Symboles

---

$\stackrel{d}{=}$	égalité en distribution
$\mathbf{1}_A$	fonction indicatrice de l'ensemble $A$
$\mathbb{I}_d$	matrice d'identité de taille $d \times d$
$\mathbb{N}$	ensemble des entiers naturels
$\mathbb{R}$	ensemble des réels
$\mathbb{Z}$	ensemble des entiers signés
$\mathbb{E}$	espérance
$\mathcal{W}$	matrice orthogonale associée à la base orthonormée de la DWT
$\Phi$	représente un dictionnaire multi-échelle
$\lambda^M$	seuil minimax
$\lambda^U$	seuil universel
$\lambda^S$	seuil 'SURE'
$\lambda^{HS}$	seuil 'SURE' hybride
$\lambda^{CV}$	seuil de la validation croisée
$\delta_\lambda$	opérateur non linéaire de seuillage avec un seuil $\lambda$
$\delta^H$	seuillage dur
$\delta^S$	seuillage doux
$\delta^F$	seuillage 'Firm'
$\delta^G$	seuillage 'nonnegative garrote'
$\delta^{SCAD}$	seuillage 'SCAD'
$\theta$	ensemble des hyperparamètres du modèle a priori
$\sigma_\epsilon$	écart-type de bruit
$\delta(\cdot)$	distribution de Dirac
Bernoulli	loi Bernoulli
$\mathcal{N}$	loi Normale
$\mathcal{L}$	loi Laplacienne
$\mathcal{E}$	loi Exponentielle
$\mathcal{DE}$	loi double exponentielle
$\Gamma(\cdot)$	fonction Gamma
$\Psi$	la dérivée de la fonction $\log \Gamma(\cdot)$
$\mathcal{P}(\cdot)$	probabilité
$\mathcal{L}[\cdot]$	transformée de Laplace
Cov	covariance
Kurt	kurtosis
Var	variance
diag	diagonale d'une matrice

$\det$	déterminant d'une matrice
$A \succ 0$	la matrice $A$ est définie positive
$A \succeq 0$	la matrice $A$ est semi-définie positive
$\ell\ell$	log de vraisemblance
$K_\nu$	fonction de Bessel modifiée de première espèce d'ordre $\nu$
$\kappa_i$	cumulant d'ordre $i$
$D_\nu$	la fonction de cylindre parabolique d'ordre partiel $\nu$

## Notations

---

$y_{mn}, g_{mn}, \epsilon_{mn}$	représentent respectivement les valeurs réelles échantillonnées de l'image bruitée $y$ , l'image à estimer $g$ et le bruit blanc gaussien $\epsilon$ .
$d_{mn}^{oj}, s_{mn}^{oj}$	représentent respectivement les coefficients de détails de la transformée d'ondelettes de l'image $y$ et $g$ à la position $(m, n)$ , l'échelle $j$ et l'orientation $o$ .
$c_{mn}, a_{mn}$	représentent respectivement les coefficients d'approximation de la transformée d'ondelettes de l'image $y$ et $g$ à la position $(m, n)$ .
$HH_j, HL_j, LH_j$	Les sous-bandes correspondent respectivement aux coefficients de détail d'orientation diagonale, horizontale et verticale.
$LL_{J_c}$	La sous-bande $LL_{J_c}$ représente les coefficients d'approximation à l'échelle la plus grossière.
$\mathcal{B}_{p,q}^r$	Espace de Besov de paramètres $r$ , $p$ et $q$ .
$M_j(p)$	Moment d'ordre $p$ de la distribution de probabilité des coefficients d'une représentation (e.g. ondelettes) à l'échelle $2^j$ .
$RB(g y, \sigma_\epsilon^2)$	signifie l'application de la règle de Bayes pour introduire l'information <i>a priori</i> sur l'image à estimer $g$ sachant l'image observée $y$ et le niveau de bruit $\sigma_\epsilon$ .
$X \sim \mathcal{N}(\mu, \sigma_\epsilon^2)$	signifie que la VA $X$ suit une loi normale de moyenne $\mu$ et de variance $\sigma_\epsilon^2$ .
$X \sim S_\alpha(\beta, \mu, \sigma_\epsilon)$	signifie que $X$ est distribuée selon une loi stable de paramètres $\alpha$ , $\beta$ , $\sigma_\epsilon$ et $\mu$ .
$X \sim S\alpha S$	signifie que $X$ suit une loi stable symétrique $X \sim S_\alpha(0, 0, \sigma_\epsilon)$ .
$g_F$	une version filtrée d'une image $g$ par un filtre passe-bande $F$ .



# Introduction Générale



# Avant-propos

---

## Contexte et problématique

L'image recueillie en sortie de tout capteur d'image subit une dégradation engendrée par la chaîne d'acquisition. Celle-ci est représentée par la fonction de transfert du système. On peut distinguer les causes dites déterministes engendrant une perte d'information provoquée par une transformation déterministe comme un filtrage (e.g. lissage par la défocalisation d'une optique) ou une déformation (e.g. mouvement apparent d'une caméra), où seule une portion de l'image est acquise. Les autres causes sont, elles, dites stochastiques puisque les observations elles-mêmes sont des mesures physiques soumises à des fluctuations aléatoires dont les sources peuvent être le bruit du capteur, les fluctuations de la source lumineuse, etc.

D'une manière générale, le signal déterministe de l'image observée est contaminé par des fluctuations stochastiques que l'on qualifie généralement de bruit. Ce dernier peut être soit additif, soit multiplicatif. Les traitements de restauration sont souvent indispensables pour améliorer la qualité des images observées. Notre principal problème, que nous allons traiter, consistera à récupérer une image de bonne qualité, proche de l'image originale, à partir d'une image bruitée de mauvaise qualité. Un exemple de débruitage d'image est présenté dans la Fig 1.

Dans la littérature du traitement d'images, différentes méthodes de débruitage ont été proposées et développées. Tout d'abord, des méthodes de filtrage spatial ont été proposées. Celles-ci consistent à réduire le bruit dans les zones qui ne présentent pas d'objets intéressants et à accentuer la perception des structures d'intérêt. Ces techniques de filtrage utilisent un filtre passe-bas pour supprimer les hautes fréquences, ce qui a pour inconvénient d'atténuer les contours de l'image. Pour parer à ces problèmes, de nouvelles techniques, plus performantes, ont vu le jour aux cours des années 80 et 90 ; citons notamment les approches variationnelles basées sur les EDPs, les approches utilisant les champs de Markov et les approches basées sur les transformées multi-échelles, notamment la transformée en ondelettes. Récemment, ces dernières ont montré leur puissance dans le cadre de l'estimation statistique. Par le biais de ces transformées parcimonieuses, l'énergie du signal utile est concentrée sur un faible nombre de coefficients, ce qui offre ainsi un cadre naturel non linéaire pour estimer ce signal. En effet, il suffit de seuiller les coefficients de l'image observée et d'inverser la transformée pour obtenir une estimée du signal utile. Par ailleurs, les transformées parcimonieuses présentent de très bonnes propriétés mathématiques permettant de couvrir une très large classe d'images.

Dans le cadre de cette thèse, nous nous sommes intéressés à l'estimation statistique bayésienne dans le domaine des transformées multi-échelles parcimonieuses orientées et non-orientées comme solution au problème de débruitage.

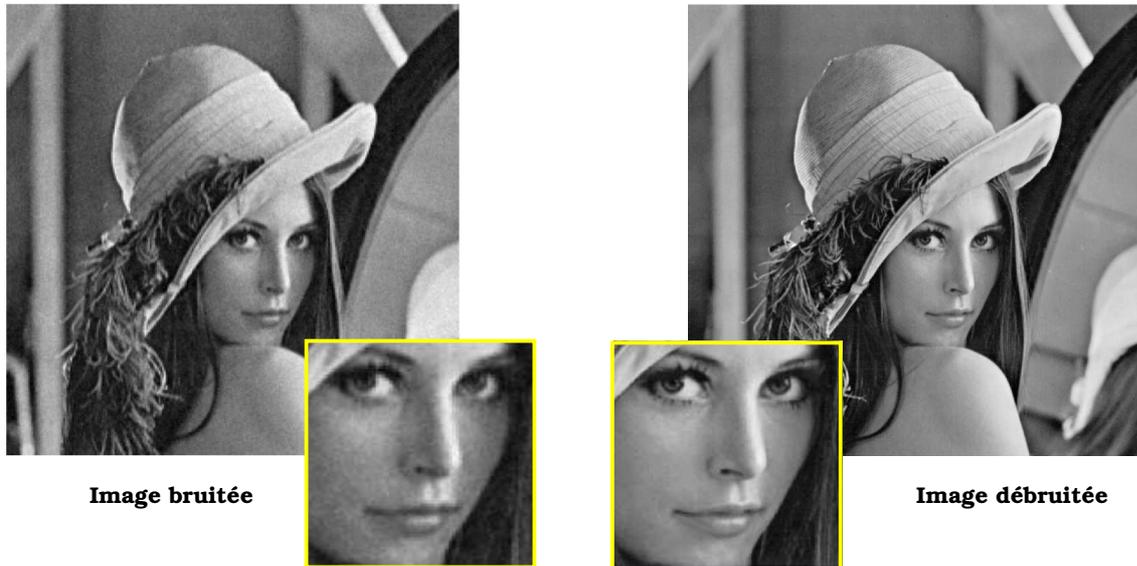


FIG. 1 – Exemple de débruitage d’images.

## Organisation et principales contributions de ce travail

Le présent manuscrit est organisé en trois parties. La première est consacrée à l’état de l’art relatif au débruitage multi-échelle des images. La deuxième et la troisième parties sont les principales contributions de notre travail. Nous y exposons des modèles statistiques univariés et multivariés dans l’espace des transformées apportant une solution à la problématique posée.

Dans le premier chapitre, après avoir exposé le formalisme général du problème envisagé dans le domaine des transformées multi-échelles, nous présentons les principales méthodes développées à ce jour s’appuyant sur les représentations multi-échelles orientées et non-orientées. Nous distinguons deux approches : une approche classique de type seuillage de coefficients et une approche bayésienne basée sur un modèle statistique *a priori* adapté à la modélisation des coefficients. Pour conclure cet état de l’art, dans la section 1.5, nous proposons un bilan comparatif entre les deux types d’approches tout en soulignant leurs limitations.

Dans le deuxième chapitre, nous nous penchons sur la modélisation statistique des images et nous décrivons les propriétés statistiques permettant de caractériser les images, dans l’espace des transformées, afin d’adapter les traitements pour le débruitage. Ensuite, nous exposons les principaux modèles généraux pour la modélisation statistique qui ont été proposés dans la littérature du traitement d’images.

Dans le contexte bayésien et afin de reconstruire l’image non bruitée, il est nécessaire d’imposer des contraintes sur la solution reconstruite, ce qui consiste à choisir un modèle *a priori*. Dans le chapitre 3, nous présentons le cadre général du modèle de mélange d’échelles de gaussiennes comme *a priori* adéquat à la modélisation des statistiques marginales des images dans l’espace des transformées multi-échelles non-orientées, et en l’occurrence les ondelettes. Plus précisément, nous introduisons deux types d’*a priori* :  $\alpha$ -stable et les Formes K de Bessel (BKF) qui tiennent compte des propriétés parcimonieuses des coefficients d’images dans le domaine multi-échelle, et les différentes techniques permettant d’estimer les hyperparamètres de ces derniers. Ceci est effectué dans un cadre univarié où les coefficients d’images dans l’espace des transformées sont supposés indépendants.

---

Dans le chapitre 4, nous proposons de nouveaux estimateurs bayésiens de type *Espérance conditionnelle a posteriori* (ECP) et *Maximum a posteriori* (MAP) pour le débruitage, basés sur les *a priori* :  $\alpha$ -stable et BKF. Deux étapes sont nécessaires pour la mise en oeuvre de ces estimateurs, la première consistant à estimer les hyperparamètres du modèle de l'*a priori* en présence du bruit, et la seconde à trouver une forme analytique pour l'estimateur bayésien correspondant. Dans la section 4.5, nous montrons des résultats sur des images naturelles pour le cas du bruit blanc gaussien.

Les résultats, que nous présentons dans la deuxième partie, ont fait l'objet de plusieurs publications [Boubchir et al., 2003, Boubchir et al., 2004, Fadili & Boubchir, 2005, Boubchir & Fadili, 2005a, Boubchir & Fadili, 2006, Boubchir, 2006].

Dans la troisième partie, au chapitre 5, nous présentons le cadre général du modèle de mélange d'échelles de gaussiennes dans un cadre multivarié permettant de prendre en compte l'information géométrique du voisinage, qui se traduit par les dépendances inter- et intra-échelles entre les coefficients d'images dans l'espace des transformées. Nous introduisons aussi le modèle multivarié de l'*a priori* des Formes K de Bessel et les estimateurs permettant d'estimer ces hyperparamètres.

Au chapitre 6, nous proposons un estimateur bayésien multivarié basé sur une extension multivariée de l'*a priori* BKF, formulés dans le domaine des transformées multi-échelles orientées (les curvelets) et non-orientées (les ondelettes non-décimées). Pour la mise en oeuvre de cet estimateur, nous proposons, dans un premier temps, un estimateur pour les hyperparamètres du modèle de l'*a priori* en présence du bruit, et ensuite, la forme analytique correspondante pour l'estimateur bayésien. Dans la section 6.3, nous montrons des résultats sur des images naturelles pour le cas du bruit gaussien.

L'étape de la modélisation statistique multivariée, que nous proposons dans le chapitre 5, a fait l'objet de deux publications [Boubchir & Fadili, 2005c, Boubchir & Fadili, 2005b]. Par ailleurs, les résultats des débruiteurs bayésiens multivariés sur les images, que nous exposons dans le chapitre 6, sont en cours de préparation pour publication.



## Première partie

---

# Représentations parcimonieuses multi-échelles pour le débruitage

---



# Chapitre 1

## État de l'art : débruitage multi-échelle des images

---

### Sommaire

---

<b>1.1</b>	<b>Régression non-paramétrique . . . . .</b>	<b>10</b>
<b>1.2</b>	<b>Transformées multi-échelles et débruitage . . . . .</b>	<b>12</b>
1.2.1	Formulation générale du problème . . . . .	12
1.2.2	Débruitage classique . . . . .	13
1.2.3	Débruitage bayésien . . . . .	14
1.2.4	Estimation du niveau de bruit . . . . .	14
<b>1.3</b>	<b>Débruitage dans le domaine des transformées multi-échelles non-orientées . . . . .</b>	<b>15</b>
1.3.1	Approches classiques sans <i>a priori</i> . . . . .	15
1.3.1.1	Méthodes terme à terme . . . . .	15
1.3.1.2	Méthodes classiques par blocs . . . . .	20
1.3.2	Approches bayésiennes . . . . .	21
1.3.2.1	Méthodes bayésiennes univariées . . . . .	21
1.3.2.2	Méthodes bayésiennes multivariées . . . . .	24
<b>1.4</b>	<b>Débruitage dans le domaine des transformées multi-échelles orientées . . . . .</b>	<b>26</b>
1.4.1	Approches classiques sans <i>a priori</i> . . . . .	26
1.4.2	Approches bayésiennes . . . . .	26
<b>1.5</b>	<b>Conclusion . . . . .</b>	<b>28</b>

---

L'obtention d'information à partir de mesures corrompues par un bruit reste un problème ouvert, que ce soit en traitement du signal ou en traitement d'image. Au cours des années 90, les ondelettes sont apparues comme un nouvel outil de débruitage, notamment sous l'influence des travaux fondateurs de Donoho & Johnstone. Un aperçu global sur les différentes approches du problème de la régression non paramétrique (débruitage) est présenté dans la première partie de ce chapitre. La deuxième partie concerne le formalisme du problème envisagé dans le domaine des transformées multi-échelles. Dans la dernière partie, nous présentons les principales approches développées à ce jour s'appuyant sur des représentations multi-échelles orientées et non-orientées.

## 1.1 Régression non-paramétrique

La régression non-paramétrique a constitué un outil fondamental au cours des vingt dernières années dans le domaine du traitement statistique du signal, et est toujours un domaine actif de recherche. L'objectif consiste à recouvrer en mettant en avant différents types d'estimateurs linéaires comme une fonction inconnue  $g$ , corrompue par du bruit, sans spécification d'un modèle explicite (linéaire ou non) sur la fonction à retrouver. Les techniques de régression non-paramétrique (ou débruitage) offrent un panel d'outils simples d'emploi permettant de récupérer une structure, sans imposer un modèle paramétrique au préalable.

Au cours des années 80 et 90, des centaines d'articles sont parus, solution à la question de la régression non-paramétrique. Parmi les plus populaires se trouvent notamment ceux reposant sur l'emploi de séries orthogonales ou de splines [Härdle, 1990, Green & Silverman, 1994, Fan & Gijbels, 1996, Eubank, 1999, Wand & Jones, 1995]. Cependant ces différents types d'estimateurs souffrent d'un comportement peu satisfaisant dans le cas d'échantillons de taille finie et aussi de beaucoup de problèmes lorsque la fonction est irrégulière ou avec des singularités isolées.

Parallèlement, différents estimateurs non linéaires, dont les plus populaires sont ceux basés sur l'emploi d'arbres de classification et de régression et les splines de régression adaptatives, ont été proposés comme alternatives aux estimateurs présentés précédemment [Härdle, 1990, Green & Silverman, 1994, Fan & Gijbels, 1996, Eubank, 1999, Wand & Jones, 1995]. Bien que certains aient atteint des performances optimales du point de vue asymptotique, leur implémentation s'avère encore complexe, et leurs domaines d'application souvent limités à des fonctions présentant certaines propriétés de régularité.

Durant les années 90, le domaine de la régression non-paramétrique a été dominé par deux types d'estimateurs non linéaires introduits par Donoho & Johnstone [Donoho & Johnstone, 1994, Donoho & Johnstone, 1995] et Donoho, Johnstone, Kerkyacharian & Picard [Donoho et al., 1995], à savoir la "contraction par ondelettes" (wavelet shrinkage) et le "seuillage par ondelettes" (wavelet thresholding). Ces estimateurs, grâce aux efforts menés par Mallat pour aller de l'analyse multi-résolution par ondelettes à la théorie des bancs de filtres, sont facilement mis en application par des algorithmes rapides, et sont ainsi très attrayants dans leur mise en oeuvre [Mallat, 1989].

En statistiques mathématiques, de nouveaux travaux théoriques ont par ailleurs montré l'optimalité de la régression non-paramétrique dans le domaine des ondelettes au sein d'espaces fonctionnels beaucoup plus larges que ceux utilisés précédemment. En d'autres termes, les différentes classes de fonctions envisagées couvrent aussi bien celles de Hölder ou de Sobolev que celles des classes de fonctions irrégulières comme celles appartenant aux espaces de Besov. Cet accroissement du domaine d'application de ces différents estimateurs se révèle être d'un grand intérêt dans le cadre d'analyse de signaux réels, comme dans le cas de la parole, d'électrocardiogrammes ou des signaux sismiques [Meyer, 1992, Wojtaszczyk, 1997, Donoho & Johnstone, 1998a, Härdle et al., 1998]. Ces signaux peuvent être caractérisés à l'aide des espaces de Besov ou de Triebel [Meyer, 1992, Donoho & Johnstone, 1998a]. Cependant, l'analyse d'un point de vue asymptotique ne permet pas de prédire le comportement d'un estimateur pour un échantillon de taille finie et ainsi de déterminer la taille minimale de l'échantillon permettant une estimation statistique. L'analyse du risque exact réalisée par Marron [Marron et al., 1998] et appliquée aux deux types de seuillage par ondelettes introduits par Donoho & Johnstone [Donoho & Johnstone, 1994], à savoir le seuillage minimax et le seuillage universel, a permis la détermination du domaine d'application des estimateurs employant la transformée en ondelettes. Bruce & Gao [Bruce & Gao, 1996, Gao & Bruce, 1997, Gao, 1998] et Antoniadis [Antoniadis et al., 2001] ont également fourni des outils permettant l'analyse du comportement de ces estimateurs en

présence d'échantillons de taille finie.

Il a été montré que ces types d'estimateurs sont asymptotiquement presque optimaux ou optimaux au sens minimax alors que les estimateurs linéaires traditionnels sont sous-optimaux en ce qui concerne les classes de fonctions appartenant aux espaces de Besov ou de Triebel (voir [Delyon & Juditsky, 1996, Abramovich et al., 2000]).

Depuis la parution des articles fondateurs par Donoho & Johnstone [Donoho & Johnstone, 1994, Donoho & Johnstone, 1995, Donoho et al., 1995], la littérature de traitement d'image a assisté à une abondance de papiers appliquant ou proposant des modifications de l'algorithme original dans des problèmes d'estimation et/ou restauration d'images. Différentes alternatives au seuillage par ondelettes ont été développées [Vidakovic, 1999, Percival & Walden, 2000]. Ainsi, Donoho & Johnstone [Donoho & Johnstone, 1995] proposèrent l'estimateur SURE basé sur l'estimation du risque sans biais de Stein. Weyrich & Warhola [Weyrich & Warhola, 1995], Nason [Nason, 1996], Jansen, Malfait & Bultheel [Jansen et al., 1997] ont considéré des estimateurs basés sur une approche de validation croisée afin de déterminer le seuil. Abramovich & Benjamini [Abramovich & Benjamini, 1995, Abramovich & Benjamini, 1996] et Ogden & Parzen [Ogden & Parzen, 1996a, Ogden & Parzen, 1996b] considérèrent le seuillage sous l'angle d'une procédure de test d'hypothèses multiples. Enfin, des modifications ont été apportées par Hall, Penev, Kerkycharian & Pickard [Hall et al., 1997], Cai [Cai, 1999], Efromovich [Efromovich, 1999, Efromovich, 2000] et Cai & Silverman [Cai & Silverman, 2000] suggérant la réalisation du seuillage par bloc, qui consiste à réaliser le seuillage non plus coefficient par coefficient mais par blocs de coefficients. Il a été montré que ce type d'approche offrait un gain de performances au sens de l'erreur quadratique moyenne.

Dans le cadre d'un paradigme bayésien, différentes approches ont également envisagé l'estimation dans le domaine des ondelettes. Citons par exemple [Simoncelli & Adelson, 1996, Chipman et al., 1997, Abramovich et al., 1998, Crouse et al., 1998, Johnstone & Silverman, 1998, Vidakovic, 1998, Moulin & Liu, 1999, Clyde & George, 1999, Clyde & George, 2000, Vannucci & Corradi, 1999, Huang & Lu, 2000, Chang et al., 2000a, Chang et al., 2000b, Achim et al., 2001]. Ces estimateurs s'avèrent plus efficaces que ceux présentés précédemment, que ce soit terme à terme ou par bloc [Antoniadis et al., 2001]. Dans le cadre d'une approche bayésienne, on suppose que la distribution des coefficients d'ondelettes est imposée comme *a priori*. Le choix d'une fonction de coût conditionne l'obtention d'une règle d'estimation bayésienne. Un choix simple est l'*a priori* gaussien [Chipman et al., 1997] ou encore un mélange d'une distribution gaussienne et d'une distribution de Dirac centrée en zéro [Clyde et al., 1998, Abramovich et al., 1998]. Dans [Vidakovic & Ruggeri, 2000], les auteurs ont proposé un modèle bayésien hiérarchique à deux niveaux, avec une densité exponentielle symétrique et une Dirac centrée en zéro, et un *a priori* exponentiel sur la variance. Mallat [Mallat, 1989] a été le premier à proposer les distributions gaussiennes généralisées (GGD). Elles sont très communément adoptées comme *a priori* à des fins d'estimation et de compression dans le domaine des ondelettes [Simoncelli & Adelson, 1996, Moulin & Liu, 1999, Chang et al., 2000a, Chang et al., 2000b]. Simoncelli [Simoncelli, 1999] a utilisé un mélange local de gaussiennes pour le débruitage. Cependant, à cause de sa décroissance exponentielle rapide, l'*a priori* GGD se révèle incapable de modéliser correctement les queues de distribution relevées, comportement typique des distributions des coefficients d'ondelettes. Achim et al. [Achim et al., 2001] ont proposé la mise en oeuvre d'un *a priori* basé sur les lois  $\alpha$ -stables [Nikias & Shao, 1995]. Cependant, leur estimateur des hyperparamètres reste extrêmement limité, particulièrement en présence de bruit. De plus, aussi bien pour l'*a priori* GGD que  $\alpha$ -stable, mais aussi pour bon nombre d'*a priori* dans la littérature, aucune forme analytique n'a été proposée pour l'estimateur bayésien. Ce qui nécessite une implémentation numérique des intégrales, impliquées dans les estimateurs, particulièrement

instable (notamment à cause des bornes d'intégration infinies) et coûteuse en temps.

De nombreux nouveaux travaux sont apparus en appliquant ces approches classiques et bayésiennes dans le domaine des transformées multi-échelles orientées, e.g. les curvelets [Candès & Donoho, 2002, Starck et al., 2002, Candès & Donoho, 2004, Candès et al., 2006], contourlets [Do & Vitterli, 2003a, Do & Vitterli, 2003b], wedgelets [Donoho, 1998], bandelets [Mallat & LePennec, 2005, LePennec & Mallat, 2005] et les ondelettes orientées "steerable wavelets" [Freeman & Adelson, 1991, Simoncelli et al., 1992]. Beaucoup de ces transformées sont redondantes et permettant de décomposer les images en sous-bandes orientées. La raison de se tourner vers ces représentations redondantes est d'avoir la propriété de presque "invariance" par translation et d'obtenir une meilleure sélectivité directionnelle. Parmi les meilleures méthodes de débruitage d'images, opérant dans le domaine des transformées multi-échelles orientées, se trouvent celles proposées dans [Portilla et al., 2003, Starck et al., 2003, Matalon et al., 2005]. Ces méthodes sont à la pointe de l'état de l'art.

Du fait de l'intérêt croissant porté à la restauration non-linéaire, nous nous sommes focalisés sur des estimateurs non-linéaires non-paramétriques formulés dans le cadre de décompositions multi-échelles. Le formalisme introduit ainsi que les principales approches développées à ce jour s'appuyant sur une représentation multi-échelle d'images font l'objet des sections suivantes.

## 1.2 Transformées multi-échelles et débruitage

### 1.2.1 Formulation générale du problème

Le problème du débruitage (connu sous le nom de "régression non-paramétrique" en statistiques mathématiques) peut se mettre sous la forme générale suivante :

$$y_{mn} = g_{mn} + \epsilon_{mn} \quad (1.1)$$

où nos observations dégradées  $y_{mn}$ ,  $m, n = 0, \dots, N-1$  représentant les valeurs réelles échantillonnées d'une image bruitée, sont modélisées comme la somme d'un signal  $g_{mn}$  à estimer et d'un bruit blanc gaussien  $\epsilon_{mn}$  de moyenne nulle et de variance  $\sigma^2$ . L'objectif est de recouvrer l'information  $g$  contenue dans le signal bruité  $y$  sans pour autant faire d'hypothèses sur une structure paramétrique de  $g$ . Étant donné le vecteur  $\mathbf{y}$  qui représente les valeurs échantillonnées de  $y$ , la transformée multi-échelle de  $\mathbf{y}$  est donnée par  $\mathbf{d} = \Phi^T \mathbf{y}$  où  $\Phi$  est une matrice (dite dictionnaire) de taille  $N^2 \times L$  (avec  $L \geq N^2$ ).

Dans le cas de la transformée d'ondelettes discrète orthogonale, la matrice  $\Phi = \mathcal{W}$  où  $\mathcal{W}$  est une matrice orthogonale associée à la base orthonormée choisie, et  $\mathbf{d}$  un vecteur comprenant d'une part les coefficients d'approximation et d'autre part les coefficients de détail de la transformée en ondelettes discrète (DWT). Du fait de l'orthogonalité de la matrice  $\mathcal{W}$ , la transformée en ondelettes discrète inverse (IDWT) est donnée par  $\mathbf{y} = \mathcal{W}^T \mathbf{d}$ . Dans le cas où la taille  $N$  peut se mettre sous la forme  $N = 2^J$  avec  $J \in \mathbb{N}$ , la DWT ainsi que sa transformée inverse peuvent être implémentées à l'aide de l'algorithme pyramidal proposé par Mallat [Mallat, 1989] employant un banc de filtres miroirs en quadrature. Dans le cas bidimensionnel, les sous-bandes  $HH_j$ ,  $HL_j$  et  $LH_j$ ,  $j = J_c, \dots, J-1$  correspondent respectivement aux coefficients de détail d'orientations diagonale, horizontale et verticale. La sous-bande  $LL_{J_c}$  représente les coefficients d'approximations à l'échelle la plus grossière. En appliquant la DWT [Mallat, 1989] à l'image bruitée  $y$  nous obtenons alors à partir de l'Eq.1.1 :

$$\begin{cases} c_{mn} &= a_{mn} + \epsilon_{mn}, \quad m, n = 0, \dots, 2^{J_c} - 1 \\ d_{mn}^{o,j} &= s_{mn}^{o,j} + \epsilon_{mn}, \quad j = J_c, \dots, J - 1; m, n = 0, \dots, 2^j - 1 \end{cases} \quad (1.2)$$

où nous notons  $a_{mn}$  (resp.  $c_{mn}$ ) le coefficient d'approximation de la DWT de l'image  $g$  (resp.  $y$ ) à la position  $(m, n)$  et  $s_{mn}^{o,j}$  (resp.  $d_{mn}^{o,j}$ ) le coefficient de détail de la DWT de l'image  $g$  (resp.  $y$ ) à la position  $(m, n)$ , l'échelle  $j$  et l'orientation  $o$ . Un exemple d'application de la DWT sur une image test est illustré par la Fig.1.2. Du fait du caractère orthogonal de la transformée en ondelettes, les coefficients  $\epsilon_{mn}$  sont des variables aléatoires indépendantes  $\mathcal{N}(0, \sigma^2)$  qui définissent aussi un bruit blanc gaussien.

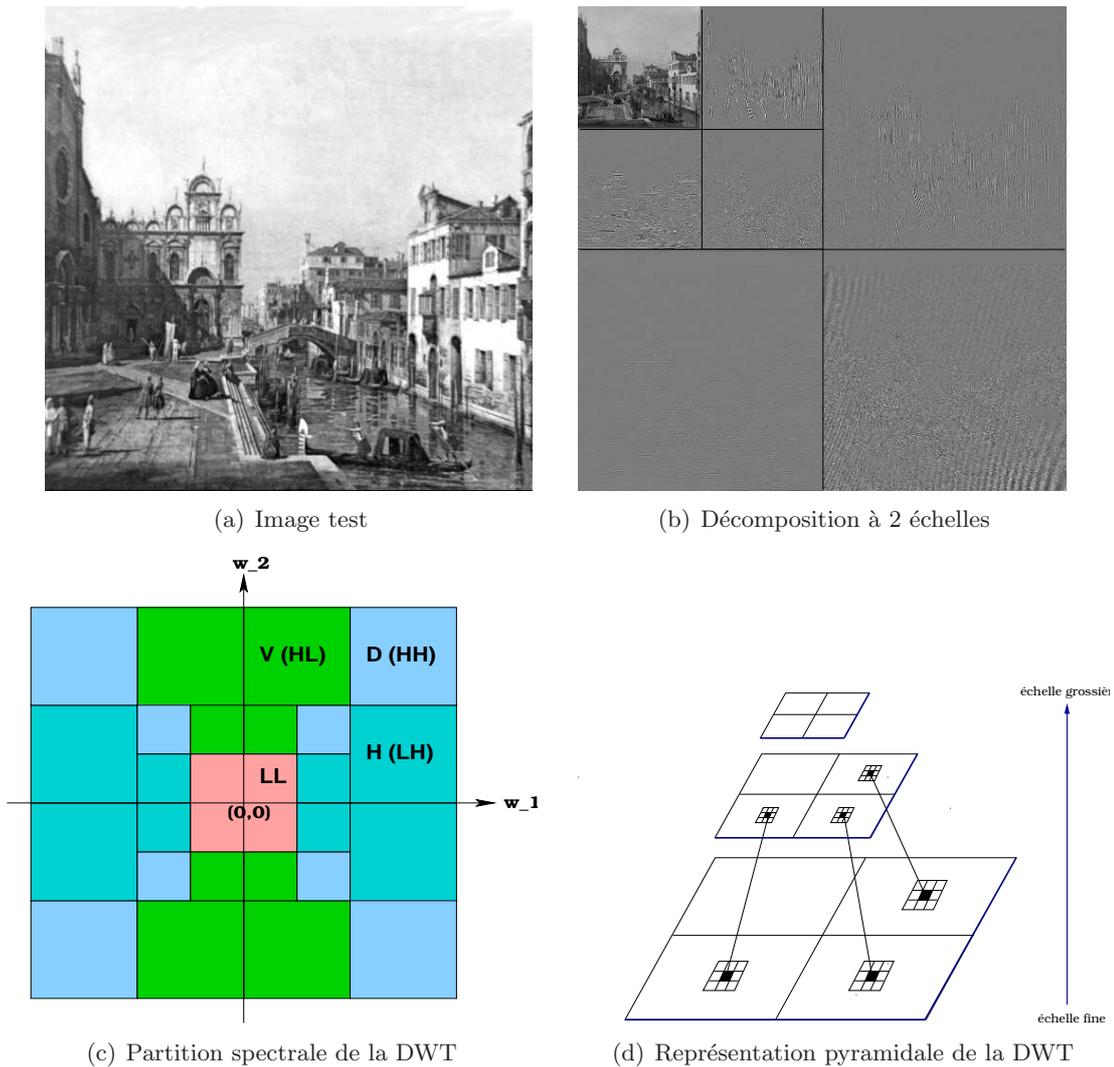


FIG. 1.1 – Application de la transformée d'ondelettes séparable 2D sur une image.

### 1.2.2 Débruitage classique

Les approches de débruitage classique sont des approches basées sur une sélection judicieuse des coefficients d'image coefficient par coefficient. Du fait du caractère creux (propriété dite de

parcimonie, "sparse" en anglais) des transformées multi-échelles [Mallat, 1999], nous pouvons supposer de façon intuitive que seuls quelques coefficients de détail  $s_{mn}^{oj}$  ayant une valeur suffisamment élevée contribuent à l'image à recouvrir  $g$ , alors que les coefficients de faibles valeurs sont dus essentiellement au bruit qui contamine de façon uniforme tous les coefficients. Il est également recommandé de conserver les coefficients d'approximation  $c_{mn}$ . Ces derniers, relatifs aux composantes de basses fréquences, sont essentiellement caractéristiques du signal original. Le diagramme suivant représente le processus complet de débruitage dans le domaine des transformées multi-échelles.

$$\mathbf{y} \xrightarrow{\Phi^T} \{c_{mn}, d_{mn}^{oj}\} \xrightarrow{\text{estimateur non-linéaire } \delta_\lambda} \{c_{mn}, \delta_\lambda(d_{mn}^{oj})\} \xrightarrow{\mathcal{R}} \hat{\mathbf{g}}$$

où  $\Phi^T$  représente une transformée multi-échelle et  $\mathcal{R}$  est l'opérateur de reconstruction (qui est  $\Phi$  dans le cas des bases ou  $\Phi^+$ , qui est l'inverse généralisée à gauche de Moor-Penrose, dans le cas des trames<sup>1</sup>).  $\delta_\lambda$  est un opérateur non linéaire de type seuillage de coefficients, reposant sur la conservation en intégralité des coefficients  $c_{mn}$  et sur une sélection judicieuse des coefficients  $d_{mn}^{oj}$ . Les coefficients ayant été traités, le signal restauré est reconstruit par la transformée inverse.

### 1.2.3 Débruitage bayésien

Dans le cadre d'une approche bayésienne, un modèle statistique *a priori* adapté à la classe des signaux à estimer est imposé aux coefficients d'ondelettes pour décrire leur distribution. L'application de la règle de Bayes (RB) pour introduire l'information *a priori* permet d'estimer le signal  $g$ . Le diagramme suivant représente le processus complet de débruitage bayésien dans le domaine des transformées multi-échelles.

$$\mathbf{y} \xrightarrow{\Phi^T} \{c_{mn}, d_{mn}^{oj}\} \longrightarrow \underbrace{a \text{ priori sur } \{a_{mn}, s_{mn}^{oj}\}}_{\text{paramétrisé par } \theta} \xrightarrow{\text{estimateur a posteriori}} \{c_{mn}, \hat{s}_{mn}^{oj}(d_{mn}^{oj} | \theta, \sigma_\epsilon^2)\} \xrightarrow{\mathcal{R}} \hat{\mathbf{g}}$$

où  $\theta$  est l'ensemble des hyperparamètres du modèle *a priori*.

Pour ces deux catégories de débruitage, une estimation du niveau de bruit  $\sigma_\epsilon$  doit être réalisée.

### 1.2.4 Estimation du niveau de bruit

Dans la littérature, la majorité des méthodes de débruitage n'aborde que le cas de bruit blanc gaussien, plus simple à traiter, bien que, en situation de données réelles, il ne soit pas spécialement facile d'estimer le niveau de bruit  $\sigma_\epsilon$ . Notons que le bruit poissonnien à forte intensité ou le bruit de mélange poissonnien-gaussien peut être stabilisé pour le ramener au cas gaussien. Avec une transformée discrète orthogonale en ondelettes, le bruit blanc se décompose en série de coefficients aléatoires normaux centrés et décorrélés  $\epsilon_{mn}$  (Eq.1.2).

En utilisant des arguments de la statistique robuste, Donoho & Johnstone ont proposé une estimation de  $\sigma_\epsilon$  dans le domaine des ondelettes en ne considérant que les coefficients de l'échelle de décomposition la plus fine. Le choix de l'échelle la plus fine repose sur l'hypothèse que les coefficients en ondelettes correspondants sont en grande majorité dus au bruit blanc. Une estimée de  $\hat{\sigma}_\epsilon$  est alors obtenue par un résultat classique en statistique robuste :

<sup>1</sup>Ceci est vrai lorsque la trame de reconstruction est la trame duale, ce qui est vrai par exemple pour un banc de filtres QMF en ondelettes.

$$\hat{\sigma}_\epsilon = \frac{\text{MAD}(\{d_{mn}^{J-1}\})}{0.6745} \quad (1.3)$$

où le MAD est la valeur médiane absolue des coefficients de l'échelle la plus fine. Le facteur 0.6745 est choisi après une calibration avec une distribution gaussienne. Cet estimateur très robuste est également très populaire pour le débruitage multi-échelle.

Après avoir présenté le formalisme général du problème de débruitage, nous allons introduire, dans ce qui suit, le formalisme de différents estimateurs de débruitage développés à ce jour s'appuyant sur des représentations parcimonieuses multi-échelles orientées et/ou non-orientées. Ces estimateurs sont classés en deux types d'approches : approches classiques sans *a priori* et approches bayésiennes.

**Remarque 1.1** *Dans tout ce document, nous employons le terme "transformée multi-échelle non-orientée" pour dénommer les transformées sans sélectivité directionnelle comme la transformée d'ondelettes à une bande. Par abus de langage, cette terminologie dénotera aussi la transformée en ondelettes standard qui présente trois orientations.*

## 1.3 Débruitage dans le domaine des transformées multi-échelles non-orientées

### 1.3.1 Approches classiques sans *a priori*

#### 1.3.1.1 Méthodes terme à terme

**1.3.1.1.1 Estimateurs par seuillage d'ondelettes** Donoho & Johnstone ont proposé un estimateur non-linéaire de  $g$  reposant sur la conservation en intégralité des coefficients  $c_{mn}$  et sur une sélection judicieuse des coefficients  $d_{mn}^{oj}$  [Donoho & Johnstone, 1994, Donoho & Johnstone, 1995, Donoho et al., 1995]. Ces auteurs suggèrent l'extraction des coefficients de détail significatifs par comparaison de ces derniers avec un paramètre de seuillage  $\lambda > 0$  dont le choix est décrit ultérieurement. Les fonctions de seuillage résultantes, se déclinent sous deux formes, seuillage dur (noté  $H$ , Fig.1.2-(a)) et seuillage doux (noté  $S$ , Fig.1.2-(b)), définis respectivement par,

$$\delta_\lambda^H(d_{mn}^{oj}) = \begin{cases} 0 & \text{si } |d_{mn}^{oj}| \leq \lambda \\ d_{mn}^{oj} & \text{si } |d_{mn}^{oj}| > \lambda \end{cases} \quad (1.4)$$

et

$$\delta_\lambda^S(d_{mn}^{oj}) = \begin{cases} 0 & \text{si } |d_{mn}^{oj}| \leq \lambda \\ d_{mn}^{oj} - \text{sign}(d_{mn}^{oj})\lambda & \text{si } |d_{mn}^{oj}| > \lambda \end{cases} \quad (1.5)$$

Le seuillage dur (fonction discontinue) correspond à une loi du type 'kill or keep'. En d'autres termes, soit le coefficient est conservé, soit il est mis à zéro.

Le seuillage doux (fonction continue) correspond à un opérateur de contraction de type 'kill or shrink'. En d'autres termes, soit le coefficient se voit retrancher la valeur du seuil, soit il est mis à zéro.

En terme de risque d'estimation, le seuillage dur aboutit à une variance plus importante de la fonction estimée que celle obtenue par seuillage doux avec un seuil identique. Du fait de sa discontinuité, il présente en outre une sensibilité importante vis-à-vis des faibles variations des

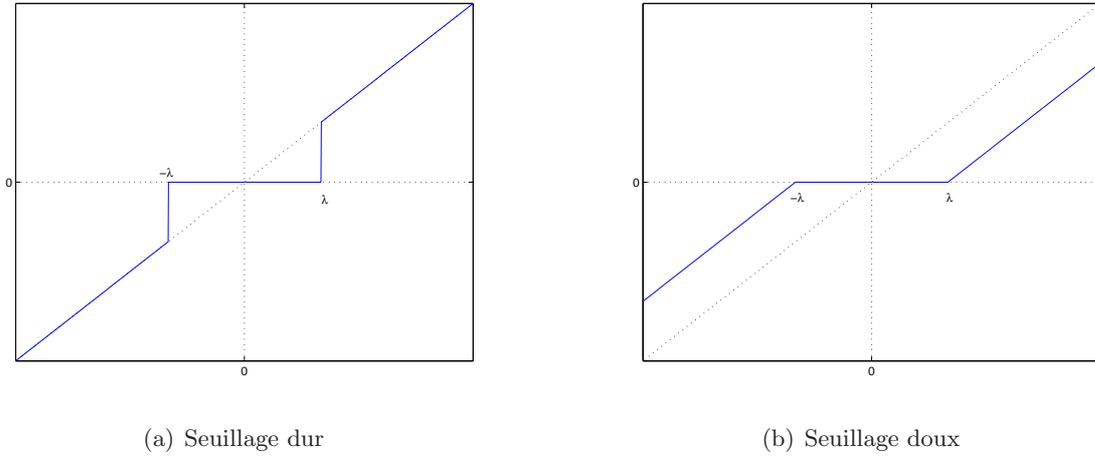


FIG. 1.2 – Estimateurs par seuillage d'ondelettes.

données. La réduction des coefficients de grande valeur, dans le cas du seuillage doux entraîne un biais d'estimation plus élevé.

Dans le cas des transformées multi-échelles avec des bases, le seuillage doux a une interprétation statistique d'un estimateur MAP (cf. annexe A) avec un *a priori* de parcimonie sur les coefficients.

**1.3.1.1.2 Autres variantes** Plusieurs méthodes ont été proposées afin de trouver un compromis entre le seuillage dur et le seuillage doux [Gao & Bruce, 1997, Gao, 1998, Vidakovic, 1999, Antoniadis et al., 2001]. Chacune des fonctions d'estimation citées précédemment est dépendante du choix du seuil  $\lambda$ .

- **Seuillage 'Firm'**

L'estimateur de seuillage de loi 'Firm' proposée par Gao & Bruce [Gao & Bruce, 1997] évite la discontinuité du seuillage dur et le biais d'estimation du seuillage doux, et se place comme cas intermédiaire entre les deux. Il est défini de la manière suivante,

$$\delta_{\lambda_1, \lambda_2}^F(d_{mn}^{oj}) = \begin{cases} 0 & \text{si } |d_{mn}^{oj}| \leq \lambda_1 \\ \text{sign}(d_{mn}^{oj}) \frac{\lambda_2(|d_{mn}^{oj}| - \lambda_1)}{\lambda_2 - \lambda_1} & \text{si } \lambda_1 < |d_{mn}^{oj}| \leq \lambda_2 \\ d_{mn}^{oj} & \text{si } |d_{mn}^{oj}| > \lambda_2 \end{cases} \quad (1.6)$$

Néanmoins, cette méthode souffre d'un important désavantage. Elle nécessite le choix de deux seuils sans aucune théorie asymptotique pour en préconiser une méthode de choix.

- **Seuillage 'nonnegative garrote'**

Afin de pallier le problème de l'estimateur précédent concernant le choix de deux seuils, Gao [Gao, 1998] propose le seuillage 'nonnegative garrote' qui conserve les avantages de la loi précédente tout en ne nécessitant que le choix d'un seul seuil. Le seuillage 'nonnegative garrote' (fonction continue) correspond à une contraction. Il est défini par,

$$\delta_{\lambda}^G(d_{mn}^{oj}) = \begin{cases} 0 & \text{si } |d_{mn}^{oj}| \leq \lambda \\ d_{mn}^{oj} - \frac{\lambda^2}{d_{mn}^{oj}} & \text{si } |d_{mn}^{oj}| > \lambda \end{cases} \quad (1.7)$$

Cet estimateur offre des résultats bien meilleurs, dans le cas des échantillons de taille

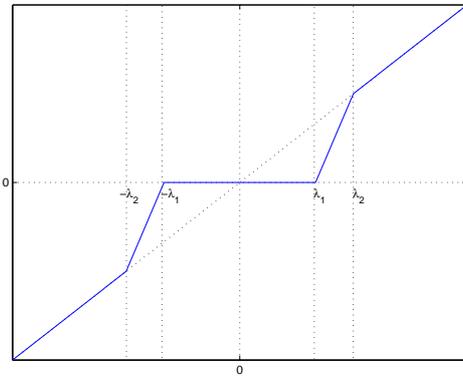
limitée, que les estimateurs de seuillages dur et doux mais reste comparable à l'estimateur de seuillage 'Firm'.

- **Seuillage 'SCAD'**

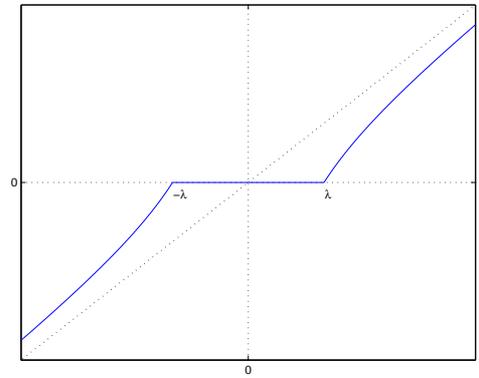
Dans le même sens, Antoniadis & Fan [Antoniadis et al., 2001] ont proposé le seuillage 'SCAD' dont la règle de seuillage est définie comme suit,

$$\delta_{\lambda}^{SCAD}(d_{mn}^{oj}) = \begin{cases} \text{sign}(d_{mn}^{oj}) \max(0, |d_{mn}^{oj}| - \lambda) & \text{si } |d_{mn}^{oj}| \leq 2\lambda \\ \frac{(\alpha-1)d_{mn}^{oj} - \alpha\lambda \text{sign}(d_{mn}^{oj})}{\alpha-2} & \text{si } 2\lambda < |d_{mn}^{oj}| \leq \alpha\lambda \\ d_{mn}^{oj} & \text{si } |d_{mn}^{oj}| > \alpha\lambda \end{cases} \quad (1.8)$$

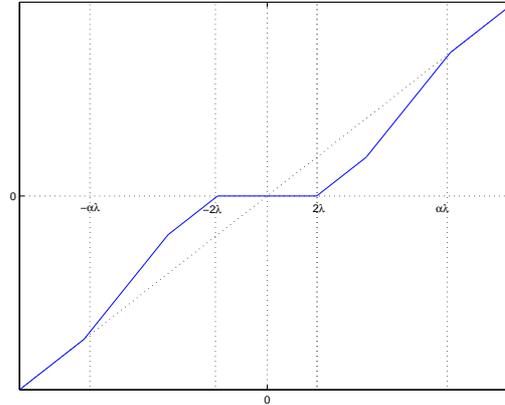
où  $\alpha = 3.7$  (recommandé par Antoniadis & Fan). Cet estimateur (fonction linéaire par morceaux) correspond à une loi de contraction.



(a) Seuillage 'Firm'



(b) Seuillage 'nonnegative garrote'



(c) Seuillage 'SCAD'

FIG. 1.3 – Autres variantes des estimateurs par seuillage d'ondelettes.

**1.3.1.1.3 Choix du seuil** Il existe de nombreuses méthodes permettant de déterminer la valeur du seuil. Ces différentes méthodes forment deux catégories distinctes : le seuillage global et le seuillage dépendant de l'échelle. Dans le premier cas, une unique valeur de  $\lambda > 0$  est appliquée à l'ensemble des coefficients d'ondelettes, alors que dans le second cas une valeur de  $\lambda > 0$

est définie pour chaque échelle de la décomposition. Nous allons nous intéresser aux différentes approches les plus communément adoptées.

• **Seuil minimax** Donoho & Johnstone ont proposé le seuillage minimax qui applique un seuil optimal (au sens minimax)[Donoho & Johnstone, 1994]. Le seuil minimax dépend de la taille de l'échantillon, et est choisi de manière à minimiser le risque maximum. En d'autres termes, le seuil minimax est défini de la manière suivante :

$$\lambda^M = \hat{\sigma}_\epsilon \lambda_N^* \quad (1.9)$$

où  $\lambda_N^*$  est correspond à la valeur de  $\lambda$  vérifiant

$$\lambda_N^* = \inf_{\lambda} \sup_{d_{mn}^{oj}} \left( \frac{R_\lambda(d_{mn}^{oj})}{N^{-2} + R_{oracle}(d_{mn}^{oj})} \right) \quad (1.10)$$

avec  $R_\lambda(d_{mn}^{oj}) = \mathbb{E} \left[ \delta_\lambda(d_{mn}^{oj}) - d_{mn}^{oj} \right]^2$  et  $R_{oracle}(d_{mn}^{oj})$  est le risque optimal obtenu à l'aide d'un oracle, qui simplifie l'estimation en fournissant de l'information normalement disponible sur l'image. Donoho & Johnstone considèrent deux oracles [Donoho & Johnstone, 1994] : le DLP (en anglais *Diagonal Linear Projection*), faisant intervenir un opérateur diagonal réalisant une projection linéaire et aboutissant à une décision de type 'kill or keep', et le DLS (en anglais *Diagonal Linear Shrinker*), oracle définissant le facteur d'atténuation à appliquer à chacun des coefficients de détail d'ondelettes.

• **Seuil universel** Une alternative à l'utilisation du seuil minimax a été proposée par Donoho & Johnstone [Donoho & Johnstone, 1994]. Elle repose sur l'utilisation d'une valeur de seuil universel :

$$\lambda^U = \hat{\sigma}_\epsilon \sqrt{2 \log(N^2)} \quad (1.11)$$

où  $N^2$  est la taille du signal en nombre d'échantillons. Une autre valeur de seuil universel a été longtemps utilisée dans la communauté astronomique :  $\lambda^U \simeq 3 - 4\hat{\sigma}_\epsilon$  pour  $N \simeq 256$  [Starck et al., 2002].

Le seuillage universel est substantiellement plus important que celui obtenu au sens minimax, mais s'avère aisément implémentable. Aussi, un nombre plus limité de coefficients est employé lors de la reconstruction ce qui a pour effet de lisser le signal en sortie par rapport au cas minimax d'où un biais d'estimation plus élevé. Il est de loin le plus répandu dans la communauté de traitement du signal et des images.

• **SureShrink** Dans [Donoho & Johnstone, 1995], Donoho & Johnstone ont proposé l'estimateur SureShrink où un seuil est défini pour les coefficients d'ondelettes de chaque échelle de décomposition. L'idée est basée sur la minimisation de l'estimation sans biais du risque de Stein 'SURE' ("*Stein's unbiased risk estimator*") [Stein, 1982]. Le seuil SureShrink est défini par

$$\lambda_j^S = \arg \min_{0 \leq \lambda \leq \lambda^U} \text{SURE} \left( \lambda; \frac{d_{mn}^{oj}}{\hat{\sigma}_\epsilon} \right) \quad (1.12)$$

où  $\lambda^U$  est le seuil universel donné par l'Eq.1.11.

L'inconvénient de cet estimateur réside dans les situations extrêmement parcimonieuses des coefficients d'ondelettes. Pour éviter ce problème, Donoho & Johnstone ont proposé une autre

alternative au seuillage SureShrink, dite le mode hybride du seuillage SURE, basé sur l'idée heuristique suivante [Donoho & Johnstone, 1995] : si un ensemble des coefficients d'ondelettes est jugé avoir une représentation parcimonieuse, alors utiliser le seuil universel  $\lambda^U = \hat{\sigma} \sqrt{2 \log(2^j)}$  ; sinon le critère SURE est utilisé pour déterminer la valeur du seuil. En termes mathématiques, le seuillage du mode hybride s'écrit sous la forme suivante, pour  $j = J_c, \dots, J - 1$ ,

$$\lambda_j^{HS} = \begin{cases} \hat{\sigma}_\epsilon \sqrt{2 \log(2^j)} & \text{si } \sum_{m,n=0}^{2^j-1} (d_{mn}^{oj})^2 \leq \hat{\sigma}_\epsilon^2 2^{j/2} (2^{j/2} + j^{3/2}) \\ \lambda_j^S & \text{sinon.} \end{cases} \quad (1.13)$$

• **Test d'hypothèses simples et multiples** Abramovich & Benjamini ont abordé le problème de débruitage sous l'angle de la théorie de décision par des tests d'hypothèses binaires [Abramovich & Benjamini, 1995]. L'idée est de traiter l'approche du seuillage par ondelettes par un test d'hypothèses multiples où chaque coefficient de détail est soumis au test d'hypothèse suivant :

$$H_0 : d_{mn}^{oj} = 0 \quad \text{contre} \quad H_1 : d_{mn}^{oj} \neq 0$$

Si le test  $H_0$  est rejeté, alors le coefficient  $d_{mn}^{oj}$  est retenu dans la reconstruction ; sinon il est rejeté.

Dans [Abramovich & Benjamini, 1995, Abramovich & Benjamini, 1996], Abramovich & Benjamini ont proposé un estimateur de seuillage pour corriger les tests multiples et maximiser le nombre de coefficients retenus dans le modèle. Leur approche est basée sur la méthode FDR (en anglais the false discovery rate) de Benjamini & Hochberg [Benjamini & Hochberg, 1995] afin de contrôler le taux d'erreur global. Les auteurs dans [Abramovich et al., 2000] ont montré l'optimalité de minimiser le rapport du FDR.

• **Test d'hypothèses récursives** L'approche de test d'hypothèses multiples décrite ci-dessus produit un seuil global. Ogden & Parzen [Ogden & Parzen, 1996a] ont développé une procédure de test d'hypothèses récursives permettant de produire un seuil  $\lambda_j$  à chaque échelle de décomposition, de même que l'estimateur de seuil SureShrink. Cette procédure n'inclut que les coefficients d'ondelettes de forte évidence nécessaires pour la reconstruction, comparée à l'approche [Abramovich & Benjamini, 1995, Abramovich & Benjamini, 1996] qui cherche à inclure le plus de coefficients possible.

• **Validation croisée** Une règle principale pour choisir le seuil  $\lambda$  est de minimiser l'erreur quadratique moyenne (EQM) entre l'image estimée par l'estimateur de seuillage d'ondelettes  $\hat{g}_\lambda$  et l'image vraie  $g$ .

Le critère de validation croisée a été utilisé comme procédure automatique pour guider l'estimation du seuil dans plusieurs situations (voir par exemple [Green & Silverman, 1994] ou [Eubank, 1999]). Cette approche a été adoptée pour le débruitage multi-échelle par [Jansen et al., 1997] et [Nason, 1994, Nason, 1996]. L'idée est de regrouper les coefficients d'ondelettes de l'image observée  $y$  (Eq.1.1) à chaque échelle  $j$  en deux sous-ensembles de même cardinalité : ensemble des coefficients d'indice pair et d'indice impair. Ensuite, appliquer un estimateur de seuillage dur (Eq.1.4) ou doux (Eq.1.5) de seuil  $\lambda$  à chaque ensemble pour estimer les coefficients d'indice pair et impair notés respectivement  $\hat{g}_\lambda^E$  et  $\hat{g}_\lambda^O$ , afin d'estimer l'EQM  $\hat{\xi}$  par

$$\hat{\xi}(\lambda) = \sum_{j=1}^{N/2} \left[ (\hat{g}_{\lambda,j}^E - y_{2j+1})^2 + (\hat{g}_{\lambda,j}^O - y_{2j})^2 \right] \quad (1.14)$$

où un unique minimum existe pour l'Eq.1.14

Dans [Nason, 1994], l'auteur a montré l'existence d'un unique minimum pour  $\hat{\xi}$ ,

$$\lambda_{\min} = \arg \min_{\lambda \geq 0} \xi(\lambda) \quad (1.15)$$

Finalement, le seuil de validation croisée est défini comme suit,

$$\lambda^{CV} = \left( 1 - \frac{\log 2}{\log N} \right)^{-1/2} \lambda_{\min} \quad (1.16)$$

La valeur du seuil minimum  $\lambda_{\min}$  dépend de  $N/2$  coefficients. Une correction d'ajustement pour déterminer la valeur de  $\lambda_{\min}$  pour  $N$  coefficients a été proposée dans [Nason, 1994].

• **Seuillage invariant par translation** Le seuillage de coefficients d'ondelettes bruités crée des oscillations près des discontinuités. Coifman & Donoho ont proposé le seuillage invariant par translation pour atténuer ces oscillations et améliorer le rapport signal sur bruit (SNR) [Coifman & Donoho, 1995]. Le seuillage des coefficients d'ondelettes des signaux traduits produit des oscillations différentes car celles-ci sont créées par des ondelettes différentes. Le moyennage réduit l'amplitude de ces oscillations car elles ne sont plus en phase. Dans le calcul de l'estimateur invariant par translation, et au lieu de décaler le signal, on peut décaler les ondelettes dans la direction inverse, ce qui est notamment réalisé par l'algorithme "à trous" développé par Mallat [Mallat, 1989]. Notons que la procédure du seuillage dans [Coifman & Donoho, 1995] est équivalente au seuillage des coefficients de la transformée d'ondelettes non-décimée si toutes les translations du cycle-spining sont appliquées.

### 1.3.1.2 Méthodes classiques par blocs

**1.3.1.2.1 Seuillage par blocs sans recouvrement** L'estimateur de seuillage par blocs sans recouvrement a été proposé par Cai [Cai, 1999]. A chaque échelle de décomposition  $j = J_c, \dots, J-1$ , les coefficients de détail  $d_{mn}^j$  sont groupés dans des blocs de coefficients sans recouvrement de taille  $B$  (où  $B$  n'est pas nécessairement divisible par  $2^j$ ). Les premiers coefficients peuvent être réutilisés pour remplir le dernier bloc (cas augmenté, Fig.1.4(b)) ou les derniers coefficients restants ne peuvent pas être utilisés (cas tronqué, Fig.1.4(c)). Soit  $(jb)$  l'ensemble des indices des coefficients dans le  $b^{\text{ème}}$  bloc à l'échelle  $j$ ,

$$(jb) = \{(j, (n, m)) : (b-1)B + 1 \leq (m, n) \leq bB\}, \quad (1.17)$$

et soit  $S_{(jb)}^2$  l'énergie du signal bruité dans le bloc  $(jb)$ . Dans chaque bloc  $(jb)$ , les coefficients d'ondelettes  $d_{mn}^j$  sont estimés par l'intermédiaire de l'estimateur du risque SURE,

$$\tilde{d}_{mn}^{(jb)} = \max \left( 0, \frac{S_{(jb)}^2 - \lambda B \sigma^2}{S_{(jb)}^2} \right) d_{mn}^j \quad (1.18)$$

Cai [Cai, 1999] a suggéré de prendre  $B = \log N$  et  $\lambda = 4.50524$  (qui représente la solution de l'équation  $\lambda - \log \lambda - 3 = 0$ ).

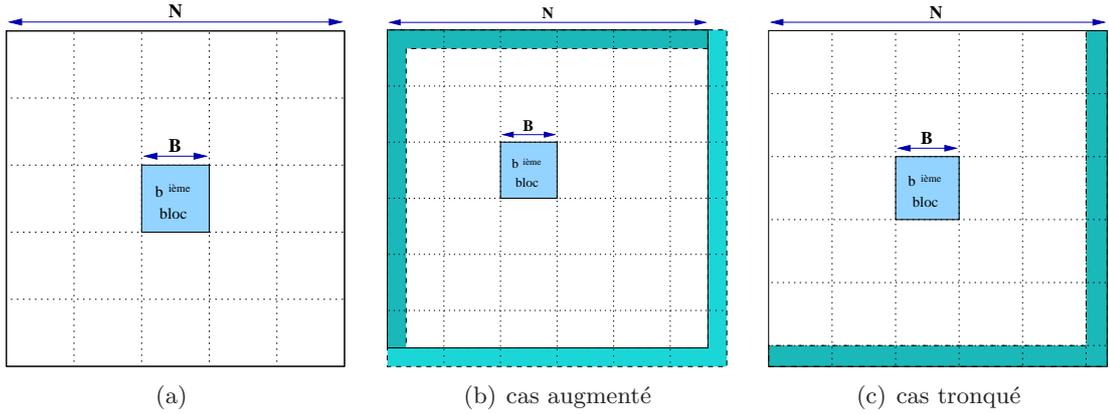


FIG. 1.4 – Exemple de représentation par blocs sans recouvrement.

**1.3.1.2.2 Seuillage par blocs avec recouvrement** Cai & Silverman ont proposé l'estimateur de seuillage par blocs avec recouvrement [Cai & Silverman, 2000], en modifiant celui sans recouvrement [Cai, 1999].

A chaque échelle de décomposition  $j = J_c, \dots, J-1$ , les coefficients de détail  $d_{mn}^j$  sont groupés dans des blocs de coefficients sans recouvrement ( $jb$ ) de taille  $B_0$ . La taille de chaque bloc est augmentée de  $B_1 = \max(1, \lfloor B_0/2 \rfloor)$  dans toutes les directions, afin de former des grands blocs avec recouvrement ( $jB$ ) de taille  $B = B_0 + 2B_1$ .

Soit  $S_{(jB)}^2$  l'énergie du signal bruité dans le grand bloc ( $jB$ ). Dans chaque bloc ( $jb$ ) les coefficients d'ondelettes sont estimés simultanément par le minimiseur du risque SURE suivant,

$$\tilde{d}_{mn}^{(jb)} = \max \left( 0, \frac{S_{(jB)}^2 - \lambda B \sigma^2}{S_{(jB)}^2} \right) d_{mn}^j \quad (1.19)$$

Cai & Silverman [Cai & Silverman, 2000] ont suggéré d'utiliser  $B_0 = \log(N)/2$  avec  $\lambda = 0.450524$  ou bien  $B_0 = B_1 = 1$  avec  $\lambda = \frac{2}{3} \log N$ .

## 1.3.2 Approches bayésiennes

Dans le cadre bayésien, plusieurs modèles statistiques *a priori* adaptés à la modélisation statistique des coefficients d'ondelettes dans le cas bruité ont été envisagés. La combinaison de l'information contenue dans les données observées (e.g  $y$ ) et celle contenue dans la loi *a priori* sur le signal à estimer  $g$  permet d'obtenir la distribution conditionnelle a posteriori. Cette dernière permet d'obtenir l'estimateur bayésien et d'estimer la fonction inconnue  $g$ . Plusieurs types de méthodes ont été proposés dans la littérature dont nous dressons ci-dessous un panorama.

### 1.3.2.1 Méthodes bayésiennes univariées

- **Méthodes basées sur l'a priori gaussien** Une approche bayésienne de type seuillage par contraction, basée sur l'*a priori* gaussien, a été proposée par Coifman [Chipman et al., 1997]. L'idée est de modéliser la distribution des coefficients d'ondelettes par un mélange de deux distributions normales de variances différentes. Les hyperparamètres de l'*a priori* sont estimés de manière adaptative à chaque niveau de résolution.

Une autre variante est celle d'un modèle Bernoulli-gaussien proposé par Clyde [Clyde et al., 1998]. Dans cette approche, les coefficients de détail  $s_{mn}^j : j = J_c, \dots, J-1; m, n = 0, \dots, 2^j - 1$

ayant une amplitude nulle sont obtenus par,

$$s_{mn}^{jo} | \gamma_{mn}^{jo} \sim \mathcal{N}(0, \gamma_{mn}^{jo} \tau_j^2) \quad (1.20)$$

et

$$\gamma_{mn}^{jo} \sim \text{Bernoulli}(\pi_j) \quad (1.21)$$

où les  $s_{mn}^{jo}$  sont des variables aléatoires (VAs) conditionnellement indépendantes sur les  $\gamma_{mn}^{jo}$ . Ces derniers sont des VAs binaires pour déterminer si la valeur du coefficient est nulle ( $\gamma_{mn}^{jo} = 0$ ) ou non ( $\gamma_{mn}^{jo} = 1$ ), et suivent une loi de Bernoulli de paramètre  $\pi_j$  (où  $\mathcal{P}(\gamma_{mn}^{jo} = 1) = \pi_j$ ). Les expressions permettant d'obtenir le seuil et l'estimateur de seuillage  $RB(s_{mn}^{oj} | d_{mn}^{oj}, \sigma_\epsilon^2)$  sont définies en détail dans [Clyde et al., 1998, Antoniadis et al., 2001].

• **Méthode avec a priori par mélange fini de gaussiennes** Deux types d'estimateurs bayésiens basés sur l'*a priori* gaussien par mélange fini ont été proposés dans [Abramovich et al., 1998, Clyde & George, 1999, Clyde & George, 2000].

Clyde & George [Clyde & George, 1999, Clyde & George, 2000] ont proposé un estimateur de seuillage par contraction en minimisant un risque bayésien  $L_2$ . L'expression de cet estimateur ECP correspond à une règle de seuillage dépendante de l'échelle où les coefficients  $d_{mn}^{oj}$  sont seuillés par un facteur non linéaire [Clyde & George, 1999, Clyde & George, 2000]. Dans [Abramovich et al., 1998], les auteurs ont proposé un estimateur de seuillage d'ondelettes en minimisant un risque bayésien  $L_1$ . Cet estimateur correspond à la médiane a posteriori.

Pour l'estimation des hyperparamètres, les auteurs ont utilisé l'algorithme EM pour l'estimation des paramètres du modèle *a priori*, et l'estimateur robuste MAD de Donoho & Johnstone pour estimer le niveau de bruit.

• **Approche bayésienne de test d'hypothèses** Vidakovic a proposé un estimateur de seuillage basé sur une approche bayésienne de test d'hypothèses [Vidakovic, 1998]. Pour chaque coefficient de détail  $d_{mn}^{oj} | s_{mn}^{oj}, \sigma_\epsilon^2 \sim \mathcal{N}(s_{mn}^{oj}, \sigma_\epsilon^2)$ , cette méthode implique le test d'hypothèse suivant :  $H_0 : s_{mn}^{oj} = 0$  contre  $H_1 : s_{mn}^{oj} \neq 0$ . Si l'hypothèse  $H_0$  est rejetée, alors  $s_{mn}^{oj}$  est estimé à partir de  $d_{mn}^{oj}$  où à chaque échelle  $j = J_c, \dots, J - 1$  la distribution *a priori* est définie par

$$s_{mn}^{oj} \sim \pi_j \zeta(s_{mn}^{oj}) + (1 - \pi_j) \delta(0) \quad (1.22)$$

où  $\delta(0)$  est une distribution de Dirac centrée en zéro.  $\zeta$  décrit le comportement de la distribution des coefficients  $s_{mn}^{oj}$  non nulle (cas de  $H_0$  est fausse). Abramovich & Sapatinas [Abramovich & Sapatinas, 1999] ont étudié le cas où  $\zeta$  est la PDF normale. En appliquant le théorème de Bayes pour le test d'hypothèses, l'estimateur de seuillage est donné par la formule suivante :

$$\hat{s}_{mn}^{oj} = d_{mn}^{oj} \mathbf{1}_{(\eta_j < 1)} \quad \text{avec} \quad \eta_j = \frac{\mathcal{P}(H_0 | d_{mn}^{oj})}{\mathcal{P}(H_1 | d_{mn}^{oj})} \quad (1.23)$$

où  $\mathbf{1}_{\eta_j}$  est l'indicatrice usuelle et  $\eta_j$  est le rapport de vraisemblance conditionnelle a posteriori. Les coefficients d'ondelettes  $d_{mn}^{oj}$  seront seuillés si  $\eta_j > 1$  ; sinon ils seront gardés. L'application de cet estimateur nécessite l'estimation des hyperparamètres  $\{\pi_j, \eta_j, \sigma_\epsilon\}$ . Cette étape est accomplie à l'aide de l'algorithme EM de façon analogue à ce qui a été proposé précédemment dans [Clyde & George, 1999, Clyde & George, 2000].

• **Méthode avec a priori gaussien sur le signal et exponentiel sur la variance du bruit** Vidakovic & Ruggeri [Vidakovic & Ruggeri, 2000] ont proposé une méthode qui impose un *a priori* gaussien sur les coefficients d'ondelettes  $s_{mn}^{oj}$  et un *a priori* exponentiel sur la variance du bruit ( $\sigma_\epsilon^2 \sim \mathcal{E}(\mu)$  où  $\mu > 0$ ).

Pour obtenir une règle de seuillage bayésien de forme analytique, les auteurs ont troqué l'*a priori* gaussien contre un *a priori* Bernoulli-Laplacien pour les coefficients de détail  $s_{mn}^{oj}$ ,

$$s_{mn}^{oj} \sim \pi_j \mathcal{L}(0, \mu) + (1 - \pi_j) \delta(0) \quad (1.24)$$

L'espérance conditionnelle a posteriori des coefficients  $s_{mn}^{oj}$  est donnée par l'expression explicite baptisée BAMS dans [Vidakovic & Ruggeri, 2000]. L'application de cette expression analytique, dans la pratique, nécessite une estimation des hyperparamètres  $\{\mu, \sigma_\epsilon\}$ . Vidakovic & Ruggeri ont estimé ces hyperparamètres en utilisant les quantiles d'ordre 1 et 3 de l'échelle la plus grossière  $J - 1$  [Vidakovic & Ruggeri, 2000].

• **Distribution gaussienne généralisée (GGD)** Mallat a été le premier à proposer la distribution gaussienne généralisée (GGD) [Mallat, 1989] à des fins de compression. Ensuite, plusieurs estimateurs ont adopté cet *a priori* à des fins d'estimation dans le domaine des ondelettes [Simoncelli & Adelson, 1996, Moulin & Liu, 1999, Chang et al., 2000a, Chang et al., 2000b]. Moulin & Liu [Moulin & Liu, 1999] ont proposé un estimateur bayésien de type MAP (cf. annexe A) basé sur cet *a priori*.

Le modèle *a priori* GGD est défini par

$$\mathcal{P}(s) = \frac{\nu \eta(\nu)}{2\Gamma(1/\nu)c} e^{-\left(\frac{\eta(\nu)|s|}{c}\right)^\nu} \quad (1.25)$$

où  $c$  est le paramètre d'échelle,  $\nu$  est le paramètre de forme et  $\eta(\nu) \triangleq \sqrt{\frac{\Gamma(3/\nu)}{\Gamma(1/\nu)}}$ ,  $\Gamma(\cdot)$  est la fonction Gamma. Deux cas spéciaux peuvent être précisés pour ce modèle : cas d'une distribution de gaussienne pour  $\nu = 2$  et cas d'une distribution Laplacienne pour  $\nu = 1$ .

La PDF de GGD a une décroissance exponentielle rapide vers l'infini pour des valeurs décroissantes de  $\nu$ . Du fait de ce comportement, l'*a priori* se révèle incapable de modéliser les queues lourdes de la distribution des coefficients d'ondelettes.

Le modèle GGD avec un paramètre de forme  $0.3 < \nu < 1$  présente un modèle raisonnable pour la distribution des coefficients d'ondelettes d'une image [Mallat, 1989]. Il peut être appliqué à chaque orientation et à chaque échelle de la décomposition sauf à l'échelle la plus grossière. Moulin & Liu [Moulin & Liu, 1999] ont montré que l'estimateur MAP par un *a priori* GGD est un seuillage par contraction avec un seuil  $\lambda$  égal à

$$\lambda = \frac{2 - \nu}{2(1 - \nu)} (2(1 - \nu))^{1/2-\nu} \eta(\nu)^{\frac{\nu}{2-\nu}} \sigma_\epsilon^{\frac{2}{2-\nu}} c^{\frac{-\nu}{2-\nu}} \quad (1.26)$$

Notons que pour le cas de la distribution Laplacienne, cet estimateur MAP est identique à un estimateur de type seuillage doux [Donoho & Johnstone, 1994], de seuil  $\lambda = \frac{\sqrt{2}\sigma_\epsilon^2}{c}$  avec  $c = \sigma_\epsilon(\log N)^{-1/2}$ . Lorsque  $\nu$  est petit ( $\nu \rightarrow 0$ ), on retrouve un estimateur de type seuillage dur [Donoho & Johnstone, 1994], de seuil  $\lambda \approx \sigma_\epsilon \sqrt{\frac{6\sqrt{3}}{e}} \nu^{-1/2}$  avec  $\nu = \frac{3\sqrt{3}/e}{\log N}$ . Cela montre que les estimateurs classiques de type seuillage d'ondelettes sont un cas particulier de cet estimateur bayésien MAP.

• **Estimateurs basés sur l'a priori  $\alpha$ -stable** Achim et al. [Achim et al., 2001] ont proposé la mise en oeuvre de l'a priori  $\alpha$ -stable capable de modéliser la distribution de coefficients de détail à queue lourde. Nous ne rentrerons pas en détail dans cet a priori puisque cela fera l'objet de la section 3.5 (chapitre 3).

Dans [Achim et al., 2001], les auteurs ont proposé un estimateur bayésien de type ECP. Toutefois, aucune forme analytique n'a été fournie pour cet estimateur bayésien sauf quelques cas particuliers, ce qui nécessite une implémentation numérique des intégrales particulièrement instable à cause des bornes d'intégration infinies et coûteuse en temps.

L'estimation des hyperparamètres du modèle a priori est une étape importante pour mettre en oeuvre le débruiteur mais elle reste une tâche très cruciale. Si cette estimation reste accessible en absence de bruit, elle devient beaucoup plus délicate en sa présence. Plusieurs méthodes ont été envisagées pour estimer les hyperparamètres en absence du bruit. Citons par exemple : la méthode de maximum de vraisemblance [Nolan, 1997], la méthode des quantiles [Fama & Roll, 1971], la méthode des moments fractionnaires [Ma & Nikias, 1995] et aussi les méthodes basées sur la fonction caractéristique [Koutrouvelis, 1980, Koutrouvelis, 1981].

Dans [Mathieu, 2002], l'auteur a proposé une implémentation très rapide et stable pour l'estimateur bayésien ECP basée sur l'implémentation numérique des intégrales de Fourier impliquées dans l'ECP à base de FFT [Press et al., 1992]. Toutefois, le problème des hyperparamètres persiste.

### 1.3.2.2 Méthodes bayésiennes multivariées

En raison de l'introduction de l'information contenue dans le voisinage des coefficients d'ondelettes, les méthodes classiques par blocs de coefficients présentent une meilleure qualité d'estimation en comparaison avec les méthodes classiques terme à terme. Il en est de même pour les estimateurs bayésiens par blocs. Cette tâche est cruciale pour préserver la géométrie locale des images.

• **Estimateur bayésien de type seuillage par blocs** Les coefficients d'ondelettes peuvent être seuillés par blocs de coefficients plutôt que terme à terme. Cette idée a été présentée par Abramovich [Abramovich et al., 2000] pour obtenir un estimateur bayésien par blocs sans recouvrement.

En considérant le modèle d'observation donné par l'équation 1.2, à chaque échelle de décomposition  $j$  ( $j = J_c, \dots, J-1$ ) les coefficients  $d_{mn}^{oj}$  sont groupés dans des blocs sans recouvrement  $b_{jK}$  de taille  $B_j = j$ , où  $B_j$  n'est pas nécessairement divisible par  $2^j$ .

Les auteurs [Abramovich et al., 2000] ont considéré le modèle a priori suivant pour chaque bloc  $b_{jK}$  et à chaque échelle  $j$  :

$$b_{jK} \sim \pi_j \mathcal{N}(0, V_j) + (1 - \pi_j) \delta(0) \quad (1.27)$$

où  $\delta(0)$  et  $V_j$  sont respectivement un vecteur de  $B_j$  de Dirac centrées en zéro et la matrice de covariance de taille  $B_j \times B_j$ .

Deux type d'estimateurs bayésiens correspondant aux risques  $L_1$  et  $L_2$  ont été proposés dans [Abramovich et al., 2000]. L'estimation des hyperparamètres  $\{\pi_j, V_j\}$  est obtenue par une méthode de Bayes empirique basée sur l'estimateur de maximum de vraisemblance marginale.

• **Seuillage basé sur une décomposition déterministe/stochastique** Les approches bayésiennes citées précédemment pour l'obtention d'un estimateur bayésien de type seuillage d'ondelettes, imposent un a priori de moyenne nulle pour les coefficients d'ondelettes  $d_{mn}^{oj}$ . Huang

& Cressie [Huang & Cressie, 2000] ont proposé une approche bayésienne avec un *a priori* gaussien multivarié (de moyenne non nulle à estimer et à introduire dans l'estimateur de seuillage). L'idée est de décomposer l'image à estimer  $g$  en deux parties : une partie dite "déterministe" qui regroupe les coefficients significatifs et une partie dite "stochastique" qui regroupe les coefficients de moyenne nulle.

Soient  $\mathbf{c}_{J_c}$  et  $\mathbf{d}_j$  (resp.  $\mathbf{a}_{J_c}$  et  $\mathbf{s}_j$ ), avec  $j = J_c, \dots, J-1$ , les vecteurs des coefficients d'approximation et de détail observés et inconnus à toutes les échelles  $j$ . Huang & Cressie ont proposé le modèle bayésien multivarié suivant,

$$\omega | (\beta, \sigma_\epsilon^2) \sim \mathcal{N}(\beta, \sigma_\epsilon^2 I) \quad \text{et} \quad \beta | (\mu, V) \sim \mathcal{N}(\mu, \Sigma(V)) \quad (1.28)$$

où  $\omega = \{\mathbf{c}_{J_c}, \mathbf{d}_{J_c}, \dots, \mathbf{d}_{J-1}\}$  et  $\beta = \{\mathbf{a}_{J_c}, \mathbf{s}_{J_c}, \dots, \mathbf{s}_{J-1}\}$ .  $\beta$ , qui correspond à la partie de composantes déterministes, peut s'écrire sous la forme suivante :

$$\beta = \mu + \eta \quad (1.29)$$

où  $\mu$  et  $\eta$  correspondent respectivement aux parties déterministe et stochastique avec  $\eta \sim \mathcal{N}(0, \Sigma(V))$ .  $\Sigma(V)$  permet de décrire la variabilité et la corrélation dans l'image.

L'estimateur ECP de  $\beta$  sachant  $\omega$  et  $\sigma_\epsilon^2$  s'écrit comme suit,

$$(\beta | \omega, \sigma_\epsilon^2) = \mu + \Sigma(V) (\Sigma(V) + \sigma_\epsilon^2 \mathbb{I})^{-1} (\omega - \mu) \quad (1.30)$$

Pour mettre en application cet estimateur, les auteurs [Huang & Cressie, 2000] ont suggéré d'estimer les hyperparamètres  $\{\sigma_\epsilon, \mu, \Sigma(V)\}$  ainsi : le niveau de bruit  $\sigma_\epsilon$  par l'estimateur MAD [Donoho & Johnstone, 1994] ou bien par une méthode basée sur le variogramme des observations [Huang & Cressie, 2000],  $\mu = \mathbf{c}_{J_c}$  et le vecteur  $V$  par la méthode de maximum de vraisemblance sur les données  $\omega$ .

• **Estimateur basé sur l'a priori  $\alpha$ -stable multivarié** Dans [Sendur & Selesnick, 2002, Achim & Kuruoglu, 2004], les auteurs ont utilisé la loi  $\alpha$ -stable multivariée pour modéliser les dépendances inter-échelle entre coefficients dans le domaine des ondelettes basée sur la distribution  $\alpha$ -stable isotrope. Cette représentation statistique permet de mieux modéliser les distributions des coefficients de détail à queue relevée aussi bien que les dépendances inter-échelle entre coefficients.

Sendur & Selesnick ont proposé un estimateur bayésien bivarié de type MAP seulement pour le cas des distributions gaussienne et Laplacienne [Sendur & Selesnick, 2002]. Dans [Achim & Kuruoglu, 2004], Achim & Kuruoglu ont proposé l'estimateur bayésien bivarié MAP correspondant au cas de la distribution de Cauchy.

Les hyperparamètres du modèle de la distribution  $\alpha$ -stable bivariée pour les observations bruitées sont estimés par la méthode de maximum de vraisemblance en utilisant les techniques d'intégration de Monte-Carlo et par l'estimateur MAD pour estimer le niveau de bruit [Donoho & Johnstone, 1994].

## 1.4 Débruitage dans le domaine des transformées multi-échelles orientées

### 1.4.1 Approches classiques sans a priori

- **Estimateurs par seuillage de X-lets géométriques**

Plusieurs extensions pour les estimateurs par seuillage d'ondelettes ont été récemment proposées dans le domaine des transformées multi-échelles orientées, e.g. les X-lets géométriques : curvelets [Candès & Donoho, 2002, Starck et al., 2002, Candès & Donoho, 2004], contourlets [Do & Vitterli, 2003a, Do & Vitterli, 2003b], bandelets [Mallat & LePenec, 2005, LePenec & Mallat, 2005], ondelettes complexes [Kingsbury, 1999] et ondelettes orientées "steerable pyramid" [Freeman & Adelson, 1991, Simoncelli et al., 1992]. Ces transformées sont des transformées redondantes qui permettent de décomposer l'image en sous-bandes orientées et aussi de bien capturer et préserver la géométrie exprimée par les agglomérats de coefficients. De plus, certaines de ces transformées montrent des propriétés de presque-invariance par translation et permettent aussi d'obtenir une meilleure sélectivité directionnelle.

- **Estimateurs par seuillage** Une extension de l'estimateur par seuillage d'ondelettes basée sur la transformée de curvelet a été proposée par Starck et al. [Starck et al., 2002]. Les curvelets 1<sup>ère</sup> génération ont été proposées par Candès & Donoho [Candès & Donoho, 1999] et constituent une nouvelle famille de frames d'ondelettes géométriques plus efficaces que les transformées multi-échelles non-orientées pour la représentation de certaines géométries des images ( $C^2$  par morceaux au delà de la discontinuité  $C^2$ ). Les auteurs dans [Starck et al., 2002] suggèrent l'extraction des coefficients de curvelet de détail significatifs par comparaison de ces derniers avec un paramètre de seuillage  $\lambda = k\tilde{\sigma}_\lambda$  où  $\sigma_\lambda$  est l'approximation de la variance calculée par des simulations de Monte-Carlo. Pour la première échelle de décomposition ( $j = 1$ )  $k = 4$ , et  $k = 3$  pour les autres échelles ( $j > 1$ ). Les fonctions de seuillage résultantes se déclinent sous deux formes, seuillages dur et doux, définis de la même façon que celles données par l'Eq.1.4 et l'Eq.1.5.

D'autres extensions pour les estimateurs par seuillage d'ondelettes sont apparues récemment. Citons par exemple, les estimateurs de seuillage basés sur la transformée de contourlets qui ont été proposés dans les travaux de Minh Do & Vitterli [Do & Vitterli, 2003a, Do & Vitterli, 2003b] et d'Eslami & Radha [Eslami & Radha, 2006]. Dans [Mallat & LePenec, 2005, LePenec & Mallat, 2005], Mallat & LePenec ont utilisé ces estimateurs de seuillage avec la transformée de bandelets. Aussi dans [Claypoole & Baraniuk, 2000], les auteurs ont proposé des estimateurs de seuillage basés sur les wedgelets.

Les résultats des simulations de débruitage ont montré que les estimateurs classiques avec des transformées multi-échelles orientées (e.g. X-lets géométriques) sont souvent meilleurs comparés aux estimateurs classiques avec les transformées multi-échelles non-orientées, en terme de qualité visuelle et en terme de PSNR.

### 1.4.2 Approches bayésiennes

- **Estimateur basé sur le mélange d'échelles de gaussiennes** Portilla et al. [Portilla et al., 2003] ont proposé une approche bayésienne basée sur un modèle statistique de voisinage de coefficients d'ondelettes, basée sur l'*a priori* de mélange d'échelles de gaussiennes [Andrews & Mallows, 1974]. Cette méthode exploite le fait que les coefficients d'ondelettes s'agglomèrent<sup>2</sup> de

<sup>2</sup>Ces agglomérations se concentrent au voisinage de la géométrie de l'image

manière cohérente à travers les échelles et à des positions proches les unes des autres.

Dans [Portilla et al., 2003], le voisinage de  $\mathbf{d}_{mn}^{oj}$  est modélisé par un mélange infini de gaussiennes dépendant uniquement de l'échelle  $j$  et une VA de mélange :

$$\forall m, n : \mathbf{d}_{mn}^{oj} \stackrel{d}{=} \sqrt{z_j} \mathbf{u} \quad (1.31)$$

où  $z_j$  est une variable aléatoire scalaire appelé le multiplicateur, indépendante de  $\mathbf{u}$ , de moyenne 1, et  $\mathbf{u}$  est un vecteur de VAs gaussiennes centrées de covariance  $\Sigma$ . Dans [Portilla et al., 2003],  $z_j$  suit un *a priori* non-informatif de Jeffrey.

Le modèle statistique proposé dans [Portilla et al., 2003], permet de modéliser les distributions marginales et jointes des coefficients d'ondelettes, et s'avère efficace pour capturer le comportement leptokurtique des coefficients d'ondelettes et de prendre en compte la corrélation entre les coefficients dans le voisinage. Cette méthode a été implémentée avec la transformée en ondelettes orientée "steerable pyramid". Les résultats expérimentaux ont montré la performance de cette approche bayésienne de débruitage en terme de qualité visuelle nettement supérieure comparée aux différentes approches classiques et bayésiennes et en terme de PSNR. Elle représente actuellement l'une des méthodes à la pointe de l'état de l'art.

• **Estimateur basé sur les modèles Markoviens** La combinaison des dépendances inter- et intra-échelle entre les coefficients d'ondelettes dans une base orthonormale décimée est connue seulement pour fournir des performances (des améliorations) mineures à l'estimateur (au débruiteur) [Liu & Moulin, 2000]. Cependant, dans une base d'ondelettes non-décimée, une telle combinaison a des avantages clairs en termes de mesures quantitatives de qualité de l'image et de la qualité visuelle des résultats. Ceci a été démontré dans [Malfait & Roose, 1997]. L'approche de [Malfait & Roose, 1997] combine les dépendances inter et intra-échelle d'une manière puissante et élégante : un modèle aléatoire de champs de Markov (MRF) à deux niveaux est employé pour coder les connaissances *a priori* de la répartition des coefficients d'ondelettes, c-à-d, coder les propriétés géométriques des images de détail. Les dépendances inter-échelle entre les coefficients d'ondelettes sont exprimées par un modèle de probabilité conditionnelle permettant de calculer les mesures significatives (rapports inter-échelle). Ces mesures significatives sont combinées avec le modèle *a priori* afin d'estimer le signal d'intérêt. Le modèle conditionnel de [Malfait & Roose, 1997] est un modèle paramétrique mais de type heuristique, ce qui complique son implémentation en pratique. Dans [Jansen & Bultheel, 1999, Jansen & Bultheel, 2001], les auteurs ont développé théoriquement l'approche de [Malfait & Roose, 1997] et ont proposé des algorithmes pratiques pour l'implémentation.

Dans [Pizurica et al., 2002], les auteurs ont proposé une approche bayésienne basée sur les MRF anisotropes. Leur méthode développe trois aspects : (1) une caractérisation statistique de différentes mesures significatives de coefficients d'ondelettes, (2) un modèle conditionnel pour combiner les propriétés statistiques inter et intra-échelle, (3) un modèle de champ de Markov anisotrope comme *a priori*. Les différentes formulations des rapports inter-échelle, citées dans [Malfait & Roose, 1997, Hsung et al., 1999], ont inspiré les auteurs [Pizurica et al., 2002]. Ils ont déterminé une densité de probabilité conditionnelle pour les rapports inter-échelle au lieu d'employer des modèles heuristiques. Cette approche s'avère plus efficace en termes de précision et de temps de calcul que celles citées précédemment.

D'autres extensions des MRF ont été proposées. A titre d'exemple, le modèle de Markov caché contextuel proposé dans [Crouse & Baraniuk, 1997] et le modèle de Markov caché contextuel local proposé dans [Fan & Xia, 2001]. Ce dernier permet d'exploiter les statistiques locales ainsi

que les dépendances inter- et/ou intra-échelle entre les coefficients d'ondelettes et de les modéliser par un mélange fini de gaussiennes.

Il est à noter que l'une des difficultés principales de ce type d'estimateurs bayésiens réside dans l'estimation des hyperparamètres du modèle markovien. Pour cela, les auteurs ont envisagé l'utilisation de l'algorithme EM pour le calcul des probabilités conditionnelles des états sachant les observations. Se rajoute à cela, un temps d'exécution important puisque des échantillonneurs stochastiques sont mis en jeu lors de l'optimisation des configurations.

- **Estimateur basé sur les arbres de Markov cachés** Crouse, Nowak & Baraniuk ont proposé le modèle d'arbre de Markov caché (HMT) pour la modélisation statistique d'images dans le domaine des ondelettes [Crouse et al., 1998]. Ce modèle de HMT permet de capturer les dépendances inter-échelle par une structure d'arbre sur les coefficients d'ondelettes (processus arborescent plutôt que séquentiel) et d'introduire les dépendances markoviennes le long de la structure d'arbre. Ce modèle a été appliqué pour le débruitage et la compression d'images dans [Romberg et al., 1999]. Il a été aussi proposé pour la déconvolution des images satellitaires et aériennes [Jalobeanu, 2001].

- **Estimateur de type filtrage combiné** Un nouveau type d'approche pour la restauration d'images a été récemment proposé dans [Starck et al., 2003], basé sur la combinaison de plusieurs transformées multi-échelles orientées et/ou non orientées afin de bénéficier des avantages de chacune d'elles (notion de diversité morphologique).

L'idée est que les ondelettes ne sont pas optimales pour l'analyse d'objets anisotropes dans l'image (e.g., les lignes, les contours, etc.) comparées aux transformées multi-échelles orientées qui intègrent la notion de directionnalité et qui permettent de représenter des objets anisotropes de manière parcimonieuse, mais restent efficaces pour la détection de structures isotropes à différentes échelles.

Il a été montré dans [Starck et al., 2002], pour le problème de débruitage, que les approches de curvelet présentent une meilleure performance en terme de PSNR comparées avec celles de la transformée d'ondelettes non-décimée. Toutefois, en terme de qualité visuelle certains d'artefacts le long des contours peuvent être observés sur l'image restaurée. Afin d'améliorer la qualité de l'image restaurée, les auteurs dans [Starck et al., 2003] ont proposé une approche de filtrage combinée, des dictionnaires de transformées intègrent les ondelettes et les curvelets.

## 1.5 Conclusion

Après cet état de l'art sur les différentes approches pour le problème de la régression non-paramétrique dans le domaine des transformées multi-échelles (débruitage multi-échelle), des remarques se dégagent sur le fait que les approches bayésiennes s'avèrent plus efficaces que celles présentées dans le cadre classique (sans *a priori*) que ce soit terme à terme (coefficient par coefficient) ou bien par bloc de coefficients. De plus, les méthodes bayésiennes multivariées présentent une meilleure qualité d'estimation pour le signal inconnu  $g$  comparées aux méthodes univariées. Cela est dû à l'introduction de l'information contenue dans les dépendances inter- et intra-échelle entre les coefficients. Cependant, les approches bayésiennes souffrent de quelques problèmes. Pour un bon nombre d'*a priori*, aucune forme analytique n'a été proposée pour l'estimateur bayésien ce qui nécessite une implémentation numérique soit des intégrales impliquées dans les estimateurs ECP, soit l'optimisation itérative pour le MAP. Vient s'ajouter à ceci le

---

problème d'estimation des hyperparamètres qui reste une étape cruciale et extrêmement critique particulièrement en présence de bruit.

Dans le cadre de cette thèse, nous proposons de nouveaux estimateurs bayésiens dans le domaine des transformées multi-échelles orientées et non-orientées comme solution au problème de la régression non-paramétrique, basés sur les lois statistiques :  $\alpha$ -stable et les Formes K de Bessel.



## Chapitre 2

# Modélisation statistique des images

---

### Sommaire

---

<b>2.1</b>	<b>Pourquoi une modélisation statistique? . . . . .</b>	<b>31</b>
<b>2.2</b>	<b>Axiomes et propriétés relatives aux images naturelles . . . . .</b>	<b>32</b>
2.2.1	Invariance par changement d'échelle . . . . .	32
2.2.2	Persistance à travers les échelles . . . . .	32
2.2.3	Dépendance intra-échelle . . . . .	32
<b>2.3</b>	<b>Modèles statistiques des images . . . . .</b>	<b>34</b>
2.3.1	Modèles génératifs parcimonieux . . . . .	34
2.3.1.1	Représentation parcimonieuse . . . . .	34
2.3.1.2	Images compressibles . . . . .	35
2.3.2	Autres modèles . . . . .	35
2.3.2.1	Modèles de superposition . . . . .	35
2.3.2.2	Les champs aléatoires de Markov . . . . .	36
<b>2.4</b>	<b>Lois dans les représentations parcimonieuses . . . . .</b>	<b>36</b>
2.4.1	Loi marginale . . . . .	36
2.4.2	Loi conditionnelle/jointe . . . . .	38
<b>2.5</b>	<b>Conclusion . . . . .</b>	<b>39</b>

---

Après avoir posé le problème de débruitage, il est nécessaire de comprendre quelle est la structure de l'image réelle que l'on observe à la sortie du capteur. La modélisation de l'image observée est une étape indispensable pour reconstruire l'image qu'on cherche à obtenir dans de bonnes conditions.

Ce chapitre a pour objectif de présenter quelques propriétés statistiques permettant de caractériser les images observées dans certains domaines transformés. Ainsi, nous dressons un état de l'art concis des principaux modèles pour la modélisation statistique des images qui ont été proposés dans la littérature.

### 2.1 Pourquoi une modélisation statistique ?

L'approche probabiliste pour la modélisation et le traitement des images a été initiée par U. Grenander (voir la revue faites dans [Grenander, 1993]), puis développée et popularisée plus amplement par D. Geman & S. Geman [Geman & Geman, 1984] dans leurs travaux primordiaux sur les champs de Markov en traitement d'images.

L'hypothèse de base pour la modélisation statistique est l'existence d'une probabilité sur l'ensemble des images, qui serait la loi des images appartenant à un corpus (e.g. images naturelles). Cette loi de probabilité illustre le comportement statistique d'un ensemble d'images particulier. Pour les textures par exemple, les champs aléatoires de Markov ont été mis à profit pour les modéliser et les synthétiser [Winkler, 1995]. Des alternatives basées sur des transformées appropriées ont aussi été avancées dans [Simoncelli & Portilla, 1998]. L'ambition est de définir un modèle permettant de générer des images de scènes appartenant à un corpus, des images naturelles par exemple.

## 2.2 Axiomes et propriétés relatives aux images naturelles

### 2.2.1 Invariance par changement d'échelle

Une des propriétés les plus marquantes, relative aux images naturelles, est l'invariance par changement d'échelle (et aussi par translation). En termes statistiques, la distribution marginale des images naturelles est inchangée si les images sont grossies ou réduites. Ceci peut être exprimé par une loi de puissance (une densité spectrale de puissance) qui évolue en  $\frac{A}{\omega^{2-\eta}}$ , où  $\omega$  est la fréquence spatiale et  $\eta$  est l'exposant caractéristique. Cette loi de puissance n'est autre que la manifestation de la nature fractale ou invariante par changement d'échelle des images naturelles [Mandelbrot & van Ness, 1968]. D'un point de vue statistique, l'invariance par changement d'échelle peut se formaliser par :

$$\mathcal{P}(X(\mathcal{F}^{-1}(n, m))) = \mathcal{P}(X(n, m)) \quad (2.1)$$

pour tout sous-ensemble mesurable  $X$  (image) de l'espace muni de la mesure de probabilité  $\mathcal{P}$  et  $\mathcal{F}(n, m) = (sn + t_n, sm + t_m)$  est une fonction de  $\mathbb{R}^2$  vers  $\mathbb{R}^2$ ,  $s$  traduit le zoom (échelle),  $(n, m)$  est la location spatiale et  $(t_n, t_m)$  la translation.

Ainsi, il est naturel de faire appel à une décomposition multi-échelle pour analyser les images lorsqu'on évoque la notion d'invariance d'échelle, pour observer éventuellement de nouvelles propriétés, se manifestant au niveau de cette décomposition.

### 2.2.2 Persistance à travers les échelles

Lorsqu'on observe les modules des coefficients de détail d'une décomposition à différentes échelles, on note immédiatement une ressemblance. Sur la Fig.2.1, on a décomposé l'image de Lena (comportant des contours, textures et zones régulières); on constate une persistance évidente de certains détails, notamment des contours et des régions homogènes.

### 2.2.3 Dépendance intra-échelle

Dans une décomposition multi-échelle, on observe une ressemblance entre les sous-bandes de détails (correspondant à différentes orientations) à une même échelle, bien que si on regarde les coefficients un par un, ils ne sont ni égaux ni nécessairement de même signe.

Sur la Fig.2.2, on constate que les coefficients de détail de forte valeur tendent à se regrouper autour des bords des objets (e.g. les contours, les lignes, etc.) quelle que soit l'orientation. Ces agglomérats de coefficients tendent ainsi à décrire un chaînage géométrique local.

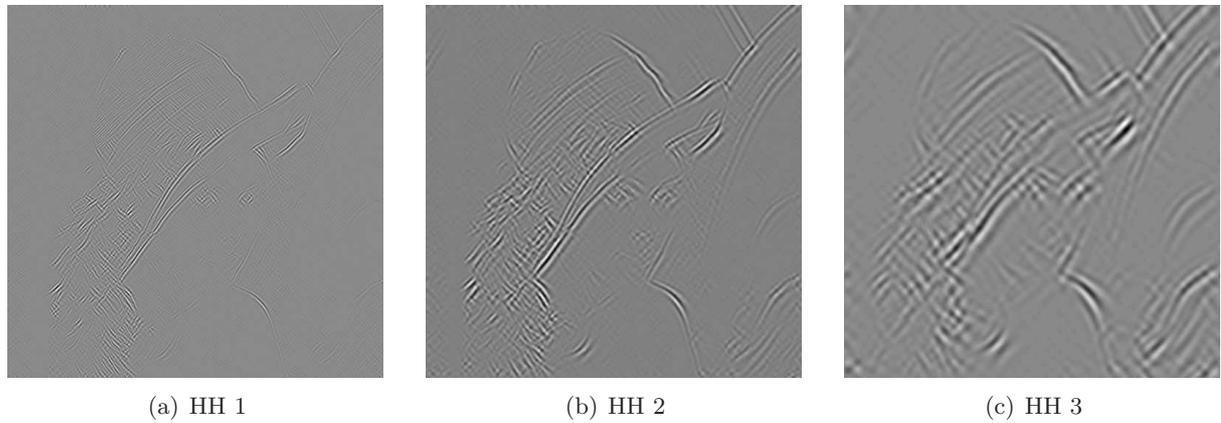


FIG. 2.1 – Sous-bandes de détails diagonaux de la DWT (ondelette Daubechies-8) appliquée sur l'image de Lena (de taille [512x512]) aux échelles : 1, 2 et 3, montrant la persistance des détails à travers les échelles.

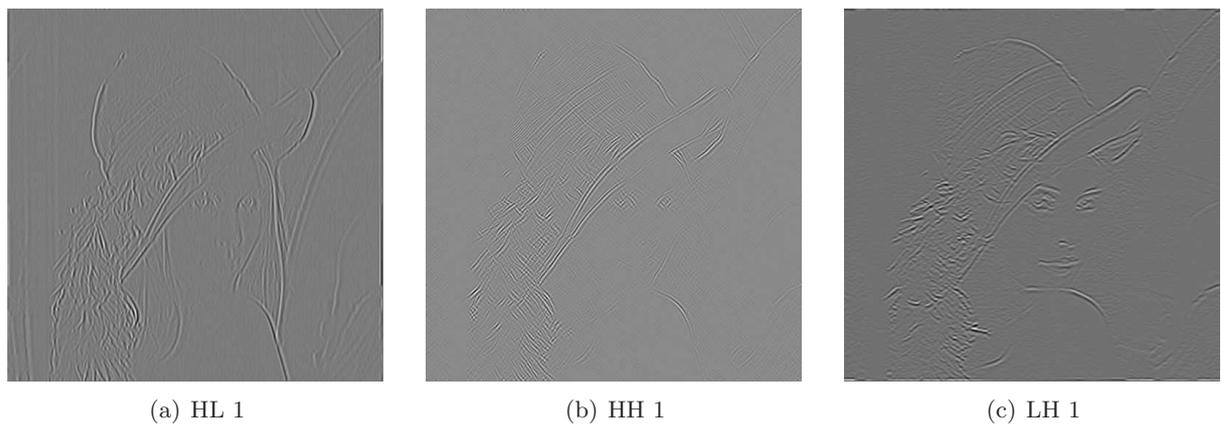


FIG. 2.2 – Sous-bandes de la DWT (ondelette Daubechies-8) de l'image de Lena à l'échelle 1 : détails verticaux ( $HL$ ), diagonaux ( $HH$ ), horizontaux ( $LH$ ), illustrant la dépendance entre les sous-bandes d'une même échelle.

## 2.3 Modèles statistiques des images

Dans la littérature, plusieurs travaux se sont penchés sur la modélisation statistique d'images d'un point de vue mathématique, essayant ainsi de caractériser les phénomènes à l'origine de la génération des images. Une approche fréquente à tous ces modèles est le placement aléatoire de formes planaires (lignes, objets, etc.) dans une image selon des lois particulières. On retrouve les modèles génératifs parcimonieux, les modèles de superposition [Mumford & Gidas, 2001, Grenander & Srivastava, 2001] et les champs aléatoires de Markov [Li, 1995, Winkler, 1995, Zhu et al., 2000].

### 2.3.1 Modèles génératifs parcimonieux

Les modèles géométriques ont été proposés pour représenter des images géométriques. L'idée est de trouver une représentation parcimonieuse (dite aussi creuse ou compacte) du contenu de ces images. Des outils classiques ont été utilisés pour construire une représentation creuse, particulièrement ces dernières années grâce au développement en analyse harmonique appliquée computationnelle (ondelettes et au delà, curvelets [Candès & Donoho, 2002, Starck et al., 2002, Candès & Donoho, 2004, Candès et al., 2006], bandelets [Mallat & LePennec, 2005, LePennec & Mallat, 2005], contourlets [Do & Vitterli, 2003a, Do & Vitterli, 2003b]). Récemment, des transformées adaptatives ont été envisagées pour se rapprocher des modèles physiologiques (voir par exemple les travaux de [Olshausen & Field, 1996, Mallat, 2006]).

#### 2.3.1.1 Représentation parcimonieuse

Soit  $X$  une image,  $X \in \mathbb{R}^N$  et  $N$  représente le nombre de pixels.  $X$  peut être représentée suivant le modèle génératif linéaire comme suit :

$$X = \sum_{k=1}^m \alpha_k \varphi_k = \Phi \alpha \quad (2.2)$$

où  $\Phi$  est une matrice (dictionnaire d'atomes) de taille  $N \times m$  représentant l'ensemble des atomes  $\{\varphi_k\}_k$ .

Un modèle parcimonieux suppose que quelques coefficients de  $\alpha$  seront retenus, ce qui est formalisé par la pseudo-norme  $\ell_0$ ,

$$\|\alpha\|_{\ell_0} \stackrel{d}{=} \text{Card} \{k \mid \alpha_k \neq 0\} \leq M \quad (2.3)$$

où  $M$  représente le nombre de coefficients retenus dans le modèle. Ce modèle suppose que l'image  $X$  appartient à l'union des espaces de  $\binom{m}{M}$  vecteurs de dimension  $M$ . Une autre représentation exacte de la parcimonie suppose que les coefficients  $\alpha$  de  $X$  dans le dictionnaire  $\Phi$  aient une norme  $\ell_p$  petite ( $0 < p \leq 1$ ),

$$\|\alpha\|_{\ell_p}^p \stackrel{d}{=} \sum_k |\alpha_k|^p \leq \zeta \quad (2.4)$$

Dans le cas où  $\Phi$  est une base orthonormée de  $\ell_2$ , l'approximation de  $X$  dans Eq.2.2 se calcule de façon simple. Il suffit d'imposer un seuil  $\lambda > 0$  et de rejeter les coefficients d'amplitude inférieure à ce seuil

$$X_M \stackrel{d}{=} \sum_{|\langle f, \varphi_k \rangle| > \lambda} \langle f, \varphi_k \rangle \varphi_k \quad \text{avec } M = \text{Card} \{k \mid |\langle f, \varphi_k \rangle| > \lambda\} \quad (2.5)$$

Le seuil  $\lambda$  contrôle la qualité de l'approximation et est lié aussi au nombre de coefficients non nuls de  $M$ . La fonction  $X_M$  obtenue est la meilleure approximation de  $X$  avec  $M$  coefficients dans la base orthonormée  $\Phi$  satisfaisant

$$\|X - X_M\|_{\ell_2}^2 \leq CM^{-\beta} \quad (2.6)$$

où  $C$  est une constante qui ne dépend que de  $X$ , et  $\beta > 0$  est un exposant dépendant de la régularité de l'image  $X$  traduisant la décroissance de l'erreur d'approximation non-linéaire (non-linéaire puisque les coefficients pris en compte pour approcher  $X$  sont choisis en fonction de  $X$ ).

### 2.3.1.2 Images compressibles

Les images réelles ne sont pas parcimonieuses au sens strict de l'Eq.2.3 mais sont plutôt compressibles. En effet, l'erreur d'approximation non-linéaire n'est pas toujours nulle mais tend vers 0 plus ou moins rapidement lorsque  $M$  est grand. Par exemple, il a été montré pour les images  $C^\alpha$  l'erreur d'approximation non-linéaire décroît en  $O(M^{-\alpha})$  dans le domaine des ondelettes pour une ondelette avec un nombre de moments nuls  $> \alpha$ . Cette propriété d'approximation optimale est perdue lorsque l'image devient  $C^\alpha$  par morceaux avec des courbes de discontinuités de longueur finie.

Plus récemment, des représentations parcimonieuses géométriques des images ont vu le jour. Ainsi, Candès & Donoho [Candès & Donoho, 1999] ont montré que l'on peut construire des représentations non-adaptatives géométriques des images en les décomposant sur des frames de curvelets. Ils ont montré que pour des images  $C^2$  en dehors de contours de courbure finie, l'erreur d'approximation non-linéaire par les  $M$  coefficients de curvelets les plus importants est en  $O((\log_2 M)^3 M^{-2})$ . LePennec & Mallat [Mallat & LePennec, 2005, LePennec & Mallat, 2005] ont aussi montré qu'une construction adaptative sur une base de bandelets est possible atteignant une vitesse de décroissance de l'erreur en  $O(M^{-2})$ . La construction des bandelets nécessite toutefois le calcul d'un flot géométrique peu aisé en présence du bruit.

Quoi qu'il en soit, en argumentant par les inégalités classiques de la théorie de l'approximation (inégalité de Bernstein), si l'erreur d'approximation non-linéaire décroît en  $O(M^{-\beta})$  pour  $\beta$  assez grand (e.g. 2), alors on sait que les coefficients ordonnés  $|\alpha_k| = O\left(k^{-\frac{1+\beta}{2}}\right)$  [Mallat, 1999, Chapitre IX]. De ce fait, on voit que les coefficients les plus forts sont ceux les moins fréquents (i.e. peu probables) et inversement pour les coefficients les plus faibles. Cette simple observation appuyée par les arguments de la théorie de l'approximation nous indique clairement que si une image est compressible dans une transformée donnée, alors ses coefficients peuvent être vus comme les réalisations d'une VA piqué en 0 avec des queues relevées.

## 2.3.2 Autres modèles

### 2.3.2.1 Modèles de superposition

**2.3.2.1.1 Approche générique** Afin de capturer l'invariance d'échelle et la non-gaussianité des images, Mumford & Gidas [Mumford & Gidas, 2001] ont mis en évidence une famille de distributions infiniment divisibles (en invoquant le théorème de Lévy-Khintchine). Ces distributions apparaissent dès lors que ces images sont modélisées par la superposition d'objets aléatoirement placés dans l'espace. En effet, en utilisant les axiomes d'invariance (e.g. échelle, translation), on peut écrire l'image comme la décomposition en série stochastique suivante :

$$X(n, m) = \sum_i \mathbf{g}_i(\mathbf{s}_i n + t_{i_n}, \mathbf{s}_i m + t_{i_m}) \quad (2.7)$$

du moment que cette somme converge (au moins au sens des distributions). Mumford & Gidas [Mumford & Gidas, 2001] ont montré que  $(\mathbf{s}_i, t_{i_n}, t_{i_m})$  suivent une loi de Poisson dans le groupe de transformations affines de densité  $c ds dt_n dt_m/s$ .  $\mathbf{g}_i$  sont les objets (patrons) aléatoires obéissant à une loi de probabilité avec la mesure de Lévy réduite. Clairement, cette équation suppose que les objets sont aléatoires (avec une certaine loi) répartis aléatoirement selon une loi de Poisson. Afin de respecter l'axiome d'invariance énoncé précédemment, la taille des objets (échelle  $s$ ) doit avoir une densité qui évolue en  $1/s^3$  [Mumford & Gidas, 2001].

**2.3.2.1.2 Le modèle du générateur déplacé** Une autre approche a été proposée par Grenander & Srivastava [Grenander & Srivastava, 2001]. Son point de départ est le *modèle du générateur déplacé* (en anglais Transported Generator Model). L'une des différences avec l'approche de Mumford est que les patrons sont remplacés par des profils 2D appelés générateurs. Ainsi, l'intensité de l'image au pixel  $(n, m)$  peut s'écrire comme étant la somme des contributions provenant des  $n$  objets (de générateurs  $\mathbf{g}_i \in \mathcal{G}$ ) générés aléatoirement selon une certaine probabilité dont est munie  $\mathcal{G}$  :

$$X(n, m) = \sum_{i=1}^n a_i \mathbf{g}_i \left( \frac{n - t_{i_n}}{s_i}, \frac{m - t_{i_m}}{s_i} \right) \quad (2.8)$$

où les  $a_i \in \mathbb{R}$  sont des VAs indépendantes gaussiennes de moyenne nulle et de variance unitaire, les  $s_i$  sont uniformes dans  $[0, N - 1]$  (grille des pixels), les  $t_i$  sont des processus de Poisson 2D homogènes d'intensité  $\rho$  sur un compact de  $\mathbb{R}^2$ . Toutes ces variables sont supposées mutuellement indépendantes. Les générateurs étant inconnus, les variables sous-jacentes sont indéterminées. Le but est donc de rechercher une loi de probabilité sur  $X$  où ses versions filtrées passe-bandes  $X_F = F * X$  en incorporant implicitement les incertitudes sur ces variables. Notons que caractériser statistiquement  $X$  ou ses combinaisons linéaires (i.e. version filtrée linéairement) revient au même par le théorème de Cramér-Wold.

### 2.3.2.2 Les champs aléatoires de Markov

Les modèles de champs de Markov ont été proposés pour pouvoir modéliser les interactions entre les coefficients de façon simple, à travers un modèle probabiliste. Ces modèles ont montré leur efficacité pour la synthèse de texture et la modélisation des images, pour plus de détail voir [Li, 1995, Winkler, 1995, Zhu et al., 2000, Grenander & Srivastava, 2001, Mumford & Gidas, 2001].

## 2.4 Lois dans les représentations parcimonieuses

### 2.4.1 Loi marginale

Comme nous l'avons vu en 2.3.1, le point de départ commun est de choisir un espace muni d'une base et de projeter l'espace des images dans un sous-espace de dimension inférieure dans lequel seules quelques composantes sont significatives. Plusieurs études ont ainsi mis en évidence que les propriétés statistiques, dans de telles représentations, sont loin d'avoir un comportement gaussien.

En effet, les PDFs dans les espaces transformés sont leptokurtiques (un pic prononcé) avec des queues lourdes (voir exemple de la Fig.2.3). Ce comportement a été observé du moment que l'image est compressible dans la représentation choisie.

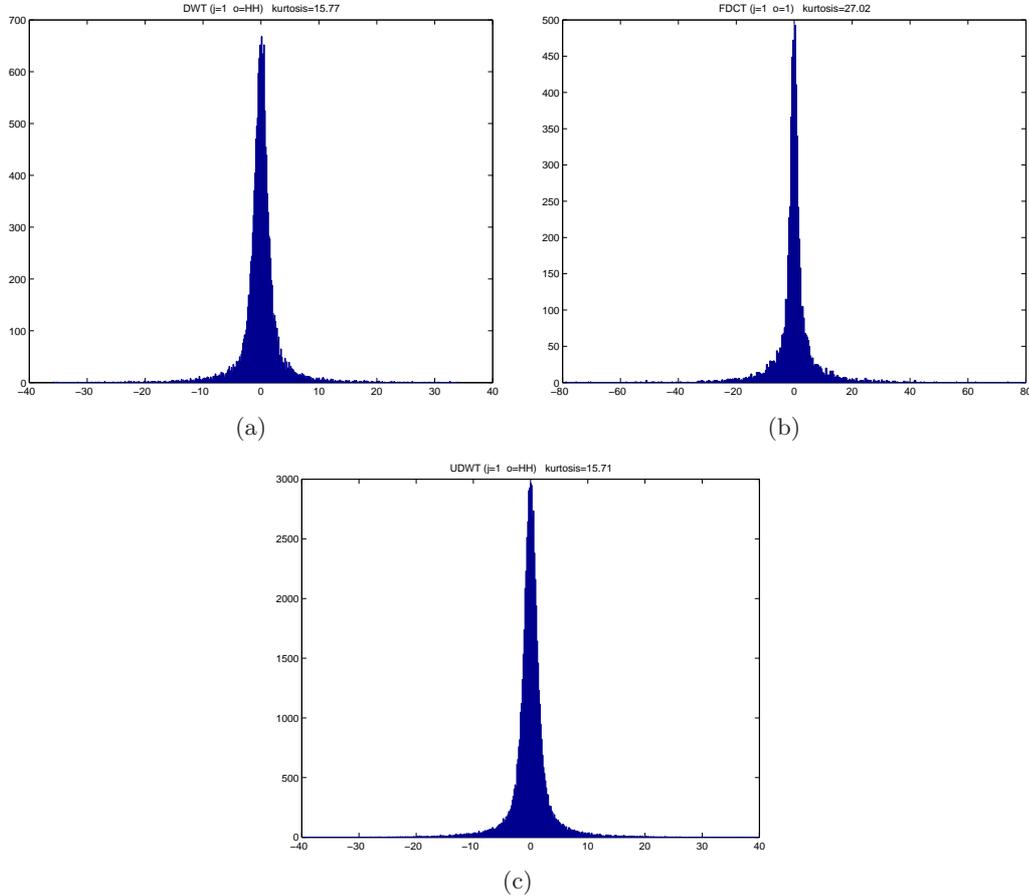


FIG. 2.3 – Distributions marginales des coefficients de l'image de Lena avec : DWT (a), FDCT (b) et UDWT (c). Ces lois marginales sont clairement non gaussiennes, caractérisées pas des densités symétriques centrées en 0 avec des queues relevées (leptokurtique). Les kurtosis ont des valeurs significativement plus élevées que la valeur gaussienne de 3.

Une manière de modéliser ce comportement est d'ajuster des modèles de distribution partant des observations précédentes (PDFs leptokurtiques à queues lourdes). A titre d'exemple, Mallat [Mallat, 1989] et ensuite Simoncelli [Simoncelli, 1999] ont proposé la distribution gaussienne généralisée (GGD).

D'autres modèles peuvent s'avérer judicieux dans un tel contexte. Citons l'exemple des distributions  $\alpha$ -stables introduites par le mathématicien Paul Lévy en 1924 au cours de ses études concernant le comportement limite des sommes de variables aléatoires indépendantes, et la distribution des Formes K de Bessel (BKF) [Grenander & Srivastava, 2001]. Nous reviendrons en détail sur ces lois dans la deuxième partie. Notons aussi les travaux récents de Wainwright et Simoncelli [Wainwright et al., 2000] sur les mélanges d'échelles de gaussiennes. Ces modèles statistiques ont pour vertu de regrouper comme cas particulier certaines des distributions que nous avons citées (notamment la distribution  $\alpha$ -stable et BKF).

Notons que la caractéristique fondamentale qui différencie les lois stables des lois normales

réside dans le fait que la queue de distribution d'une loi stable décroît selon une loi de puissance, alors que la queue d'une distribution gaussienne décroît selon une loi exponentielle. Ceci signifie que les distributions stables présentent une décroissance asymptotique plus lente que celle de la distribution normale. Cependant, en dépit de ce comportement, les formes explicites des PDFs des lois  $\alpha$ -stables ne sont pas connues en général sauf cas particuliers comme la loi de Cauchy ou gaussienne. Ceci limite leur mise en oeuvre pratique puisque des méthodes d'intégration numérique lourdes sont nécessaires. Un autre problème conceptuel dans l'application de ce type de modèles statistiques aux images réside dans la pathologie suivante : les moments d'ordre  $k$  d'une VA  $\alpha$ -stable ne sont pas finis. Ces deux difficultés rendent certaines techniques d'estimation usuelles (maximum de vraisemblance, méthodes des moments) inutilisables en pratique. Des astuces judicieuses ont été proposées dans la littérature pour contourner ces difficultés [Nikias & Shao, 1995, Mathieu, 2002]. Dans la seconde partie de ce document, nous reviendrons en détail sur la loi  $\alpha$ -stable et nous proposons des solutions originales pour la mise en oeuvre un modèle des statistiques marginales ainsi qu'en débruitage.

## 2.4.2 Loi conditionnelle/jointe

Dire qu'une image est une collection de VA indépendantes est une simplification qui enlève beaucoup à l'information portée par l'image et que nous percevons par notre système visuel. En effet, ce qui fait l'image n'est pas une simple adjonction quelconque de pixels mais bien un agencement particulier de ces pixels qui décrit une géométrie, une texture, etc. Ainsi, si un pixel est sur le contour d'un objet, on a de fortes présomptions de penser qu'au moins un de ses voisins est sur le contour aussi. Une façon de traduire cette observation en termes statistiques est de dire que les pixels sont inter-dépendants. Les champs de Markov sont une manière de prendre en compte ces dépendances.

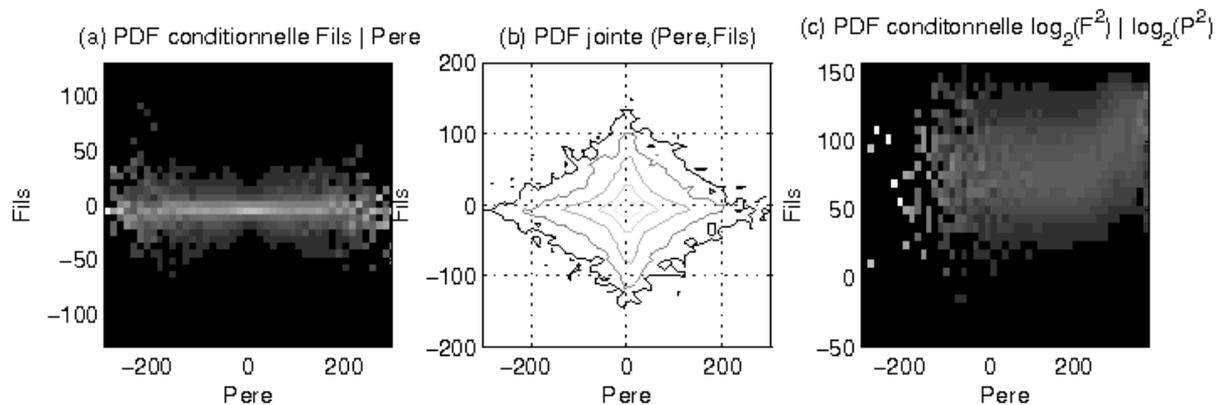


FIG. 2.4 – Exemple de PDFs (observées et ajustées par des modèles) jointe et conditionnelle de deux coefficients d'ondelettes à deux échelles différentes (le père est à une échelle plus grossière par rapport au fils). Les coefficients d'ondelettes sont ceux de l'image de Lena. L'intensité sur chaque figure reflète l'amplitude de la PDF.

Les Fig.2.4.(a)-(b) donnent une illustration des PDFs jointe et conditionnelle de deux coefficients d'ondelettes à deux échelles différentes. Deux remarques importantes sont à émettre. Premièrement, les PDFs sont centrées en (0,0) (valeurs les plus probables) mais le comportement statistique n'est absolument pas gaussien comme en témoigne la PDF jointe. Deuxièmement, bien

que les deux coefficients soient décorrélés (voir figure de gauche où la moyenne du fils ne dépend pas du père), ils sont dépendants (la dispersion des coefficients du fils est plus importante à grandes valeurs absolues du père). En effet, la PDF conditionnelle du log du fils sachant le log du père est unimodale et concentrée (à forte valeur) autour d'une (presque) droite, ce qui révèle que la variance conditionnelle  $\mathbb{E}[F^2|P^2]$  est approximativement proportionnelle à  $F^2$  (voir (c)). Ce raisonnement peut être poussé pour englober non seulement les relations père-fils, mais aussi la relation aux voisins directs de chaque coefficient.

Une approche pour modéliser ces PDFs jointes consiste à étendre les PDFs marginales à queues lourdes au cas multivarié. C'est cette démarche que nous adaptons en troisième partie de ce document. Nous montrons que ces modèles de PDF jouissent de plusieurs propriétés intéressantes. Nous discutons aussi tous les détails nécessaires à leur mise en oeuvre, ainsi bien pour la modélisation que pour le débruitage.

## 2.5 Conclusion

Nous avons évoqué quelques propriétés et lois statistiques nécessaires pour l'analyse statistique d'images. Nous avons aussi passé en revue quelques une des principales approches pour la modélisation statistique d'images proposées dans la littérature. Il apparaît donc évident que la modélisation statistique est une étape indispensable pour définir les propriétés de l'image que l'on cherche à recouvrir et pour traduire les connaissances *a priori* que l'on possède sur cette image.

Pour formuler ces contraintes, nous proposons, dans la suite, des modèles statistiques adaptés pour capter le caractère parcimonieux des coefficients de représentation de ces images. Ceux-ci permettent ainsi de traduire avec flexibilité des *a priori* sur les images dans le domaine des transformées multi-échelles. Le mariage fructueux entre ces modèles *a priori* pour les représentations parcimonieuses d'une part, et de la théorie de l'estimation bayésienne, nous permettra d'atteindre l'objectif fixé : le débruitage.



Deuxième partie

---

Statistiques univariées pour la restauration

---



# Chapitre 3

## Modélisation des statistiques marginales

---

### Sommaire

---

<b>3.1</b>	<b>Quel <i>a priori</i>?</b>	<b>44</b>
<b>3.2</b>	<b>Modèle de mélange d'échelles de gaussiennes : cadre général</b>	<b>44</b>
3.2.1	Définition	44
3.2.2	Propriétés	44
<b>3.3</b>	<b>L'<i>a priori</i> SMG et l'espace de Besov</b>	<b>47</b>
<b>3.4</b>	<b>Estimation des hyperparamètres</b>	<b>49</b>
<b>3.5</b>	<b>Cas 1 : <i>a priori</i> <math>\alpha</math>-stable</b>	<b>49</b>
3.5.1	Définition	49
3.5.2	Propriétés	50
3.5.2.1	Variables aléatoires $\alpha$ -stables symétriques	50
3.5.2.2	Quelques opérateurs arithmétiques	50
3.5.2.3	Comportement asymptotique des queues de distribution	51
3.5.2.4	PDFs $\alpha$ -stable connues	51
3.5.3	Approximation analytique de la PDF $\alpha$ -stable	52
3.5.4	Du mélange d'échelles au mélange fini	54
3.5.4.1	Modèle de mélange de gaussiennes	54
3.5.4.2	Approximation analytique des PDFs $S\alpha S$	55
3.5.5	Sélection de modèles	56
3.5.5.1	Le critère MDL (" <i>minimum description length</i> ")	57
3.5.5.2	Simulations Monte-Carlo et divergence KL	57
3.5.6	Estimation des hyperparamètres	59
3.5.6.1	Méthodes des quantiles	59
3.5.6.2	Méthodes basées sur la fonction caractéristique	60
3.5.6.3	Maximum de vraisemblance	62
3.5.6.4	Méthodes des moments fractionnaires	62
3.5.6.5	Méthodes basées sur le comportement asymptotique des queues	62
3.5.7	Performance des différents estimateurs	62
<b>3.6</b>	<b>Cas 2 : <i>a priori</i> des Formes K de Bessel</b>	<b>65</b>
3.6.1	Définition	65
3.6.2	Propriétés	65
3.6.2.1	Quelques propriétés fondamentales	65

3.6.2.2	Moments d'ordre $p$ . . . . .	67
3.6.3	L' <i>a priori</i> BKF et l'espace de Besov . . . . .	68
3.6.4	Estimation des hyperparamètres . . . . .	68
3.6.4.1	Méthode des cumulants . . . . .	68
3.6.4.2	Algorithme EM . . . . .	69
3.6.5	Performance des différents estimateurs . . . . .	72
<b>3.7</b>	<b>Qualité de l'<i>a priori</i> : application à la modélisation des statistiques marginales</b> . . . . .	<b>73</b>
3.7.1	Objectifs de l'expérience . . . . .	73
3.7.2	Protocole expérimental . . . . .	73
3.7.3	Discussion des résultats . . . . .	73

### 3.1 Quel *a priori* ?

Choisir un espace généralement hilbertien ( $\ell_2(\mathbb{Z}^2)$ ) muni d'une trame ou d'une base et projeter une image dans un sous-espace de dimension inférieure dans lequel seules quelques composantes sont significatives facilite l'obtention d'un *a priori* [Mallat, 1989, Simoncelli & Portilla, 1998, Grenander & Srivastava, 2001]. En effet, les PDFs de ces composantes dans de telles représentations sont loin d'avoir un comportement gaussien mais plutôt un comportement leptokurtique (un pic prononcé) avec des queues lourdes (queues relevées) (voir Fig.2.3).

Dans ce chapitre, nous introduisons un modèle statistique comme *a priori* bayésien adéquat à la modélisation statistique d'une grande classe d'images et adapté pour capturer ce comportement leptokurtique avec queues lourdes. Il s'agit du modèle de mélange d'échelles de gaussiennes (SMG) introduit en statistiques par Andrews et Mallows [Andrews & Mallows, 1974].

### 3.2 Modèle de mélange d'échelles de gaussiennes : cadre général

#### 3.2.1 Définition

Soit  $X$  une VA à valeurs dans  $\mathbb{R}$ . Sous le modèle SMG, il existe deux VAs indépendantes  $U \geq 0$  et  $Z \sim \mathcal{N}(0, 1)$  telles que :

$$X \stackrel{d}{=} Z\sqrt{U} \quad (3.1)$$

$\stackrel{d}{=}$  est l'égalité en distribution.

#### 3.2.2 Propriétés

Il est aisé de prouver les propriétés suivantes :

- La PDF de  $X$  est donnée par

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} u^{-1/2} e^{-\frac{x^2}{2u}} f_U(u) du \quad (3.2)$$

- La fonction caractéristique de  $X$  est donnée par

$$\psi_X(\omega) = \mathbb{E} [\mathbb{E} [e^{-i\omega u} | U]] = \mathbb{E} [\psi_Z(\omega; u)] = \int_0^{+\infty} e^{-\frac{u\omega^2}{2}} f_U(u) du = \mathcal{L}[f_U] \left( \frac{\omega^2}{2} \right) \quad (3.3)$$

où  $\psi_Z$  est la fonction caractéristique de la distribution normale de la VA  $Z$ , et  $\mathcal{L}[f_U]$  est la transformée de Laplace de  $f_U$ .

- La PDF est unimodale, symétrique et dérivable à gauche et à droite du mode 0.
- La PDF de  $X$  existe en 0 si et seulement si  $\mathbb{E}[U^{-1/2}] < +\infty$ .
- La PDF de  $U$  est reliée à la transformée de Laplace inverse de  $f_X$  :

**Proposition 3.1** *Si  $f_U$  est dérivable sur  $\mathbb{R}^+$ , alors :*

$$f_U(u) = \sqrt{\frac{\pi}{2}} u^{-3/2} \mathcal{L}[f_X]((2u)^{-1}) \quad (3.4)$$

où  $\mathcal{L}[f_X](\omega)$  est la transformée de Laplace inverse de  $f_X(\sqrt{y})$ .

**Preuve:**

Posons tout d'abord la VA  $V = U^{-1/2}$ . En opérant le changement de variable  $v = (2w)^{-1}$ , on peut alors écrire à partir de l'Eq.3.2 :

$$\begin{aligned} f_X(\sqrt{y}) &= \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} v e^{-\frac{yv^2}{2}} f_V(v) dv \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} f_V\left((2w)^{1/2}\right) e^{-yw} dw \\ &= \mathcal{L}\left[(2\pi)^{-1/2} f_V\left((2w)^{1/2}\right)\right] \end{aligned} \quad (3.5)$$

Ainsi, la transformée de Laplace inverse de  $f_X(\sqrt{y})$  est :

$$\mathcal{L}[f_X](\omega) = (2\pi)^{-1/2} f_V\left((2\omega)^{1/2}\right) \Rightarrow f_V(v) = \sqrt{2\pi} \phi(v^2/2) \quad (3.6)$$

où  $\phi$  est la PDF de la loi normale.

Finalement, en appliquant le changement de variable inverse  $U = V^{-2}$ , on obtient le résultat annoncé <sup>3</sup>. ■

- Le résultat suivant donne une caractérisation nécessaire et suffisante pour que  $X$  puisse se mettre sous la forme d'un modèle de mélange d'échelles de gaussiennes.

**Proposition 3.2** *La VA  $X$  possède une décomposition de l'Eq.3.1 ssi les dérivées  $k^{\text{ème}}$  de  $f_X(\sqrt{y})$  sont de signe alterné, i.e. :*

$$\left(-\frac{d}{dy}\right)^k f_X(\sqrt{y}) \geq 0 \quad \forall y > 0 \quad (3.7)$$

**Preuve:**

La condition nécessaire est évidente par dérivation directe de l'Eq.3.2. La suffisance est une conséquence du théorème d'Andrews et Mallows ([Andrews & Mallows, 1974]), en posant  $V = U^{-1/2}$ . ■

Cette propriété révèle qu'une telle famille de loi est très bien adaptée pour capturer le caractère creux des représentations parcimonieuses, et est donc légitime comme *a priori* pour les coefficients de ces représentations. En effet, du fait du caractère creux des décompositions, les PDFs empiriques des coefficients sont unimodales, centrées autour du mode, leptokurtiques et à queues lourdes. Un *a priori* adéquat devrait typiquement posséder une

<sup>3</sup>Ce résultat peut être facilement vérifié pour la loi des Formes K de Bessel par exemple.

PDF avec au moins sa dérivée première négative et sa dérivée seconde positive sur  $\mathbb{R}^+$ . Si on choisit un *a priori* dont la PDF (presque dérivable partout sauf éventuellement en 0) possède des dérivées de signe alterné, alors la condition suffisante de la proposition 3.2 est vérifiée, et cette PDF fera forcément partie du modèle de mélange de gaussiennes.

- **Moments absolus d'ordre  $p$**

Soient  $Z \sim \mathcal{N}(0, \sigma^2)$  et  $U > 0$  de loi  $\mathbb{L}_\theta$ , où  $\theta$  est le vecteur des hyperparamètres. Sans perte de généralité, nous supposons que  $\mathbb{E}[U] = 1$ .

**Proposition 3.3** *Le moment absolu d'ordre  $p$  de la VA  $X$  est donné par :*

$$M_X(p) = \frac{2^{\frac{p}{2}}}{\sqrt{\pi}} \sigma^p \Gamma\left(\frac{p+1}{2}\right) M_U\left(\frac{p}{2}\right) \quad (3.8)$$

**Preuve:**

On a,

$$\mathbb{E}[|X|^p] = \mathbb{E}[\mathbb{E}[|X|^p|U]] = \nu_p \mathbb{E}\left[U^{\frac{p}{2}}\right] = \nu_p M_U\left(\frac{p}{2}\right) \quad (3.9)$$

où  $\nu_p = \mathbb{E}[|Z|^p]$ .

Pour calculer  $\nu_p$ , on a par calcul direct, après le changement de variable  $\omega = z^2$

$$\nu_p = M_Z(p) = \frac{2^{\frac{p}{2}}}{\sqrt{\pi}} \sigma^p \Gamma\left(\frac{p+1}{2}\right) < +\infty \quad (3.10)$$

En substituant les différent termes, le résultat suit. ■

La famille de modèle SMG est aussi forcément leptokurtique pour tout  $U$  aléatoire. Ceci est résumé dans la proposition suivante :

**Proposition 3.4** *Soit  $X = \sqrt{U}Z$  avec  $U \geq 0$  et  $Z \sim \mathcal{N}(0, \sigma^2)$ . Pour tout  $U$  aléatoire, le kurtosis de la VA  $X$  est toujours strictement positif.*

**Preuve:**

Écrivons le kurtosis d'une VA  $X$  :

$$\begin{aligned} \text{Kurt}[X] &= \frac{M_4(X)}{\text{Var}[X]^2} - 3 \\ &= \mathbb{E}[U^2] \frac{\nu_4}{\sigma^4} - 3 \\ &= 3(\mathbb{E}[U^2] - 1) \end{aligned} \quad (3.11)$$

car par définition, pour une variable gaussienne  $\frac{\nu_4}{\sigma^4} = 3$ .

Finalement par l'inégalité de Jensen :

$$\mathbb{E}[U^2] > \mathbb{E}[U]^2 = 1 \quad (3.12)$$

Si l'inégalité ci-dessus n'est pas stricte, cela voudrait dire que  $\text{Var}[U] = \mathbb{E}[U^2] - 1 = 0$ , ou encore que  $U \equiv cte$  avec une probabilité de 1. Or,  $U$  est aléatoire. ■

- **Cas particuliers du modèle de mélange d'échelles de gaussiennes**

Deux lois se présentent comme cas particulier pour le modèle de mélange d'échelles de gaussiennes. D'après l'Eq.3.1,

- (i) Si  $U$  suit une loi  $\alpha$ -stable positive, alors on trouve l'a priori dit  $\alpha$ -*stable*.
- (ii) Si  $U$  suit une loi de Gamma, on trouve l'a priori de **Formes  $K$  de Bessel**.

Notons aussi que la GGD est un cas particulier de SMG où la PDF de  $U$  ne possède pas de forme analytique simple.

### 3.3 L'a priori SMG et l'espace de Besov

Les espaces de Besov forment une classe d'espaces intéressante dans le domaine du traitement d'image. En effet, d'une part les fonctions régulières par morceaux peuvent être classées dans de tels espaces et d'autre part l'appartenance à des espaces de Besov peut fournir des informations sur la régularité des images.

Ici, on négligera la définition théorique des espaces de Besov par le module de continuité et on se contentera d'une caractérisation pratique par les coefficients d'ondelettes [Meyer, 1992].

**Théorème 3.1** ([Meyer, 1992]) *Soit  $g = \sum_{j,k} d_{j,k} \psi_{j,k}$  où  $d_{j,k}$  sont les coefficients d'ondelettes et  $\psi$  est l'ondelette à l'échelle  $2^j$  décalée de  $k$ . On a l'équivalence suivante :*

$$g \in \mathcal{B}_{p,q}^r \iff \|d\|_{\mathcal{B}_{p,q}^r} = \begin{cases} |d_{0,0}| + \left\{ \sum_{j=0}^{\infty} 2^{jr'q} \left( \sum_{k=0}^{2^j-1} |d_{j,k}|^p \right)^{\frac{q}{p}} \right\}^{\frac{1}{q}} & \text{si } 1 \leq q < \infty \\ |d_{0,0}| + \sup_{j \geq 0} \left\{ 2^{jr'} \left( \sum_{k=0}^{2^j-1} |d_{j,k}|^p \right)^{\frac{1}{p}} \right\} & \text{si } q = \infty \end{cases} \quad (3.13)$$

où  $r' = (r + \frac{1}{2} - \frac{1}{p})$  et  $1 \leq p < \infty$ .

$p$  est le paramètre de régularité de l'image  $g$ .

L'équivalence des normes pour les espaces Besov se relie à la distribution *a priori* des coefficients d'ondelettes aux différents niveaux de détail. Une relation explicite, entre les paramètres du modèle SMG et ceux de l'espace de Besov auquel appartient l'image (p.s.), est donnée par le théorème suivant :

**Théorème 3.2** *Soient  $X_{j,k} \stackrel{d}{=} Z_j \sqrt{U}$  des VAs iid à chaque échelle, où  $M_U(p) < +\infty$   $1 \leq p < +\infty$  et  $\sqrt{\text{Var}[X_{j,k}]} = \sigma_j = \sigma_0 2^{-j\beta}$  (propriété d'invariance d'échelle), avec  $(0 < \sigma_0 < +\infty, \beta \geq 0)$ . On a alors,*

$$g \in \mathcal{B}_{p,q}^r \text{ (p.s.) ssi } \beta > (r + \frac{1}{2}) \text{ pour } 1 \leq p < \infty \text{ et } 1 \leq q \leq \infty.$$

**Preuve:**

On a,

$$M_j(p) = \mathbb{E} [\|X_j\|_p^p] = \sum_k \mathbb{E} [|X_{j,k}|^p] = \nu_p \sigma_0^p 2^{j(1-\beta p)} M_U \left( \frac{p}{2} \right) \quad (3.14)$$

Notons

$$M_0(p) = \nu_p \sigma_0^p M_U \left( \frac{p}{2} \right) < +\infty \text{ par hypothèse.} \quad (3.15)$$

Par l'inégalité de Tchebychev :

$$\begin{aligned}
\sum_{j=0}^{+\infty} \mathcal{P} \left\{ |2^{-j(1-\beta p)} \|X_j\|_p^p - M_0(p)| > \epsilon \right\} &\leq \epsilon^{-2} \sum_{j=0}^{+\infty} 2^{-2j(1-\beta p)} \text{Var}(\|X_j\|_p^p) \\
&\leq \epsilon^{-2} \sum_{j=0}^{+\infty} 2^{-2j(1-\beta p)} \sum_{k=0}^{2^j-1} \text{Var} [|X_{j,k}|^p] \\
&\leq \epsilon^{-2} \sum_{j=0}^{+\infty} 2^{-2j(1-\beta p)} \sum_{k=0}^{2^j-1} \mathbb{E} [|X_{j,k}|^{2p}] \\
&\text{on a } M_j(2p) = \nu_{2p} \sigma_0^{2p} 2^{j(1-2\beta p)} M_U(p) \\
&\leq \epsilon^{-2} M_0(2p) \sum_{j=0}^{+\infty} 2^{-j} \\
&< +\infty \text{ pour } p \text{ fini}
\end{aligned} \tag{3.16}$$

Par le premier lemme de Borel-Cantelli, il s'ensuit alors que :

$$2^{-j(1-\beta p)} \|X_j\|_p^p \xrightarrow{p.s.} M_0(p) \text{ lorsque } j \longrightarrow +\infty \tag{3.17}$$

Ceci induit grâce aux équivalences des normes pour les espaces de Besov  $\mathcal{B}_{p,q}^r$

(i)  $1 \leq p < +\infty$  et  $1 \leq q < +\infty \implies M_0(p)$  est fini.

$$\begin{aligned}
g \in \mathcal{B}_{p,q}^r &\iff \sum_{j=0}^{+\infty} 2^{jr'q} \|X_j\|_p^q < +\infty \\
&\iff \sum_{j=0}^{+\infty} 2^{jr'q} 2^{j(1-\beta p)\frac{q}{p}} (M_0(p))^{\frac{q}{p}} < +\infty \\
&\iff \sum_{j=0}^{+\infty} 2^{j(r'q + \frac{q}{p} - q\beta)} < +\infty \\
&\text{avec } r' = r + \frac{1}{2} - \frac{1}{p}
\end{aligned} \tag{3.18}$$

la série ci-dessus converge ssi

$$\begin{aligned}
r'q + \frac{q}{p} - q\beta < 0 &\iff qr + \frac{q}{2} - \frac{q}{p} + \frac{q}{p} - q\beta < 0 \\
&\iff r + \frac{1}{2} - \beta < 0 \\
&\iff \beta > r + \frac{1}{2}
\end{aligned} \tag{3.19}$$

(ii)  $q = +\infty$  et  $1 \leq p < +\infty \implies M_0(p)$  est encore fini.

L'équivalence des normes nécessite pour que  $g \in \mathcal{B}_{p,q}^r$

$$\begin{aligned}
\sup_{j \geq 0} \left[ 2^{jr'} 2^{j \left( \frac{1-\beta p}{p} \right)} \right] < +\infty &\iff r' + \frac{1}{p} - \beta < 0 \\
&\iff r + \frac{1}{2} - \frac{1}{p} + \frac{1}{p} - \beta < 0 \\
&\iff \beta > r + \frac{1}{2}
\end{aligned} \tag{3.20}$$

et donc puisque  $M_0(p)$  est fini, l'image  $g$  est dans  $\mathcal{B}_{p,q}^r$  p.s. dès lors que  $\beta > r + \frac{1}{2}$ .  $\blacksquare$

### 3.4 Estimation des hyperparamètres

Dans le cadre d'une approche bayésienne, il est nécessaire de caractériser entièrement le modèle de distribution, i.e. d'estimer les différents paramètres permettant sa caractérisation.

La distribution de  $U$ , dans le modèle SMG donné par l'Eq.3.1, dépend d'un ensemble d'hyperparamètres  $\theta$  (paramètres caractérisant l'*a priori*). Dans le domaine des transformées multi-échelles, ces hyperparamètres sont estimés à partir des coefficients observés à chaque orientation et à chaque échelle. Il existe diverses méthodes permettant d'estimer les hyperparamètres  $\theta$ . Parmi les plus usuels, citons par exemple : le maximum de vraisemblance, l'algorithme EM, la méthode des moments et des cumulants, la méthode des quantiles, les méthodes basées sur la fonction caractéristique, les méthodes basées sur le comportement asymptotique des queues, etc.

Pour le débruitage bayésien d'images, l'étape d'estimation des hyperparamètres est cruciale et conditionne les performances finales de l'algorithme de débruitage. Notons aussi que cette estimation reste accessible en absence de bruit et devient plus délicate en sa présence. Les différentes méthodes envisagées en absence de bruit seront développées dans ce qui suit pour le cas de l'*a priori*  $\alpha$ -stable et les Formes K de Bessel. En présence de bruit, nous y reviendrons en détail dans le chapitre suivant.

### 3.5 Cas 1 : a priori $\alpha$ -stable

La notion de loi de probabilité stable a été introduite par le mathématicien Paul Lévy au cours de ses études concernant le comportement des sommes de variables aléatoires indépendantes [Lévy, 1954]. La notion de stabilité émane du fait que la somme de deux variables aléatoires indépendantes, suivant une loi  $\alpha$ -stable de paramètre  $\alpha$ , suit également une loi  $\alpha$ -stable. La distribution stable est définie de plusieurs manières équivalentes liées essentiellement à la propriété de stabilité, au théorème central limite et à la fonction caractéristique [Samorodnitsky & Taqqu, 1994, Zolotarev, 1986]. Dans ce qui suit, nous nous contenterons d'une définition décrite entièrement à l'aide de la fonction caractéristique.

#### 3.5.1 Définition

Une distribution stable est caractérisée par quatre paramètres : l'exposant de stabilité  $\alpha$ , le paramètre d'échelle  $\sigma$ , le paramètre d'asymétrie  $\beta$  et le paramètre de position  $\mu$ . Dans le cas où  $\alpha = 2$ , la distribution stable est gaussienne ;  $\sigma$  est proportionnel à l'écart-type de la distribution  $\sigma$ ,  $\beta = 0$  et  $\mu$  est la moyenne.

**Définition 3.1** Une variable aléatoire  $X$  est appelée  $\alpha$ -stable s'il existe des paramètres  $0 < \alpha \leq 2$ ,  $\sigma \geq 0$ ,  $-1 \leq \beta \leq 1$ , et  $\mu \in \mathbb{R}$  telles que sa fonction caractéristique s'écrive de la façon suivante : (formulation de Zolotarev [Zolotarev, 1986])

$$\psi_X(\omega) = \begin{cases} \exp(i\mu\omega - \sigma^\alpha |\omega|^\alpha (1 - i\beta \text{sign}(\omega) \tan(\frac{\pi\alpha}{2}))), \omega \in \mathbb{R} & \text{si } \alpha \neq 1 \\ \exp(i\mu\omega - \sigma^\alpha |\omega|^\alpha (1 - i\beta \text{sign}(\omega) - \frac{2}{\pi} \log |\omega|)), \omega \in \mathbb{R} & \text{si } \alpha = 1 \end{cases} \quad (3.21)$$

Les quatre paramètres qui caractérisent une loi stable s'interprètent comme suit :

- L'exposant caractéristique  $\alpha$ ,  $0 < \alpha \leq 2$ . Il décrit la forme de la distribution ou son degré de "lepto-kurticité".
- Le paramètre d'asymétrie  $\beta$ ,  $\beta \in [-1, 1]$ . Si  $\beta = -1$ , on dit que la distribution est totalement asymétrique à gauche. Si  $\beta = 0$ , on dit que la distribution est symétrique. Si  $\beta = 1$ , la distribution est dite alors totalement asymétrique à droite.
- Le paramètre de position  $\mu$ ,  $\mu \in \mathbb{R}$ . Si  $\alpha > 1$ , ce paramètre est égal à la moyenne.
- Le paramètre d'échelle  $\sigma = \gamma^{\frac{1}{\alpha}}$ ,  $\sigma \in \mathbb{R}^+$ . Si  $\alpha = 2$ , l'écart-type de la distribution est égal à  $\sqrt{2}\sigma$ .

### 3.5.2 Propriétés

Les lois stables vérifient certaines propriétés très utiles dans la pratique. Nous en présentons ici quelques unes.

#### 3.5.2.1 Variables aléatoires $\alpha$ -stables symétriques

Une VA stable est appelée  $\alpha$ -stable *symétrique* ( $S\alpha S$ ) si sa distribution est symétrique.

**Proposition 3.5** Soit  $X \sim S_\alpha(\beta, \mu, \sigma)$ .  $X$  suit une distribution  $\alpha$ -stable symétrique si  $\beta = 0$  et  $\mu = 0$ .

$X$  est symétrique par rapport à  $\mu$  si  $\beta = 0$ .

D'après la Définition 3.1 et la Proposition 3.5, on voit que si  $X$  est  $S\alpha S$ , alors sa fonction caractéristique s'écrit sous une forme très simple :

$$\mathbb{E}[\exp(i\omega X)] = \exp(-\sigma^\alpha |\omega|^\alpha), \omega \in \mathbb{R} \quad (3.22)$$

Par ailleurs, une VA  $X$  est dite  $S\alpha S$  *réduite* si  $\sigma = 1$ .

#### 3.5.2.2 Quelques opérateurs arithmétiques

**Propriété 3.1** Si  $X_1$  et  $X_2$  sont deux VAs indépendantes avec  $X_i \sim S_\alpha(\beta_i, \mu_i, \sigma_i)$ ,  $i = 1, 2$ , alors  $X_1 + X_2 \sim S_\alpha(\beta, \mu, \sigma)$  avec :

$$\beta = \frac{\beta_1 \sigma_1^\alpha + \beta_2 \sigma_2^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha}, \sigma = (\sigma_1^\alpha + \sigma_2^\alpha)^{\frac{1}{\alpha}}, \mu = \mu_1 + \mu_2. \quad (3.23)$$

**Propriété 3.2** Si  $X \sim S_\alpha(\beta, \mu, \sigma)$ , et  $a, b \in \mathbb{R}$ , alors

$$\begin{cases} aX + b \sim S_\alpha(\text{sign}(a)\beta, a\mu + b, |a|\sigma) & \text{si } \alpha \neq 1 \\ aX + b \sim S_\alpha(\text{sign}(a)\beta, a\mu - \frac{2}{\pi}a(\log |a|)\sigma\beta, |a|\sigma) & \text{si } \alpha = 1 \end{cases} \quad (3.24)$$

**Propriété 3.3** Si  $X \sim S_\alpha(\beta, \mu, \sigma)$ , avec  $0 < \alpha < 2$ , alors

$$\begin{cases} \mathbb{E}[|X|^p] < \infty & \text{si } p \in (0, \alpha) \\ \mathbb{E}[|X|^p] = \infty & \text{si } p \in [\alpha, 2) \end{cases} \quad (3.25)$$

Cette propriété pose un problème théorique pour l'équivalence des normes pour les espaces de Besov.

**Propriété 3.4** Si  $X \sim S_\alpha(0, \mu, \sigma)$ , avec  $1 < \alpha \leq 2$ , alors

$$\mathbb{E}[X] = \mu \quad (3.26)$$

### 3.5.2.3 Comportement asymptotique des queues de distribution

**Propriété 3.5** ([Nikias & Shao, 1995, Nolan, 1997]) Si  $X \sim S_\alpha(0, \mu, \sigma)$ , avec  $0 < \alpha \leq 2$ , alors

$$\begin{cases} \lim_{x \rightarrow \infty} x^\alpha \mathcal{P}[X > x] & = C_\alpha \frac{1+\beta}{2} \sigma^\alpha \\ \lim_{x \rightarrow \infty} x^\alpha \mathcal{P}[X < -x] & = C_\alpha \frac{1-\beta}{2} \sigma^\alpha \end{cases} \quad (3.27)$$

où

$$C_\alpha = \left( \int_0^\infty x^{-\alpha} \sin(x) dx \right)^{-1} \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha) \cos(\frac{\pi\alpha}{2})} & \text{si } \alpha \neq 1 \\ \frac{2}{\pi} & \text{si } \alpha = 1 \end{cases} \quad (3.28)$$

Cette propriété constitue la caractéristique fondamentale qui différencie les lois stables des lois normales. En effet, la queue de distribution d'une loi stable décroît selon une loi de puissance, alors que la queue d'une distribution gaussienne décroît selon une loi exponentielle. Ceci signifie que les distributions stables présentent une décroissance asymptotique plus lente que celle de la distribution normale.

### 3.5.2.4 PDFs $\alpha$ -stable connues

Les formes explicites des fonctions de densité de probabilité (PDFs) des lois  $\alpha$ -stables ne sont pas connues en général, l'exception est faite pour quelques cas particuliers que nous citons :

- Loi Normale  $S_2(0, \mu, \sigma)$  de densité

$$f_X(x) = \frac{1}{2\sigma\sqrt{\pi}} \exp\left(-\frac{(x-\mu)^2}{4\sigma^2}\right) \quad (3.29)$$

- Loi de Cauchy  $S_1(0, \mu, \sigma)$  de densité

$$f_X(x) = \frac{2\sigma}{\pi((x-\mu)^2 + 4\sigma^2)} \quad (3.30)$$

- Loi de Lévy  $S_{\frac{1}{2}}(1, \mu, \sigma)$  de densité

$$f_X(x) = \left(\frac{\sigma}{2\pi}\right)^{\frac{1}{2}} (x-\mu)^{-\frac{3}{2}} \exp\left(-\frac{\sigma}{2(x-\mu)}\right) \times \mathbf{1}_{(\mu, \infty)}(x) \quad (3.31)$$

### 3.5.3 Approximation analytique de la PDF $\alpha$ -stable

La PDF d'une VA stable existe et est continue, mais à quelques exceptions près, citées ci-dessus, aucune forme analytique n'est connue. Dans [McCulloch, 1986], McCulloch a développé des algorithmes efficaces pour approcher la densité des lois  $S\alpha S$  pour  $\alpha > 0.85$ . D'autres auteurs suggèrent d'approcher la densité en inversant l'expression de la fonction caractéristique avec la transformée rapide de Fourier [Mathieu, 2002], mais le problème réside dans le calcul des intégrales impropres.

En appliquant la transformée de Fourier inverse (IFFT) à la fonction caractéristique  $\psi_X(\omega)$  d'une VA  $\alpha$ -stable donnée par l'expression (3.21),

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-i\omega x) \psi_X(\omega) d\omega \quad (3.32)$$

on obtient une représentation intégrale des fonctions PDFs  $\alpha$ -stable :

$$f(x; \alpha, \beta, \mu, \sigma = \gamma^{\frac{1}{\alpha}}) = \begin{cases} \frac{1}{\pi} \int_0^{\infty} \exp(-\sigma^\alpha \omega^\alpha) \cos \left[ (x - \mu)\omega + \beta\omega^\alpha \tan\left(\frac{\alpha\pi}{2}\right) \right] d\omega & \text{si } \alpha \neq 1 \\ \frac{1}{\pi} \int_0^{\infty} \exp(-\sigma^\alpha \omega^\alpha) \cos \left[ (x - \mu)\omega + \beta\omega^\alpha \frac{2}{\pi} \log|\omega| \right] d\omega & \text{si } \alpha = 1 \end{cases} \quad (3.33)$$

Cette intégrale peut être évaluée analytiquement seulement pour  $\alpha = 2$ . Dans ce cas la distribution est Gaussienne (voir Eq.3.29). Pour  $\alpha = 1$  et  $\beta = 0$ , la distribution est une distribution de Cauchy (voir Eq.3.30). Et finalement pour  $\alpha = \frac{1}{2}$  et  $\beta = 1$ , la distribution est dite de Lévy (voir Eq.3.31).

Dans Nolan [Nolan, 1997], l'auteur a prouvé l'existence d'une intégrale exacte équivalente à celle de l'Eq.3.33, où cette fois-ci, les bornes d'intégration sont finies. Cette approche exacte, bien que plus stable numériquement, reste très lente en calcul. Elle n'offre cependant pas d'expression analytique ni pour le débruiteur bayésien ni pour la distribution marginale des coefficients bruités. Elle sera néanmoins utilisée comme référence lors de nos comparaisons par la suite.

Les résultats du calcul de ces densités par une intégration numérique par la quadrature de Gauss pour différentes valeurs du paramètre  $\alpha$ ,  $\beta$  et  $\sigma$  sont illustrés pour les Fig 3.1 et 3.2. Les Fig 3.1(a) et 3.1(b) montrent les diverses formes prises par les PDFs en faisant varier le paramètre  $\alpha$ . La Fig.3.2(a) illustre la densité lorsqu'on fait varier  $\beta$  entre -1 et 1. La figure Fig3.2(b) montre la dépendance des densités vis-à-vis du paramètre  $\sigma$ .

**Remarque** Comme nous l'avons déjà remarqué, du fait du caractère creux des représentations parcimonieuses comme les ondelettes, la plupart des coefficients de détail d'un signal analysé sont nuls et seuls quelques uns sont significatifs. Ainsi, la répartition de ces coefficients est caractérisée par une densité centrée en 0 avec des queues relevées [Mallat, 1989, Simoncelli, 1999, Chang et al., 2000a, Chang et al., 2000b, Achim et al., 2001, Portilla et al., 2003]. C'est exactement cette propriété qui est révélée par la densité de probabilité  $S\alpha S$   $S_\alpha(0, 0, \gamma = \sigma^\alpha)$  (Fig.3.2).

Nous présentons maintenant une méthode, qui est rapide et numériquement très stable, pour obtenir l'expression analytique de la PDF avec des valeurs de paramètres arbitraires.

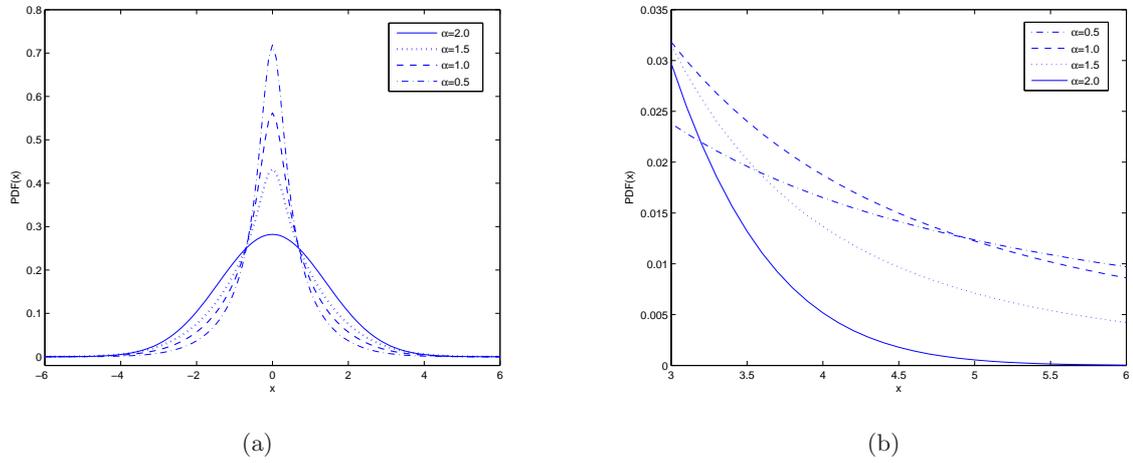


FIG. 3.1 – (a) Densités des lois stables (PDFs) de paramètres  $\alpha \in \{2, 1.5, 1, 0.5\}$ ,  $\beta = 0$ ,  $\sigma = 1$  et  $\mu = 0$ . (b) Évolution de la queue de distribution avec le paramètre  $\alpha$ .

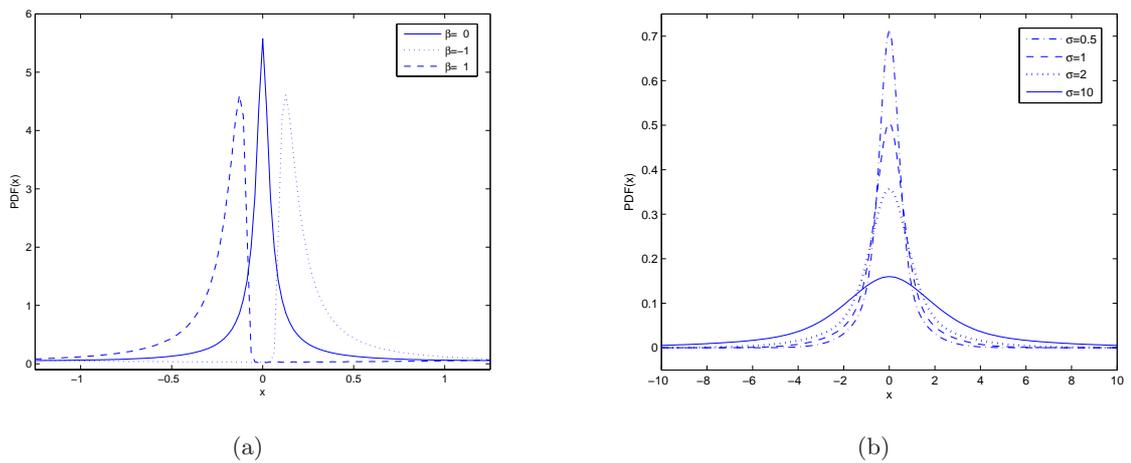


FIG. 3.2 – (a) L'effet du paramètre d'asymétrie  $\beta$ , avec  $\alpha = 1.5$ ,  $\sigma = 1$  et  $\mu = 0$ . (b) L'effet du paramètre d'échelle  $\sigma$ , avec  $\alpha = 1$ ,  $\beta = 0$  et  $\mu = 0$ .

### 3.5.4 Du mélange d'échelles au mélange fini

#### 3.5.4.1 Modèle de mélange de gaussiennes

Le modèle de mélange de gaussiennes est basé sur le corollaire de la propriété de mélange d'échelles des variables aléatoires  $\alpha$ -stables [Kuruoglu, 1998]. Ce corollaire est en fait un cas particulier du modèle SMG comme le met en lumière le corollaire 3.1.

La notion de mélange, donnée par le théorème suivant, émane du fait qu'une variable aléatoire  $\alpha$ -stable peut être représentée en fonction de deux variables aléatoires indépendantes, l'une suit une loi  $\alpha$ -stable symétrique et l'autre suit une loi  $\alpha$ -stable positive [Andrews & Mallows, 1974, Samorodnitsky & Taqqu, 1994].

**Théorème 3.3 (Théorème de mélange)** *Soit la VA stable  $Z \sim S_{\alpha_z}(0, 0, \gamma_z)$  avec  $0 < \alpha_z \leq 2$  et  $0 < \alpha_x < \alpha_z$ .*

*Il existe une VA positive  $U \sim S_{\frac{\alpha_x}{\alpha_z}}\left(-1, 0, \left(\cos\left(\frac{\pi\alpha_x}{2\alpha_z}\right)\right)^{\frac{\alpha_z}{\alpha_x}}\right)$  indépendante de  $Z$ , telle que :*

$$X = U^{\frac{1}{\alpha_z}} Z \sim S_{\alpha_x}(0, 0, \gamma_x) \quad (3.34)$$

Un cas particulier du théorème 3.3, donné par le corollaire suivant, montre que chaque VA  $S_{\alpha S}$  peut être représentée en fonction d'une VA gaussienne et d'une VA  $\alpha$ -stable [Andrews & Mallows, 1974, Samorodnitsky & Taqqu, 1994].

**Corollaire 3.1 (Corollaire du théorème de mélange d'échelles)** *Soit  $Z$  une VA gaussienne ( $\alpha_z = 2$ )*

$$Z \sim \mathcal{N}(0, 2\gamma_z)$$

*Il existe une VA  $\alpha$ -stable positive  $U \sim S_{\frac{\alpha_x}{2}}\left(-1, 0, \left(\cos\left(\frac{\pi\alpha_x}{4}\right)\right)^{\frac{2}{\alpha_x}}\right)$  indépendante de  $Z$ , telle que :*

$$X = U^{\frac{1}{2}} Z \sim S_{\alpha_x}(0, 0, \gamma_x) \quad (3.35)$$

Nous allons à présent exploiter ce résultat pour fournir une approximation précise pour la PDF. Cette approximation s'avère par la suite tout aussi intéressante en présence de bruit.

Si on définit  $V = U^{\frac{1}{2}} \geq 0$ , la PDF de  $X$  est déduite par la propriété de marginalisation des densités de probabilités :

$$f_X(x) = \frac{1}{\sqrt{4\pi\gamma_x}} \int_0^{+\infty} \exp\left(-\frac{x^2}{4\gamma_x v^2}\right) f_V(v) v^{-1} dv \quad (3.36)$$

où  $f_V(v)$  représente la PDF de la VA de mélange  $V$ . L'échantillonnage de  $f_X(x)$  sur un ensemble fini de  $N$  points permet d'obtenir une approximation de la PDF du modèle de mélange d'échelles de gaussiennes :

$$\hat{f}_X(x) \approx \frac{\sum_{j=1}^N v_j^{-1} \exp\left(-\frac{x^2}{4\gamma_x v_j^2}\right) f_V(v_j)}{\sqrt{4\pi\gamma_x} \sum_{j=1}^N f_V(v_j)} \quad (3.37)$$

Cette expression analytique de la PDF  $S_{\alpha S}$  est seulement une approximation, puisque l'intégrale continue a été approchée par une somme finie de Riemann. Pour une bonne approximation, l'Eq.3.36 doit être échantillonnée sur un grand ensemble de points. Pour réduire la complexité du

modèle dans Eq.3.37 nous employons un nombre restreint de composantes et l'échantillonnage de l'Eq.3.36 se fait pour quelques points seulement. Dans ce cas, l'affinage de cette approximation brute, en utilisant l'algorithme EM [Dempster et al., 1977], permet d'obtenir une meilleure estimation.

### 3.5.4.2 Approximation analytique des PDFs $S\alpha S$

Comme nous l'avons expliqué, les formes explicites des PDFs  $\alpha$ -stables ne sont pas connues en général, et trouver une expression analytique est une tâche délicate. Le modèle de mélange fini de gaussiennes, retenu dans la section précédente (Eq.3.37), permet d'aboutir à une estimation de la PDF  $S\alpha S$ . L'algorithme d'estimation se présente sous la forme suivante :

---

#### Algorithme 1 Approximation analytique des PDFs $S\alpha S$ par SMG

---

**Entrées:** soient  $\mathbf{x} = \{x_1, \dots, x_m\}$  les échantillons des observations.

- 1: Initialisation des paramètres de la distribution  $\alpha$ -stable :  $\{\frac{\alpha}{2}, \beta = -1, \mu = 0, \gamma = (\cos(\frac{\pi\alpha}{4}))^{\frac{2}{\alpha}}\}$

La fonction caractéristique de  $U$  est donnée par :

$$\psi_U(\omega) = \begin{cases} \exp\left[-\left(\cos\left(\frac{\pi\alpha}{4}\right)\right)^{\frac{2}{\alpha}}|\omega|^\alpha\left[1 + j\text{sign}(\omega)\tan\left(\frac{\pi\alpha}{2}\right)\right]\right] & \text{si } \alpha \neq 1 \\ \exp\left[-\left(\cos\left(\frac{\pi\alpha}{4}\right)\right)^{\frac{2}{\alpha}}|\omega|^\alpha\left[1 + j\text{sign}(\omega)\left(-\frac{2}{\pi}\log|\omega|\right)\right]\right] & \text{si } \alpha = 1 \end{cases} \quad (3.38)$$

- 2: Évaluer la PDF stable positive  $f_U$  en  $N$  points en appliquant la transformée de Fourier rapide inverse à la fonction caractéristique  $\psi_U(\omega)$ , où  $N$  correspond au nombre de gaussiennes dans le modèle de mélange.
- 3: La PDF de la VA  $V = U^{\frac{1}{2}}$ , dite fonction de mélange, est obtenue par :

$$h(v_j) = 2v_j f_U(v_j^2) \quad (3.39)$$

- 4: La substitution de la fonction de mélange, calculée dans l'étape (3), par l'équation Eq.3.37 permet d'obtenir l'approximation analytique de la PDF  $S\alpha S$  :

$$\hat{f}_X(x_m) = \frac{\sum_{j=1}^N \phi\left(x_m; 0, 2\gamma v_j^2\right) v_j f_U(v_j^2)}{\sum_{j=1}^N v_j f_U(v_j^2)} \quad (3.40)$$

où  $\phi(\cdot)$  est la PDF normale des observations.

- 5: Affinage de l'approximation par l'algorithme EM [Dempster et al., 1977, McLachlan & Peel, 2000]. L'étape précédente est utilisée comme étape d'initialisation pour l'algorithme EM. Pour un mélange de gaussiennes, nous cherchons à estimer

$$\hat{f}_X(\hat{x}_m) = \sum_{j=1}^N \mathcal{P}(x_m|j) P_j \quad (3.41)$$

où les  $P_j = \frac{h(v_j)}{\sum_{j=0}^N h(v_j)}$  sont les proportions de mélange et  $\mathcal{P}(x_m|j) = \phi(x_m; 0, \sigma_j)$  sont des PDFs normales cachées avec  $\sigma_j^2 = 2\gamma v_j^2$ .

---

### 3.5.5 Sélection de modèles

Bien que l'approximation que nous proposons soit assez bonne et très stable, il reste toutefois à définir le nombre de composantes  $N$  dans le mélange de gaussiennes.

Du point de vue graphique, la Fig.3.3 montre qu'il est possible de trouver une bonne estimation pour la PDF  $S_{\alpha S}$  en utilisant le modèle de mélange fini de gaussiennes donné par l'algorithme défini précédemment. De plus, nous constatons, suite à l'observation des courbes, que la PDF de mélange de gaussiennes est très proche de la PDF exacte évaluée à partir des intégrales de bornes finies (Eq.3.33) pour différentes valeurs du paramètre  $\alpha$ . Ceci est confirmé par les divergences de "Kullback-Leibler" (KL) calculées entre les deux PDFs, pour différentes valeurs du paramètre  $\alpha$ .

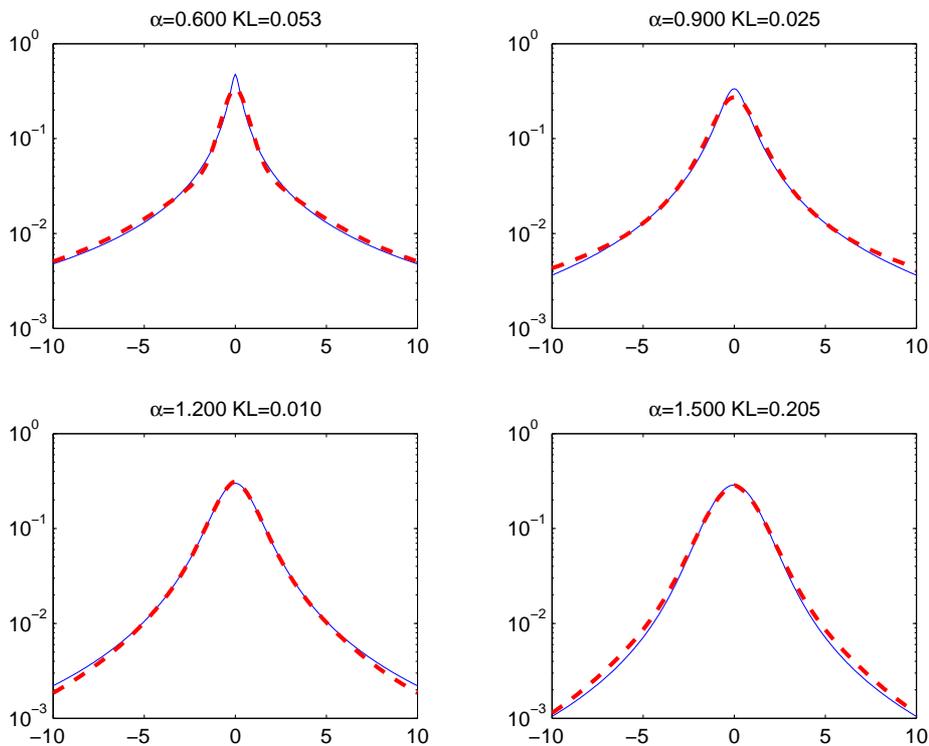


FIG. 3.3 – Comparaison sur une échelle log-log de la PDF exacte, calculée à partir de l'Eq.3.33 (trait plein), et la PDF approchée par un mélange de 8 gaussiennes (- -) pour différentes valeurs du paramètre  $\alpha$ . La divergence de "Kullback-Leibler" (KL) entre les deux PDFs est notée au dessus de chaque graphe.

Afin d'approfondir l'analyse, nous avons étudié l'influence du nombre de gaussiennes dans le mélange pour en proposer un choix. Plusieurs solutions s'offrent à nous. Les critères de sélection de modèles tels que les critères BIC [Schwarz, 1978], AIC [Akaike et al., 1973], MDL [Rissanen, 1996] peuvent être utilisés. Deux techniques ont été développées, permettant d'estimer le nombre de gaussiennes optimal  $N$ .

### 3.5.5.1 Le critère MDL ("minimum description length")

L'estimation du nombre de gaussiennes optimal est défini par le minimum de la fonction de coût  $C_{\text{MDL}}$  [Figueiredo et al., 1999],

$$\hat{N} = \arg \min_N \{C_{\text{MDL}}(\hat{\Theta}_N, N); N = N_{\min}, \dots, N_{\max}\} \quad (3.42)$$

où  $\hat{\Theta}_N$  est l'ensemble des paramètres  $P_j$  et  $\sigma_j$  estimés par notre algorithme. La fonction  $C_{\text{MDL}}$  est définie par le  $-\log$  de la fonction de vraisemblance (fonction décroissante de  $N$ ) plus une fonction croissante de  $N$ , dont le rôle est de pénaliser les grandes valeurs de  $N$ . Cette formulation générale définit le critère MDL mesurant la complexité du modèle mis en jeu, et donné sous la forme suivante :

$$C_{\text{MDL}}(\hat{\Theta}_N, m) = -\ell\ell(\hat{\Theta}_N, \mathbf{x}) + \frac{\mathbf{K}(N)}{2} \log(m) \quad (3.43)$$

où  $\mathbf{x} = \{x_1, \dots, x_m\}$  sont les échantillons de nos observations,  $\ell\ell$  est le log de vraisemblance mis en jeu

$$\ell\ell(\hat{\Theta}_N, \mathbf{x}) = \sum_{i=1}^m \log \left[ \sum_{j=1}^N \phi(x_i; 0, \sigma_j^2) P_j \right] \quad (3.44)$$

$\mathbf{K}(N)$  représente le nombre de paramètres effectifs.

Dans le cas d'un mélange de gaussiennes multivariées centrées avec des moyennes et des covariances arbitraires,  $\mathbf{K}(N) = (N - 1) + N(d + d(d + 1)/2)$  où  $d$  est la dimension du vecteur aléatoire  $X = [X_1, \dots, X_d]$ . Dans notre cas,  $d = 1$  et seulement les proportions de mélange et les variances sont considérées. Il s'en suit que  $\mathbf{K}(N) = 2N - 1$  et donc,

$$C_{\text{MDL}}(N) = - \sum_{i=1}^m \log \left( \sum_{j=1}^N \phi(x_i; 0, \sigma_j^2) P_j \right) + \frac{2N - 1}{2} \log(m) \quad (3.45)$$

La Fig.3.4 montre l'évolution du critère MDL en fonction de  $N$  pour différentes valeurs de  $\alpha$ . Hormis le cas  $\alpha = 2$  (cas gaussien), on constate que le nombre de gaussiennes optimal, qui est le minimum du MDL, se situe dans l'intervalle  $[4, 8]$ . Pour une meilleure approximation, le choix de  $N = 8$  nous paraît suffisant. De plus, ce choix nous a été confirmé par un autre critère que nous présentons maintenant.

### 3.5.5.2 Simulations Monte-Carlo et divergence KL

Une autre alternative consiste à effectuer des simulations "Monte-Carlo" et à mesurer les différences entre les PDFs approchées et les PDFs exactes pour chaque valeur de  $N$ . Nous avons adopté la divergence de KL comme mesure d'écartement entre les PDFs exactes et approchées. La Fig3.5 indique l'évolution de la divergence de KL en fonction de  $N$ . On peut clairement constater que cette distance converge vers 0 très rapidement à partir de  $N \geq 8$  pour différentes valeurs de  $\alpha$  excepté le cas gaussien ( $\alpha=2$ ) où  $N = 1$ .

De ces résultats, ces différents critères plaident en faveur d'une valeur de  $N = 8$  pour assurer un compromis entre la complexité du modèle assurant une très bonne qualité de l'approximation et un temps de calcul raisonnable.

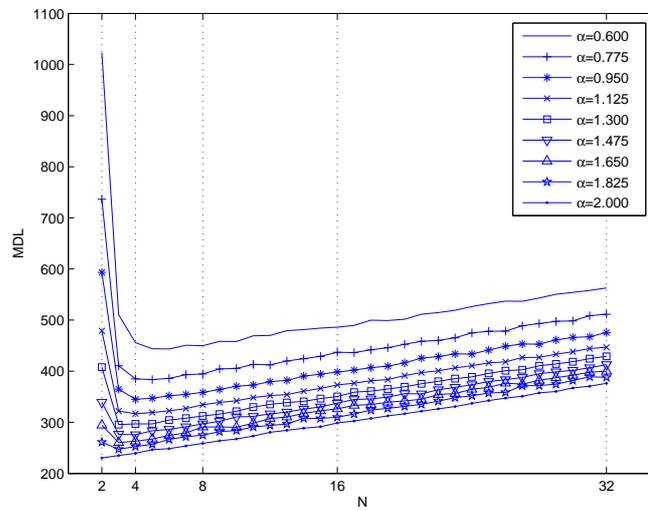


FIG. 3.4 – Évolution du critère MDL en fonction du nombre de gaussiennes  $N$  pour différentes valeurs du paramètre  $\alpha$ .

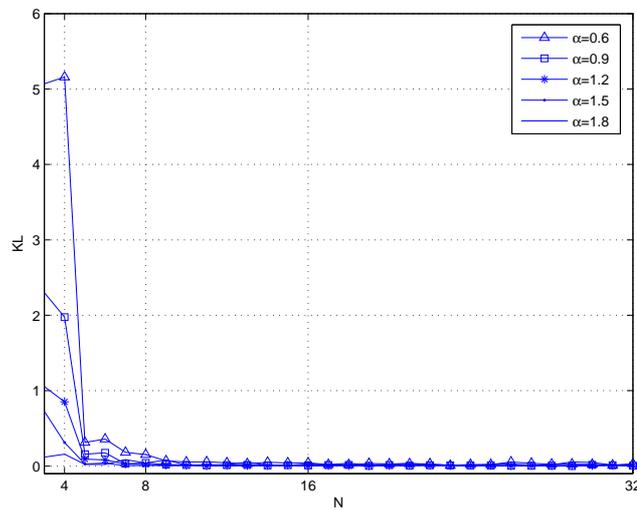


FIG. 3.5 – Évolution de la divergence de KL, calculée entre la PDF exacte et la PDF approchée par un mélange de gaussiennes, en fonction du nombre de gaussiennes pour différentes valeurs du paramètre  $\alpha$ .

### 3.5.6 Estimation des hyperparamètres

Après avoir établi une forme analytique pour la PDF  $S\alpha S$ , il reste encore à estimer ses différents paramètres lorsque l'on dispose d'échantillon de cette loi.

Dans le cadre du modèle SMG, l'estimation des paramètres  $P_j$  et  $\sigma_j$  dépend des paramètres arbitraires  $\alpha$  et  $\sigma = \gamma^{\frac{1}{\alpha}}$ . Diverses méthodes existent permettant d'estimer les paramètres  $\alpha$  et  $\sigma$  qui caractérise cet *a priori*. On peut les classer suivant les catégories suivantes :

#### 3.5.6.1 Méthodes des quantiles

Les méthodes des quantiles partent d'observations empiriques sur les queues relevées des distributions  $\alpha$ -stable. Nous introduisons ici deux méthodes pour l'estimation des paramètres.

- **Méthode de Fama-Roll** Fama & Roll [Fama & Roll, 1971] ont proposé une méthode basée sur les quantiles des échantillons empiriques. Cette méthode permet une estimation des paramètres pour une distribution  $S\alpha S$  ( $\beta = 0$  et  $\mu = 0$ ) avec  $1 < \alpha \leq 2$ . Elle propose une estimation de  $\sigma$  à l'aide de l'expression suivante :

$$\hat{\sigma} = \frac{\hat{x}_{0.72} - \hat{x}_{0.28}}{1.654} \quad (3.46)$$

où  $x_\eta$  est le quantile<sup>4</sup> à  $\eta$ . Cette estimation repose sur l'observation que  $(\hat{x}_{0.72} - \hat{x}_{0.28})/\sigma$  présente une valeur proche de 1.654 à 0.4% pour  $1 < \alpha \leq 2$  et  $\beta = 0$ . Cette expression permet une estimation de  $\sigma$  sans aucune connaissance préalable concernant  $\alpha$ . L'exposant caractéristique  $\alpha$  peut être estimé à partir de l'observation du comportement de la queue de distribution. Fama et Roll proposent de choisir  $\hat{\sigma}$  satisfaisant la relation,

$$\hat{\alpha} = \frac{\hat{x}_f - \hat{x}_{1-f}}{2\hat{\sigma}} \quad (3.47)$$

On montre que  $f$  prenant les valeurs de 0.95, 0.96, 0.97 permet une bonne estimation de  $\alpha$ . L'inconvénient de cette méthode repose sur la propagation du biais réalisé sur l'estimation de  $\sigma$  lors de l'estimation de  $\alpha$ .

- **Méthode de McCulloch** La méthode Fama-Roll, basée sur les quantiles des échantillons empiriques, n'est valide que pour les distributions symétriques et souffre d'un biais asymptotique élevé; de plus, les conditions imposées aux paramètres sont très restrictives. McCulloch [McCulloch, 1986] a développé une technique d'estimation sans biais, dérivée de celle de Fama-Roll, permettant une estimation rapide des paramètres  $\alpha$  et  $\sigma$  sous la restriction  $0.6 \leq \alpha \leq 2$  en utilisant cinq quantiles de l'échantillon.

L'estimateur de McCulloch définit :

$$\nu_\alpha = \frac{x_{0.95} - x_{0.05}}{x_{0.75} - x_{0.25}} \quad (3.48)$$

et

$$\nu_\beta = \frac{x_{0.95} + x_{0.05} - 2x_{0.50}}{x_{0.95} - x_{0.05}} \quad (3.49)$$

<sup>4</sup>Soit  $F$  la fonction de répartition d'une VA stable  $X \sim S_\alpha(\sigma, \beta, \mu)$ , et soit  $x_\eta$  le quantile d'ordre  $\eta$ , c'est-à-dire,  $F(x_\eta) = \eta$ , et  $\hat{x}_\eta$  est le quantile empirique correspondant.

où  $x_\eta$  représente le quantile d'ordre  $\eta$  et  $\hat{x}_\eta$  est le quantile empirique correspondant.

Ces deux définitions sont entièrement indépendantes de  $\sigma$  et  $\mu$ . De plus, ils sont respectivement des fonctions décroissante et croissante de  $\alpha$  et  $\beta$ . Cette relation peut être inversée et les paramètres  $\alpha$  et  $\beta$  peuvent être considérés comme des fonctions dépendant de  $\nu_\beta$  et  $\nu_\alpha$ . Nous pouvons alors écrire,

$$\hat{\alpha} = \vartheta_1(\nu_\alpha, \nu_\beta), \quad \hat{\beta} = \vartheta_2(\nu_\alpha, \nu_\beta) \quad (3.50)$$

où  $\vartheta_1$  et  $\vartheta_2$  sont des fonctions définies dans l'annexe B (cf. Tableau B-1 et B-2).

En substituant  $\nu_\alpha$  et  $\nu_\beta$  par leurs valeurs estimées à partir de l'échantillon, nous obtenons par interpolation les estimations de  $\hat{\alpha}$  et  $\hat{\beta}$  (cf. tableau B-1 et B-2 dans l'annexe B).

Nous définissons alors,

$$\nu_\sigma = \frac{x_{0.75} - x_{0.25}}{\sigma} \quad (3.51)$$

Le Tableau B-3, fourni en annexe B, caractérise le comportement de  $\nu_\sigma$  en fonction de  $\vartheta_3(\alpha, \beta)$ . Nous obtenons alors une estimation de  $\sigma$ ,

$$\hat{\sigma} = \frac{\hat{x}_{0.75} - \hat{x}_{0.25}}{\vartheta_3(\hat{\alpha}, \hat{\beta})} \quad (3.52)$$

où  $\vartheta_3$  est la fonction définie dans l'annexe B (cf. tableau B-3).

### 3.5.6.2 Méthodes basées sur la fonction caractéristique

L'expression de la fonction caractéristique empirique est la suivante

$$\hat{\psi}(\omega) = \frac{1}{m} \sum_{j=1}^m \exp\{i\omega X_j\} \quad (3.53)$$

A partir de la loi des grands nombres,  $\hat{\psi}(\omega)$  est un estimateur consistant de la fonction caractéristique théorique  $\psi(\omega)$ . Nous introduisons deux méthodes d'estimation basées sur cette expression.

• **Méthode  $M_1$**  La méthode des moments a été proposée par Press [Press, 1972] et repose sur des manipulations de la fonction caractéristique.

Pour tout  $\alpha$ ,

$$|\psi(\omega)| = \exp(-\sigma^\alpha |\omega|^\alpha) \quad (3.54)$$

nous obtenons alors,

$$-\log |\psi(\omega)| = \sigma^\alpha |\omega|^\alpha \quad (3.55)$$

Deux cas sont à considérer,  $\alpha \neq 1$  et  $\alpha = 1$ . Dans le premier cas, nous choisissons deux valeurs de  $\omega$  non nulles telles que  $\omega_1 \leq \omega_2$ . Nous obtenons alors, pour  $k = 1, 2$  :

$$-\log |\psi(\omega_k)| = \sigma^\alpha |\omega_k|^\alpha \quad (3.56)$$

On résout le système d'équation en  $\alpha$  et  $\sigma$ , où l'on remplace  $\psi(\omega)$  par  $\hat{\psi}(\omega)$  pour obtenir :

$$\hat{\alpha} = \frac{\log \left( \frac{\log |\hat{\psi}(\omega_1)|}{\log |\hat{\psi}(\omega_2)|} \right)}{\log \left| \frac{\omega_1}{\omega_2} \right|} \quad (3.57)$$

et

$$\log(\hat{\sigma}) = \frac{\log |\omega_1| \log \left( -\log |\hat{\psi}(\omega_2)| \right) - \log |\omega_2| \log \left( -\log |\hat{\psi}(\omega_1)| \right)}{\log \left| \frac{\omega_1}{\omega_2} \right|} \quad (3.58)$$

Dans le cas où  $\alpha = 1$ , nous obtenons alors une expression très simple de l'estimée de  $\sigma$ ,

$$\hat{\sigma} = -\frac{\log |\hat{\psi}(\omega_1)|}{|\omega_1|} \quad (3.59)$$

La méthode des moments est facile à implémenter et elle est très efficace en temps de calcul mais sensible au nombre d'échantillons et au choix de la plage de  $\omega$ .

• **Méthode par Régression** Koutrouvelis [Koutrouvelis, 1980] a présenté une méthode de type régression qui construit une expression linéaire à partir de la fonction caractéristique et les paramètres  $\alpha$  et  $\sigma$ . L'estimation des paramètres  $\alpha$  et  $\sigma$  est réalisée par régression de  $x = \log \left( -\log |\psi(\omega)|^2 \right)$  sur  $w = \log |\omega|$  en utilisant le modèle suivant,

$$x_k = m + \alpha w_k + \epsilon_k, \quad k = 1, \dots, K \quad (3.60)$$

où  $m = \log(2\sigma^\alpha)$ , et  $\epsilon_k$  correspond à un terme d'erreur. Koutrouvelis propose l'utilisation de  $\omega_k = k\pi/25$  pour  $k = 1, 2, \dots, K$ ;  $K$  prenant ses valeurs entre 9 et 134 pour différentes valeurs estimées de  $\alpha$  et de taille de l'échantillon. Weron [Weron, 1995] propose une simplification du choix du coefficient  $K$  sans perte significative de performance (cf. tableau suivant).

TAB. 3.1 – Choix de  $K$  en fonction de  $\alpha$ .

$\alpha$	$K$
1.5-2.0	10
0.6-1.5	20
0.4-0.6	60
< 0.4	120

L'algorithme suivant présente une procédure d'estimation des paramètres  $\alpha$  et  $\sigma$  pour une distribution  $S\alpha S$  sous forme récursive.

---

**Algorithme 2** Procédure d'estimation des paramètres  $\alpha$  et  $\sigma$  pour une distribution  $S\alpha S$

---

- 1: Un niveau d'erreur admissible est fixé ( $0 < \epsilon < 1$ ) ainsi que le nombre d'itérations maximum.
  - 2: Les valeurs initiales de  $\alpha$  et  $\sigma$ , soient respectivement  $\hat{\alpha}_0$  et  $\hat{\sigma}_0$ , sont établies à l'aide de la méthode de McCulloch ou de Fama-Roll.
  - 3: Nous posons :  $\hat{\alpha} = \hat{\alpha}_0$  et  $\hat{\sigma} = \hat{\sigma}_0$ .
  - 4: Fixer la valeur de  $K$  (choisie à partir de la table 3.1), et pour tous les points  $\omega_k, k = 1, 2, \dots, K$  calculer la fonction caractéristique empirique  $\hat{\psi}(\omega_k)$
  - 5: Calculer  $x_k = \log \left( -\log |\hat{\psi}(\omega_k)|^2 \right)$
  - 6: Ajuster la régression linéaire  $x_i = m + \alpha w_k + \epsilon_k$ , pour obtenir les estimés de  $\hat{\alpha}$  et  $\hat{m}$
  - 7: Obtenir l'estimateur de  $\hat{\sigma}$  à partir de  $m = \log(2\hat{\sigma}^{\hat{\alpha}})$
  - 8: Si  $|\hat{\sigma} - 1| \geq \epsilon$  et le nombre d'itérations est inférieur au nombre d'itérations maximum, alors répéter les étapes 4 à 8.
-

Cette méthode est très facile à implémenter et elle est très efficace en temps de calcul.

### 3.5.6.3 Maximum de vraisemblance

Le point clé pour appliquer la méthode de MV est le calcul de la densité d'une variable aléatoire stable. Cependant, aucune forme analytique n'est disponible pour l'estimateur de MV et on a recours à des procédures de minimisation numérique extrêmement pénibles et coûteuses [Nolan, 1997]. Ces inconvénients majeurs rendent cette possibilité inutile en pratique.

### 3.5.6.4 Méthodes des moments fractionnaires

Il est clair que d'après la propriété 3.25, il n'est malheureusement pas possible d'avoir recours à des estimateurs des hyperparamètres basés sur les moments, puisque ces derniers ne sont finis que pour  $0 < p < \alpha$ . Ma & Nikias [Ma & Nikias, 1995] ont proposé une classe d'estimateurs basée sur les moments fractionnaires d'ordre faible ( $< \alpha$ ). Tsihrintzis et Nikias [Tsihrintzis & Nikias, 1996] ont aussi proposé une technique d'estimation basée sur les moments d'ordre extrême (grand). Cependant, ces estimateurs souffrent de sérieuses limitations (biais important) à faible nombre d'échantillons.

### 3.5.6.5 Méthodes basées sur le comportement asymptotique des queues

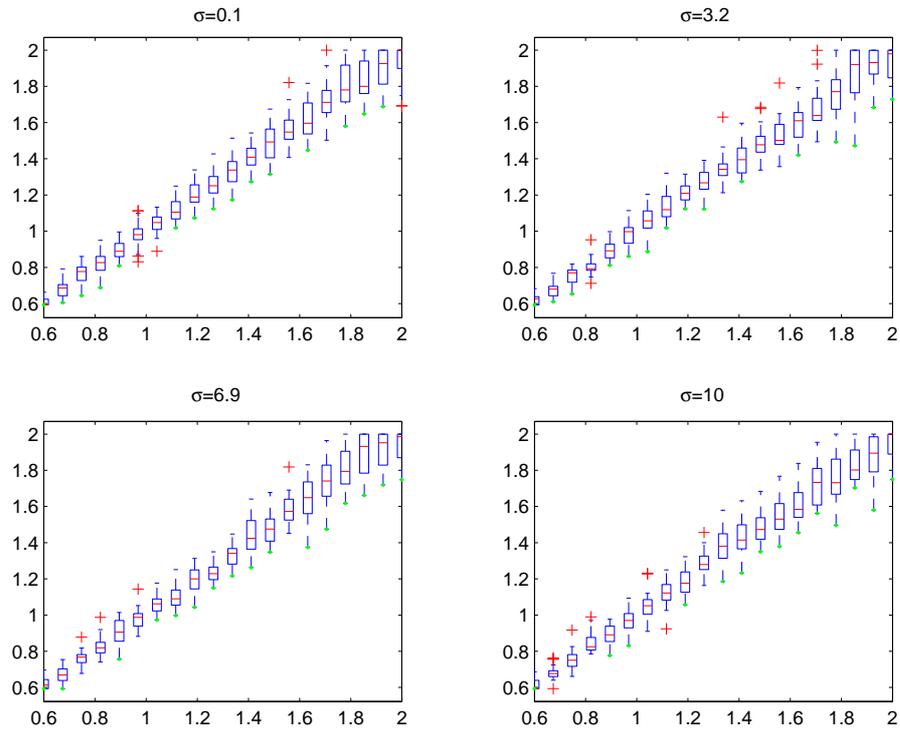
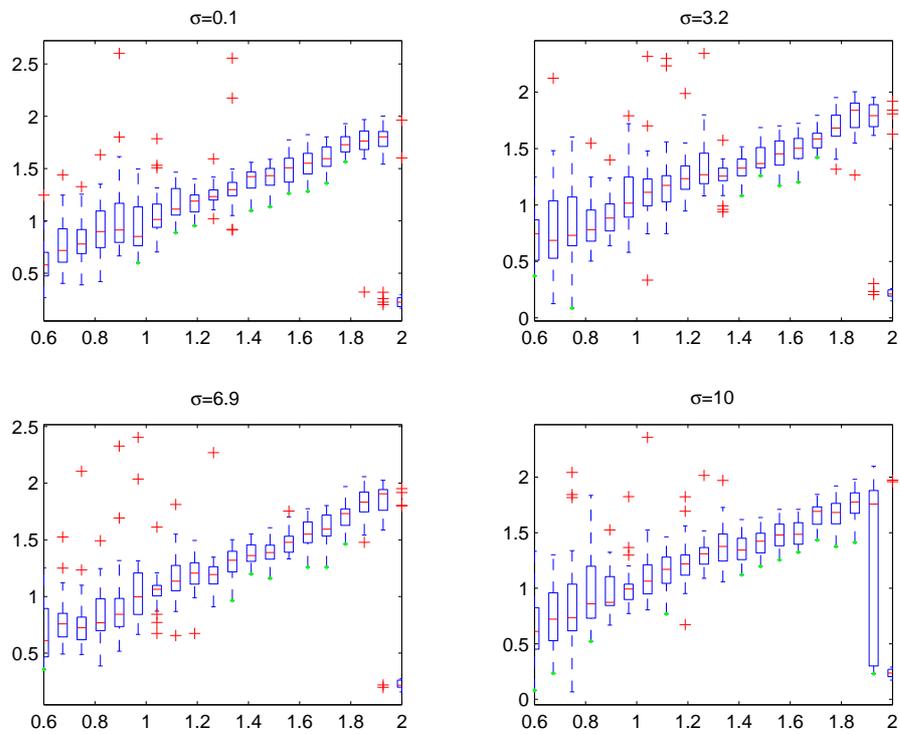
Celles-ci sont fondées sur la décroissance géométrique des distributions  $\alpha$ -stables lorsque  $|x| \rightarrow +\infty$ . Une estimation de  $\alpha$  est directement déduite de la pente de ces asymptotes en échelle  $\log - \log$ . De nombreuses difficultés s'opposent à l'implémentation de cette méthode. Citons par exemple la plage des  $x$  suffisamment grands (infinis!) que l'on doit utiliser.

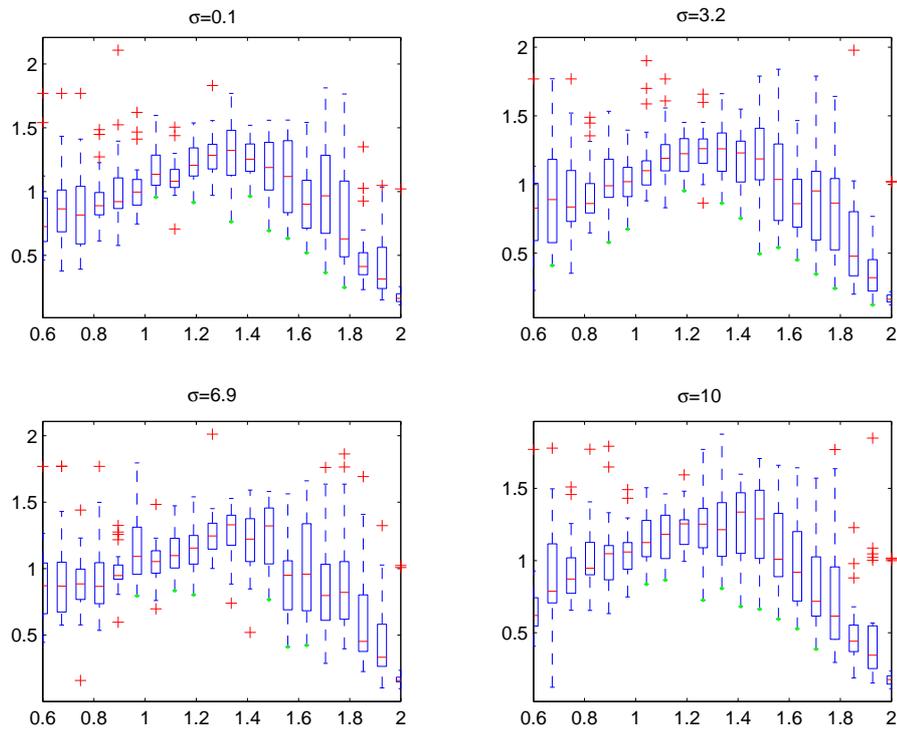
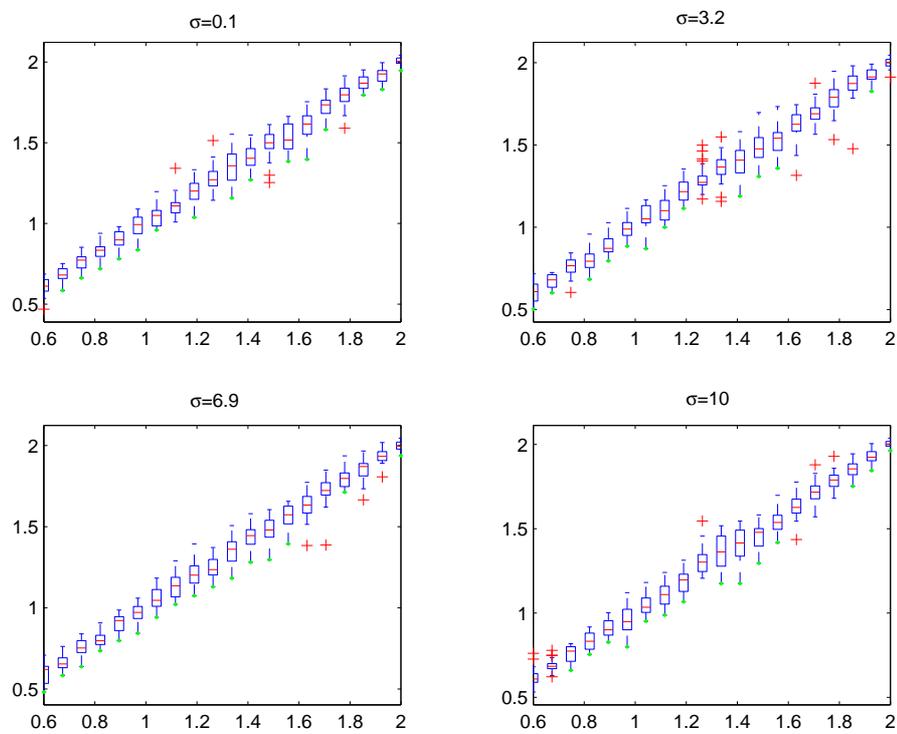
## 3.5.7 Performance des différents estimateurs

Certaines méthodes d'estimation des hyperparamètres ont été testées à l'aide de simulations numériques dans [Weron, 1995, Bates & McLaughlin, 2000]. Les méthodes des quantiles et celles basées sur la fonction caractéristique présentent des performances similaires et s'avèrent supérieures en pratique aux méthodes des moments fractionnaires.

Nous avons effectué des simulations numériques avec plus d'estimateurs afin de tester leurs performances, notamment les estimateurs des quantiles et ceux basés sur la fonction caractéristique citées précédemment en détail avec leurs procédures d'estimation. L'échantillon a été simulé à l'aide de l'algorithme de Weron [Weron, 1995]. Les résultats sont présentés par les Fig.3.10,3.11, 3.8 et 3.9.

Les méthodes des quantiles de Fama-Roll et des moments s'avèrent inférieures aux méthodes par régression et de McCulloch. Les méthodes de McCulloch et par régression présentent des performances similaires à l'exception des valeurs de  $\alpha$  proches de 2. Pour de telles valeurs, où la queue de distribution est relativement réduite, la méthode par régression s'avère légèrement plus robuste que celle de McCulloch du fait de la correction réalisée dans le contexte récursif.

FIG. 3.6 – Estimation du paramètre  $\alpha$  par la méthode de McCulloch pour  $\sigma = 0.1 \ 3.2 \ 6.9 \ 10$ FIG. 3.7 – Estimation du paramètre  $\alpha$  par la méthode de Fama-Roll pour  $\sigma = 0.1 \ 3.2 \ 6.9 \ 10$

FIG. 3.8 – Estimation du paramètre  $\alpha$  par la méthode  $M_1$  pour  $\sigma = 0.1 \ 3.2 \ 6.9 \ 10$ FIG. 3.9 – Estimation du paramètre  $\alpha$  par la méthode par régression pour  $\sigma = 0.1 \ 3.2 \ 6.9 \ 10$

## 3.6 Cas 2 : a priori des Formes K de Bessel

### 3.6.1 Définition

Soit  $X = g_F$  une version filtrée d'une image  $g$  par un filtre passe-bande  $F$ . En utilisant le modèle du générateur déplacé (voir 2.3.2.1.2), il a été montré par [Grenander & Srivastava, 2001] que la PDF de  $X = g_F$  peut s'exprimer grâce la Forme K de Bessel (BKF) comme suit,

$$f_X(x) = \frac{1}{\sqrt{\pi}\Gamma(\alpha)} \left(\frac{c}{2}\right)^{-\frac{\alpha}{2}-\frac{1}{4}} \left|\frac{x}{2}\right|^{\alpha-\frac{1}{2}} K_{\alpha-\frac{1}{2}}\left(\sqrt{\frac{2}{c}}|x|\right) \quad (3.61)$$

où  $K_{\alpha-\frac{1}{2}}$  est la fonction de Bessel modifiée de première espèce d'ordre  $\alpha-\frac{1}{2}$  (voir [Abramowitz & Stegun, 1972] et [Gradshteyn & Ryzhik, 1980])

$$K_\nu(zx) = \frac{\Gamma(\nu + \frac{1}{2})(2x)^\nu}{\sqrt{\pi}z^\nu} \int_0^{+\infty} \frac{\cos(zt)dt}{(t^2 + x^2)^{\nu+\frac{1}{2}}} \quad (3.62)$$

$$\left(Re(\nu) > -\frac{1}{2}, z > 0, |\arg x| < \frac{\pi}{2}\right)$$

$\alpha > 0$  et  $c > 0$  représentent respectivement le paramètre de forme et d'échelle. La fonction caractéristique correspondante est donnée par

$$\psi(\omega) = \left(1 + \frac{c\omega^2}{2}\right)^{-\alpha} \quad (3.63)$$

Cette PDF est caractérisée par deux paramètres : un paramètre de forme  $\alpha$  et un paramètre d'échelle  $c$ . Pour  $\alpha = 1$ ,  $f$  est une PDF double exponentielle. Plus généralement,  $f_X$  est la  $\alpha^{\text{ième}}$  convolution des PDFs doubles exponentielles. Si  $\alpha > 1$ , on retrouve le cas gaussien (particulièrement quand  $\alpha \gg 1$ ) ce qui est intuitivement acceptable en utilisant l'argument du théorème central limite (une forme générale sera donnée plus tard dans la proposition 5.2). Si  $\alpha < 1$ , la forme de la PDF aura un pic plus prononcé avec des queues très lourdes.

### 3.6.2 Propriétés

Nous présentons ici quelques propriétés qui caractérisent la loi BKF, et qui sont utiles dans la pratique.

#### 3.6.2.1 Quelques propriétés fondamentales

- La PDF est unimodale et symétrique autour du mode.
- Les cumulants d'une VA BKF ont la forme suivante :

$$\kappa_{2i} = \alpha \left(\frac{c}{2}\right)^i \frac{(2i)!}{i}, i \geq 1 \quad (3.64)$$

Tous les cumulants d'ordre impair sont nuls, alors que ceux d'ordre pair sont non nuls.

- Soit  $X \sim BKF(\alpha, c)$ , la variance et le kurtosis de  $X$  sont donnés par

$$\text{Var}(X) = \kappa_2 = \alpha c \quad (3.65)$$

$$\text{Kurt}(X) = \frac{\kappa_4}{\kappa_2^2} + 3 = \frac{3}{\alpha} + 3 > 0 \quad (3.66)$$

où  $\kappa_i$  est le cumulatif d'ordre  $i$ .

- si  $X \sim BKF(\alpha, c)$  alors  $\frac{X}{\sigma} \sim BKF\left(\alpha, \frac{c}{\sigma^2}\right)$ . Ceci peut être démontré facilement.

Les résultats du calcul numérique des PDFs BKF pour différentes valeurs du paramètre de forme  $\alpha$  et d'échelle  $c$  sont illustrés par les Fig 3.10 et Fig 3.11. Les Fig 3.10(a) et Fig 3.10(b) montrent les diverses formes prises par les PDFs en faisant varier le paramètre  $\alpha$ . La Fig.3.11 illustre la densité lorsqu'on fait varier le paramètre  $c$ .

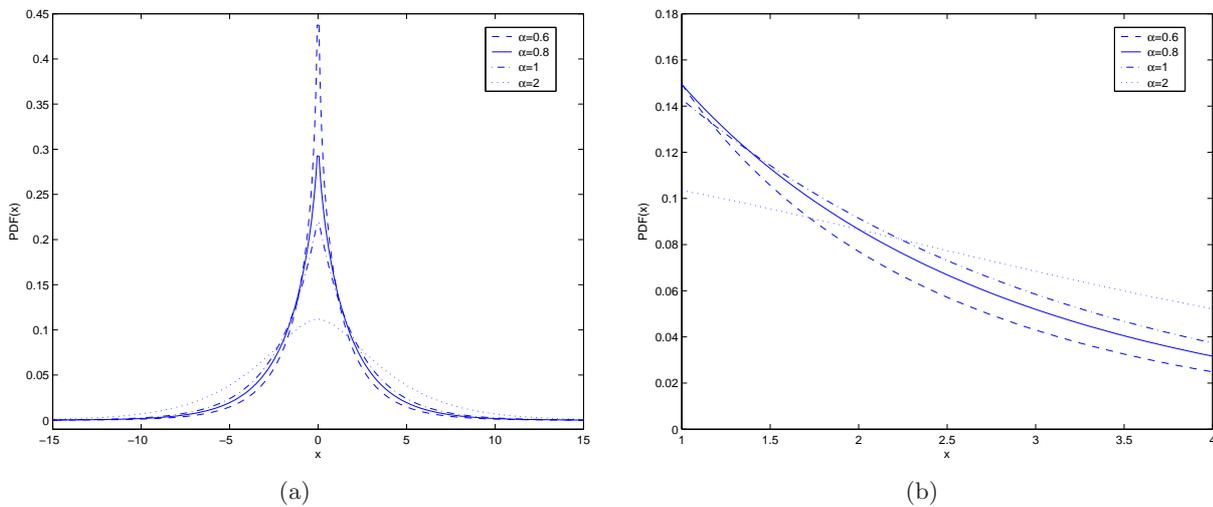


FIG. 3.10 – (a) PDFs des lois BKF de paramètres  $\alpha \in \{0.5, 0.75, 1, 2\}$  avec  $c = 10$ . (b) Évolution de la queue de distribution avec le paramètre de forme  $\alpha$ .

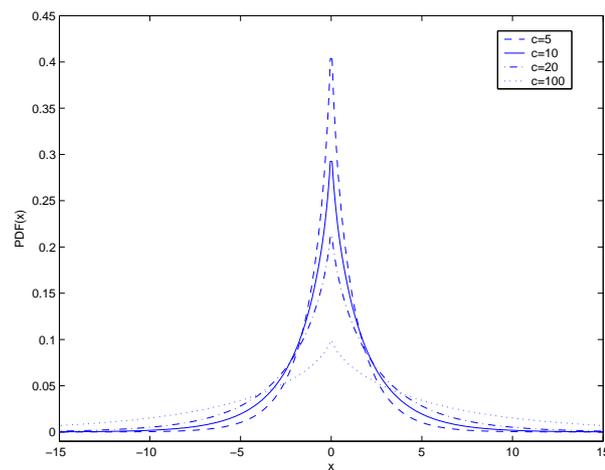


FIG. 3.11 – L'effet du paramètre d'échelle  $c$  avec  $\alpha = 0.8$ .

### 3.6.2.2 Moments d'ordre $p$

**Proposition 3.6** Soit  $Z \sim \mathcal{N}(0, \sigma^2)$  et  $U \sim \Gamma(\alpha)$ . Il existe une VA BKF  $X \sim BKF(\alpha, \frac{1}{\alpha})$  telle que :

$$X \stackrel{d}{=} Z\sqrt{U} \quad (3.67)$$

où  $X|u \sim \mathcal{N}(0, u)$ . Rappelons que la loi BKF est un cas particulier du modèle SMG. Le moment d'ordre  $p$  de la VA  $X$  est donnée par :

$$M_X(p) = \frac{\sigma^p}{\sqrt{\pi}} \left( \sqrt{\frac{2}{\alpha}} \right)^p \frac{\Gamma(\frac{p}{2} + \alpha)\Gamma(\frac{p}{2} + \frac{1}{2})}{\Gamma(\alpha)} \quad (3.68)$$

**Preuve:**

On a

$$\mathbb{E}[|X|^p] = \mathbb{E}_U[\mathbb{E}[X^p|U]] = \nu_p \mathbb{E}_U[U^{\frac{p}{2}}] = \nu_p M_U\left(\frac{p}{2}\right) \quad (3.69)$$

et donc,

$$M_X(p) = \nu_p M_U\left(\frac{p}{2}\right) \quad (3.70)$$

Pour calculer  $M_U\left(\frac{p}{2}\right)$ , on a par définition

$$U \sim \Gamma(\alpha) \implies f_U(u) = \frac{1}{c\Gamma(\alpha)} \left(\frac{u}{c}\right)^{\alpha-1} e^{-\frac{u}{c}} \quad (3.71)$$

avec  $c = \frac{1}{\alpha} > 0$ ,  $0 < \alpha < +\infty$  où

$$\begin{aligned} M_U\left(\frac{p}{2}\right) &= \frac{1}{c\Gamma(\alpha)} \int_0^{+\infty} |u|^{\frac{p}{2}} \left(\frac{u}{c}\right)^{\alpha-1} e^{-\frac{u}{c}} du \\ &= \frac{1}{c^\alpha \Gamma(\alpha)} \int_0^{+\infty} |u|^{\frac{p}{2} + \alpha - 1} e^{-\frac{u}{c}} du \end{aligned} \quad (3.72)$$

En opérant le changement de variable  $v = \frac{u}{c}$ , on peut écrire à partir de l'Eq.3.72 :

$$\begin{aligned} M_U\left(\frac{p}{2}\right) &= \frac{c^{\frac{p}{2} + \alpha}}{c^\alpha \Gamma(\alpha)} \int_0^{+\infty} |v|^{\frac{p}{2} + \alpha - 1} e^{-v} dv \\ &= \frac{c^{\frac{p}{2}}}{\Gamma(\alpha)} \Gamma\left(\frac{p}{2} + \alpha\right) \\ &= \left(\frac{1}{\sqrt{\alpha}}\right)^p \frac{\Gamma\left(\frac{p}{2} + \alpha\right)}{\Gamma(\alpha)} \end{aligned} \quad (3.73)$$

De plus,

$$\nu_p = \frac{2^{\frac{p}{2}}}{\sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right) \sigma^p \quad (3.74)$$

La substitution de l'Eq.3.73 et Eq.3.74 dans l'Eq.3.70 permet d'obtenir le résultat énoncé. ■

### 3.6.3 L'a priori BKF et l'espace de Besov

Une relation explicite, entre les paramètres du modèle BKF et l'espace de Besov, est donnée par la proposition suivante :

**Corollaire 3.2** Soit  $x_{j,k} \sim BKF(\alpha, \sigma_j)$  (nouvelle paramétrisation avec  $\sigma^2 = \alpha c$ ), où  $0 < \alpha \leq 1$  et  $\sigma_j = \sigma_0 2^{-j\beta}$  (propriété d'invariance d'échelle), avec  $(\sigma_0 > 0, \beta \geq 0)$ . On a alors,  $g \in B_{p,q}^s$  ssi  $\beta > (s + \frac{1}{2})$  pour  $1 \leq p < \infty$  et  $1 \leq q \leq \infty$ .

**Preuve:**

Ce corollaire s'élargit du théorème 3.2 car le cas particulier de la loi BKF est obtenu lorsque la variable  $U$  suit une loi Gamma.

Notons que l'Eq. 3.15 dans la preuve devient comme suit :

$$M_0(p) = \frac{1}{\sqrt{\pi}} \left( \sqrt{\frac{2}{\alpha}} \sigma_0 \right)^p \frac{\Gamma(\frac{p}{2} + \alpha) \Gamma(\frac{p}{2} + \frac{1}{2})}{\Gamma(\alpha)} \quad (3.75)$$

et donc, le moment d'ordre  $p$  de la distribution de probabilité est donné par

$$M_j(p) = 2^{j(1-\beta p)} M_0(p) \quad (3.76)$$

■

Finalement, le modèle BKF est adapté pour modéliser la classe des images appartenant à l'espace de Besov.

### 3.6.4 Estimation des hyperparamètres

La PDF des lois BKF est caractérisée par deux paramètres : un paramètre de forme  $\alpha$  et un paramètre d'échelle  $c$ . Nous avons proposé deux estimateurs pour l'estimation de ces paramètres, basés sur la méthode de cumulants et l'algorithme EM.

#### 3.6.4.1 Méthode des cumulants

L'idée de cette approche est basée sur les statistiques de cumulants d'ordre  $i$  d'une VA BKF, définis par l'Eq.3.64, pour l'estimation des hyperparamètres de la loi *a priori* BKF.

Soit  $X \sim BKF(\alpha, c)$ . A partir de l'Eq.3.65 et l'Eq.3.66, les paramètres  $\alpha$  et  $c$  peuvent être estimés par les cumulants d'ordre 2 et 4 de la VA  $X$  selon

$$\hat{\alpha} = \frac{3}{\text{Kurt}(X) - 3} \quad (3.77)$$

et

$$\hat{c} = \frac{\text{Var}(X)}{\hat{\alpha}} \quad (3.78)$$

où le kurtosis et la variance peuvent être aussi estimés par des formules statistiques classiques. Cependant, l'estimation de ces quantités statistiques pour un faible nombre d'échantillons  $m$  provoque un biais important. Ce cas de figure se produit par exemple pour les coefficients de la transformée d'ondelettes décimée d'une image vers l'échelle de décomposition grossière. Afin de pallier ce problème, nous proposons un estimateur sans biais pour estimer ces cumulants en utilisant les statistiques- $k$  [Fisher, 1928, Rose & Smith, 2002]. Les statistiques- $k$  de la variance et du Kurtosis sont définis comme suit,

$$\hat{\kappa}_2 = \frac{m}{m-1} \hat{M}_2 \quad (3.79)$$

et

$$\hat{\kappa}_4 = \frac{m^2 \left[ (m+1) \hat{M}_4 - 3(m-1) \hat{M}_2^2 \right]}{(m-1)(m-2)(m-3)} \quad (3.80)$$

où  $\hat{M}_i$  sont les moments estimés d'ordre  $i$ .

### 3.6.4.2 Algorithme EM

Le principe de cette méthode est de reconstruire les données manquantes aux observations  $X \sim BKF(\alpha, c)$  et ensuite d'estimer les hyperparamètres  $\theta = \{\alpha, c\}$ .

Dans l'Eq.3.1, nous supposons  $U$  les données cachées et  $(X, U)$  les données complètes. L'algorithme suivant présente une procédure d'estimation des paramètres  $\alpha$  et  $c$  alternant entre deux étapes.

---

**Algorithme 3** Algorithme EM pour l'estimation des hyperparamètres d'une distribution BKF

---

**Entrées:** on part d'une estimation initiale  $\theta^{(0)}$

- 1: **répéter**
- 2: **Étape E :**  
On calcule,

$$M_U^{(t)}(p) = \mathbb{E}[U^p | x] = \left( \frac{cx^2}{2} \right)^{\frac{p}{2}} \frac{K_{p+\alpha-\frac{1}{2}} \left( \sqrt{\frac{2}{c}} |x| \right)}{K_{\alpha-\frac{1}{2}} \left( \sqrt{\frac{2}{c}} |x| \right)} \quad (3.81)$$

$$\varphi_1^{(t)} = \mathbb{E}[\log(U) | x] = \log(c) + \frac{1}{2} \log \left( \frac{x^2}{2c} \right) + \frac{\partial_\alpha K_{\alpha-\frac{1}{2}} \left( \sqrt{\frac{2}{c}} |x| \right)}{K_{\alpha-\frac{1}{2}} \left( \sqrt{\frac{2}{c}} |x| \right)} \quad (3.82)$$

- 3: **Étape M :**  
Les estimés de  $\alpha$  et  $c$  sont comme suit,

$$c^{(t+1)} = \frac{1}{\alpha^{(t+1)}} M_U^{(t)}(1) \quad (3.83)$$

où  $\alpha^{(t+1)}$  est la solution de l'équation suivante :

$$\Psi(\alpha) - \log(\alpha) = \frac{1}{m} \sum_{i=1}^m \varphi_1^{(t)}(i) - \log \left( M_U^{(t)}(1) \right) \quad (3.84)$$

$m$  est le nombre d'échantillons.

- 4: **jusqu'à** Convergence.
- 

#### Preuve:

Soit  $X \sim BKF(\alpha, c)$ . Nous rappelons que sous le modèle SMG, il existe deux VAs indépendantes  $U \sim \Gamma(\alpha, c)$  et  $Z \sim \mathcal{N}(0, 1)$  telles que  $X \stackrel{d}{=} Z\sqrt{U}$ .

Le calcul du score de vraisemblance

$$Q(\theta|\theta^{(t)}) = \mathbb{E}[\ell\ell(X|U; \theta)|X, \theta^{(t)}] + \mathbb{E}[\ell\ell(U|\theta)|X, \theta^{(t)}] \quad (3.85)$$

nécessite le calcul de l'anti-log de vraisemblance suivant :

$$\ell\ell(X|U) + \ell\ell(U) = \frac{1}{2} \sum_{i=1}^m \log u_i + \frac{1}{2} \sum_{i=1}^m \frac{x_i^2}{u_i} + m \log c + m \log \Gamma(\alpha) + (1 - \alpha) \sum_{i=1}^m \log \frac{u_i}{c} + \sum_{i=1}^m \frac{u_i}{c}$$

où  $m$  est le nombre d'échantillons.

Et donc,  $Q$  peut s'écrire comme suit,

$$\begin{aligned} Q(\theta|\theta^{(t)}) &= \frac{1}{2} \sum_{i=1}^m \varphi_{1(i)} + \frac{1}{2} \sum_{i=1}^m \varphi_{2(i)} x_i^2 + m \log c + N \log \Gamma(\alpha) + (1 - \alpha) \sum_{i=1}^m \varphi_{1(i)} \\ &\quad - (1 - \alpha) m \log c + \sum_{i=1}^m \frac{\varphi_{3(i)}}{c} \\ &= \left( \frac{3}{2} - \alpha \right) \sum_{i=1}^m \varphi_{1(i)} + \frac{1}{2} \sum_{i=1}^m \varphi_{2(i)} x_i^2 + \sum_{i=1}^m \frac{\varphi_{3(i)}}{c} + \alpha m \log c + m \log \Gamma(\alpha) \end{aligned} \quad (3.86)$$

où nous notons les statistiques conditionnelles suffisantes des données manquantes

$$\varphi_{1(i)} = \mathbb{E}[\log(U)|x_i]; \quad \varphi_{2(i)} = \mathbb{E}[U^{-1}|x_i]; \quad \varphi_{3(i)} = \mathbb{E}[U|x_i] \quad \text{pour } i = 1, \dots, m \quad (3.87)$$

Le calcul de  $\varphi_2$  et  $\varphi_3$  est effectué par le moment d'ordre  $p$  de la distribution de probabilité de  $U$  défini comme suit,

$$\begin{aligned} M_U(p) = \mathbb{E}[U^p|x] &= \frac{\int_0^\infty u^p f_{X|U}(x|u) f_U(u) du}{\int_0^\infty f_{X|U}(x|u) f_U(u) du} \\ &= \frac{\int_0^\infty u^{p+\alpha-\frac{3}{2}} \exp\left(\frac{x^2}{2u} - \frac{u}{c}\right) du}{\int_0^\infty u^{\alpha-\frac{3}{2}} \exp\left(\frac{x^2}{2u} - \frac{u}{c}\right) du} \end{aligned} \quad (3.88)$$

où  $\varphi_3$  (resp.  $\varphi_2$ ) est le moment d'ordre 1 de la distribution de probabilité de  $U$  (resp. de  $U^{-1}$ ).

$\varphi_1$  est calculé comme suit,

$$\begin{aligned} \varphi_1 = \mathbb{E}[\log(U)|x] &= \frac{\int_0^\infty \log(u) f_{X|U}(x|u) f_U(u) du}{\int_0^\infty f_{X|U}(x|u) f_U(u) du} \\ &= \frac{\int_0^\infty \log(u) u^{\alpha-\frac{3}{2}} \exp\left(\frac{x^2}{2u} - \frac{u}{c}\right) du}{\int_0^\infty u^{\alpha-\frac{3}{2}} \exp\left(\frac{x^2}{2u} - \frac{u}{c}\right) du} \end{aligned} \quad (3.89)$$

Pour établir le résultat final de ces intégrales, nous avons besoin de la formule d'intégrale de [Gradshteyn & Ryzhik, 1980] (page 340, Eq.9).

Après calcul, nous obtenons

$$M_U(p) = \left(\frac{cx^2}{2}\right)^{\frac{p}{2}} \frac{K_{p+\alpha-\frac{1}{2}}\left(\sqrt{\frac{2}{c}}|x|\right)}{K_{\alpha-\frac{1}{2}}\left(\sqrt{\frac{2}{c}}|x|\right)} \quad (3.90)$$

et

$$\varphi_1 = \log(c) + \frac{1}{2} \log\left(\frac{x^2}{2c}\right) + \frac{\partial_\alpha K_{\alpha-\frac{1}{2}}\left(\sqrt{\frac{2}{c}}|x|\right)}{K_{\alpha-\frac{1}{2}}\left(\sqrt{\frac{2}{c}}|x|\right)} \quad (3.91)$$

d'où les résultats donnés par l'étape E de l'algorithme.

Notons que ces intégrales convergent et peuvent être calculées rapidement par des méthodes de quadrature.

L'étape M de l'algorithme permet d'estimer les hyperparamètres  $\theta = \{\alpha, c\}$  par

$$\theta^{(t+1)} = \arg \min_{\theta \in \Theta} Q(\theta|\theta^{(t)}) \quad (3.92)$$

tout en respectant la propriété fondamentale  $Q(\theta|\theta^{(t+1)}) < Q(\theta|\theta^{(t)})$ .

Ce qui implique que

$$\nabla_\theta Q(\theta|\theta^{(t)}) = 0 \iff \begin{cases} \partial_\alpha Q(\theta|\theta^{(t)}) = 0 \\ \partial_c Q(\theta|\theta^{(t)}) = 0 \end{cases} \quad (3.93)$$

Le calcul de la dérivée partielle de  $Q$  en fonction de  $\alpha$  et  $c$  est comme suit :

$$\partial_\alpha Q = m\Psi(\alpha) - \sum_{i=1}^m \varphi_{1(i)} + m \log c = 0 \quad (3.94)$$

$$\partial_c Q = \frac{\alpha m}{c} - \frac{1}{c^2} \sum_{i=1}^m \varphi_{3(i)} = 0 \quad (3.95)$$

où  $\Psi$  est la dérivée de la fonction  $\log \Gamma(\cdot)$ .

En évaluant Eq.3.95, nous obtenons

$$c^{(t+1)} = \frac{1}{m\alpha^{(t)}} \sum_{i=1}^m \varphi_{3(i)}^{(t)} = \frac{1}{\alpha^{(t)}} M_U^{(t)}(1) \quad (3.96)$$

d'où le résultat de l'Eq.3.83.

En substituant l'Eq.3.96 dans l'Eq.3.94, nous obtenons

$$\partial_\alpha Q = \Psi(\alpha) - \frac{1}{m} \sum_{i=1}^m \varphi_{1(i)}^{(t)} - \log(\alpha) + \log\left(M_U^{(t)}(1)\right) = 0 \quad (3.97)$$

où  $\alpha^{(t+1)}$  est la solution de l'équation suivante

$$\Psi(\alpha) - \log(\alpha) = \frac{1}{m} \sum_{i=1}^m \varphi_{1(i)}^{(t)} - \log\left(M_U^{(t)}(1)\right) \quad (3.98)$$

d'où le résultat énoncé par l'Eq.3.84 de l'algorithme.

Notons que la solution de l'Eq.3.84 est unique puisque la fonction  $\Psi(\alpha) - \log(\alpha)$  est bijective et est strictement négative sur  $\mathbb{R}^+$ , et le membre de droite de l'Eq.3.98 est toujours négatif par l'inégalité de Jensen. ■

**Proposition 3.7** *L'algorithme ci-dessus converge en norme vers un point stationnaire.*

**Preuve:**

Il suffit de remarquer la continuité du score de vraisemblance  $Q$  par rapport à  $\theta$  et  $\theta^{(t)}$ . De plus, la solution de l'Eq.3.84 est unique puisque  $\Psi(\alpha) - \log(\alpha)$  est bijective sur  $\mathbb{R}^+$  et strictement négative. Le résultat suit le théorème classique de convergence de l'algorithme EM cyclique [Meng & Rubin, 1993]. ■

### 3.6.5 Performance des différents estimateurs

Nous avons effectué une simulation numérique afin de tester les performances de ces deux types d'estimateurs : méthodes de cumulants et l'algorithme EM.

Dans la Fig.3.12, nous constatons suite à l'observation des courbes que les estimés des paramètres  $\alpha$  et  $c$  par l'algorithme EM sont beaucoup plus proches et moins dispersés par rapport aux valeurs initiales de la simulation comparés aux estimés de la méthode de cumulants. Ceci montre que l'estimation par l'algorithme EM est meilleure comparée à celle de la méthode de cumulants, les différences deviennent très notables lorsque  $\alpha$  et  $c$  sont grands.

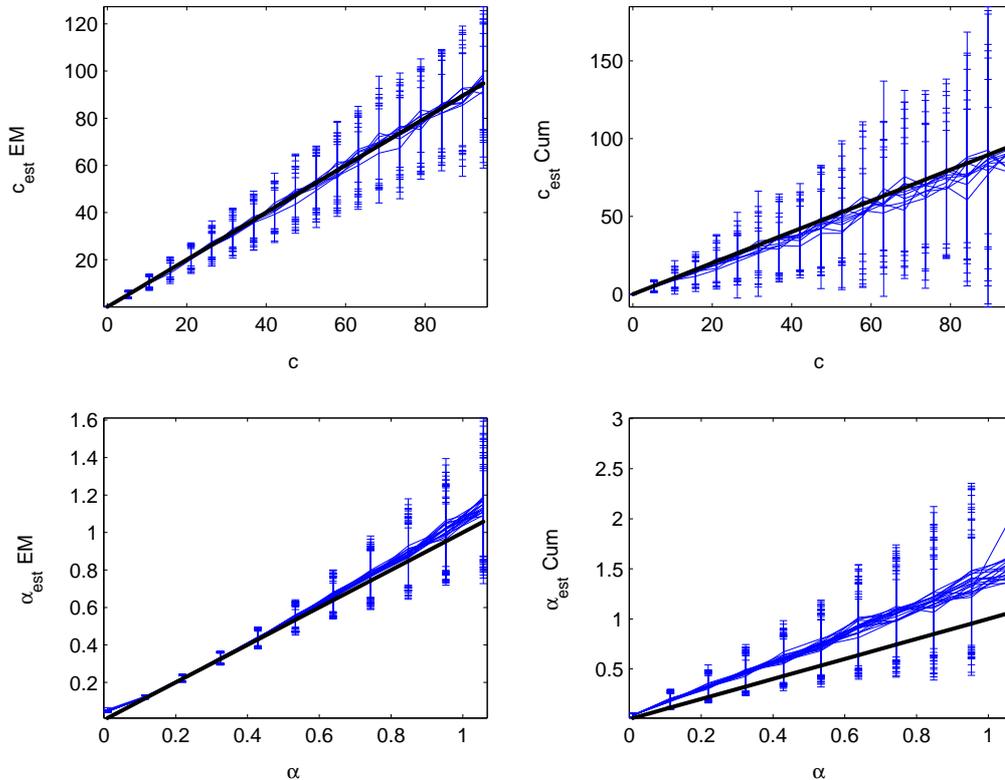


FIG. 3.12 – Comparaison entre l'estimation des hyperparamètres  $\alpha$  et  $c$  par la méthode des cumulants et l'algorithme EM. Les estimés de  $\alpha$  et  $c$  par l'algorithme EM sont plus proches des valeurs initiales comparés aux estimés par la méthode des cumulants.

## 3.7 Qualité de l'a priori : application à la modélisation des statistiques marginales

### 3.7.1 Objectifs de l'expérience

L'objectif de cette simulation numérique est d'évaluer la qualité des *a priori* présentés ci-dessus, à savoir l'*a priori*  $\alpha$ -stable approximé par le modèle des mélanges finis (dénomé  $\alpha$ -stable mixture) et l'*a priori* BKF. Nous comparons nos PDFs marginales à une alternative très utilisée dans la littérature, à savoir le modèle GGD [Mallat, 1989]. Une comparaison des PDFs estimées et observées des coefficients de détails d'ondelettes des images naturelles, est illustrée sur les Fig.3.13 et 3.14, et le Tab.3.3.

### 3.7.2 Protocole expérimental

TAB. 3.2 – Protocole expérimental

Grandeur	Choix	
	Fig.3.13	Fig.3.14
<i>a priori</i> univarié	$\alpha$ -stable mixture, BKF, GGD, $\alpha$ -stable	
Transformées	ondelettes à support compact de Daubechies-4	
DWT	2 échelles et 3 orientations	3 échelles et 3 orientations
Estimateur des hyperparamètres	algorithme EM	cumulants et algorithme EM
Nombre de gaussiennes optimal	16	
Monte-Carlo		50
Métrique		divergence KL
Base d'image	trois images tests	100 images

### 3.7.3 Discussion des résultats

Dans la Fig.3.13 et suite à l'observation des courbes, nous constatons que les PDFs estimées par l'*a priori* BKF et  $\alpha$ -stable mixture sont plus proches de la PDF marginale observée, comparées aux PDFs estimées par l'*a priori*  $\alpha$ -stable (version originale de [Achim et al., 2001]) et GGD [Mallat, 1989].

Nous avons effectué, dans la Fig3.14 et le Tab.3.3, une étude quantitative en calculant la divergence KL entre la PDF marginale observée et les PDFs estimées sur par diverses méthodes sur une base de 100 images [url : base d'images, ]. Cette divergence de KL a été moyennée pour chaque orientation à chaque échelle sur toute la base d'images. Nous constatons tout d'abord que le modèle *a priori* avec l'algorithme EM présente des performances meilleures comparé au modèle BKF avec les cumulants. Ceci est du à la qualité de l'estimateur des hyperparamètres avec l'algorithme EM qui est meilleur comparé à celui avec la méthode des cumulants. Aussi, ce modèle présente des performances légèrement au dessous du modèle  $\alpha$ -stable mixture pour les échelles les plus fines. De plus, le modèle  $\alpha$ -stable mixture présente des performances bien meilleures comparé au modèle  $\alpha$ -stable [Achim et al., 2001] surtout pour les échelles les plus fines. Par ailleurs, le modèle GGD [Mallat, 1989] même s'il présente des bonnes performances il reste inférieur à nos modèles.

Notons en plus que l'écart-type de la divergence de KL (noté entre parenthèse dans Tab.3.3) peut parfois être relativement grand. Ceci signifie que le meilleur estimateur moyen pour une grande base d'images peut engendrer des erreurs non négligeables ponctuellement sur certaines images. Il n'est par conséquent pas forcément optimal pour une classe d'image donnée.

Les modèles *a priori* BKF et  $\alpha$ -stable mixture sont adaptés pour capturer le comportement des queues de distribution relevées de coefficients de détail où les *a priori* GGD et  $\alpha$ -stable sont moins bons pour modéliser correctement ce comportement typique de ces coefficients dans certains cas. Finalement, ceci confirme notre pressentiment dicté par la théorie que ces modèles sont adéquats pour la modélisation des distributions à comportement leptokurtique et queues lourdes.

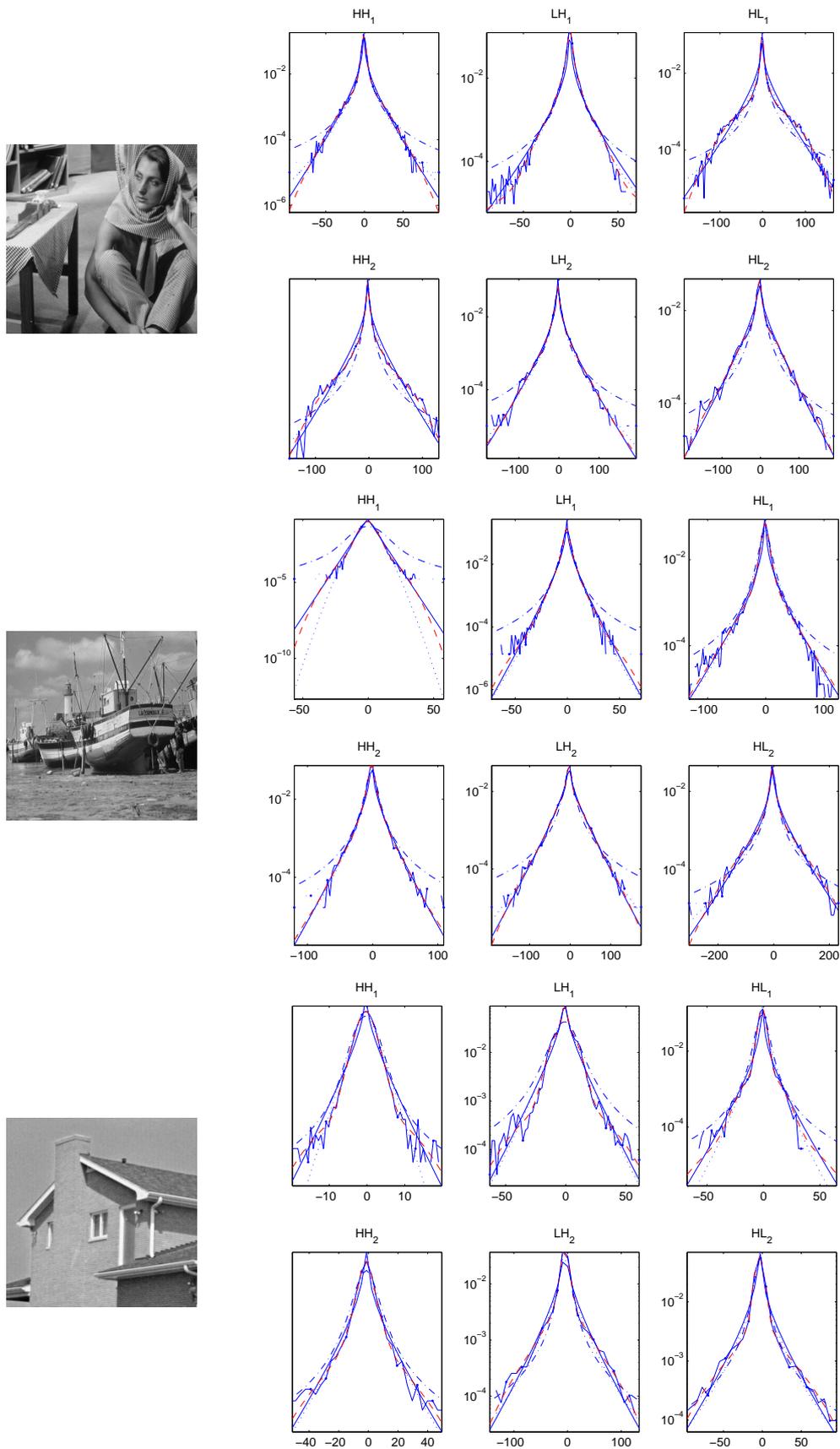


FIG. 3.13 – Comparaison sur une échelle log-log entre la PDF marginale observée et les PDFs estimées des coefficients de détail d'ondelettes pour trois d'images test, à deux échelles et trois orientations  $HH$ ,  $HL$  et  $LH$ . L'histogramme observé (-●-) a été ajusté par trois modèles d'a priori : BKF (trait plein),  $\alpha$ -stable mixture (trait pointillé),  $\alpha$ -stable (alternance de points et de traits) et GGD (points).

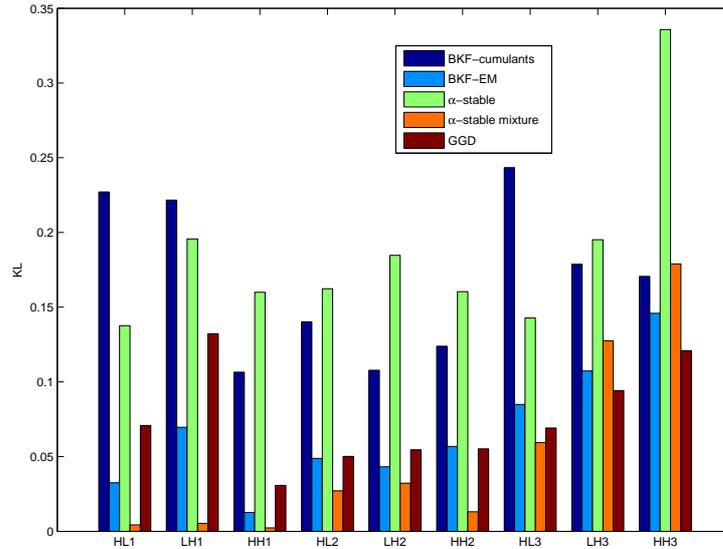


FIG. 3.14 – Évolution de la divergence KL, calculée entre la PDF observée et les PDFs estimées par l'*a priori* BKF (avec la méthode des cumulants et l'algorithme EM),  $\alpha$ -stable [Achim et al., 2001],  $\alpha$ -stable mixture et GGD. La divergence KL a été moyennée pour chaque orientation à chaque échelle sur une base de 100 images.

TAB. 3.3 – La moyenne (et l'écart-type) de la divergence KL entre la PDF marginale observée et les PDFs estimées par l'*a priori* : BKF,  $\alpha$ -stable mixture,  $\alpha$ -stable et GGD, à trois échelles et trois orientations sur une base de 100 images.

HH			
<i>a priori</i>	$j = 1$	$j = 2$	$j = 3$
BKF	0.1114 (0.5445)	0.1236 (0.2916)	0.1616 (0.4234)
$\alpha$ -stable mixture	0.1310 (0.8347)	0.1171 (0.7355)	0.0956 (0.5540)
$\alpha$ -stable	1.3088 (5.7553)	0.4692 (1.2683)	0.5198 (1.2900)
GGD	0.3734 (2.8552)	0.1218 (0.2118)	0.3063 (0.8608)
HL			
BKF	0.1956 (0.8126)	0.1255 (0.2681)	0.1151 (0.2005)
$\alpha$ -stable mixture	0.1218 (0.6790)	0.0918 (0.4900)	0.0758 (0.2692)
$\alpha$ -stable	1.0059 (3.9762)	0.6530 (1.3354)	0.6589 (1.1550)
GGD	0.2496 (1.0691)	0.0738 (0.1224)	0.0713 (0.0913)
LH			
BKF	0.1906 (0.6381)	0.1231 (0.2954)	0.1146 (0.1784)
$\alpha$ -stable mixture	0.2014 (1.0001)	0.0795 (0.2581)	0.1071 (0.3432)
$\alpha$ -stable	0.9227 (1.9900)	0.7872 (1.5418)	0.7968 (1.3823)
GGD	0.1280 (0.4248)	0.0768 (0.1347)	0.0821 (0.1105)

## Chapitre 4

# Débruitage bayésien avec *a priori* univarié

---

### Sommaire

---

<b>4.1</b>	<b>Rappels sur l'<i>a priori</i></b> . . . . .	<b>77</b>
<b>4.2</b>	<b>Estimation bayésienne : cas univarié</b> . . . . .	<b>78</b>
<b>4.3</b>	<b>Estimateur de l'<i>espérance conditionnelle a posteriori</i> (ECP)</b> . . . . .	<b>79</b>
4.3.1	Estimateur ECP avec l' <i>a priori</i> $\alpha$ -stable . . . . .	79
4.3.1.1	La PDF marginale des coefficients bruités . . . . .	79
4.3.1.2	Estimation des hyperparamètres . . . . .	79
4.3.1.3	Débruiteur bayésien terme-à-terme . . . . .	80
4.3.2	Estimateur ECP avec l' <i>a priori</i> BKF . . . . .	82
4.3.2.1	La PDF marginale des coefficients d'ondelettes . . . . .	82
4.3.2.2	Estimation des hyperparamètres . . . . .	85
4.3.2.3	Débruiteur bayésien terme-à-terme . . . . .	87
<b>4.4</b>	<b>Estimateur <i>maximum a posteriori</i> (MAP)</b> . . . . .	<b>91</b>
4.4.1	Estimateur MAP avec l' <i>a priori</i> BKF . . . . .	91
<b>4.5</b>	<b>Expérimentation et résultats</b> . . . . .	<b>93</b>
4.5.1	Objectifs de l'expérience . . . . .	93
4.5.2	Protocole expérimental . . . . .	94
4.5.3	Discussion des résultats . . . . .	94
<b>4.6</b>	<b>Conclusion</b> . . . . .	<b>101</b>

---

### 4.1 Rappels sur l'*a priori*

Dans la première partie de cette thèse, nous avons effectué une classification méthodique de l'état de l'art sur le problème de débruitage. Les méthodes présentées cherchent pour la plupart à reconstruire une solution qui présente une certaine régularité. Celle-ci peut être imposée de différentes manières, et réalisée au moyen des transformées multi-échelles, dans un cadre classique et bayésien.

Dans le contexte bayésien, afin de reconstruire l'image non bruitée et tout en conservant les détails, il est nécessaire d'imposer des contraintes sur la solution reconstruite, ce qui consiste à choisir un modèle *a priori*. Dans le chapitre précédent, nous avons présenté le modèle SMG

comme *a priori* adéquat à la modélisation statistique des images dans l'espace des transformées parcimonieuses, et en l'occurrence les ondelettes. En pratique, nous avons introduit un cadre général pour les modèles *a priori* :  $\alpha$ -stable et BKF qui tient compte des propriétés parcimonieuses des coefficients d'images dans le domaine multi-échelle, et les différentes techniques permettant d'estimer les hyperparamètres de ces derniers. Ceci a été effectué dans un cadre univarié où les coefficients d'images dans l'espace des transformées sont supposés indépendants.

Dans ce chapitre, nous proposons de nouveaux estimateurs bayésiens ECP et MAP (cf. annexe A) pour le débruitage, basés sur les *a priori* :  $\alpha$ -stable et BKF. Deux étapes sont mises en jeu pour la mise en oeuvre de ces estimateurs, la première consistant à estimer les hyperparamètres du modèle de l'*a priori* en présence de bruit, et la seconde à trouver une forme analytique pour l'estimateur bayésien correspondant.

## 4.2 Estimation bayésienne : cas univarié

Dans le processus du débruitage bayésien décrit dans l'état de l'art (1.2.3), on rappelle qu'il est recommandé de conserver les coefficients d'approximation qui sont caractéristiques de l'image originale, et d'estimer les coefficients de détail de l'image à recouvrer à partir de ceux de l'image bruitée.

Le modèle de dégradation<sup>5</sup> choisi par la suite est le suivant :

$$d = s + \epsilon \quad (4.1)$$

où le bruit blanc gaussien est  $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ .

Dans le contexte bayésien, l'information *a priori* est imposée aux coefficients d'ondelettes pour décrire leur distribution. Les coefficients de détail à chaque échelle et orientation de l'image à estimer  $s$  suivent une loi de probabilité du modèle *a priori* SMG pour traduire la connaissance initiale sur  $s$ , paramétrée par un certain nombre de paramètres à estimer  $\theta_1$

$$s \sim \text{SMG}(\theta_1) \quad (4.2)$$

Le modèle probabiliste associé à  $d$  sachant  $s$  est gaussien et permet de traduire l'incertitude due au bruit sur les données bruitées  $d$ ,

$$d|s \sim \mathcal{N}(s, \theta_2) \quad (4.3)$$

Ce modèle dépend d'un certain nombre de paramètres à estimer  $\theta_2$ .

Une fois les coefficients  $d$  déterminés, nous cherchons la distribution *a posteriori* de  $s$  en utilisant la règle de Bayes pour combiner l'information contenue dans les données d'observation et celle contenue dans la loi *a priori*.

Cette loi *a posteriori* dépend d'un ensemble d'hyperparamètres  $\theta = \{\theta_1, \theta_2\}$ , et contient toute l'information disponible sur  $s$ . Leur calcul complet peut devenir très délicat, mais généralement, on se contente de définir un estimateur ponctuel à partir de cette loi de type ECP ou bien MAP (cf. Annexe A).

Rappelons aussi qu'en pratique, le calcul d'estimateurs ECP nécessite le calcul d'intégrales de dimension très élevée, alors que le calcul de l'estimateur MAP nécessite une optimisation<sup>6</sup>.

<sup>5</sup>Notons que  $d$  (resp.  $s$ ) correspond aux coefficients  $d_{mn}^{oj}$  (resp.  $s_{mn}^{oj}$ ).

<sup>6</sup>Dans le cas d'un *a priori* gaussien les estimateurs bayésiens ECP et MAP sont identiques (estimateur de Wiener).

## 4.3 Estimateur de l'espérance conditionnelle a posteriori (ECP)

### 4.3.1 Estimateur ECP avec l'a priori $\alpha$ -stable

#### 4.3.1.1 La PDF marginale des coefficients bruités

Pour le cas de l'a priori  $\alpha$ -stable et à partir de l'Eq.4.2 et l'Eq.4.3, les coefficients de détail à chaque échelle et à chaque orientation de l'image à estimer  $s$  suivent une distribution SaS,

$$s \sim S_\alpha(0, 0, \gamma = \sigma^\alpha) \quad (4.4)$$

et le modèle probabiliste associé à  $d$  sachant  $s$  est gaussien

$$d|s \sim \mathcal{N}(0, \sigma_\epsilon^2) \quad (4.5)$$

où  $\theta_1 = \{\alpha, \sigma\}$  et  $\theta_2 = \{\sigma_\epsilon\}$ .

La PDF marginale des observations  $d$  peut s'écrire sous la forme suivante,

$$\mathcal{P}(d; \theta) = \int \phi(d - s; \theta_2) \mathcal{P}(s; \theta_1) ds \quad (4.6)$$

où  $\mathcal{P}(s; \theta_1)$  est la PDF SaS approximée par le mélange fini de gaussiennes, donnée par l'algorithme 3.5.4.2 [page 55], avec l'ensemble des hyperparamètres  $\theta_1$ <sup>7</sup>

$$f(s; \{P_j, \sigma_j\}) = \frac{1}{\sqrt{2\pi}} \sum_{j=1}^N P_j \sigma_j^{-1} \exp\left(-\frac{s^2}{2\sigma_j^2}\right) \quad (4.7)$$

où  $N$  est le nombre de gaussiennes.

Et  $\phi(d - s; \theta_2)$  est la PDF normale du bruit d'écart-type  $\theta_2 = \{\sigma_\epsilon\}$

$$\phi(d - s; \sigma_\epsilon) = \frac{1}{\sqrt{2\pi}\sigma_\epsilon} \exp\left(-\frac{(d - s)^2}{2\sigma_\epsilon^2}\right) \quad (4.8)$$

Après l'évaluation de l'intégrale dans l'Eq.4.6, l'expression analytique de la PDF marginale SaS est

$$\tilde{f}(d; \theta) = \frac{1}{\sqrt{2\pi}} \sum_{j=1}^N P_j (\sigma_j^2 + \sigma_\epsilon^2)^{-\frac{1}{2}} \exp\left(-\frac{d^2}{2(\sigma_j^2 + \sigma_\epsilon^2)}\right) \quad (4.9)$$

où  $N$  est le nombre de gaussiennes.

#### 4.3.1.2 Estimation des hyperparamètres

Dans le contexte de débruitage d'images, les hyperparamètres sont estimés à partir des coefficients d'ondelettes observés à chaque orientation et à chaque échelle. Pour mettre en application la formule de l'Eq.4.9, on doit alors estimer l'ensemble des paramètres  $\theta = \{P_j, \sigma_j, \sigma_\epsilon\}$  qui, à leur tour, mèneront à une procédure de débruitage adaptatif à chaque sous-bande. L'étape d'estimation de ces hyperparamètres est cruciale. Il est clair que celle-ci conditionne les performances finales de l'algorithme de débruitage.

Dans le cadre du modèle  $\alpha$ -stable, l'estimation des hyperparamètres  $P_j$  et  $\sigma_j$  dépend des paramètres arbitraires  $\alpha$  et  $\sigma$  (l'étape EM de l'algorithme 3.5.4.2). Si cette estimation reste accessible en absence de bruit, elle devient beaucoup plus délicate en sa présence.

<sup>7</sup>Rappelons que les hyperparamètres de  $\theta_1 = \{\alpha, \sigma\}$  dépendent des paramètres  $\{P_j, \sigma_j\}$ .

Dans le contexte du débruitage, les observations dont on dispose sont corrompues par un bruit blanc gaussien. Il s'en suit une difficulté majeure lors de l'estimation des hyperparamètres de la loi *a priori*. L'estimation des paramètres  $\alpha$  et  $\sigma$ , dans le cas de notre modèle, ne sert que d'initialisation. Nous avons alors, suite aux discussions précédentes, choisi l'estimateur basé sur les quantiles de McCulloch partant du fait qu'à SNR raisonnable, les queues de la distribution marginale des coefficients d'ondelettes observés sont peu influencées par la présence du bruit.

La méthode de McCulloch permet une estimation initiale et rapide des paramètres  $\alpha$  et  $\sigma$  sous la restriction  $0.6 \leq \alpha \leq 2$ , on en déduit alors le paramètre  $v_j$  qui à son tour nous permet l'initialisation des paramètres  $P_j$  et  $\sigma_j$  en appliquant l'étape M de l'algorithme 3.5.4.2 [55]. Par ailleurs, le niveau de bruit  $\sigma_\epsilon$  est estimé à partir de l'orientation  $HH$  de l'échelle la plus fine en utilisant l'estimateur robuste de Donoho & Johnstone [Donoho & Johnstone, 1994], donné par l'Eq.1.3. L'utilisation des valeurs d'initialisation des paramètres  $\alpha$ ,  $\sigma$  et  $\sigma_\epsilon$  dans l'algorithme 3.5.4.2 [55] permet d'obtenir une forme analytique stable pour la PDF marginale des coefficients d'ondelettes observés. Ceci peut être récapitulé comme suit,

- (a) Utilisation de la méthode de McCulloch [McCulloch, 1986] pour obtenir  $\hat{\alpha}$  et  $\hat{\sigma}$ , et l'estimateur MAD pour obtenir  $\hat{\sigma}_\epsilon$  [Donoho & Johnstone, 1994].
- (b) Appliquer les étapes 1-4 de l'algorithme 3.5.4.2 en utilisant  $\hat{\alpha}$  et  $\hat{\sigma}$ , pour obtenir  $\hat{\sigma}_j$  et  $\hat{P}_j$ .
- (c) Définir  $\hat{\sigma}'_j = \sqrt{\hat{\sigma}_j^2 + \hat{\sigma}_\epsilon^2}$
- (d) Appliquer l'algorithme EM pour affiner les estimés  $\hat{\sigma}'_j$  et  $\hat{P}_j$  associés au modèle analytique de la PDF marginale donné par l'Eq.4.9.

Dans la Fig.4.1, et suite à l'observation des courbes, nous constatons clairement que la PDF marginale estimée par l'Eq.4.9 est très proche de la PDF marginale observée. Ceci est confirmé par la distance KL calculée entre les deux PDFs.

### 4.3.1.3 Débruiteur bayésien terme-à-terme

La forme analytique de l'estimateur ECP (cf. annexe A) S $\alpha$ S des coefficients  $s$ , conditionnellement sur l'ensemble des hyperparamètres  $\theta$ , permet de réaliser le débruiteur bayésien.

Le calcul de l'estimateur ECP nécessite l'évaluation des intégrales du numérateur et du dénominateur. L'expression du dénominateur ECP a été établie par l'Eq.4.9, et l'expression du numérateur est la suivante

$$\int_{-\infty}^{+\infty} s \mathcal{P}(d|s; \theta_2) \mathcal{P}(s; \theta_1) ds = \frac{1}{\sqrt{2\pi}} \sum_{j=1}^N P_j \frac{d \sigma_j^2}{(\sigma_j^2 + \sigma_\epsilon^2)^{\frac{3}{2}}} \exp\left(-\frac{d^2}{2(\sigma_j^2 + \sigma_\epsilon^2)}\right) \quad (4.10)$$

Après quelques simplifications de calcul, l'expression analytique de l'estimateur ECP S $\alpha$ S devient :

$$\hat{s}_{ECP}(d; \theta) = \frac{\sum_{j=1}^N P_j \frac{\sigma_j^2 d}{\sigma_j^2 + \sigma_\epsilon^2} \phi(d; \sigma_j^2 + \sigma_\epsilon^2)}{\sum_{j=1}^N P_j \phi(d; \sigma_j^2 + \sigma_\epsilon^2)} \quad (4.11)$$

Ce résultat est une somme pondérée d'estimateurs de Wiener (un mélange fini de Wiener).

La Fig.4.2 montre l'évolution de la fonction  $s(d)$  de cet estimateur bayésien ECP. Les graphes de la Fig.4.2(a) (resp. Fig.4.2(b)) montrent l'influence du rapport  $\frac{\sigma}{\sigma_\epsilon}$  (resp. du paramètre  $\alpha$ ) sur

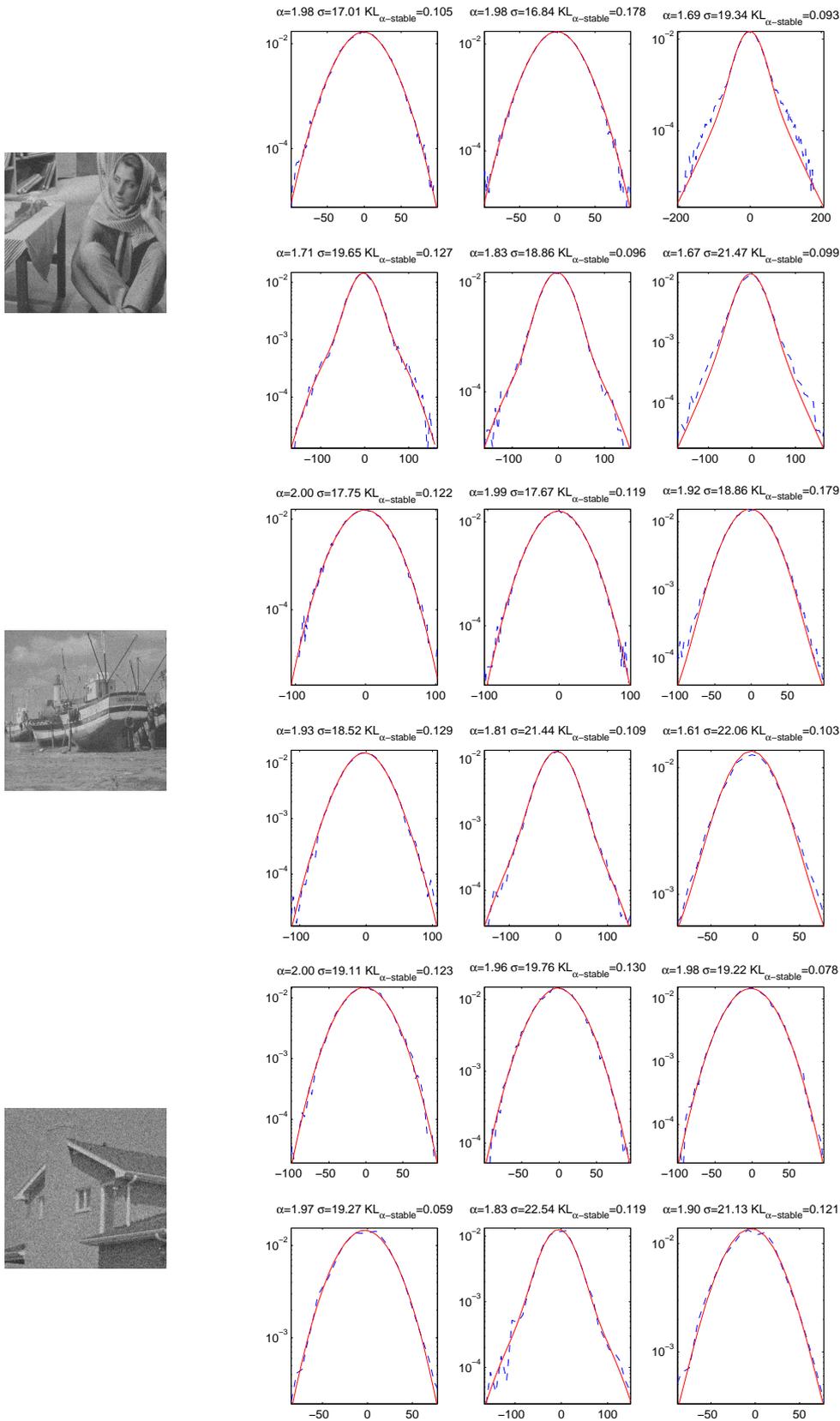


FIG. 4.1 – Comparaison entre l'histogramme marginal observé (trait en pointillé) et la PDF  $\alpha$ -stable estimée (trait plein) des coefficients de détail d'ondelettes pour trois d'images tests corrompues par un bruit blanc gaussien (SNR= 15dB). Seulement, deux niveaux de décomposition sont montrés pour chacune des trois images test : Barbara, Boat et House. Les trois colonnes correspondent aux orientations  $HH$ ,  $HL$  et  $LH$ .

les courbes de  $s(d)$  en fonction du paramètre  $\alpha \in [0.6, 2]$  (resp. du rapport  $\frac{\sigma}{\sigma_\epsilon}$ ). Le rapport  $\frac{\sigma}{\sigma_\epsilon}$  peut être intuitivement interprété comme une forme de mesure du SNR. Nous constatons que la contraction (shrinkage) diminue d'autant plus que le rapport  $\frac{\sigma}{\sigma_\epsilon}$  augmente. Intuitivement cela peut aisément s'expliquer par le fait que la contribution du signal/image est d'autant plus grande que le rapport  $\frac{\sigma}{\sigma_\epsilon}$  est important.

De plus, nous constatons que la contraction diminue quand  $\alpha$  diminue. Cela peut aussi se comprendre en considérant le fait que plus la valeur de  $\alpha$  est faible et plus la queue de la distribution *a priori* est importante, d'où une plus grande probabilité qu'une grande valeur mesurée (située en queue de distribution) soit due à la présence de signal.

### 4.3.2 Estimateur ECP avec l'*a priori* BKF

#### 4.3.2.1 La PDF marginale des coefficients d'ondelettes

Pour le cas de l'*a priori* BKF et à partir de l'Eq.4.2 et l'Eq.4.3, les coefficients de détail à chaque échelle et à chaque orientation à estimer  $s$  suivent une loi BKF,

$$s \sim BKF(\alpha, c) \quad (4.12)$$

et le modèle probabiliste associé à  $d$  sachant  $s$  est gaussien

$$d|s \sim \mathcal{N}(0, \sigma_\epsilon) \quad (4.13)$$

où  $\theta_1 = \{\alpha, c\}$  et  $\theta_2 = \{\sigma_\epsilon\}$ .

La forme analytique de la PDF marginale de  $d$  sachant les hyperparamètres  $\theta$  (Eq.4.6) n'est pas facile à obtenir sauf par approximation.

Le lemme suivant présente une forme analytique approximée plus pratique pour la PDF marginale en approximant la fonction de Bessel modifiée  $K_\nu$  par l'approximation donnée dans [Gradshteyn & Ryzhik, 1980], voir [Fadili & Boubchir, 2005].

**Lemme 4.1** *Pour  $0 < \alpha \leq 1$  et  $c > 0$ , la PDF marginale des observations  $d$  peut être bien approximée par :*

$$\begin{aligned} \tilde{f}(d; \theta) &= \left(\frac{\sigma_\epsilon^2}{2c}\right)^{\frac{\alpha}{2}} \phi(d; 0, \sigma_\epsilon^2) (I_+ + I_-) \\ \text{où } I_\pm &= \exp\left(\frac{\pm \frac{d}{\sigma_\epsilon} + \sigma_\epsilon \sqrt{\frac{2}{c}}}{2}\right)^2 D_{-\alpha}\left(\pm \frac{d}{\sigma_\epsilon} + \sigma_\epsilon \sqrt{\frac{2}{c}}\right) \end{aligned} \quad (4.14)$$

où  $D_\nu(\cdot)$  est la fonction de cylindre parabolique d'ordre partiel  $\nu$  [Abramowitz & Stegun, 1972, Gradshteyn & Ryzhik, 1980] :

$$D_\nu(x) = \frac{e^{-\frac{x^2}{4}}}{\Gamma(-\nu)} \int_0^{+\infty} e^{-xt - \frac{t^2}{2}} t^{-\nu-1} dt, \text{ pour } \text{Re}(\nu) < 0 \quad (4.15)$$

$\tilde{f}(d)$  est une fonction paire de mode zéro.

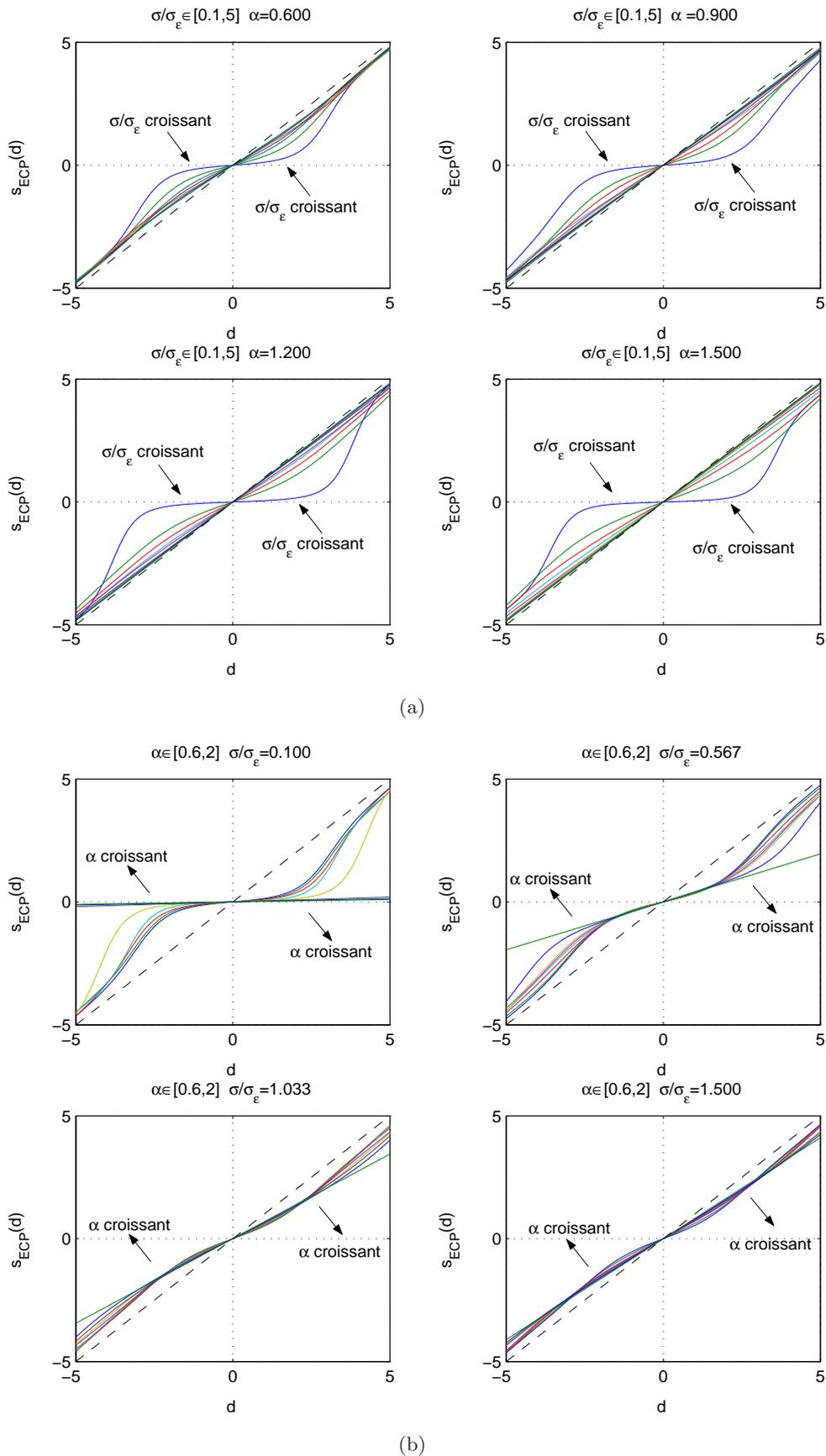


FIG. 4.2 – Évolution de la fonction  $s(d)$  de l'estimateur bayésien ECP  $\alpha$ -stable. (a) En fonction du rapport  $\frac{\sigma}{\sigma_\epsilon}$ . (b) En fonction du paramètre  $\alpha$ .

**Preuve:**

Nous employons le résultat d'approximation de la fonction  $K_\nu$  donné dans [Abramowitz & Stegun, 1972, Gradshteyn & Ryzhik, 1980]. Pour  $|d|$  grande et  $\nu = \alpha - \frac{1}{2}$ , l'expression asymptotique d'ordre 1 de  $K_\nu$  est la suivante

$$K_\nu(d) \sim \sqrt{\frac{\pi}{2d}} e^{-d} \quad (4.16)$$

avec un reste donné par,

$$|R| < \left| \frac{\Gamma(\nu + 3/2)}{(2d)\Gamma(\nu - 1/2)} \right|, \quad \text{pour } \nu - \frac{1}{2} \leq 1 \quad (4.17)$$

Ceci signifie que pour des valeurs réelles positives de  $\nu$  et  $d$ , le reste  $R$  est inférieur en valeur absolue à un terme ignoré de même signe, à condition que  $\alpha = \nu + \frac{1}{2} \leq 2$ . En appliquant ce résultat à la PDF BKF, on obtient :

$$\tilde{f}(d; \alpha, c) = \frac{(2c)^{-\frac{\alpha}{2}}}{\Gamma(\alpha)} |d|^{\alpha-1} \exp\left(-\sqrt{\frac{2}{c}}|d|\right) \quad (4.18)$$

Bien évidemment, pour  $\alpha > 1$ , la PDF dans l'Eq.4.18 est bimodale avec un mode non nul  $\pm(\alpha - 1)\sqrt{\frac{2}{c}}$ . Par conséquent, l'Eq.4.18 est valable seulement pour  $p \leq 1$ . En utilisant cette expression dans Eq.4.6 et après décomposition de l'intégrale en deux parties pour  $s$  positif et négatif, nous pouvons écrire :

$$\begin{aligned} \tilde{f}(d; \alpha, c, \sigma_\epsilon^2) &= \frac{(2c)^{-\frac{\alpha}{2}}}{\Gamma(\alpha)} \phi(d; 0, \sigma_\epsilon^2) [I_+ + I_-] \quad (4.19) \\ \text{où } I_\pm &= \int_0^{+\infty} s^{\alpha-1} \exp\left(-\frac{s^2}{2\sigma_\epsilon^2} - s\left(\pm\frac{d}{\sigma_\epsilon^2} + \sqrt{\frac{2}{c}}\right)\right) ds \end{aligned}$$

Pour établir le résultat final, nous aurons besoin de la formule d'intégrale de [Gradshteyn & Ryzhik, 1980] à la page 337 :

$$\begin{aligned} \int_0^{+\infty} s^{\nu-1} \exp(-\beta s^2 - \gamma s) ds &= (2\beta)^{-\frac{\nu}{2}} \Gamma(\alpha) \exp\left(\frac{\gamma^2}{8\beta}\right) D_{-\nu}\left(\frac{\gamma}{\sqrt{2\beta}}\right) \\ &\quad (Re(\nu) > 0, Re(\beta) > 0) \quad (4.20) \end{aligned}$$

où  $D_\nu$  est la fonction de cylindre parabolique d'ordre partiel  $\nu$  [Gradshteyn & Ryzhik, 1980]. Plusieurs représentations intégrales existent pour  $D_\nu$ . Dans notre cas l'argument  $\nu$  est toujours strictement négatif et donc nous choisissons la représentation suivante [Gradshteyn & Ryzhik, 1980] :

$$D_\nu(x) = \frac{\exp\left(-\frac{x^2}{4}\right)}{\Gamma(-\nu)} \int_0^{+\infty} \exp\left(-xt - \frac{t^2}{2}\right) t^{-\nu-1} dt, \quad \text{pour } Re(\nu) < 0 \quad (4.21)$$

où  $\nu = \alpha$  et  $\beta = \frac{1}{2\sigma_\epsilon^2}$  sont strictement positives par définition. En combinant l'intégrale dans l'Eq.4.20 avec celle de l'Eq.4.21 et après quelques arrangements algébriques, le résultat dans l'Eq.4.14 suit. ■

Du point de vue pratique, il existe plusieurs algorithmes numériques pour calculer la fonction  $D_\nu$  impliquée dans l'Eq.4.14. Dans [Zhang & Jin, 1996], les auteurs ont développé des méthodes numériques pour calculer  $D_\nu(x)$  d'ordre  $\nu$  réel et des arguments réels. Leur méthode récursive est basée sur les séries de MacLaurin-Taylor et des expressions asymptotiques. Elles sont stables et rapides à calculer.

**Remarque 4.1** *Comme c'est précisé par Grenander et al. [Grenander & Srivastava, 2001, Srivastava et al., 2002], la plupart des images naturelles filtrées ont une valeur de  $\alpha \leq 1$ , alors notre condition dans le lemme 4.1 n'est pas restrictive et l'approximation est très bonne.*

#### 4.3.2.2 Estimation des hyperparamètres

Pour mettre en application la formule de l'Eq.4.14, on doit alors estimer les hyperparamètres  $\theta = \{\alpha, c, \sigma_\epsilon\}$  qui, à leur tour, mèneront à une procédure de débruitage adaptatif à chaque sous-bande.

Pour les observations bruitées, nous avons proposé deux méthodes pour estimer les paramètres  $\alpha$  et  $\sigma$  :

**Méthode des cumulants** Celle-ci possède la vertu d'être facilement adaptable en présence de bruit gaussien additif. En effet, les paramètres  $\alpha$  et  $\sigma$  sont estimés, en utilisant les cumulants d'ordre 2 et 4, comme suit :

$$\begin{aligned}\hat{\alpha} &= \frac{3 \max(\hat{\kappa}_2 - \hat{\sigma}_\epsilon^2, 0)^2}{\hat{\kappa}_4} \\ \hat{c} &= \frac{\max(\hat{\kappa}_2 - \hat{\sigma}_\epsilon^2, 0)}{\hat{\alpha}}\end{aligned}\tag{4.22}$$

$\hat{\kappa}_i$  sont les estimées de  $\kappa_i$  par les  $k$ -statistiques.

**Algorithme EM** Soit  $U$  les données cachées et  $(d, U)$  les données complètes. La procédure d'estimation est donnée par l'algorithme suivant :

---

**Algorithme 4** Algorithme EM pour l'estimation des paramètres  $\alpha$  et  $\sigma$  pour une distribution BKF en présence de bruit

---

**Entrées:** On part d'une estimation initiale  $\theta^{(0)}$

- 1: **répéter**
- 2: **Étape E :**  
On calcule,

$$M_U^{(t)}(1) = \mathbb{E}[U|d] = \frac{\int_0^\infty \frac{u^{\alpha^{(t)}}}{\sqrt{u+\sigma_\epsilon^2}} \exp\left(-\frac{d^2}{2(u+\sigma_\epsilon^2)} - \frac{u}{c^{(t)}}\right) du}{\int_0^\infty \frac{u^{\alpha^{(t)}-1}}{\sqrt{u+\sigma_\epsilon^2}} \exp\left(-\frac{d^2}{2(u+\sigma_\epsilon^2)} - \frac{u}{c^{(t)}}\right) du} \quad (4.23)$$

$$\varphi_1^{(t)} = \mathbb{E}[\log(U)|d] = \frac{\int_0^\infty \log(u) \frac{u^{\alpha^{(t)}-1}}{\sqrt{u+\sigma_\epsilon^2}} \exp\left(-\frac{d^2}{2(u+\sigma_\epsilon^2)} - \frac{u}{c^{(t)}}\right) du}{\int_0^\infty \frac{u^{\alpha^{(t)}-1}}{\sqrt{u+\sigma_\epsilon^2}} \exp\left(-\frac{d^2}{2(u+\sigma_\epsilon^2)} - \frac{u}{c^{(t)}}\right) du} \quad (4.24)$$

Ces intégrales convergent et peuvent être calculées assez rapidement par des méthodes de quadrature, en les décomposant de  $[0, a[$  et de  $[a, +\infty[$  (e.g.  $a = 1$ ) et en procédant à un changement de variable  $v = \frac{1}{u}$  dans  $[a, +\infty[$ .

- 3: **Étape M :**  
Les estimés de  $\alpha$  et  $c$  sont comme suit,

$$c^{(t+1)} = \frac{1}{\alpha^{(t+1)}} M_U^{(t)}(1) \quad (4.25)$$

où  $\alpha^{(t+1)}$  est la solution de l'équation suivante :

$$\Psi(\alpha) - \log(\alpha) = \frac{1}{m} \sum_{i=1}^m \varphi_1^{(t)}(i) - \log\left(M_U^{(t)}(1)\right) \quad (4.26)$$

- 4: **jusqu'à** Convergence.
- 

**Preuve:**

La preuve de l'algorithme ci-dessus est similaire à la preuve de l'algorithme EM en absence de bruit (page 69).

Le calcul du score de vraisemblance  $Q$  nécessite le calcul de l'antilog de vraisemblance suivant :

$$\begin{aligned} \ell\ell(d|U) + \ell\ell(U) &= -\frac{1}{2} \sum_{i=1}^m \log(u_i + \sigma_\epsilon^2) - \frac{1}{2} \sum_{i=1}^m \frac{d_i^2}{u_i + \sigma_\epsilon^2} - m \log c - m \log \Gamma(\alpha) \\ &\quad - (\alpha - 1) \sum_{i=1}^m \log \frac{u_i}{c} - \sum_{i=1}^m \frac{u_i}{c} \end{aligned}$$

où  $m$  est le nombre d'échantillons.

Et donc,  $Q$  peut s'écrire comme suit,

$$\begin{aligned}
Q(\theta|\theta^{(t)}) &= \frac{1}{2} \sum_{i=1}^m \varphi_{1(i)} + \frac{1}{2} \sum_{i=1}^m \varphi_{2(i)} d_i^2 + m \log c + m \log \Gamma(\alpha) + (1 - \alpha) \sum_{i=1}^m \varphi_{1(i)} \\
&\quad - m(1 - \alpha) \log c + \sum_{i=1}^m \frac{\varphi_{3(i)}}{c} \\
&= \left( \frac{3}{2} - \alpha \right) \sum_{i=1}^m \varphi_{1(i)} + \frac{1}{2} \sum_{i=1}^m \varphi_{2(i)} d_i^2 + \sum_{i=1}^m \frac{\varphi_{3(i)}}{c} + \alpha m \log c + m \log \Gamma(\alpha) \quad (4.27)
\end{aligned}$$

où nous notons

$$\varphi_{1(i)} = \mathbb{E} [\log(U + \sigma_\epsilon^2) | d_i] ; \varphi_{2(i)} = \mathbb{E} [(U + \sigma_\epsilon^2)^{-1} | d_i] ; \varphi_{3(i)} = \mathbb{E} [U | d_i] , \text{ pour } i = 1, \dots, m \quad (4.28)$$

Nous précisons que  $\varphi_1$  et  $\varphi_2$  ne sont pas à calculer numériquement car  $\sigma_\epsilon^2$  est estimé par l'estimateur MAD de Donoho & Johnstone [Donoho & Johnstone, 1994].

Suivant les mêmes démarches de démonstration dans 3.6.4.2, le calcul de  $\varphi_2 = \mathbb{E} [U^{-1} | d]$  et  $\varphi_3 = \mathbb{E} [U | d]$  est effectué par le moment d'ordre  $p$  de la distribution de probabilité de  $U$  où  $d|U \sim \mathcal{N}(0, u + \sigma_\epsilon^2)$

La preuve de l'étape M de l'algorithme est identique à celle dans 3.6.4.2.

Notons aussi que la solution de l'Eq.4.26 est unique puisque le facteur  $\Psi(\hat{\alpha}) - \log(\hat{\alpha})$  est bijectif sur  $\mathbb{R}^+$  et strictement négatif. ■

Par ailleurs, le paramètre  $\sigma_\epsilon$  est estimé par l'estimateur robuste de Donoho & Johnstone [Donoho & Johnstone, 1994].

Dans la Fig.4.3, nous constatons suite à l'observation des courbes que la PDF marginale estimée par l'Eq.4.14 est très proche des histogrammes empiriques observés. Ceci est confirmé par la distance KL calculée entre les deux PDFs.

### 4.3.2.3 Débruiteur bayésien terme-à-terme

La forme analytique de l'estimateur ECP pour l'*a priori* BKF est donnée par le théorème suivant :

**Théorème 4.1** *Pour  $0 < \alpha \leq 1$  et  $c$  strictement positif, l'expression analytique de l'estimateur ECP est comme suit :*

$$\hat{s}_{ECP}(d) = \alpha \sigma_\epsilon \frac{e^{\left(\frac{-\frac{d}{\sigma_\epsilon} + \sigma_\epsilon \sqrt{\frac{2}{c}}}{2}\right)^2} D_{-\alpha-1}\left(-\frac{d}{\sigma_\epsilon} + \sigma_\epsilon \sqrt{\frac{2}{c}}\right) - e^{\left(\frac{\frac{d}{\sigma_\epsilon} + \sigma_\epsilon \sqrt{\frac{2}{c}}}{2}\right)^2} D_{-\alpha-1}\left(\frac{d}{\sigma_\epsilon} + \sigma_\epsilon \sqrt{\frac{2}{c}}\right)}{e^{\left(\frac{-\frac{d}{\sigma_\epsilon} + \sigma_\epsilon \sqrt{\frac{2}{c}}}{2}\right)^2} D_{-\alpha}\left(-\frac{d}{\sigma_\epsilon} + \sigma_\epsilon \sqrt{\frac{2}{c}}\right) + e^{\left(\frac{\frac{d}{\sigma_\epsilon} + \sigma_\epsilon \sqrt{\frac{2}{c}}}{2}\right)^2} D_{-\alpha}\left(\frac{d}{\sigma_\epsilon} + \sigma_\epsilon \sqrt{\frac{2}{c}}\right)} \quad (4.29)$$

L'estimateur ECP est une fonction non-linéaire impaire.

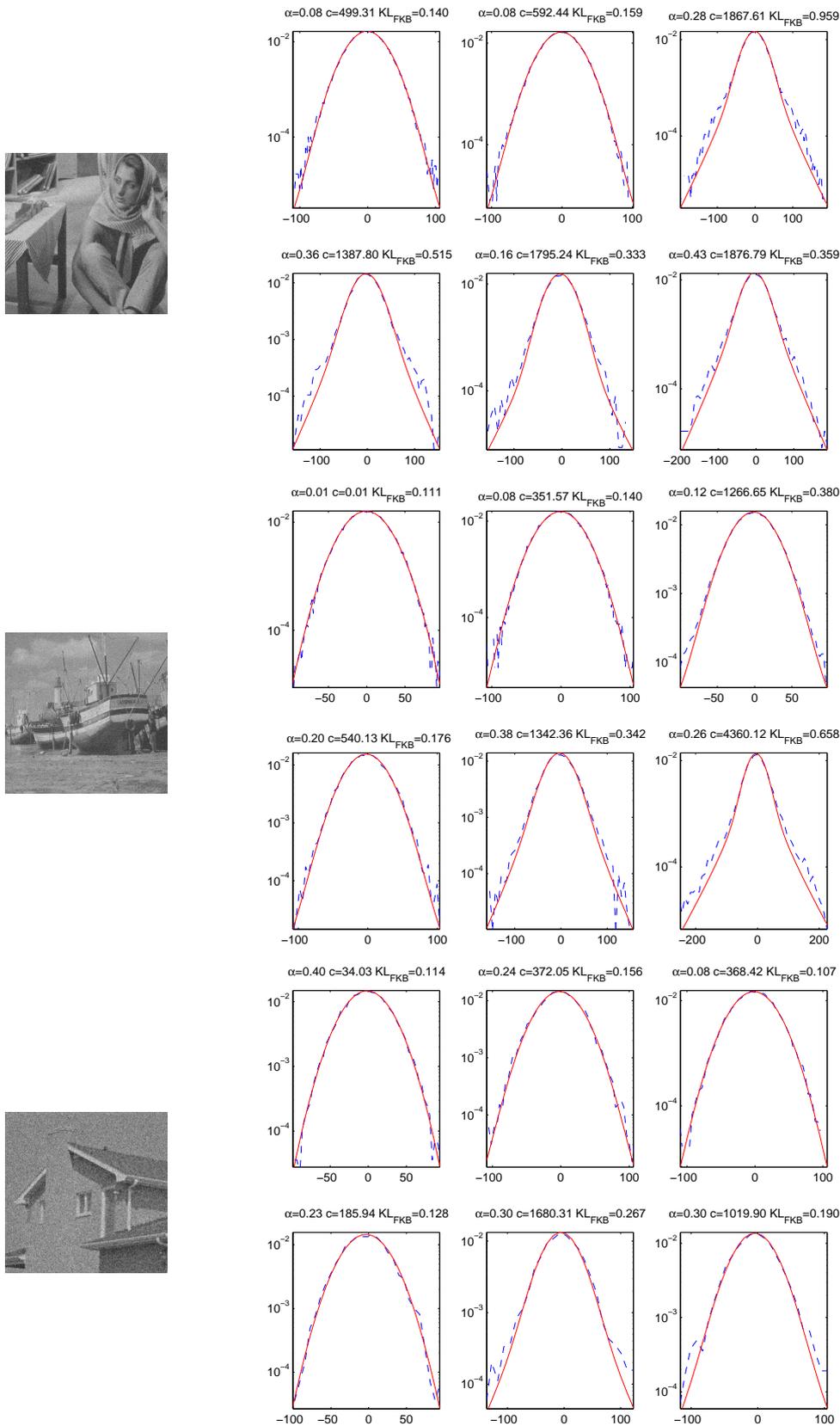


FIG. 4.3 – Comparaison entre l'histogramme marginal observé (trait en pointillé) et la PDF BKF estimée (trait plein) des coefficients de détail d'ondelettes pour trois d'images tests corrompues par un bruit blanc gaussien ( $SNR = 15dB$ ). Seulement, deux niveaux de décomposition sont montrés pour les trois images test : Barbara, Boat et House. Les trois colonnes correspondent aux orientations  $HH$ ,  $HL$  et  $LH$ .

**Preuve:**

L'expression du dénominateur de l'estimateur ECP (cf. annexe A) a été établie dans le lemme 4.1. Pour déterminer l'expression du numérateur, on suivra les mêmes démarches que dans le lemme 4.1.

En décomposant l'intégrale dans le numérateur pour  $s$  positif et négatif, nous pouvons montrer que

$$\int_{-\infty}^{+\infty} s\mathcal{P}(d-s)\mathcal{P}(s)ds = \alpha\sigma_\epsilon \left(\frac{\sigma_\epsilon^2}{2c}\right)^{\frac{\alpha}{2}} \phi(d; 0, \sigma_\epsilon^2) (I'_- - I'_+) \quad (4.30)$$

$$\text{où } I'_\pm = \int_0^{+\infty} s^\alpha \exp\left(-\frac{s^2}{2\sigma_\epsilon^2} - s\left(\pm\frac{d}{\sigma_\epsilon} + \sqrt{\frac{2}{c}}\right)\right) ds$$

A partir de la formule d'intégrale donnée par l'Eq.4.20, et pour  $\alpha + 1 > 0$  l'intégrale de  $I'_\pm$  s'écrit sous la forme suivante :

$$I'_\pm = \exp\left(\left(\frac{\pm\frac{d}{\sigma_\epsilon} + \sigma_\epsilon\sqrt{\frac{2}{c}}}{2}\right)^2\right) D_{-p-1}\left(\pm\frac{d}{\sigma_\epsilon} + \sigma_\epsilon\sqrt{\frac{2}{c}}\right) \quad (4.31)$$

Cette expression est valide pour  $\alpha > 0$ .

Finalement, en combinant le lemme 4.1, l'Eq.4.30 et l'Eq.4.31, on obtient le résultat du théorème 4.1. ■

La Fig.4.4 montre l'évolution de la fonction  $s(d)$  de l'estimateur bayésien ECP donné par le théorème 4.1. Les graphes de la Fig.4.4(a) (resp. Fig.4.4(b)) montrent l'influence du rapport  $\frac{\sqrt{c}}{\sigma_\epsilon}$  (resp. du paramètre de forme  $\alpha$ ) sur les courbes du taux de contraction en fonction du paramètre  $\alpha \in [0, 1]$  (resp. du rapport  $\frac{\sqrt{c}}{\sigma_\epsilon}$ ) où le rapport  $\frac{\sqrt{c}}{\sigma_\epsilon}$  peut être vu comme une forme de mesure du SNR.

Nous constatons, suite à l'observation des courbes, que la règle bayésienne proposée contracte fortement les coefficients d'observations  $d$  ayant une valeur trop petite et légèrement les coefficients de valeur grande. De plus, la contraction diminue et s'approche de la ligne d'identité quand  $|d| \rightarrow \infty$ . Ce comportement asymptotique peut être prouvé en utilisant l'expansion asymptotique de  $D_\nu(x)$  pour  $|x|$  très grand [Abramowitz & Stegun, 1972]. Aussi, lorsque la valeur du paramètre de forme  $\alpha$  augmente, la contraction diminue d'autant plus que le rapport  $\frac{\sqrt{c}}{\sigma_\epsilon}$  augmente.

De plus, et dans la Fig.4.4(b), nous constatons que la contraction diminue quand  $\alpha$  augmente. Cela peut s'expliquer par le fait que plus la valeur de  $\alpha$  est grande et plus la queue de la distribution *a priori* est importante, d'où une plus grande probabilité qu'une grande valeur mesurée (située en queue de distribution) soit due à la présence de signal <sup>8</sup>.

Cet estimateur bayésien (Eq.4.29) est seulement valide pour  $\alpha \leq 1$  ce qui n'est pas restrictif pour la plupart des images naturelles filtrées ayant une valeur de  $\alpha \leq 1$ . Néanmoins, dans la pratique, si  $\alpha$  est grand (i.e., la PDF tend vers une gaussienne), alors l'estimateur de Wiener  $\left(\frac{\alpha c}{\alpha c + \sigma_\epsilon^2}\right) d$  peut être utilisé.

<sup>8</sup>Un comportement analogue a été observé avec l'estimateur bayésien ECP  $\alpha$ -stable.

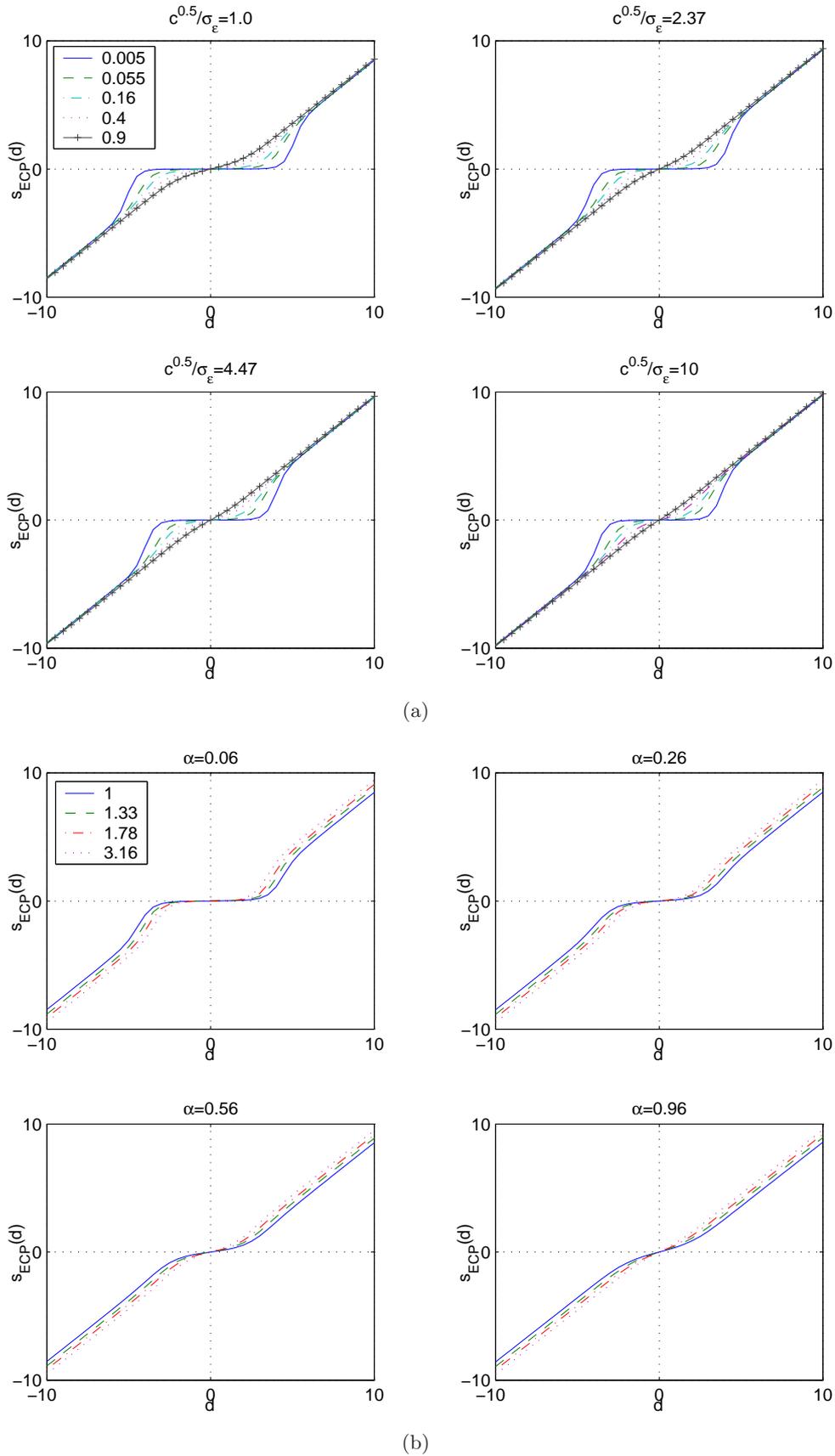


FIG. 4.4 – Évolution de la fonction  $s(d)$  de l'estimateur bayésien ECP BKF. (a) En fonction du rapport  $\frac{\sqrt{c}}{\sigma_\epsilon}$ . (b) En fonction du paramètre  $\alpha$ .

## 4.4 Estimateur maximum a posteriori (MAP)

### 4.4.1 Estimateur MAP avec l'a priori BKF

Après avoir établi l'expression analytique de la PDF marginale BKF des coefficients d'observations  $d$  (Eq.4.14) ainsi que les estimateurs des hyperparamètres  $\{\alpha, c\}$  (Eq.4.22 et algorithme 4.3.2.2 [page 86]), l'expression analytique de l'estimateur MAP (cf. annexe A) est donnée par le théorème suivant :

**Théorème 4.2** *Pour  $0 < \alpha \leq 1$  et  $c$  strictement positif.*

- *L'expression analytique de l'estimateur MAP BKF s'écrit comme suit :*

$$\hat{s}_{MAP}(d) = \begin{cases} 0 & |d| \leq \lambda \\ \frac{\text{sign}(d)}{2} \left( (|d| - \sqrt{\frac{2}{c}\sigma_\epsilon^2}) + \sqrt{(|d| - \sqrt{\frac{2}{c}\sigma_\epsilon^2})^2 + 4\sigma_\epsilon^2(\alpha - 1)} \right) & |d| > \lambda \end{cases} \quad (4.32)$$

$$\text{où } \lambda = \sqrt{2}\sigma_\epsilon \left( \sqrt{2(1 - \alpha)} + \frac{\sigma_\epsilon}{\sqrt{c}} \right)$$

- *Pour  $d \rightarrow \infty$ ,*

$$\hat{s}_{MAP}(d) = d \left( 1 - \sqrt{\frac{2}{c} \frac{\sigma_\epsilon^2}{|d|}} + \mathcal{O}(|d|^{-2}) \right) \quad (4.33)$$

#### Preuve:

Nous rappelons l'expression de l'estimateur MAP des coefficients  $s$  conditionnellement sur l'ensemble des hyperparamètres  $\theta = \{\alpha, c, \sigma_\epsilon\}$ ,

$$\begin{aligned} \hat{s}_{MAP}(d) &= \arg \max_s \log [\mathcal{P}(s|d)] \\ &= \arg \max_s [\log \phi(d - s; \sigma_\epsilon^2) + \log \mathcal{P}(s; \alpha, c)] \end{aligned} \quad (4.34)$$

où  $\phi(\cdot)$  est la PDF normale de variance  $\sigma_\epsilon^2$ , et  $\mathcal{P}(s; \alpha, c)$  est la PDF BKF donnée par l'Eq.4.18. Notons tout d'abord que la solution est impaire ; i.e,  $s(-d) = -s(d)$

En dérivant l'Eq.4.34, nous obtenons une équation du second degré :

$$s^2 - \left( |d| - \sqrt{\frac{2}{c}\sigma_\epsilon^2} \right) s - \sigma_\epsilon^2(\alpha - 1) = 0 \quad (4.35)$$

Pour qu'une solution réelle existe, le déterminant doit être positif ou nul.

$$\Delta = \left( |d| - \sqrt{\frac{2}{c}\sigma_\epsilon^2} \right)^2 + 4\sigma_\epsilon^2(\alpha - 1) > 0 \quad (4.36)$$

Ceci implique que

$$|d| > \sqrt{2}\sigma_\epsilon \left( \sqrt{2(1 - \alpha)} + \frac{\sigma_\epsilon}{\sqrt{c}} \right) \quad (4.37)$$

En notant  $\lambda = \sqrt{2}\sigma_\epsilon \left( \sqrt{2(1 - \alpha)} + \frac{\sigma_\epsilon}{\sqrt{c}} \right)$ , l'Eq.4.35 admet donc deux solutions pour  $|d| > \lambda$  :

$$\hat{s} = \frac{\left(|d| - \sqrt{\frac{2}{c}}\sigma_\epsilon^2\right) \pm \sqrt{\left(|d| - \sqrt{\frac{2}{c}}\sigma_\epsilon^2\right)^2 + 4\sigma_\epsilon^2(\alpha - 1)}}{2} \quad (4.38)$$

Seule la solution positive préserve le signe (i.e,  $s(d) > 0$  si  $d > 0$ ) et les valeurs de la solution possible (e.g.,  $\alpha = 1 \implies \hat{s}(d) = 0 \forall d$ ); d'où le résultat de l'Eq.4.32 du théorème.

L'Eq.4.33 du théorème peut être démontrée à partir de l'Eq.4.32 comme suit

$$\begin{aligned} \hat{s}_{MAP}(d) &= \frac{\text{sign}(d)}{2} \left( \left(|d| - \sqrt{\frac{2}{c}}\sigma_\epsilon^2\right) + \sqrt{\left(|d| - \sqrt{\frac{2}{c}}\sigma_\epsilon^2\right)^2 + 4(\alpha - 1)\sigma_\epsilon^2} \right) \\ &= \frac{\text{sign}(d)}{2} \left( \left(|d| - \sqrt{\frac{2}{c}}\sigma_\epsilon^2\right) + \sqrt{d^2 - 2|d|\sqrt{\frac{2}{c}}\sigma_\epsilon^2 + \frac{2\sigma_\epsilon^4}{c} + 4(\alpha - 1)\sigma_\epsilon^2} \right) \\ &= \frac{\text{sign}(d)}{2} \left( \left(|d| - \sqrt{\frac{2}{c}}\sigma_\epsilon^2\right) + \sqrt{d^2 \left(1 - \frac{2}{|d|}\sqrt{\frac{2}{c}}\sigma_\epsilon^2 + \frac{2\sigma_\epsilon^4}{cd^2} + \frac{4(\alpha - 1)\sigma_\epsilon^2}{d^2}\right)} \right) \\ &= \frac{\text{sign}(d)}{2} \left( |d| \left(1 - \sqrt{\frac{2}{c}}\frac{\sigma_\epsilon^2}{|d|}\right) + |d|\sqrt{1 - \frac{2}{|d|}\sqrt{\frac{2}{c}}\sigma_\epsilon^2 + \mathcal{O}(d^{-2})} \right) \\ &= \frac{d}{2} \left( 1 - \sqrt{\frac{2}{c}}\frac{\sigma_\epsilon^2}{|d|} + \sqrt{1 - \frac{2}{|d|}\sqrt{\frac{2}{c}}\sigma_\epsilon^2 + \mathcal{O}(d^{-2})} \right) \\ &= \frac{d}{2} \left( 1 - \sqrt{\frac{2}{c}}\frac{\sigma_\epsilon^2}{|d|} + 1 - \sqrt{\frac{2}{c}}\frac{\sigma_\epsilon^2}{|d|} + \mathcal{O}(d^{-2}) \right) \\ &= d \left( 1 - \sqrt{\frac{2}{c}}\frac{\sigma_\epsilon^2}{|d|} + \mathcal{O}(|d|^{-2}) \right) \end{aligned}$$

■

**Proposition 4.1** *L'estimateur MAP est de type seuillage doux pour  $\frac{\sigma_\epsilon^2}{c} = \log m$  lorsque  $\alpha \rightarrow 1$  (a priori Laplacien) ou bien  $m$  grand, où  $m$  est le nombre d'échantillons.*

**Preuve:**

Il suffit lorsque cet estimateur MAP est de type seuillage doux, que  $\lambda = \lambda_U$ , où  $\lambda_U$  est le seuil défini par l'Eq.1.11. Et donc,

$$2\sigma_\epsilon \left( \sqrt{2(1 - \alpha)} + \frac{\sigma_\epsilon}{\sqrt{c}} \right) = \sigma_\epsilon \sqrt{2 \log m} \implies \frac{\sigma_\epsilon^2}{c} = \log m$$

■

La Fig.4.5 montre l'évolution de la fonction  $s(d)$  de notre estimateur bayésien MAP BKF. Nous constatons que l'estimateur MAP est bien une contraction, impaire et continue en  $d$  et  $\lambda$ . Il est toujours en dessous de la ligne d'identité et s'en approche quand  $d \rightarrow \infty$  comme le prévoit le théorème 4.2. De plus, nous constatons que l'estimateur MAP est de type seuillage doux pour des SNRs faibles. La même remarque a été soulevée par Moulin [Moulin & Liu, 1999] pour l'*a priori* GGD.

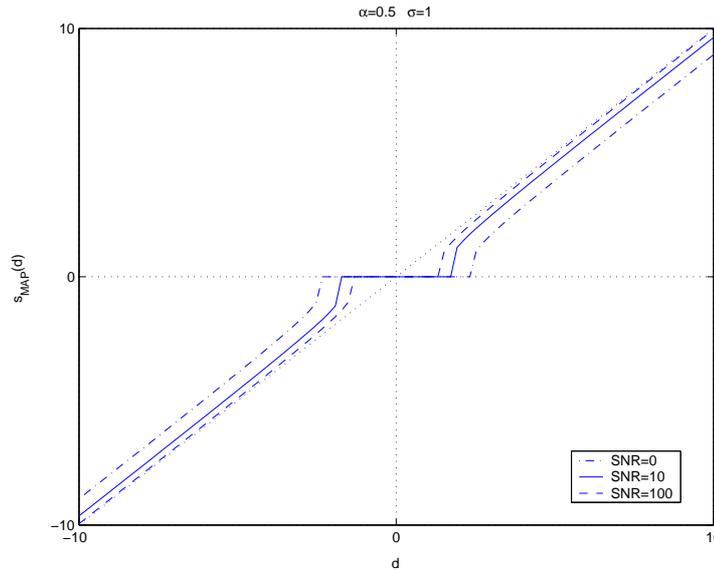


FIG. 4.5 – Évolution de la fonction  $s(d)$  de l'estimateur bayésien MAP BKF en fonction de  $d$  pour différentes valeurs de SNR (en faisant varier  $c$ ).

## 4.5 Expérimentation et résultats

### 4.5.1 Objectifs de l'expérience

Nous évaluons maintenant les performances de nos différents débruiteurs à savoir l'estimateur ECP  $\alpha$ -stable (dénomé dans la suite " $\alpha$ -stable mixture"), l'estimateur ECP BKF et l'estimateur MAP BKF, en les comparant à diverses méthodes de débruitage univariée développées dans un contexte classique et bayésien.

Six algorithmes de débruitage sont considérés : les seuillages universels dur (Hard universal) et doux (Soft universal) [Donoho & Johnstone, 1994], l'estimateur SURE basé sur l'estimation du risque sans biais de Stein [Donoho & Johnstone, 1995], l'estimateur VISU (Oracle Threshold) [Donoho & Johnstone, 1998b, Chang et al., 2000a], l'estimateur MAP GGD [Moulin & Liu, 1999] et une version d'un débruiteur bayésien avec un *a priori*  $\alpha$ -stable proposé dans [Achim et al., 2001] et développé dans [Mathieu, 2002]. Aucune forme analytique n'étant disponible pour ce dernier, celui-ci était basé sur l'implémentation numérique des intégrales de Fourier à base de FFT [Press et al., 1992], car les intégrations numériques directes posent de sérieuses difficultés, telles que des bornes infinies, qui les rendent très lentes et numériquement instables.

### 4.5.2 Protocole expérimental

TAB. 4.1 – Protocole expérimental

Grandeur	Choix			
	Fig.4.6-4.7-4.8	Fig.4.9	Fig.4.10	Fig.4.11
<i>a priori</i> univarié	$\alpha$ -stable, BKF, GGD			
Nos débruiteurs	ECP $\alpha$ -stable mixture, ECP BKF, MAP BKF			
Estimateur des hyperparamètres	Algorithme EM			
Nombre de gaussiennes optimal	16			
Méthodes de la littérature	$\alpha$ -stable, MAP GGD, Soft&Hard universal, VISU, SURE			
Transformée	ondelettes			
DWT	5 échelles et 3 orientations			
qmf	Daubechies-8			
Bruit	additif blanc gaussien			
SNR	15 dB ( $\sigma_\epsilon = 20$ )		[5,20] dB	
Critère d'évaluation	SNR			
Monte-Carlo			50	
Base d'image	Lena Barbara Mandrill	Barbara	Lena Barbara Mandrill	100 images

### 4.5.3 Discussion des résultats

Afin de quantifier les performances de ces différents estimateurs, nous avons employé le rapport SNR. Ce rapport est défini en décibel comme suit

$$\text{SNR} = 10 \log_{10} \frac{\sum_{mn} g_{mn}^2}{\sum_{mn} (\hat{g}_{mn}^2 - g_{mn}^2)^2} \quad (4.39)$$

où  $\hat{g}$  est l'estimée de  $g$ . Certaines méthodes dans la littérature rapportent leurs résultats en termes de mesure du PSNR défini par

$$\text{PSNR} = 10 \log_{10} \frac{255^2}{\sum_{mn} \frac{1}{N} (\hat{g}_{mn}^2 - g_{mn}^2)^2} \quad (4.40)$$

où  $N \times N$  est la taille de l'image bruitée.

Dans ce chapitre, nous avons comparé à des méthodes de la littérature qui elles reportent leurs performances en terme de SNR et non pas de PSNR. Ceci dit, si la finalité est la comparaison des débruiteurs, ces critères aboutissent à des résultats semblables.

Les figures ci-dessous, Fig.4.6-4.7-4.8, présentent les performances obtenues pour des images de test d'une base de 100 images digitalisées [url : base d'images, ]. Nous avons utilisé l'ondelette à support compact de Daubechies de régularité 8. L'échelle la plus grossière de la décomposition est choisie pour être  $\log_2 (\log(N^2) + 1)$  à partir des considérations asymptotiques [Antoniadis et al., 2001].

La Fig.4.6 montre les images résultantes, pour chaque méthode de débruitage, pour l'image de Lena avec un SNR=15dB en entrée. On voit que la qualité visuelle des débruiteurs bayésiens  $\alpha$ -stable mixture et ECP BKF est supérieure à celles des autres méthodes de débruitage, mais reste comparable au MAP BKF. Ce comportement a tendance à se reproduire pour les autres images tests : Barbara (Fig.4.7) et Mandrill (Fig.4.8). Le zoom sur une région texturée du pantalon de Barbara montre que les débruiteurs que nous proposons réalisent un bon compromis entre le rejet du bruit et la conservation des détails fins de l'image (e.g. les rayures du pantalon). Nos débruiteurs affichent des performances bien au delà de ce que peut offrir le débruiteur bayésien  $\alpha$ -stable dans sa version numérique dite "exacte" développée dans [Mathieu, 2002, Achim et al., 2001].

Dans la Fig.4.9, nous avons évalué les performances de débruiteurs BKF de type ECP et MAP basés sur la méthode de cumulants et l'algorithme EM pour l'estimation des hyperparamètres. Finalement, le débruiteur BKF ECP (resp. MAP) avec algorithme EM dépasse légèrement en terme de SNR le débruiteur BKF ECP (resp. MAP) avec cumulants, mais reste comparable.

Les moyennes du SNR (sur 50 simulations), données par les diverses méthodes pour les trois images tests, sont comparées dans la Fig4.10. Les SNRs ont été calculés pour chaque valeur d'entrée du SNR dans la gamme de [5, 20]dB. On constate que les débruiteurs  $\alpha$ -stable mixture et ECP BKF présentent des performances similaires, et ils dépassent l'estimateur MAP BKF, mais restent meilleurs comparés aux autres débruiteurs surtout pour les bas SNR. L'estimateur  $\alpha$ -stable dans sa version originale affiche des performances médiocres pour les bas SNR à cause de la faiblesse de l'estimateur des hyperparamètres. On constate aussi que l'estimateur VISU (seuil Oracle) s'avère d'autant plus performant comparé à l'estimateur SURE et aux seuillages universels que le rapport SNR est important.

Dans la Fig.4.11, on compare les SNRs moyennés sur les 50 simulations et les 100 images de la base, de toutes les méthodes de débruitage. Le comportement général décrit avant est confirmé par ce graphe.



FIG. 4.6 – Comparaison des différents débruiteurs sur l'image test "Lena". Cette image a été corrompue par un bruit additif gaussien de 15dB en entrée.

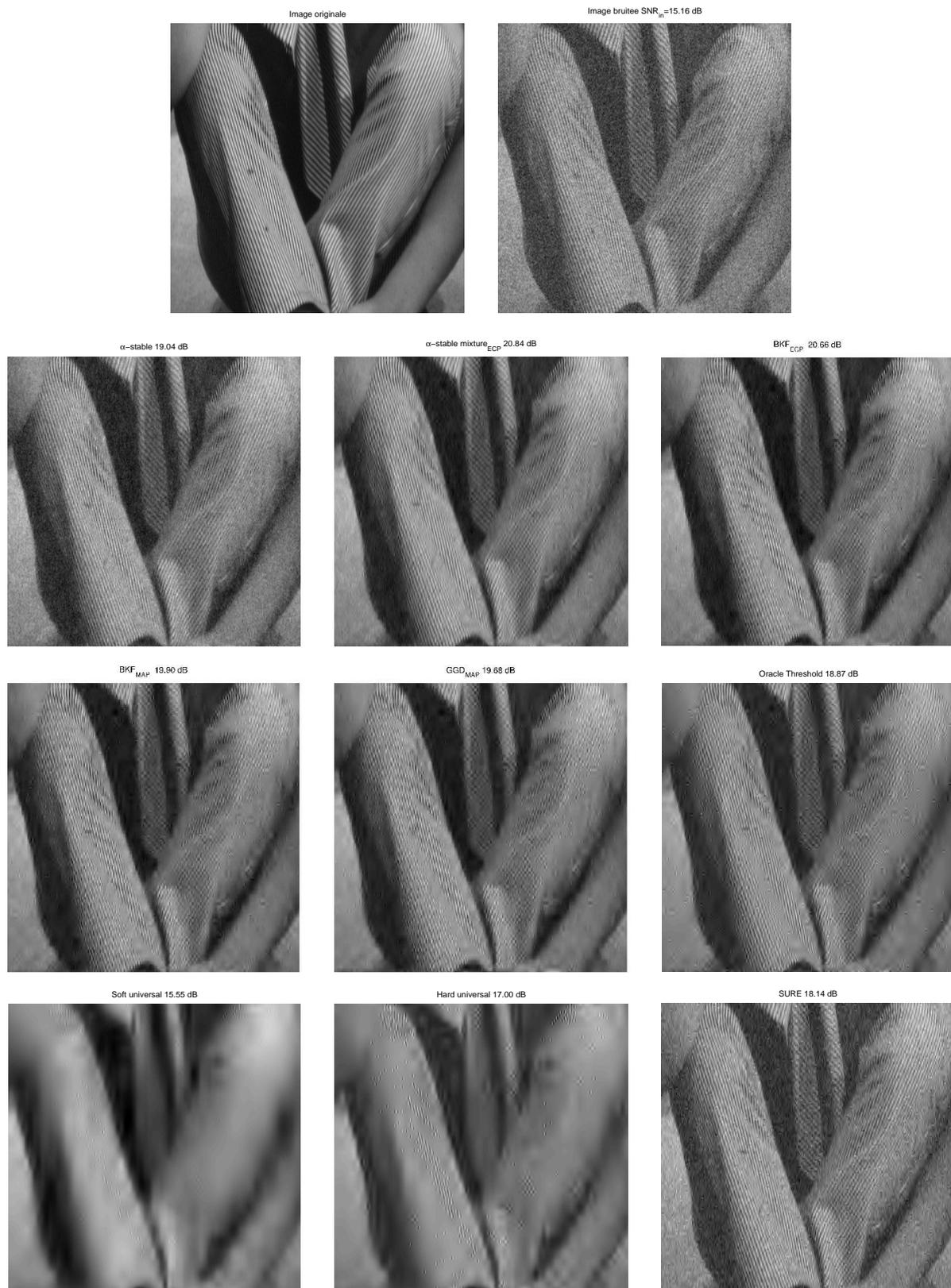


FIG. 4.7 – Comparaison des différents débruiteurs sur l'image test "Barbara" zoomée sur une région texturée du pantalon. Cette image a été corrompue par un bruit additif gaussien de 15dB en entrée. Le zoom montre que les débruiteurs que nous proposons réalise un bon compromis entre le rejet du bruit et la conservation des détails fins de l'image.

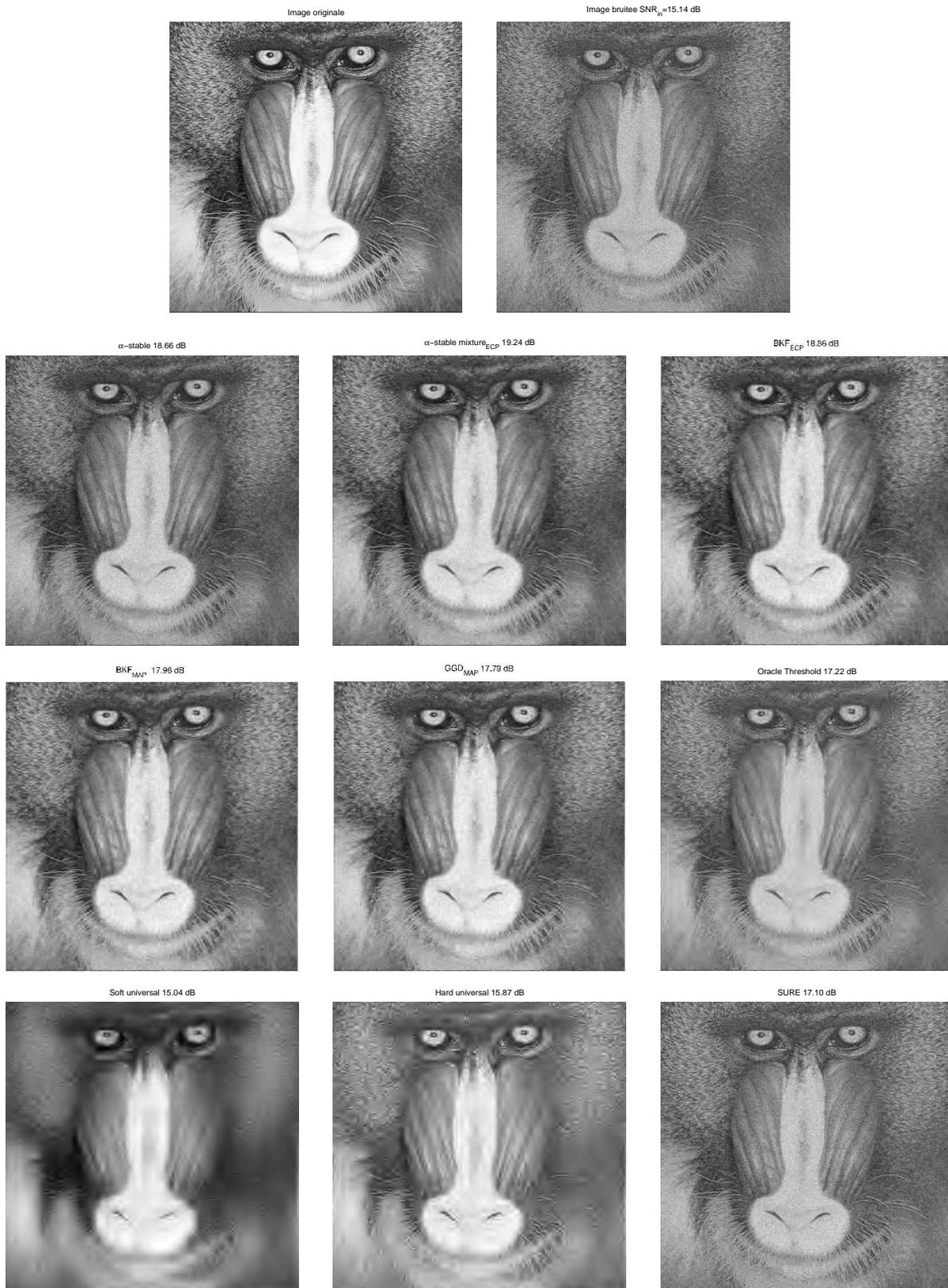


FIG. 4.8 – Autre comparaison des différents débruiteurs sur l'image test "Mandrill". Cette image a été corrompue par un bruit additif gaussien de 15dB en entrée.

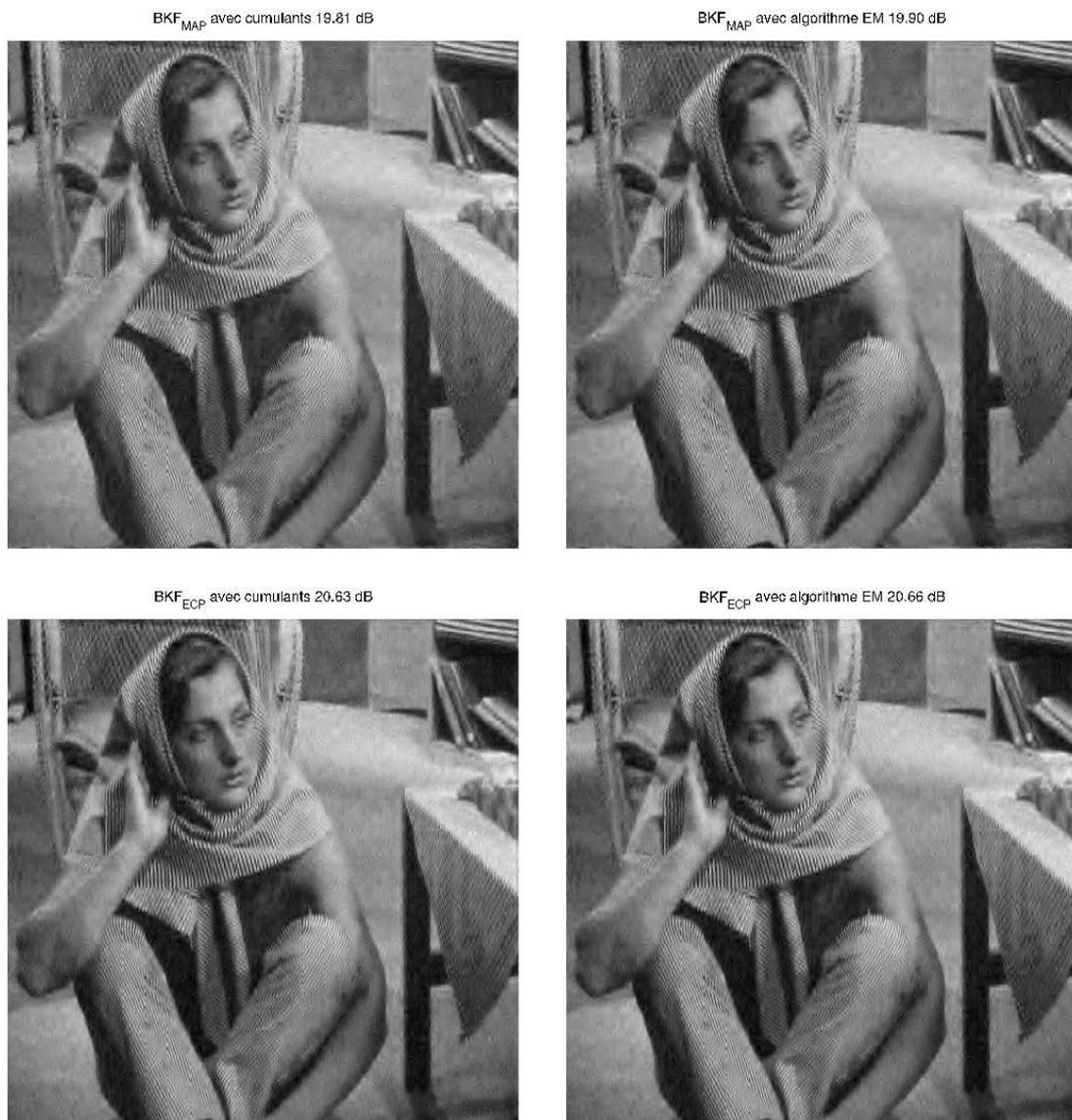
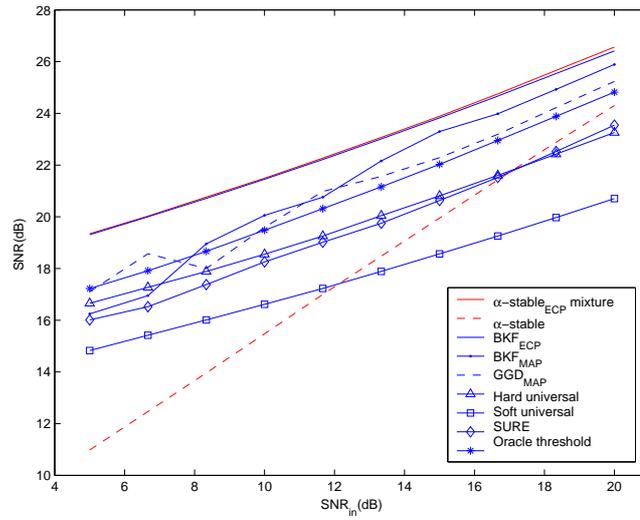
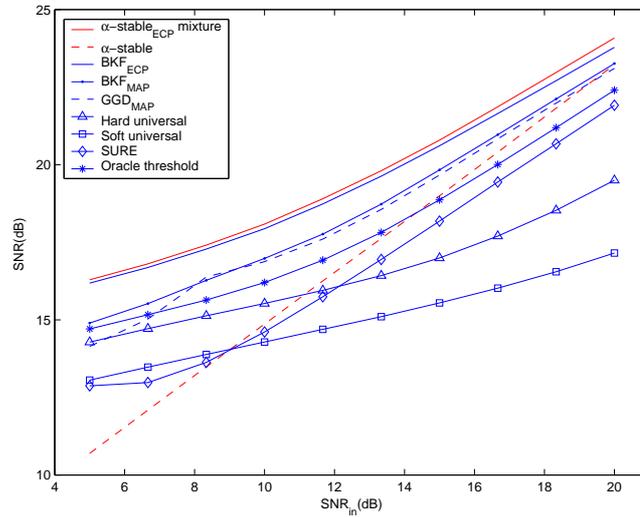


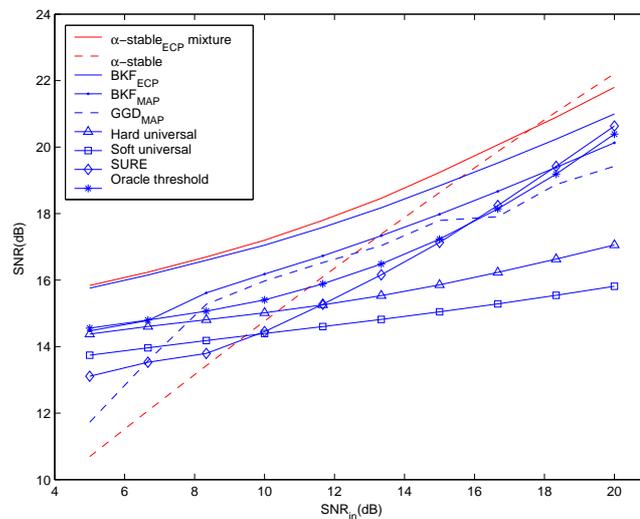
FIG. 4.9 – Évaluation des performances de débruiteurs BKF de type ECP et MAP basés sur la méthode de cumulants et l'algorithme EM pour l'estimation des hyperparamètres.



(a) Lena



(b) Barbara



(c) Mandrill

FIG. 4.10 – Comparaison de différents estimateurs de débruitage pour les trois images test (Lena, Barbara et Mandrill), pour un SNR à l'entrée de 5 à 20dB.

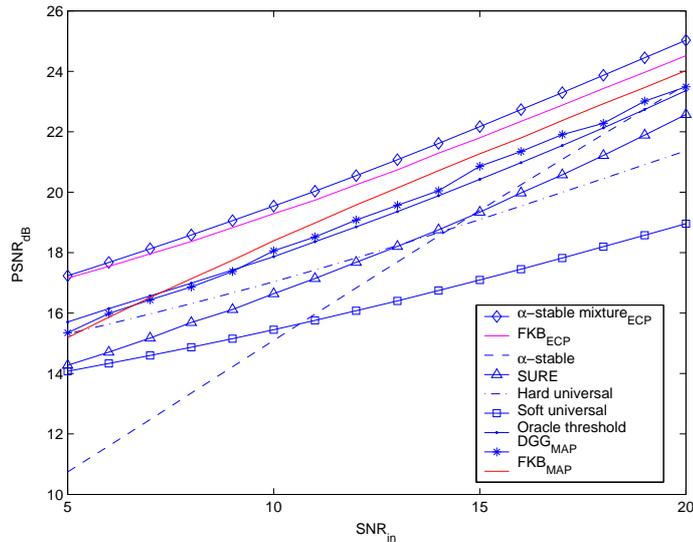


FIG. 4.11 – Comparaison de différents estimateurs de débruitage pour une base de 100 images.

## 4.6 Conclusion

Nous avons proposé trois nouveaux estimateurs bayésiens non-linéaires non-paramétriques de type ECP et MAP, basés sur le modèle statistique *a priori* SMG, formulés dans le cadre de la transformée en ondelettes. Ils peuvent néanmoins très bien s'appliquer à d'autres transformées.

Du fait du caractère creux des représentations parcimonieuses (e.g. ondelettes), la plupart des coefficients de détail de l'image bruitée sont nuls et seuls quelques uns sont significatifs. Ainsi, la répartition de ces coefficients est caractérisée par une densité centrée en zéro avec des queues relevées. C'est exactement cette propriété qui est révélée par la PDF SaS et BKF que nous avons pleinement caractérisée, et par lesquels nous avons proposé des méthodes pour estimer les hyperparamètres en présence de bruit. Ensuite, nous avons appliqué la règle de Bayes pour combiner l'information contenue dans les données d'observation (image bruitée) et celle contenue dans la loi *a priori* afin d'obtenir l'expression analytique de la distribution marginale *a posteriori* ainsi que celle de débruiteur bayésien terme à terme.

Les résultats expérimentaux, obtenus avec une base d'ondelettes orthogonale décimée, ont montré que les performances de nos débruiteurs s'avèrent supérieures à celles des débruiteurs développés dans un contexte bayésien et classique pour une grande classe d'images. Cette performance reste liée à la qualité des estimateurs des hyperparamètres.

Dans ce chapitre, nous n'avons pas comparé nos méthodes en utilisant les transformées multi-échelles orientées (curvelets) et les ondelettes non-décimées, pour une simple raison ; c'est que le bruit dans ce cas n'est plus blanc. En effet, ces transformées correspondent à des trames voire des trames ajustées. Toutefois, une comparaison de nos estimateurs univariés basés sur l'*a priori* BKF est associée à ces transformées sera effectuée au chapitre 6.

Enfin, il reste un aspect qui doit être étudié pour améliorer la performance de nos débruiteurs. Il s'agit de l'introduction de l'information géométrique. Pour ce faire, nous avons proposé un cadre statistique bayésien multivarié permettant de modéliser la distribution *a priori* des coefficients dans le domaine des transformées multi-échelles orientées et non-orientées (e.g. les curvelets [Candès & Donoho, 1999, Starck et al., 2002]) et de prendre en compte leurs dépendances inter et intra-échelle. Ceci fera l'objet de la troisième partie de cette thèse.



Troisième partie

---

Statistiques multivariées pour la restauration

---



## Chapitre 5

# Modélisation des statistiques jointes

---

### Sommaire

---

<b>5.1</b>	<b>Étude et analyse statistique . . . . .</b>	<b>106</b>
5.1.1	Curvelets et notations . . . . .	106
5.1.2	Étude statistique . . . . .	106
5.1.2.1	Distribution jointe . . . . .	106
5.1.3	Étude quantitative . . . . .	108
5.1.3.1	Information mutuelle . . . . .	108
<b>5.2</b>	<b>Modèle SMG multivarié : cadre général . . . . .</b>	<b>109</b>
5.2.1	Définition . . . . .	109
5.2.2	Propriétés . . . . .	109
5.2.3	Moments absolus d'ordre $p$ . . . . .	110
5.2.4	Cas particuliers du SMG multivarié . . . . .	110
<b>5.3</b>	<b>Estimation des hyperparamètres . . . . .</b>	<b>110</b>
<b>5.4</b>	<b>Cas 1 : modèle BKF multivarié . . . . .</b>	<b>111</b>
5.4.1	Définition . . . . .	111
5.4.2	Propriétés . . . . .	111
5.4.3	Estimation des hyperparamètres . . . . .	113
5.4.3.1	Méthode des moments . . . . .	113
5.4.3.2	Algorithme EM . . . . .	113
<b>5.5</b>	<b>Qualité de l'a priori : application à la modélisation des statis- tiques jointes . . . . .</b>	<b>115</b>
5.5.1	Objectifs de l'expérience . . . . .	115
5.5.2	Protocole expérimental . . . . .	115
5.5.3	Discussion des résultats . . . . .	115
<b>5.6</b>	<b>Conclusion . . . . .</b>	<b>116</b>

---

Nous avons observé, au chapitre 2, que les coefficients des images dans l'espace des transformées tendent à se regrouper autour des bords des objets dans l'image (e.g. les contours, les lignes, etc). Ce phénomène persiste également à d'autres orientations et à d'autres échelles. Dans la partie de modélisation des statistiques marginales, nous avons supposé que ces coefficients sont indépendants. Cette hypothèse n'est qu'une approximation qui doit être affinée en prenant en compte les dépendances inter- et intra-échelles existant entre les coefficients d'images.

Dans ce chapitre, nous présentons une étude théorique et empirique pour caractériser ces dépendances inter- et intra-échelles des coefficients de détails d'images dans le domaine des transformées multi-échelles orientées (e.g. transformée de curvelets) et non-orientées (e.g. UDWT). Ensuite, en vue de caractériser ces dépendances, nous introduisons le modèle de mélange d'échelles de gaussiennes multivarié. Ce modèle *a priori* multivarié est adéquat pour la modélisation statistique d'une grande classe de PDFs à queues lourdes.

## 5.1 Étude et analyse statistique

Dans cette section, nous présentons une analyse statistique qualitative et quantitative des dépendances en position, échelle et orientation des coefficients des images dans le domaine des curvelets de 2<sup>ème</sup> génération<sup>9</sup>, basée sur les histogrammes estimés de distributions marginales et conjointes, et sur la mesure de l'information mutuelle [Cover & Thomas, 1991] pour caractériser ces dépendances. Ceci afin de pouvoir modéliser ces interactions entre les coefficients de façon simple à travers un modèle statistique multivarié, à savoir le modèle de mélange d'échelles de gaussiennes multivarié (MSMG), que nous décrivons pleinement par la suite.

### 5.1.1 Curvelets et notations

La transformée de curvelets est une transformée multi-échelle multi-directionnelle avec des atomes indexés par un paramètre de position, d'échelle et de direction [Candès & Donoho, 1999, Candès & Donoho, 2002] (cf. annexe D). La Fig.5.1 montre une représentation de la transformée de curvelets pour l'image test Bateau, en employant la transformée de curvelets discrète de 2<sup>ème</sup> génération (FDCT) [Candès et al., 2006]. Trois niveaux d'échelles et six orientations ont été montrés.

Pour chaque coefficient de curvelet  $X$ , on définit ses *voisins* ( $VX$ ) dans la même orientation, qui représentent les huit coefficients adjacents. Ensuite, le coefficient à la même localisation spatiale dans l'échelle supérieure correspond à son *père* ( $PX$ ), et les coefficients à la même localisation spatiale et à la même échelle mais dans une autre orientation correspondent à ses *cousins* ( $CX$ ). Aussi, y-a-t'il plus d'orientations dans la représentation de curvelets comparée à la représentation en ondelettes séparable où il n'y a seulement trois directions cardinales (horizontale, verticale et diagonale). La Fig.5.2 récapitule ces relations importantes entre coefficients.

### 5.1.2 Étude statistique

Nous présentons ici une étude statistique pour apprécier les propriétés des coefficients de curvelet d'images et caractériser les dépendances inter- et intra-échelles de ces derniers.

#### 5.1.2.1 Distribution jointe

Nous avons étudié les statistiques conjointes des coefficients de curvelet. Dans la Fig.5.1, nous observons que les coefficients ayant une valeur élevée tendent à se grouper autour des bords des objets dans l'image test Bateau. Ceci persiste également à d'autres orientations et à d'autres échelles. La Fig.5.3 montre les distributions conjointes des coefficients de curvelet pour l'image Bateau dans le cas bivarié,  $\mathcal{P}(X, \cdot)$ , où  $(\cdot)$  veut dire parent, voisins ou cousins.

<sup>9</sup>cf. Annexe D pour plus de détails couvrant la FDCT.

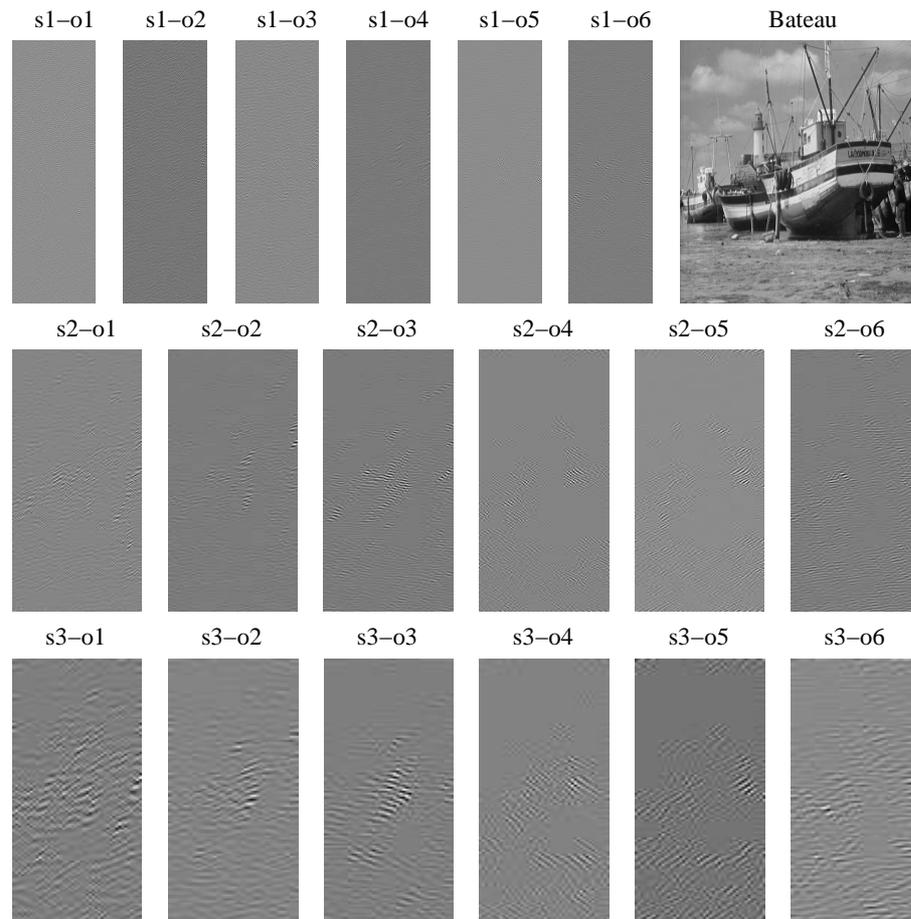


FIG. 5.1 – Transformée de curvelets discrète de 2<sup>ème</sup> génération (FDCT) de l'image test Bateau. Trois niveaux d'échelles et six orientations sont montrés ( $j$  : indique l'échelle,  $o$  : indique l'orientation).

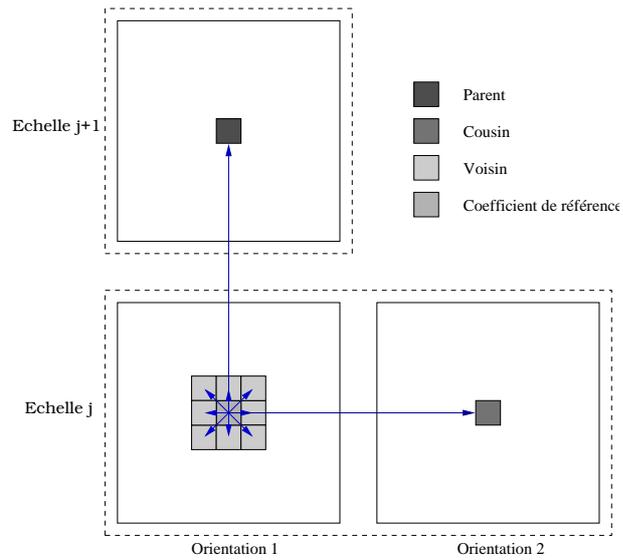


FIG. 5.2 – Dépendances inter- et intra- individuelles en position, échelle et orientation entre les coefficients de curvelet.

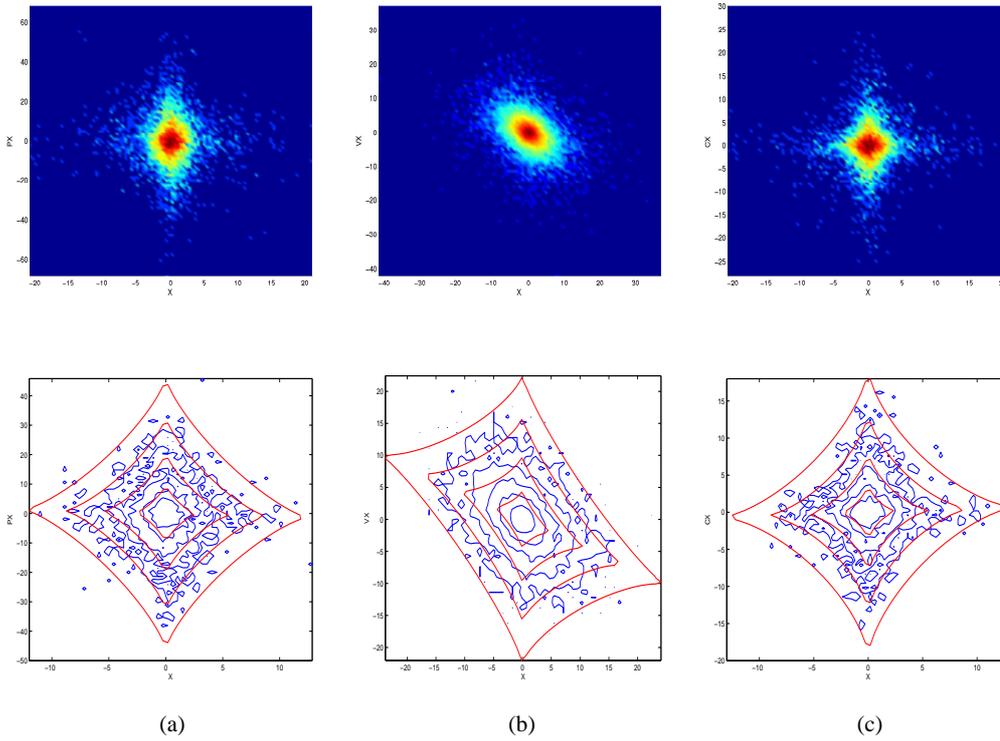


FIG. 5.3 – Distributions conjointes (en log) des coefficients de curvelet pour l’image test Bateau dans le cas bivarié,  $\mathcal{P}(X, \cdot)$ , avec leurs parents  $PX$  (a), voisins  $VX$  (b) et cousins  $CX$  (c).

### 5.1.3 Étude quantitative

Nous proposons une étude quantitative des distributions jointes pour comparer les interactions entre les coefficients de curvelet. Ceci est quantifié par le biais de l’information mutuelle (IM) [Cover & Thomas, 1991], qui est une mesure quantitative des dépendances entre les coefficients [Liu & Moulin, 2001, Po & Do, 2003].

#### 5.1.3.1 Information mutuelle

L’information mutuelle (IM) est classiquement définie comme la différence de l’entropie jointe et du produit des entropies marginales. elle est donnée par l’équation suivante :

$$\begin{aligned}
 IM(X; Y) &= D(\mathcal{P}(x, y) \| \mathcal{P}(x)\mathcal{P}(y)) \\
 &= \int_x \int_y \mathcal{P}(x, y) \log \frac{\mathcal{P}(x, y)}{\mathcal{P}(x)p(y)} dx dy
 \end{aligned} \tag{5.1}$$

Le Tab.5.1 montre l’information mutuelle moyennée sur une base de 100 images [url : base d’images, ]. Nous observons que les dépendances les plus significatives proviennent des voisins, suivis des parents et ensuite des cousins. L’influence des voisins est plus prépondérante compte tenu du fait que les coefficients de la FDCT ont tendance à s’agglomérer. La dépendance entre échelles (parent) est révélatrice d’une persistance des coefficients significatifs à travers les échelles.

Finalement, il semble que les dépendances entre orientations (cousins) soient les moins prononcées du fait du partitionnement spectral de la FDCT. Cependant, le système des curvelets

TAB. 5.1 – L'information mutuelle moyennée sur une base d'images (100 images) [url : base d'images, ], calculée pour deux échelles successives  $J$  et  $J - 1$  où  $J$  est l'échelle la plus grossière.

	échelle $J$	échelle $J - 1$
$IM(X; PX)$	0.164	0.194
$IM(X; VX)$	0.374	0.555
$IM(X; CX)$	0.142	0.151

ne forme pas une base orthogonale mais plutôt une frame, il subsiste ainsi des dépendances entre orientations.

## 5.2 Modèle SMG multivarié : cadre général

### 5.2.1 Définition

Soit  $\mathbf{X}$  un vecteur de VA (VVA) à valeurs dans  $\mathbb{R}^d$ . Sous le modèle de mélange d'échelles de gaussiennes multivarié (MSMG), il existe une VA  $U \geq 0$  et un VVA dans  $\mathbb{R}^d$   $\mathbf{Z} \sim \mathcal{N}(0, \Sigma)$  ( $U$  et  $\mathbf{Z}$  sont mutuellement indépendants) tels que :

$$\mathbf{X} \stackrel{d}{=} \mathbf{Z}\sqrt{U} \quad (5.2)$$

où  $\Sigma$  est par définition une matrice symétrique définie positive.

### 5.2.2 Propriétés

Il est assez aisé de déduire les propriétés suivantes :

- La PDF de  $\mathbf{X}$  est donnée par :

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \int_0^{+\infty} u^{-\frac{d}{2}} \exp\left(-\frac{\mathbf{x}^T \Sigma^{-1} \mathbf{x}}{2u}\right) f_U(u) du \quad (5.3)$$

- La fonction caractéristique de  $\mathbf{X}$  est donnée par :

$$\begin{aligned} \psi_{\mathbf{X}}(\boldsymbol{\omega}) &= \mathbb{E} [\mathbb{E} [e^{-i\boldsymbol{\omega}u} | U]] = \mathbb{E} [\psi_{\mathbf{Z}}(\boldsymbol{\omega}; u)] \\ &= \int_0^{+\infty} \exp\left(-\frac{u\boldsymbol{\omega}^T \Sigma \boldsymbol{\omega}}{2}\right) f_U(u) du \\ &= \mathcal{L}[f_U]\left(\frac{\boldsymbol{\omega}^T \Sigma \boldsymbol{\omega}}{2}\right) \end{aligned} \quad (5.4)$$

où  $\mathcal{L}[f_U]$  est la transformée de Laplace de  $f_U$ .

- La PDF est unimodale et différentiable presque partout sauf peut-être en 0.
- La loi de  $\mathbf{X}$  est une sous-classe des familles elliptiques multivariées [Kotz, 2004]. Tous les résultats concernant cette famille s'appliquent naturellement au VVA  $\mathbf{X}$ .
- La PDF de  $\mathbf{X}$  existe en 0 si et seulement si  $\mathbb{E} \left[ U^{-\frac{1}{2}} \right] < +\infty$ .
- La PDF de  $U$  est étroitement liée à la transformée de Laplace inverse de  $f_{\mathbf{X}}$ .

### 5.2.3 Moments absolus d'ordre $p$

Soient  $\mathbf{Z} \sim \mathcal{N}(0, \Sigma)$  et  $U \geq 0$  de loi  $\mathcal{L}_\theta$ , où  $\theta$  est le vecteur des hyperparamètres avec  $\mathbb{E}[U] = 1$ . En s'appuyant sur le cas univarié, il n'est pas difficile de prouver que le moment absolu d'ordre  $p$  de chaque composante  $i$  du VVA  $\Sigma^{-\frac{1}{2}} \mathbf{X}$ , où  $\Sigma^{-\frac{1}{2}}$  est l'inverse de la racine principale de  $\Sigma$ , est donné par :

$$M_{\mathbf{x}_i}(p) = \frac{2^{\frac{p}{2}}}{\sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right) M_U\left(\frac{p}{2}\right) \quad (5.5)$$

Par ailleurs, la Proposition 3.4 [page 46] peut être facilement généralisée pour le cas multivarié. En effet, la famille du modèle SMG est aussi forcément leptokurtique pour tout  $U$  aléatoire.

**Proposition 5.1** *Pour tout  $U$  aléatoire, la mesure du kurtosis multivarié au sens de Mardia [Mardia, 1970] du VVA  $\mathbf{X}$  est toujours strictement positive.*

**Preuve:**

Notons tout d'abord que  $\Sigma = \text{Cov}[\mathbf{X}]$ .

Écrivons ensuite la mesure du kurtosis d'un VVA  $\mathbf{X}$  proposée par Mardia :

$$\begin{aligned} \text{Kurt}[\mathbf{X}] &= \mathbb{E}\left[\left(\mathbf{X}^T \Sigma^{-1} \mathbf{X}\right)^2\right] - d(d+2) \\ &= \mathbb{E}[U^2] \mathbb{E}\left[\left(\mathbf{X}^T (u^2 \Sigma)^{-1} \mathbf{X}\right)^2 \mid U\right] - d(d+2) \\ &= d(d+2) (\mathbb{E}[U^2] - 1) > 0 \end{aligned} \quad (5.6)$$

car par définition, pour un VVA gaussiennes, le kurtosis multivarié est  $d(d+2)$ . Il est évident de voir que le cas univarié est obtenu pour  $d = 1$ . Finalement, en utilisant le même raisonnement qu'à la fin de la preuve de la Proposition 3.4, le résultat suit. ■

### 5.2.4 Cas particuliers du SMG multivarié

Trois lois se présentent comme des cas particuliers pour le modèle MSMG :

- *L'a priori* des Formes K de Bessel multivarié (si la VA  $U$  dans l'Eq.5.2 suit une loi de Gamma). Ce modèle sera au coeur de la suite de cette partie.
- *L'a priori*  $\alpha$ -stable multivarié (si la VA  $U$  suit une distribution  $\alpha$ -stable) [Achim & Kuruoglu, 2004].
- *L'a priori* non-informatif de Jeffrey utilisé dans [Portilla et al., 2003].

## 5.3 Estimation des hyperparamètres

Dans le cadre d'une approche bayésienne et afin de caractériser entièrement le modèle MSMG, il est nécessaire d'estimer les hyperparamètres qui lui sont associés.

La distribution de  $U$  dans le modèle MSMG, donné par l'Eq.5.2, dépend d'un ensemble d'hyperparamètres  $\theta$  réunissant les hyperparamètres de la loi de  $U$  et la matrice de covariance  $\Sigma$ . Dans le domaine des transformées multi-échelles, ces hyperparamètres  $\theta$  sont estimés à partir des coefficients observés à chaque orientation et à chaque échelle. Notons que les différentes méthodes permettant d'estimer les hyperparamètres, citées dans la deuxième partie pour le cas

de la modélisation univariée, peuvent être étendues au cas multivarié. Notamment, l'algorithme EM et les méthodes des moments et des cumulants.

L'étape de l'estimation des hyperparamètres reste accessible en absence du bruit et devient plus délicate en sa présence. Les différentes méthodes envisagées en absence du bruit seront développées dans ce qui suit pour le cas de l'*a priori* BKF multivarié. En présence de bruit, nous y reviendrons en détail dans le chapitre suivant.

## 5.4 Cas 1 : modèle BKF multivarié

### 5.4.1 Définition

Soit  $\mathbf{X} = \sqrt{U}\mathbf{Z}$  donné par l'Eq.5.2. Si  $U \sim \Gamma\left(\alpha, \frac{1}{\alpha}\right)$  où  $\alpha > 0$ , alors  $\mathbf{X}$  suit une distribution BKF multivariée (MBKF) dont la PDF s'écrit comme suit,

$$f_{\mathbf{X}}(\mathbf{x}; \alpha, \Sigma) = \frac{2^{1-\alpha}}{\sqrt{(2\pi)^d |\Sigma|}} (2\alpha)^{\frac{\alpha}{2} + \frac{d}{4}} \|\mathbf{x}\|_{\Sigma^{-1}}^{\alpha - \frac{d}{2}} K_{\alpha - \frac{d}{2}}\left(\sqrt{2\alpha} \|\mathbf{x}\|_{\Sigma^{-1}}\right) \quad (5.7)$$

où  $K_{\alpha - \frac{d}{2}}$  est la fonction de Bessel modifiée de premier espèce d'ordre  $\alpha - \frac{d}{2}$  (voir [Abramowitz & Stegun, 1972, Gradshteyn & Ryzhik, 1980]).  $\alpha$  est le paramètre de forme et  $\Sigma$  est une matrice symétrique définie positive.  $\|\mathbf{x}\|_{\Sigma^{-1}} = (\mathbf{x}^T \Sigma^{-1} \mathbf{x})^{\frac{1}{2}}$ .

### 5.4.2 Propriétés

Quelques propriétés de la loi MBKF qui nous seront utiles sont résumées dans la proposition suivante :

#### Proposition 5.2

Soit  $\mathbf{X} \sim MBKF(\alpha, \Sigma)$  avec  $\alpha > 0$ .

(i)

$$\mathbf{X} \xrightarrow[\alpha \rightarrow +\infty]{d} \mathcal{N}(0, \Sigma)$$

(ii)

$$\mathbb{E} \left[ (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^i \right] = 2^i \frac{\Gamma\left(\frac{d}{2} + i\right) \Gamma(\alpha + i)}{\alpha \Gamma\left(\frac{d}{2}\right) \Gamma(\alpha)} \quad (5.8)$$

(iii) Les cumulants ont la forme suivante :

$$\kappa_j(2i) = \alpha^{1-i} \frac{(2i)!}{2^i i!} \sum_{l=1}^d \sigma^{2i}(j, l) \text{ avec } i \geq 1, j = 1, \dots, d \quad (5.9)$$

où  $\sigma(j, l)$  est l'élément à la position  $(j, l)$  de la matrice  $\Sigma^{\frac{1}{2}}$ .

**Preuve:**

(i) Il suffit de noter que la fonction caractéristique d'un VVA MBKF se met sous la forme suivante :

$$\Psi_{\mathbf{X}}(\boldsymbol{\omega}) = \frac{1}{\left(1 + \frac{\boldsymbol{\omega}^T \Sigma \boldsymbol{\omega}}{2\alpha}\right)^\alpha} = e^{-\alpha \log\left(1 + \frac{\boldsymbol{\omega}^T \Sigma \boldsymbol{\omega}}{2\alpha}\right)} \quad (5.10)$$

En prenant le développement en série de Taylor de l'argument de l'exponentiel

$$\begin{aligned} \Psi_{\mathbf{X}}(\boldsymbol{\omega}) &= e^{-\alpha \sum_{i=1}^{+\infty} \frac{(-1)^{i+1}}{i} \left(\frac{\boldsymbol{\omega}^T \Sigma \boldsymbol{\omega}}{2\alpha}\right)^i} \\ &= e^{-\frac{\boldsymbol{\omega}^T \Sigma \boldsymbol{\omega}}{2}} e^{-\sum_{i \geq 2}^{+\infty} \frac{(-1)^{i+1} (\boldsymbol{\omega}^T \Sigma \boldsymbol{\omega})^i}{i 2^i \alpha^{i-1}}} \xrightarrow{\alpha \rightarrow +\infty} e^{-\frac{\boldsymbol{\omega}^T \Sigma \boldsymbol{\omega}}{2}} \quad \forall \boldsymbol{\omega} \end{aligned} \quad (5.11)$$

On conclut en appliquant le théorème de continuité de Lévy.

(ii) Le résultat annoncé par la deuxième propriété est démontré comme suit :

Soit  $\mathbf{R} = (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{\frac{1}{2}}$ . On a

$$\mathbb{E} [\mathbf{R}^i] = \mathbb{E} [\mathbb{E} [\mathbf{R}^i | U]] \quad (5.12)$$

Or  $\mathbf{X} | U \sim \mathcal{N}(0, u\Sigma)$ . Si  $U$  suit une loi de Gamma de paramètre de forme  $\alpha$  et d'espérance unitaire  $\mathbb{E} [U] = 1$ , on a

$$M_U(p) = \mathbb{E} [U^p] = \frac{\Gamma(\alpha + p)}{\alpha^p \Gamma(\alpha)} \quad (5.13)$$

et donc

$$\mathbb{E} [\mathbf{R}^i | U] = u^{\frac{i}{2}} 2^{\frac{i}{2}} \frac{\Gamma\left(\frac{d}{2} + \frac{i}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \quad (5.14)$$

Ceci implique que

$$\begin{aligned} \mathbb{E} [\mathbf{R}^i] &= \mathbb{E} \left[ U^{\frac{i}{2}} \right] 2^{\frac{i}{2}} \frac{\Gamma\left(\frac{d}{2} + \frac{i}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \\ &= \left(\frac{2}{\alpha}\right)^{\frac{i}{2}} \frac{\Gamma\left(\frac{d}{2} + \frac{i}{2}\right) \Gamma\left(\alpha + \frac{i}{2}\right)}{\Gamma\left(\frac{d}{2}\right) \Gamma(\alpha)} \end{aligned} \quad (5.15)$$

(iii) Il existe  $\mathbf{Y} \in \mathbb{R}^d$  un VVA BKF i.i.d de même paramètre  $\alpha$  avec  $\mathbf{Y} = \sqrt{U} \tilde{\mathbf{Z}}$  où  $\tilde{\mathbf{Z}} \sim \mathcal{N}(0, \mathbb{I}_d)$ , tel que :

$$\mathbf{Y} = \Sigma^{\frac{1}{2}} \tilde{\mathbf{Z}} \quad (5.16)$$

où  $\Sigma^{\frac{1}{2}}$  est la racine carrée principale de  $\Sigma$ .

Ainsi, nous définissons  $\kappa_j^{\mathbf{X}}(p)$  (resp.  $\kappa_j^{\mathbf{Y}}(p)$ ) le cumulatif d'ordre  $p$  de la  $j^{\text{ième}}$  composante de  $\mathbf{X}$  (resp.  $\mathbf{Y}$ ) par les propriétés usuelles des cumulants :

$$\kappa_j^{\mathbf{X}}(p) = \sum_{l=1}^d \sigma^p(j, l) \kappa_j^{\mathbf{Y}}(p) \quad (5.17)$$

où  $\sigma(j, l)$  est la composante de  $\Sigma^{\frac{1}{2}}$  à la ligne  $j$  et la colonne  $l$ .

D'après la partie sur les VA BKF univariés (Eq.3.64) :

$$\kappa_j^{\mathbf{Y}}(2i) = \frac{\alpha^{i-1} (2i)!}{2^i i} \quad (5.18)$$

et pour  $p$  impair,  $\kappa_j^{\mathbf{Y}}(p) = 0$  d'où le résultat annoncé. ■

### 5.4.3 Estimation des hyperparamètres

La PDF du modèle MBKF est caractérisée par deux paramètres : un paramètre de forme  $\alpha$  et  $\Sigma$ . Pour mettre en application notre modèle multivarié, nous avons utilisé deux estimateurs pour les hyperparamètres associés, à savoir la méthode de moments (tient profit de la proposition 5.2(ii)) et le maximum de vraisemblance.

#### 5.4.3.1 Méthode des moments

Soit  $\mathbf{X} \sim MBKF(\alpha, \Sigma)$ . D'après la proposition 5.2 et en prenant  $i = 2$  dans (ii), les paramètres  $\alpha$  et  $\Sigma$  sont estimés comme suit :

$$\hat{\Sigma} = \frac{1}{n} \sum_i \mathbf{x}_i \mathbf{x}_i^T \quad (5.19)$$

$$\hat{\alpha} = \left( \frac{\overline{(\mathbf{x}^T \hat{\Sigma}^{-1} \mathbf{x})^2}}{d(d+2)} - 1 \right)^{-1} \quad (5.20)$$

où  $\overline{(\mathbf{x}^T \hat{\Sigma}^{-1} \mathbf{x})^2} = \frac{1}{n} \sum_i (\mathbf{x}_i^T \Sigma^{-1} \mathbf{x}_i)^2$  est la moyenne empirique de  $\mathbb{E}[(\mathbf{x}^T \Sigma^{-1} \mathbf{x})^2]$ .

#### 5.4.3.2 Algorithme EM

De manière analogue au cas univarié, le principe ici est de reconstruire les données manquantes (variable cachée  $U$ ) et ensuite d'estimer les hyperparamètres associés :  $\alpha$  et  $\Sigma$ .

Nous supposons  $U$ , dans l'Eq.5.2, les données cachées et  $(\mathbf{X}, U)$  les données complètes. L'algorithme suivant présente une procédure d'estimation du paramètre  $\alpha$ . Par ailleurs, la matrice de covariance  $\Sigma$ , dans ce cas, est estimé une fois pour toutes par l'Eq.5.19.

---

**Algorithme 5** Algorithme EM pour l'estimation du paramètre  $\alpha$  du modèle MBKF

---

**Entrées:** on part d'une estimation initiale de  $\alpha$

1: **répéter**

2: **Étape E :**

On calcule,

$$M_U^{(t)}(1) = \mathbb{E}[U|\mathbf{x}] = \left( \frac{\mathbf{x}^T \Sigma^{-1} \mathbf{x}}{2\alpha} \right)^{\frac{1}{2}} \frac{K_{\alpha - \frac{d}{2} + 1} \left( \sqrt{2\alpha} (\mathbf{x}^T \Sigma^{-1} \mathbf{x})^{\frac{1}{2}} \right)}{K_{\alpha - \frac{d}{2}} \left( \sqrt{2\alpha} (\mathbf{x}^T \Sigma^{-1} \mathbf{x})^{\frac{1}{2}} \right)} \quad (5.21)$$

$$\varphi_1^{(t)} = \mathbb{E}[\log(U)|\mathbf{x}] = -\log(\alpha) + \frac{1}{2} \log \left( \frac{\alpha (\mathbf{x}^T \Sigma^{-1} \mathbf{x})^2}{2} \right) + \frac{\partial K_{\alpha - \frac{d}{2}} \left( \sqrt{2\alpha} (\mathbf{x}^T \Sigma^{-1} \mathbf{x})^{\frac{1}{2}} \right)}{K_{\alpha - \frac{d}{2}} \left( \sqrt{2\alpha} (\mathbf{x}^T \Sigma^{-1} \mathbf{x})^{\frac{1}{2}} \right)} \quad (5.22)$$

3: **Étape M :**

L'estimé de  $\alpha^{(t+1)}$  est la solution de l'équation suivante :

$$\Psi(\alpha^{(t)}) - \log(\alpha^{(t)}) = \overline{\varphi_1^{(t)}} - \overline{M_U^{(t)}(1)} + 1 \quad (5.23)$$

où  $\overline{\varphi_1}$  et  $\overline{\varphi_3}$  sont respectivement les moyennes empiriques de  $\varphi_1$  et  $M_U(1)$

$$\overline{\varphi_1^{(t)}} = \frac{1}{m} \sum_{i=1}^m \varphi_{1(i)}^{(t)} \quad , \quad \overline{M_U^{(t)}(1)} = \frac{1}{m} \sum_{i=1}^m \left( M_U^{(t)}(1) \right)_{(i)}$$

4: **jusqu'à** Convergence.

---

**Preuve:**

Notons tout d'abord que le résultat de l'algorithme EM présenté ci-dessus peut être démontré de manière similaire au cas BKF univarié [page 69].

Pour le cas de MBKF, le score de vraisemblance  $Q$  s'écrit comme suit,

$$\begin{aligned} Q(\theta|\theta^{(t)}) &= \frac{d}{2} \sum_{i=1}^m \varphi_{1(i)} + \frac{1}{2} \sum_{i=1}^m \varphi_{2(i)} (\mathbf{x}^T \Sigma^{-1} \mathbf{x}) - m \log \alpha + m \log \Gamma(\alpha) + (1 - \alpha) \sum_{i=1}^m \varphi_{1(i)} \\ &\quad + m(1 - \alpha) \log \alpha + \alpha \sum_{i=1}^m \varphi_{3(i)} \end{aligned} \quad (5.24)$$

où nous notons

$$\varphi_{1(i)} = \mathbb{E}[\log(U)|\mathbf{x}_i] ; \varphi_{2(i)} = \mathbb{E}[U^{-1}|\mathbf{x}_i] ; \varphi_{3(i)} = \mathbb{E}[U|\mathbf{x}_i] \quad \text{pour } i = 1, \dots, m$$

En appliquant l'Eq.3.89 et l'Eq.3.88, pour calculer  $\varphi_1$  et  $\varphi_3$ , le résultat de l'étape E suit.

Ainsi, le calcul de la dérivée partielle de  $Q$  en fonction de  $\alpha$  est :

$$\partial_\alpha Q = m\Psi(\alpha) - m \log \alpha - \sum_{i=1}^m \varphi_{1(i)}^{(t)} + \sum_{i=1}^m \varphi_{3(i)}^{(t)} - m = 0 \quad (5.25)$$

$\alpha$  est donc la solution de l'équation suivante :

$$\begin{aligned}\Psi(\alpha) - \log(\alpha) &= \frac{1}{m} \sum_{i=1}^m \varphi_{1(i)}^{(t)} - \frac{1}{m} \sum_{i=1}^m \varphi_{3(i)}^{(t)} + m \\ &= \overline{\varphi_1^{(t)}} - \overline{\varphi_3^{(t)}} + 1\end{aligned}\quad (5.26)$$

d'où le résultat de l'étape M de l'algorithme.

Contrairement au cas univarié, on peut montrer que l'unicité de la solution de l'Eq.5.26 n'est pas toujours négative. Elle est en revanche pour  $\alpha \in ]0, 1]$  par exemple. ■

## 5.5 Qualité de l'a priori : application à la modélisation des statistiques jointes

### 5.5.1 Objectifs de l'expérience

Afin d'évaluer la qualité de l'a priori MBKF présenté ci-dessus, une comparaison des PDFs jointes estimées et observées des coefficients de détails d'ondelettes et de curvelets des images naturelles, est illustrée sur les figures 5.4, 5.5 et 5.6. Deux modèles d'a priori multivariés ont été mis en comparaison avec notre modèle MBKF à savoir : l'a priori de Jeffrey [Portilla et al., 2003] et l'AMGGD (cf. annexe C). L'objectif est de tester et évaluer l'adéquation de l'a priori MBKF aux statistiques jointes observées (dépendances inter- et intra-échelles entre coefficients).

### 5.5.2 Protocole expérimental

TAB. 5.2 – Protocole expérimental

Grandeur	Choix		
	Fig.5.4	Fig.5.5	Fig.5.6
<i>a priori</i> multivarié	MBKF, AMGGD, Jeffrey		
Estimateur des hyperparamètres pour MBKF	Algorithme EM		
Transformées	ondelettes	curvelets	ondelettes et curvelets
Voisinage	coefficient-voisins direct coefficient-voisins droite/gauche coefficient-parent (inter-échelle)		
UDWT	3 échelles et 3 orientations		
FDCT	3 échelles et toutes orientations		
Métrique	divergence KL		
Images tests	image test de Barbara		100 images

### 5.5.3 Discussion des résultats

Sur la Fig.5.4(a), la PDF jointe observée des coefficients de détail de l'UDWT de l'image test de Barbara (en rouge) a été ajustée par trois modèles d'a priori : MBKF (en bleu), AMGGD

(en vert) et Jeffrey (en bleu clair), dans le cas bivarié. Les trois colonnes correspondent respectivement aux dépendances entre : les coefficients et leurs voisins directs gauche/droite, les coefficients et leurs voisins diagonaux, les coefficients et leurs parents à l'échelle supérieure. Nous constatons, suite à l'observation des courbes, que la PDF MBKF est en général plus proche de la PDF observée comparée aux PDFs du modèle AMGGD et Jeffrey. Ce comportement est aussi observé, dans la Fig.5.4(b), pour le cas trivarié. Ceci est confirmé par la divergence KL calculée entre la PDF observée et estimée.

Dans la Fig.5.5, nous illustrons cette même comparaison d'*a priori* sur les coefficients de détails de la FDCT de Barbara. Nous constatons que la PDF MBKF est plus proche de la PDF observée pour le voisinage "coefficient-parent". En revanche, le MBKF fournit des résultats moins bons comparé au modèle AMGGD pour capturer les dépendances intra-échelles, mais reste meilleur comparés au modèle de Jeffrey. Ceci est confirmé par la divergence KL calculée entre la PDF observée et estimée.

Ce comportement observé pour le cas des coefficients de détail d'ondelettes (sur 3 échelles et 3 orientations) et de curvelet (sur 3 échelles et toutes les orientations) est confirmé, dans la Fig.5.6, par la divergence KL calculée sur une base de 100 images [url : base d'images, ]. Nous constatons que le modèle MBKF dépasse le modèle de Jeffrey que ce soit pour le cas des ondelettes ou des curvelets. Par ailleurs, il est en général meilleur comparé au modèle AMGGD dans le cas des ondelettes. Cette tendance s'inverse pour la cas des curvelets où le modèle AMGGD devient meilleur. Finalement, ce résultat plaide en faveur de nos modèles de PDF, le MBKF en particulier, comme un modèle d'*a priori* adéquat pour les statistiques jointes observées.

## 5.6 Conclusion

Nous avons étudié les propriétés des coefficients de curvelets et d'ondelettes non-décimées, et nous avons montré que les coefficients ayant des valeurs élevées tendent à se grouper autour des bords des objets dans l'image, ce qui est intuitif car les coefficients dépendent de leurs parents et voisins, aussi bien que leurs cousins à différentes échelles et orientations. Ces dépendances ont été vérifiées quantitativement en mesurant l'information mutuelle, où nous avons constaté que le niveau le plus élevé de la dépendance des coefficients est à l'égard des voisins, suivi des parents, qui finalement des cousins.

Ensuite, nous avons proposé un modèle statistique multivarié de PDF permettant de capturer ces dépendances inter- et intra-échelle. Ce modèle, que nous avons pleinement caractérisé, est basé sur une extension multivariée de l'*a priori* BKF. Ainsi, nous avons proposé des estimateurs pour ses hyperparamètres en absence de bruit.

L'objectif du chapitre suivant se focalisera sur la mise à profit de ce modèle *a priori* multivarié MBKF pour le débruitage bayésien d'images dans le domaine des transformées multi-échelles orientées (e.g. curvelets) et non-orientées (e.g. ondelettes non-décimées).

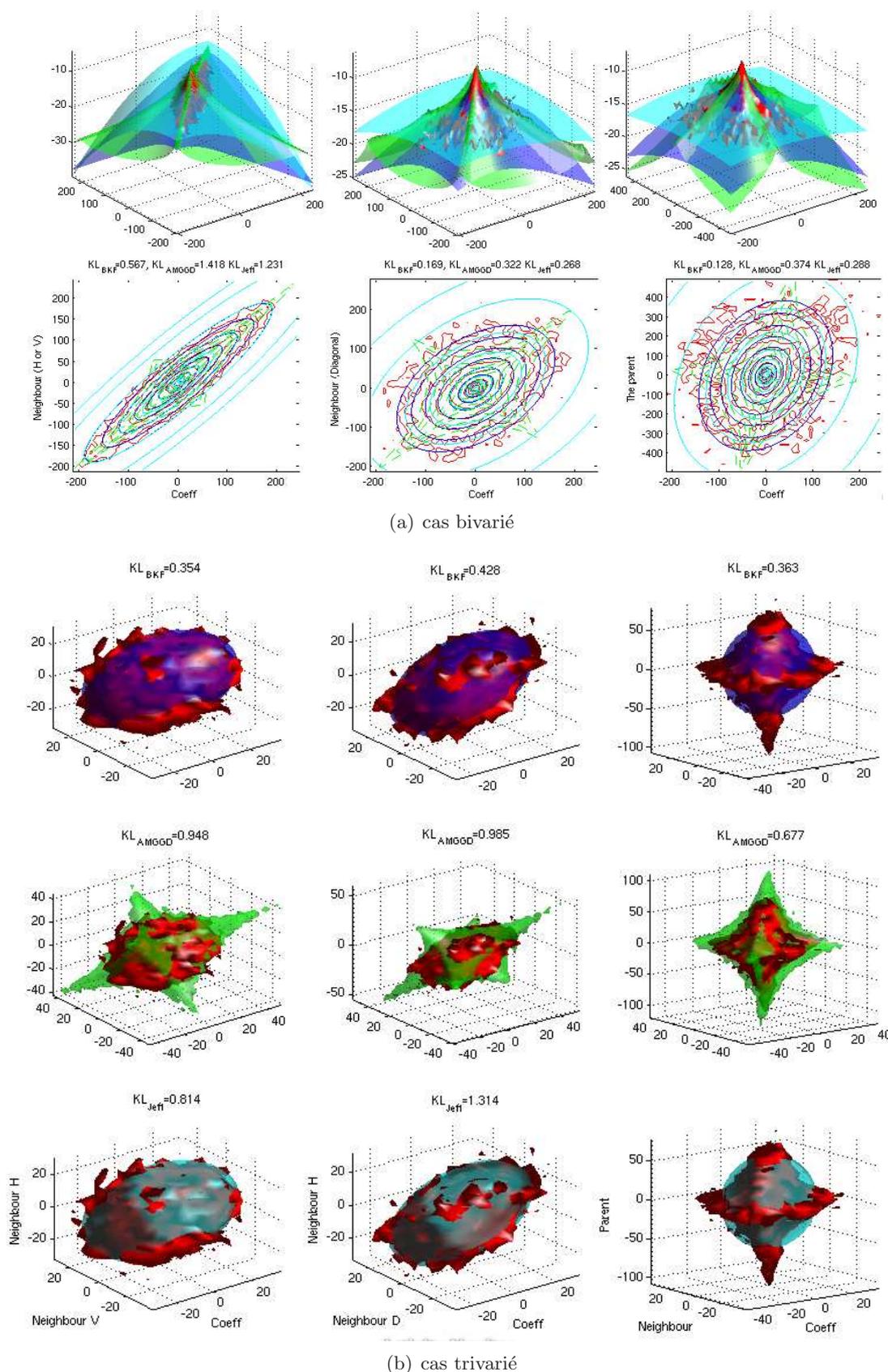


FIG. 5.4 – Exemple de comparaison sur une échelle log – log entre la PDF jointe observée et les PDFs estimées des coefficients de détail d’ondelettes pour l’image Barbara dans le cas : bivarié ( $d = 2$ ) et trivarié ( $d = 3$ ). La distribution jointe observée (en rouge) a été ajustée par trois modèles d’a priori : MBKF (en bleu), AMGGD (en vert) et Jeffrey (en bleu clair) [Portilla et al., 2003]. Les trois colonnes correspondent aux dépendances inter- et intra-échelles entre coefficients.

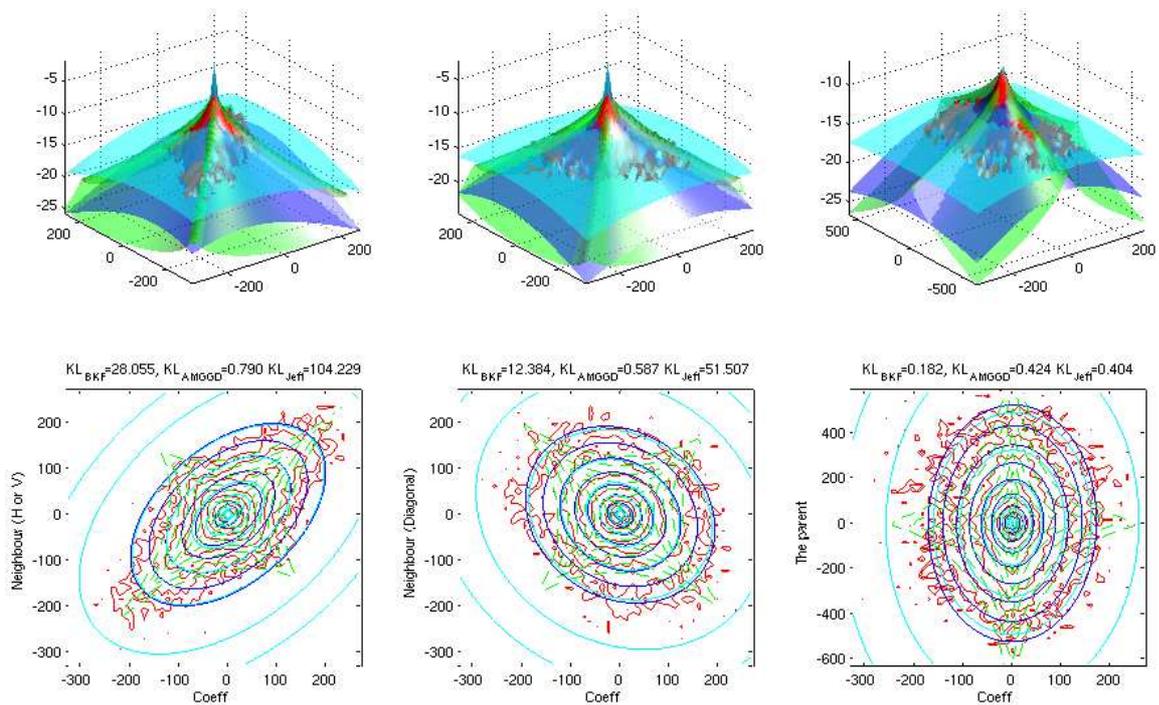


FIG. 5.5 – Exemple de comparaison sur une échelle log – log entre la PDF jointe observée et les PDFs estimées des coefficients de détail de curvelet pour l'image Barbara dans le cas bivarié ( $d = 2$ ). La distribution jointe observée (en rouge) a été ajustée par trois modèles d'*a priori* : MBKF (en bleu), AMGGD (en vert) et Jeffrey (en bleu clair) [Portilla et al., 2003]. Les trois colonnes correspondent aux dépendances inter- et intra-échelles entre coefficients.

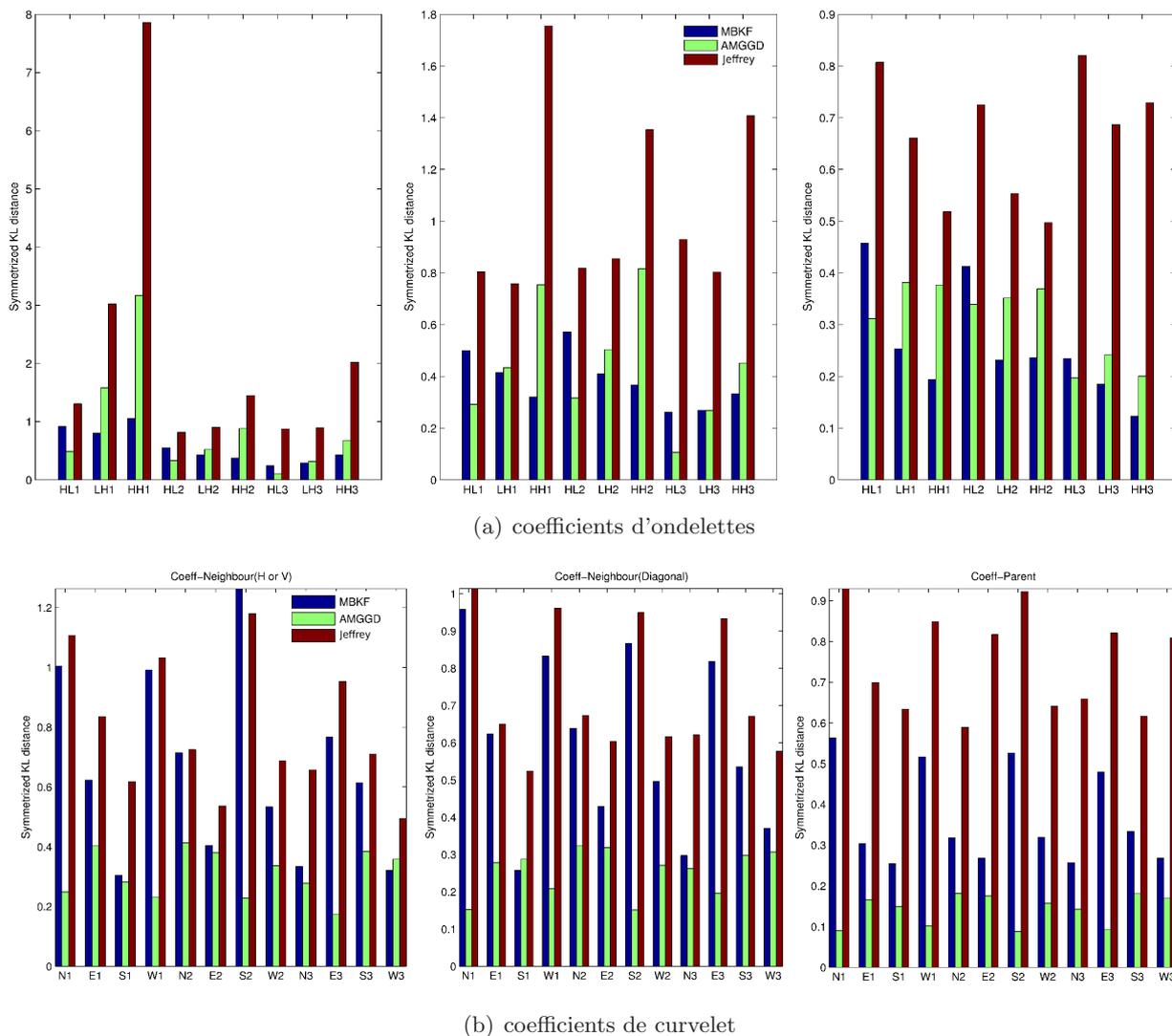


FIG. 5.6 – Évaluation de la divergence de KL, calculée entre la PDF observée et les PDF estimées par les *a priori* multivariés : MBKF, AMGGD et Jeffrey [Portilla et al., 2003]. La divergence KL a été moyennée pour chaque orientation et à chaque échelle sur une base de 100 images. Trois orientations ont été considérées pour le UDWT. Pour la FDCT et pour mieux synthétiser les résultats, nous avons aussi moyenné les valeurs KL sur toutes les orientations faisant partie d'un même quadrant : Est, Ouest, Nord, Sud.



## Chapitre 6

# Débruitage bayésien avec *a priori* multi-varié

---

### Sommaire

<b>6.1</b>	<b>Rappels sur l'estimation bayésienne : cas multivarié . . . . .</b>	<b>121</b>
<b>6.2</b>	<b>Estimateur ECP avec l'<i>a priori</i> MBKF . . . . .</b>	<b>122</b>
6.2.1	La PDF jointe des coefficients bruités . . . . .	122
6.2.2	Estimation des hyperparamètres . . . . .	123
6.2.3	Débruiteur bayésien ECP . . . . .	124
<b>6.3</b>	<b>Expérimentation et résultats . . . . .</b>	<b>125</b>
6.3.1	Objectifs de l'expérience . . . . .	125
6.3.2	Protocole expérimental . . . . .	125
6.3.3	Discussion des résultats . . . . .	125
<b>6.4</b>	<b>Conclusion . . . . .</b>	<b>126</b>

---

Dans ce chapitre, nous proposons un nouvel estimateur bayésien multivarié de type ECP pour le débruitage des images, basés sur le modèle *a priori* MBKF. Ce dernier, que nous avons déjà caractérisé dans le chapitre précédent, s'est révélé très efficace comme modèle des statistiques jointes des images dans le domaine des transformées multi-échelles parcimonieuses, en l'occurrence les curvelets. Deux étapes sont mises nécessaires pour la mise en oeuvre de ces estimateurs, la première consistant à estimer les hyperparamètres du modèle de l'*a priori* en présence de bruit, et la seconde à trouver une forme pour l'estimateur bayésien correspondant.

### 6.1 Rappels sur l'estimation bayésienne : cas multivarié

Le modèle de dégradation dans le cas multivarié s'écrit comme suit :

$$\mathbf{d} = \mathbf{s} + \epsilon \tag{6.1}$$

où chaque  $\mathbf{d} = \{d_1, \dots, d_d\}$  (resp.  $\mathbf{s} = \{s_1, \dots, s_d\}$ ) est un vecteur de coefficients de détails de l'image bruitée (resp. l'image à estimer), contenant le coefficient avec ses coefficients voisins à la même orientation et éventuellement son coefficient parent à l'échelle supérieure.  $\epsilon \sim \mathcal{N}(0, \Sigma_\epsilon)$  est un bruit gaussien de covariance  $\Sigma_\epsilon$ . On peut remarquer contrairement au cas des bases comme la DWT, ici le bruit n'est pas blanc puisque les transformées que nous manipulons dans ce chapitre

(UDWT et FDCT) sont des frames. Nous reviendrons sur le calcul de la matrice de covariance  $\Sigma_\epsilon$  dans la suite.

Dans le contexte bayésien, les coefficients de détails de chaque vecteur  $\mathbf{s}$  suivent une loi de probabilité du modèle *a priori* multivarié MSMG, paramétrée par les hyperparamètres à estimer  $\theta_1$ ,

$$\mathbf{s} \sim \text{MSMG}(\theta_1) \quad (6.2)$$

Nous rappelons que le modèle probabiliste associé à  $\mathbf{d}$  sachant  $\mathbf{s}$  est gaussien de matrice de covariance  $\Sigma_\epsilon$

$$\mathbf{d}|\mathbf{s} \sim \mathcal{N}(\mathbf{s}, \Sigma_\epsilon) \quad (6.3)$$

## 6.2 Estimateur ECP avec l'a priori MBKF

### 6.2.1 La PDF jointe des coefficients bruités

Pour le cas de l'a priori MBKF et à partir de l'Eq.6.2 et l'Eq.6.3, les coefficients à estimer  $\mathbf{s}$  suivent une loi MBKF,

$$\mathbf{s} \sim \text{MBKF}(\alpha, \Sigma) \quad (6.4)$$

et le modèle probabiliste associé à  $\mathbf{d}$  sachant  $\mathbf{s}$  est gaussien multivarié

$$\mathbf{d}|\mathbf{s} \sim \mathcal{N}(0, \Sigma_\epsilon) \quad (6.5)$$

où  $\theta_1 = \{\alpha, \Sigma\}$ .

La PDF de  $\mathbf{d}$  sachant les hyperparamètres  $\theta = \{\alpha, \Sigma, \Sigma_\epsilon\}$  n'est pas facile à obtenir (très compliqué d'avoir des formes analytiques dans le cas multivarié) sauf par approximation. Nous avons utilisé l'intégration de Monte-Carlo pour évaluer numériquement l'intégrale.

$$\begin{aligned} f(\mathbf{d}; \theta) &= \int_0^{+\infty} \phi(\mathbf{d}; 0, u\Sigma + \Sigma_\epsilon) f_U(u) du \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_0^{+\infty} \frac{1}{|u\Sigma + \Sigma_\epsilon|} \exp\left(-\frac{\mathbf{d}^T (u\Sigma + \Sigma_\epsilon)^{-1} \mathbf{d}}{2u}\right) f_U(u) du \\ &\approx \frac{1}{M} \sum_{i=1}^M \phi(\mathbf{d}; 0, u_i \Sigma + \Sigma_\epsilon) \end{aligned} \quad (6.6)$$

où  $U \sim \Gamma(\alpha)$  et  $M$  est le nombre de réalisations dans l'intégration Monte-carlo. Nous y reviendrons plus en détails dans la section 6.2.3.

Toutefois, pour la matrice  $u\Sigma + \Sigma_\epsilon$ , nous ne voulons pas avoir à l'inverser pour chaque réalisation de  $U$ . Or, le théorème suivant nous permet de simplifier complètement les calculs.

**Théorème 6.1** ([Magnus & Neudecker, 1998]) *Soit deux matrices  $\mathbf{A} \succ 0$  et  $\mathbf{B} \succeq 0$ , il existe une matrice non-singulière  $\mathbf{R}$  et une matrice diagonale  $\mathbf{V} \succeq 0$  telles que :*

$$\mathbf{A} = \mathbf{R}\mathbf{R}^T \quad \text{et} \quad \mathbf{B} = \mathbf{R}\mathbf{V}\mathbf{R}^T$$

avec  $\mathbf{R} = \mathbf{S}\Delta^{\frac{1}{2}}\mathbf{T}$  et  $\mathbf{S}^T\mathbf{A} = \Delta\mathbf{S}^T$  où  $\mathbf{S}$  et  $\Delta$  correspondent aux vecteurs et valeurs propres de  $\mathbf{A}^T$ , et  $\mathbf{T}$  et  $\mathbf{V}$  aux vecteurs et valeurs propres de  $\Delta^{-\frac{1}{2}}\mathbf{S}^T\mathbf{B}^T\mathbf{S}\Delta^{-\frac{1}{2}}$ .

Ainsi, par application directe (en identifiant  $\Sigma$  à  $\mathbf{B}$  et  $\Sigma_\epsilon$  à  $\mathbf{A}$ ) :

$$u\Sigma + \Sigma_\epsilon = \mathbf{R} (u\mathbf{V} + \mathbb{I}_d) \mathbf{R}^T \quad (6.7)$$

ce qui implique

$$(u\Sigma + \Sigma_\epsilon)^{-1} = \mathbf{R}^{-T} (u\mathbf{V} + \mathbb{I}_d)^{-1} \mathbf{R}^{-1} \quad (6.8)$$

avec  $\text{diag} (u\mathbf{V} + \mathbb{I}_d)^{-1} = \left( \frac{1}{uv_i+1} \right)_i$ .

Ce résultat théorique nous dicte directement l'algorithme suivant pour un calcul efficace et rapide de la matrice  $(u\Sigma + \Sigma_\epsilon)^{-1}$

---

**Algorithme 6** Algorithme pour calculer la matrice  $(u\Sigma + \Sigma_\epsilon)^{-1}$

---

- 1: Calculer les valeurs et les vecteurs propres  $\Delta$  et  $\mathbf{S}$  de  $\Sigma_\epsilon$ .
  - 2: Construire  $\left( \Delta^{-\frac{1}{2}} \mathbf{S}^T \mathbf{B}^T \mathbf{S} \Delta^{-\frac{1}{2}} \right)$  où  $\Sigma$  est symétrique.
  - 3: Calculer  $\mathbf{T}$  et  $\mathbf{V}$  ses vecteurs et valeurs propres.
  - 4: Construire  $\mathbf{R} = \mathbf{S} \Delta^{\frac{1}{2}} \mathbf{T}$  et son inverse.
  - 5:  $(u\Sigma + \Sigma_\epsilon)^{-1} = \mathbf{R}^{-T} (u\mathbf{V} + \mathbb{I}_d)^{-1} \mathbf{R}^{-1}$
- 

En appliquant ce résultat à l'Eq.6.6, l'approximation analytique de la PDF de  $\mathbf{d}$  est la suivante :

$$\tilde{f}(\mathbf{d}; \theta) = \frac{1}{M} \sum_{i=1}^M \frac{|\det \mathbf{R}^{-1}|}{(2\pi)^{\frac{d}{2}}} \prod_{j=1}^d \frac{1}{(u_i v_j + 1)^{\frac{1}{2}}} \exp \left( -\frac{1}{2} \sum_{j=1}^d \frac{(r_j^T \mathbf{d})^2}{(u_i v_j + 1)} \right) \quad (6.9)$$

où  $r_j$  est une colonne de la matrice  $\mathbf{R}^{-1}$ .

### 6.2.2 Estimation des hyperparamètres

Pour mettre en application la formule de l'Eq.6.9, on doit alors estimer les hyperparamètres  $\theta = \{\alpha, \Sigma, \Sigma_\epsilon\}$ .

La méthode des cumulants utilisée pour le cas de l'a priori BKF univarié (page 85) peut être appliquée facilement pour le cas de l'a priori MBKF en s'appuyant sur le résultat de la proposition 5.2(iii). En effet, les paramètres  $\alpha$  et  $\Sigma$  sont estimés comme suit :

---

**Algorithme 7** Estimation des paramètres :  $\{\alpha, \Sigma\}$

---

- 1: Estimer  $\Sigma$  :

$$\hat{\Sigma} = \frac{1}{N} \sum_{i=1}^N \mathbf{d}_i \mathbf{d}_i^T - \hat{\Sigma}_\epsilon \quad (6.10)$$

- 2: Estimer  $\alpha$  (à partir de l'Eq.5.9 avec  $i = 1$  et 2) :

$$\hat{\alpha} = \frac{3}{d} \sum_{i=1}^d \frac{\sum_{j=1}^d \sigma^4(i, j)}{(\hat{\kappa}_i)(4)} \quad (6.11)$$

où  $\sigma(i, j)$  est la composante à la position  $(i, j)$  de la matrice  $\hat{\Sigma}^{\frac{1}{2}}$ .

---

Par ailleurs,  $\Sigma_\epsilon$  est estimé dans le cas des bases orthonormées par

$$\hat{\Sigma}_\epsilon = \sigma_\epsilon^2 \mathbb{I}_d \quad (6.12)$$

où  $\sigma_\epsilon$  est le niveau de bruit estimé par le MAD [Donoho & Johnstone, 1994]. Pour le cas des frames, où le bruit est corrélé,  $\Sigma_\epsilon$  est estimé comme suit :

---

**Algorithme 8** Estimation du paramètre  $\Sigma_\epsilon$ 


---

- 1:  $\hat{\sigma}_\epsilon = \frac{\text{MAD}}{0.6745}$
  - 2: Prendre la transformée (FDCT ou UDWT par exemple) d'une impulsion de Dirac pour obtenir une approximation des atomes  $\varphi_{j,o}(\tau)$ .
  - 3: A chaque échelle  $j$  et orientation  $o$ , calculer la fonction d'auto-corrélation  $\text{FAC}_{j,o}(\tau) = \sigma_\epsilon^2 \text{IFFT}[|\hat{\varphi}_{j,o}|^2]$  où  $\hat{\varphi}_{j,o}$  est la FFT de l'atome  $\varphi_{j,o}$  à l'échelle  $j$ , orientation  $o$ .
  - 4: Construire la matrice de covariance  $\Sigma_\epsilon$  à chaque échelle et orientation à partir de la  $\text{FAC}_{j,o}$ .
- 

### 6.2.3 Débruiteur bayésien ECP

L'estimateur bayésien ECP correspondant à  $\mathbf{s}$  sachant  $\mathbf{d}$  et  $\theta$  s'écrit comme suit,

$$\hat{\mathbf{s}}_{ECP}(\mathbf{d}; \theta) = \frac{\int_0^{+\infty} \underbrace{u \Sigma (u \Sigma + \Sigma_\epsilon)^{-1} \mathbf{d}}_{\text{estimé de Wiener}} \phi(\mathbf{d}; 0, u \Sigma + \Sigma_\epsilon) f_U(u) du}{\int_0^{+\infty} \phi(\mathbf{d}; 0, u \Sigma + \Sigma_\epsilon) f_U(u) du} \quad (6.13)$$

Cet estimateur correspond à un mélange d'échelles d'estimés de Wiener locaux. Sa forme analytique n'est pas facile à obtenir sauf approximation. Cependant, plusieurs solutions alternatives s'offrent à nous grâce à la théorie de l'intégration bayésienne. Nous listons ici les principales d'entre elles en soulignant leurs avantages et leurs inconvénients :

- **Approximation analytique** : dans ce cas une approximation analytique de l'intégrale est opérée. Parmi les plus connues est l'approximation de Riemann et l'approximation de Laplace. L'une des vertus de la méthode de Laplace est qu'elle substitue le problème d'intégration par un problème d'optimisation pour la recherche du mode conditionnel *a posteriori*. Toutefois, elle nécessite pour son application des conditions de régularités de l'*a priori* qui ne sont pas vérifiées dans notre cas.
- **Intégration par des méthodes de quadrature** : parmi les méthodes de quadrature on retrouve la quadrature de Gauss et la quadrature de Lobatto. Celles-ci sont très précises mais peuvent souffrir d'un temps de calcul très lent. De plus, un soin particulier doit être porté aux bornes d'intégration dont l'une est infinie dans notre cas. Finalement, ces méthodes ne s'appliquent que pour des fonctions lisses où les intégrales sont libres de toute singularité.
- **Intégration de Monte-Carlo** : les techniques Monte-Carlo s'affranchissent des hypothèses de régularité sur la la fonction à intégrer. Elles consistent à remplacer les points de quadrature fixes utilisés dans les méthodes d'intégration au-dessus par des points aléatoires générés selon une loi de probabilité muni d'une PDF appropriée. Par exemple, lors du passage de la deuxième à la troisième ligne dans l'Eq.6.6, des points aléatoires  $(u_i)_{i=1,\dots,M}$  ont été générés à partir de la PDF  $f_U$  (loi Gamma dans le cas de l'*a priori* MBKF). Par la loi des grands nombres, cette approximation tend en loi vers la vraie intégrale. Cette approche se traduit à la fois par sa grande simplicité et sa précision. C'est pour ces raisons que nous l'avons adopté pour le calcul des intégrales apparaissant dans l'Eq.6.13. A noter

que d'autres variantes existent dans la littérature d'analyse numérique statistique; voir [Winkler, 1995] pour plus de détails.

## 6.3 Expérimentation et résultats

### 6.3.1 Objectifs de l'expérience

L'objectif est d'évaluer la performance de notre estimateur bayésien multivarié MBKF, combiné avec la FDCT et l'UDWT, en le comparant à diverses méthodes de débruitage univariées et multivariées.

Six méthodes de débruitage sont considérées : le seuillage universel dur avec UDWT et FDCT (dénomé respectivement dans la suite "UDWT Hard" et "FDCT Hard"), notre estimateur bayésien univarié BKF avec UDWT et FDCT, le filtrage combiné de [Starck et al., 2003] et le BLS GSM de [Portilla et al., 2003].

### 6.3.2 Protocole expérimental

TAB. 6.1 – Protocole expérimental

Grandeur	Choix		
	Fig.6.1	Fig.6.3	Tab.6.3
<i>a priori</i> multivarié	MBKF, Jeffrey (BLS [Portilla et al., 2003])		
Débruiteurs	Seuillage dur, BKF, MBKF, BLS GSM, Filtrage combiné		
Transformée	UDWT, FDCT		
Échelle de décomposition $J_c$	5 échelles		
qmf	Symmlet de régularité 6		
Bruit	additif blanc gaussien		
PSNR	$\sigma_\epsilon = 40$ (16 dB)	$\sigma_\epsilon = \{5, 20, 40, 60, 100\}$ ([8,35] dB)	
Créaire d'évaluation	PSNR		
Monte-Carlo			10
Filtrage combiné	curvelets + UDWT ou LDCT		
BLS GSM	"Full steerable wavelets" avec 8 orientations		
Images tests	Lena	Barbara	5 images {Lena, Barbara, Boat, Mandrill, Peppers}

### 6.3.3 Discussion des résultats

Afin de quantifier les performances de ces différentes méthodes, nous avons employé le PSNR donné par l'Eq.4.40. Les Figures 6.1, 6.2, 6.3 et 6.4 montrent les images résultantes pour chaque méthode de débruitage pour l'image de Lena (peu texturée) et Barbara (texturée), avec un PSNR = 16dB en entrée.

Pour l'image **peu texturée** de Lena, on voit que le PSNR des débruiteurs avec l'UDWT est supérieur à ceux avec la FDCT. En revanche, le zoom sur une région texturée du chapeau de Lena montre que les débruiteurs avec la FDCT conservent bien les détails fins de l'image (e.g. les rayures du chapeau), même si la FDCT a tendance à produire certains artefacts. Cette

contradiction apparente entre l'impression visuelle et le PSNR pose la question éternelle en débruitage de la pertinence du PSNR comme mesure de performance.

Nous remarquons aussi que le débruiteur MBKF multivarié dépasse celui de BKF univarié ce qui montre l'intérêt d'introduire l'information géométrique dans le voisinage des coefficients. Le débruiteur UDWT MBKF dépasse le débruiteur de filtrage combiné, mais il reste également inférieur au débruiteur BLS GSM qui dépasse tous les débruiteurs.

Pour l'image **texturée** de Barbara, on voit clairement que la qualité visuelle des débruiteurs UDWT Hard et FDCT Hard est inférieure à celles des autres débruiteurs. De plus, nous constatons que les débruiteurs avec FDCT dépassent ceux avec UDWT en terme de PSNR que ce soit pour la cas univarié ou multivarié. Cette différence se traduit aussi en terme de qualité visuelle particulièrement sur les parties texturées (e.g. pantalon). Nous constatons aussi que notre débruiteur FDCT MBKF présente des performances similaires (une différence de PSNR de 0.04 dB) au débruiteur BLS GSM et dépasse le reste des débruiteurs. La FDCT avec l'*a priori* MBKF semblent particulièrement adaptés pour des structures linéaires fines comme on peut l'observer sur le zoom de l'image de Barbara.

Dans le Tab.6.3, nous avons comparé les moyennes et les écart-type du PSNR (sur 10 simulation), donnés par les divers débruiteurs, pour cinq images tests. Les PSNRs ont été calculés pour les cinq valeurs d'entrée de  $\sigma_\epsilon = \{5, 20, 40, 60, 100\}$ , ce qui correspond à un PSNR dans la gamme de [8, 35]dB. Tout d'abord, nous constatons que les débruiteurs MBKF présentent des performances supérieures comparées aux débruiteurs BKF. Ceci montre encore une fois que, grâce à l'introduction de l'information géométrique du voisinage, les estimateurs multivariés s'avèrent performant comparés aux estimateurs univariés. Par ailleurs, les débruiteurs MBKF que nous proposons restent comparables aux débruiteurs BLS GSM de [Portilla et al., 2003], et sont mêmes meilleurs que le filtrage combiné de [Starck et al., 2003] dans certaines situations.

Une comparaison des temps de calcul des différents débruiteurs sur l'image test de Barbara, montrée Fig.6.3, a été menée comme il est montré dans Tab.6.2. En analysant ces temps de calcul, nous constatons tout d'abord que le débruiteur de filtrage combiné est extrêmement consommateur en temps de calcul du fait qu'il soit un débruiteur itératif. Ensuite, les débruiteurs BKF présentent un temps de calcul supérieur comparé à ceux des débruiteurs MBKF. Ceci est dû principalement au calcul de la fonction de cylindre parabolique (voir Théorème 4.1). Aussi, nous constatons que nos débruiteurs BKF et MBKF avec l'UDWT présentent un temps de calcul deux fois supérieur à ceux avec la FDCT. Ceci peut s'expliquer par la redondance de l'UDWT qui est de  $3J_c + 1$  ( $J_c = 5$  dans notre simulation d'où une redondance de 16), deux fois supérieure à la redondance de la FDCT qui est au plus 7.8. Finalement, notre débruiteur FDCT MBKF s'avère le plus rapide comparé aux débruiteur BLS GSM, BKF et MBKF UDWT.

## 6.4 Conclusion

Dans ce chapitre, nous avons mis en place un nouvel estimateur bayésien multivarié basé sur une extension de l'*a priori* BKF, formulé dans la cadre des transformées multi-échelles orientées (curvelets) et non-orientées (ondelettes non-décimées).

Tout d'abord, trouver une forme analytique simple pour la PDF jointe des coefficients bruités est une tâche très compliquée. Pour cela, nous avons employé la technique d'intégration de Monte-Carlo afin de trouver une forme permettant d'évaluer numériquement les intégrales dans l'expression de la PDF jointe *a posteriori* ainsi que celle de l'estimateur ECP. De plus, nous avons proposé un estimateur des hyperparamètres pour mettre en application cette forme. Ensuite, nous avons décrit les étapes nécessaires pour l'implémentation du débruiteur bayésien multivarié

TAB. 6.2 – Comparaison en temps de calcul de différents débruiteurs sur l'image test de barbara de taille  $512 \times 512$  (Fig.6.3), pour un PSNR=16 dB en entrée ( $\sigma_\epsilon = 40$ ). Tous les débruiteurs sauf le filtrage combiné ont été codés en Matlab avec des routines accélérées écrites en C, et ont été testés sur une station Linux équipée d'un processeur Intel(R) Xeon(TM) à 2.8 GHz. Le filtrage combiné était implémenté en C++. Nous avons obtenu son exécutable par [Starck et al., 2003] compilé uniquement pour Solaris que nous avons fait tourner sur une station Sun Blade 2000/1000 avec un processeur UltraSparc-III+ à 1.8 Ghz.

	Temps de calcul
UDWT Hard	2 sec
FDCT Hard	13 sec
UDWT BKF	5 min 56 sec
FDCT BKF	2 min 04 sec
UDWT MBKF	1 min 58 sec
FDCT MBKF	58 sec
Filtrage combiné	13 min 39 sec
BLS GSM	1 min 30 sec

MBKF.

Les résultats expérimentaux ont montré que les performances de notre débruiteur, avec la FDCT et l'UDWT, s'avèrent supérieures à celles du débruiteur univarié BKF. Cette performance a été constatée surtout sur les images texturées, et reste liée à l'introduction de l'information géométrique. En outre, notre débruiteur soutient donc tout à fait la comparaison avec les débruiteurs qui sont aujourd'hui à la pointe de l'état de l'art, comme l'estimateur BLS [Portilla et al., 2003] et le filtrage combiné [Starck et al., 2003].

Finalement, avec le MBKF aucune transformée multi-échelle n'est meilleure pour toutes les images, d'où la nécessité de prendre en compte la diversité morphologique. Nous avons annoncé une première tentative pour mettre un débruiteur ECP combinant plusieurs transformées [Fadili et al., 2007], à l'image du filtrage combiné de [Starck et al., 2003] qui peut être vue comme un débruiteur MAP.



FIG. 6.1 – Comparaison des différents débruiteurs sur l'image test de Lena (peu texturée).



FIG. 6.2 – Comparaison des différents débruiteurs sur l'image test de Lena zoomée sur une région texturée du chapeau. Le zoom montre que la qualité visuelle des débruiteurs avec la FDCT est meilleure à ceux avec l'UDWT.



FIG. 6.3 – Comparaison des différents débruiteurs sur l'image test de Barbara (texturée).

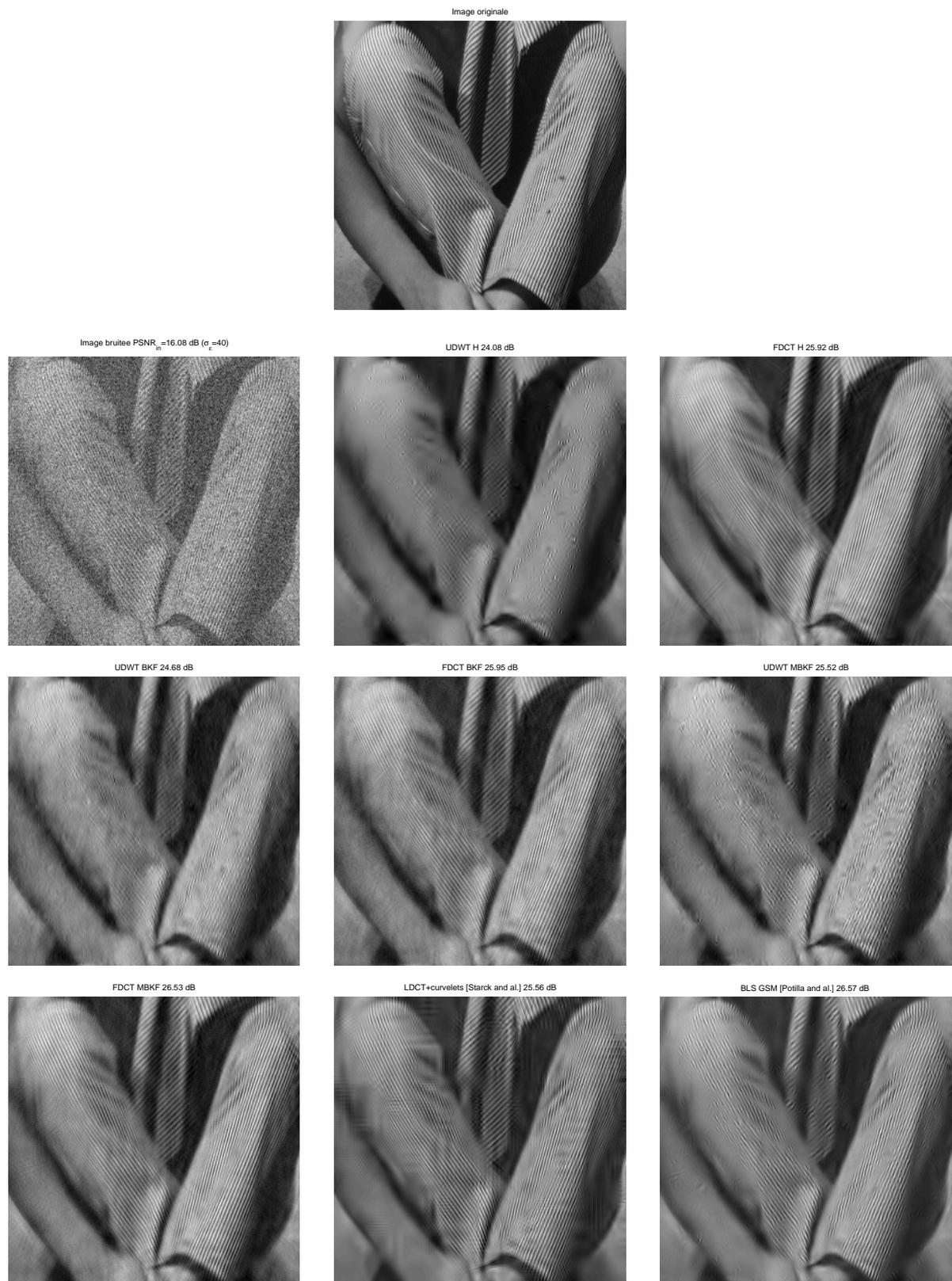


FIG. 6.4 – Comparaison des différents débruiteurs sur l'image test de Barbara zoomée sur une région texturée du pantalon.

TAB. 6.3 – Les moyennes et les écart-type (sur 10 simulations) du PSNR données par les divers débruiteurs. Pour chaque image test, nous montrons en gras les deux meilleurs PSNR correspondant à chaque valeur de  $\sigma$ .

Lena 512 × 512					
$\sigma_\epsilon$	5	20	40	60	100
Image bruitée (SNR)	34.15	22.11	16.09	12.57	8.13
UDWT Hard	37.13 (0.01)	31.40 (0.03)	28.31 (0.04)	26.53 (0.04)	24.38 (0.04)
FDCT Hard	36.96 (0.01)	31.36 (0.03)	28.32 (0.04)	26.55 (0.05)	24.28 (0.05)
UDWT BKF	38.19 (0.01)	31.36 (0.03)	28.20 (0.04)	26.51 (0.04)	24.55 (0.05)
FDCT BKF	38.00 (0.01)	31.32 (0.04)	27.84 (0.16)	25.96 (0.12)	23.17 (0.26)
UDWT MBKF	<b>38.46</b> (0.01)	<b>32.12</b> (0.07)	<b>28.84</b> (0.08)	<b>27.00</b> (0.06)	<b>24.95</b> (0.06)
FDCT MBKF	38.29 (0.01)	31.61 (0.05)	28.03 (0.03)	25.92 (0.05)	23.10 (0.07)
Filtrage combiné	37.55 (0.02)	32.00 (0.02)	28.57 (0.03)	26.47 (0.05)	23.54 (0.04)
BLS GSM	<b>38.36</b> (0.01)	<b>32.36</b> (0.03)	<b>29.22</b> (0.05)	<b>27.43</b> (0.06)	<b>25.33</b> (0.08)
Barbara 512 × 512					
UDWT Hard	32.20 (0.02)	27.09 (0.02)	24.00 (0.03)	22.90 (0.01)	21.71 (0.03)
FDCT Hard	36.12 (0.01)	28.84 (0.02)	25.76 (0.01)	24.00 (0.03)	21.94 (0.05)
UDWT BKF	36.05 (0.02)	28.21 (0.02)	24.57 (0.03)	23.14 (0.02)	21.85 (0.03)
FDCT BKF	37.14 (0.02)	29.43 (0.03)	25.80 (0.06)	23.79 (0.07)	21.29 (0.21)
UDWT MBKF	37.28 (0.01)	29.36 (0.03)	25.54 (0.04)	23.72 (0.03)	<b>22.09</b> (0.05)
FDCT MBKF	<b>37.70</b> (0.02)	<b>30.09</b> (0.09)	<b>26.41</b> (0.08)	<b>24.23</b> (0.07)	21.62 (0.04)
Filtrage combiné	35.72 (0.01)	28.57 (0.03)	25.52 (0.03)	23.72 (0.06)	21.51 (0.08)
BLS GSM	<b>37.75</b> (0.01)	<b>30.21</b> (0.02)	<b>26.50</b> (0.02)	<b>24.48</b> (0.03)	<b>22.47</b> (0.03)
Boat 512 × 512					
UDWT Hard	32.52 (0.01)	23.75 (0.02)	21.95 (0.02)	21.06 (0.01)	20.18 (0.02)
FDCT Hard	32.76 (0.01)	24.07 (0.03)	22.17 (0.03)	21.22 (0.02)	20.16 (0.03)
UDWT BKF	34.91 (0.01)	25.07 (0.04)	22.60 (0.03)	21.50 (0.03)	20.44 (0.03)
FDCT BKF	34.97 (0.01)	25.23 (0.04)	22.81 (0.05)	21.48 (0.07)	19.94 (0.11)
UDWT MBKF	<b>35.12</b> (0.01)	<b>26.33</b> (0.02)	<b>23.19</b> (0.03)	<b>21.82</b> (0.03)	<b>20.59</b> (0.03)
FDCT MBKF	<b>35.19</b> (0.01)	<b>26.37</b> (0.02)	23.15 (0.03)	21.62 (0.01)	19.89 (0.03)
Filtrage combiné	34.22 (0.01)	25.17 (0.01)	22.36 (0.02)	21.14 (0.01)	19.74 (0.04)
BLS GSM	35.02 (0.01)	26.25 (0.02)	<b>23.20</b> (0.02)	<b>21.83</b> (0.02)	<b>20.59</b> (0.02)
Mandrill 512 × 512					
UDWT Hard	35.31 (0.01)	28.99 (0.02)	26.18 (0.02)	24.61 (0.04)	22.78 (0.04)
FDCT Hard	35.11 (0.01)	28.99 (0.02)	26.21 (0.03)	24.68 (0.03)	22.79 (0.06)
UDWT BKF	36.88 (0.01)	29.51 (0.03)	26.34 (0.02)	24.74 (0.04)	22.93 (0.07)
FDCT BKF	36.65 (0.03)	29.54 (0.03)	26.29 (0.09)	24.44 (0.11)	21.92 (0.25)
UDWT MBKF	<b>37.09</b> (0.03)	<b>30.22</b> (0.03)	<b>26.96</b> (0.04)	<b>25.20</b> (0.04)	<b>23.32</b> (0.04)
FDCT MBKF	36.93 (0.01)	29.80 (0.05)	26.45 (0.03)	24.55 (0.03)	22.08 (0.05)
Filtrage combiné	35.65 (0.01)	29.65 (0.02)	26.48 (0.03)	24.64 (0.03)	22.02 (0.03)
BLS GSM	<b>36.94</b> (0.01)	<b>30.22</b> (0.02)	<b>27.08</b> (0.02)	<b>25.42</b> (0.04)	<b>23.62</b> (0.05)
Peppers 256 × 256					
UDWT Hard	36.36 (0.04)	29.17 (0.04)	25.45 (0.02)	23.14 (0.04)	21.05 (0.06)
FDCT Hard	35.43 (0.03)	28.78 (0.05)	25.72 (0.06)	23.60 (0.10)	21.08 (0.08)
UDWT BKF	37.40 (0.05)	29.44 (0.04)	25.49 (0.04)	23.14 (0.07)	21.10 (0.06)
FDCT BKF	36.77 (0.04)	29.12 (0.05)	25.74 (0.10)	23.42 (0.11)	20.69 (0.18)
UDWT MBKF	<b>37.68</b> (0.03)	<b>30.25</b> (0.06)	26.21 (0.09)	23.77 (0.06)	21.56 (0.08)
FDCT MBKF	36.92 (0.03)	29.27 (0.04)	25.77 (0.05)	23.49 (0.08)	20.64 (0.08)
Filtrage combiné	36.73 (0.02)	<b>29.92</b> (0.02)	<b>26.46</b> (0.05)	<b>24.47</b> (0.07)	<b>21.92</b> (0.09)
BLS GSM	<b>37.10</b> (0.02)	29.80 (0.05)	<b>26.36</b> (0.05)	<b>24.23</b> (0.08)	<b>21.82</b> (0.05)

## Conclusion et perspectives



# Conclusion et perspectives

---

Dans ce document, nous avons étudié le problème du débruitage qui a engendré une importante littérature en pré-traitement des images. D'où l'idée de restaurer une image de bonne qualité à partir de sa version dégradée. Nous nous sommes focalisés sur l'estimation bayésienne dans le domaine des transformées multi-échelles parcimonieuses orientées et non-orientées comme solution au problème.

Dans la première partie de cette thèse, nous avons exposé le formalisme général du problème de débruitage dans le domaine des transformées multi-échelle en présence d'un bruit additif gaussien. Ainsi, nous avons effectué un état de l'art méthodique sur le sujet, où nous avons classé les principales méthodes s'appuyant sur les représentations multi-échelles orientées et non-orientées en deux types d'approches : approche classique de type seuillage de coefficients et une approche bayésienne basé sur des modèles statistiques adaptés à la modélisation des coefficients. Pour conclure cet état de l'art, nous avons proposé un bilan comparatif entre ces deux types d'approches tout en soulignant leurs limitations. Ensuite, nous nous sommes penché sur la modélisation statistique des images afin de définir les propriétés de l'image qu'on cherche à recouvrer et pour traduire les connaissances *a priori* qu'on possède sur cette image. Nous avons donc évoqué quelques propriétés et lois statistiques permettant de caractériser les images dans l'espace des transformées. Nous avons aussi exposé les principaux modèles statistiques adaptés à capturer le caractère parcimonieux des coefficients de représentation des images.

Dans la deuxième partie, nous avons proposé un cadre statistique univarié comme *a priori* bayésien adéquat à la modélisation des statistiques marginales des coefficients des images dans le domaine des transformées multi-échelles non-orientées, et en l'occurrence les ondelettes. Les lois marginales ont été analytiquement modélisées par le biais des distributions  $\alpha$ -stable et les Formes K de Bessel. Ensuite, nous avons mis en place des estimateurs bayésiens univariés en mettant à profit les statistiques marginales. Plus précisément, nous avons proposé deux estimateurs bayésiens de type ECP pour l'*a priori* des Formes K de Bessel et un estimateur ECP pour l'*a priori*  $\alpha$ -stable. La mise en oeuvre des ces estimateurs a été effectuée en deux étapes, la première consistant à estimer les hyperparamètres du modèle de l'*a priori* en présence du bruit et la seconde à trouver une forme analytique pour l'estimateur bayésien.

Dans la troisième partie, nous nous sommes intéressés à améliorer les performances de nos estimateurs bayésiens univariés en introduisant l'information géométrique dans le voisinage des coefficients. Cette dernière est traduite par les dépendances inter- et intra-échelles des coefficients. Nous avons donc proposé un cadre statistique bayésien multivarié permettant de prendre en compte les statistiques jointes des coefficients dans le domaine des transformées multi-échelles non-orientées (les ondelettes non-décimées) et orientées (les curvelets). Ensuite, nous avons mis en place un nouvel estimateur bayésien multivarié basé sur une extension multivariée de la

distribution des Formes K de Bessel. La réalisation de notre estimateur a été une tâche très compliquée car aucune forme analytique simple pour la PDF jointe n'est disponible. Pour palier à ce problème, nous avons employé la technique d'intégration de Monte-carlo afin de trouver une forme simple pour le PDF. Ensuite, nous avons proposé un estimateur des hyperparamètres pour mettre en application cette forme, et réaliser le débruiteur bayésien multivarié.

Pour conclure, les débruiteurs bayésiens que nous avons réalisés, notamment le débruiteur multivarié, ont présenté des performances comparables à certains débruiteurs qui sont aujourd'hui à la pointe de l'état de l'art. Ces performances restent liées à la qualité du modèle *a priori* choisi et à son estimateur des hyperparamètres. De plus, nous avons remarqué qu'aucune transformée n'est tout le temps meilleure qu'une autre pour toutes les images. Ce constat est particulièrement vrai lorsque le débruiteur multivarié est utilisé.

## PERSPECTIVES

Plusieurs voies peuvent être envisagées dans le sillage de ce travail. Tout d'abord, nous souhaitons étudier d'autres *a priori* pour qui leurs estimateurs bayésiens peuvent être calculés analytiquement. Nous désirons aussi élargir notre étude sur les performances de notre débruiteur MBKF en le comparant à d'autres méthodes récemment développées, notamment les moyennes non-locales développées par [Buadès et al., 2005], les méthodes à base de patches [Dabov et al., 2006, Kervrann & Boulanger, 2006] et ceux liées à d'autres transformées adaptatives comme les bandelets [LePennec & Mallat, 2005] ou la DCT adaptative [Foi et al., 2007]. Par ailleurs, des réflexions doivent être engagées pour savoir comment adapter les estimateurs bayésiens que nous avons proposés au problème de la déconvolution et à d'autres types de bruit. Nous envisageons aussi d'approfondir notre étude sur les statistiques jointes dans le cadre bayésien multivarié avec plus d'une transformée. Plus précisément, sur la nécessité de prendre en compte la diversité morphologique des structures incluses dans une image. Dans cette direction, nous avons amorcé une première tentative pour mettre en application un estimateur bayésien combinant plusieurs transformées multi-échelle, chacune adaptée pour représenter de façon parcimonieuse une partie de l'image.

# Annexes



## Annexe A

# Éléments de la théorie de l'estimation

---

Dans le monde réel, l'image observée  $Y$  issue d'un capteur est une représentation dégradée d'une image non-observée  $X$ . Ainsi, cette image  $Y$  est considérée comme une fonction déterministe et/ou aléatoire de l'image  $X$ . Dans le paradigme bayésien, attribuer une loi de probabilité  $\mathcal{P}_Y$  aux données observées  $Y$  et une loi a priori  $\mathcal{P}_X$  aux données non-observées  $X$  permettent de traduire des informations sur l'image à estimer  $X$ . Choisir une loi  $\mathcal{P}(\cdot)$  dépend d'un certain nombre de paramètres  $\theta$ . Plusieurs méthodes sont utilisées pour déterminer les valeurs de ces paramètres. Parmi les plus adoptées dans le domaine de restauration bayésienne on trouve : la méthode du maximum de vraisemblance, l'algorithme EM et la méthode des moments. Dans ce qui suit, nous introduisons des définitions théoriques ainsi quelques propriétés pour ces estimateurs, ainsi que, le cadre théorique de l'analyse bayésienne dans le cas continu.

### A.1 Le maximum de vraisemblance

#### A.1.1 Définition générale

Soit  $Y$  une VA munie de sa densité de probabilité  $\mathcal{P}_Y$ . Considérons que cette PDF est paramétrée par un vecteur d'hyperparamètres  $\theta \in \Theta$ . L'estimation du vecteur  $\theta$  à partir d'observations  $\mathbf{y} = \{y_i\}_{i=1, \dots, m}$  de  $Y$  par la méthode maximum de vraisemblance consiste à choisir l'estimateur d'hyperparamètres  $\theta$  pour lequel la probabilité des observations est la plus forte.

**Définition A.1** *L'estimateur du maximum de vraisemblance (MV) de  $\theta$  est la valeur maximisant la fonction de vraisemblance :*

$$\hat{\theta}_{\text{MV}} = \arg \max_{\theta \in \Theta} \underbrace{\prod_{i=1}^m \mathcal{P}_Y(y_i; \theta)}_{\ell(\mathbf{y}; \theta)} \quad (\text{A.1})$$

ou encore celui minimisant  $-\ell(\mathbf{y}; \theta)$  ; le  $-\log$  de la fonction de vraisemblance  $\ell(\mathbf{y}; \theta)$  appelé score de vraisemblance :

$$\hat{\theta}_{\text{MV}} = \arg \max_{\theta \in \Theta} \ell(\theta) = \arg \min_{\theta \in \Theta} - \sum_{i=1}^m \log \mathcal{P}_Y(y_i; \theta) \quad (\text{A.2})$$

### A.1.2 Quelques propriétés

Cet estimateur jouit de plusieurs propriétés intéressantes notamment :

#### Propriété A.1

- (i) *S'il existe un estimateur sans biais à variance minimale, il sera donné par la méthode du maximum de vraisemblance. La solution de l'équation de vraisemblance est unique dans ce cas.*
- (ii) *Sous certaines conditions,  $\hat{\theta}_{MV}$  est asymptotiquement gaussien.*

Le maximum de vraisemblance est l'estimateur dominant en statistiques. Cependant, les solutions analytiques simples aux équations de vraisemblance sont une exception plutôt qu'une règle. Pour cette raison, les méthodes numériques pour trouver ces estimées du MV sont d'une importance primordiale. La plus populaire d'entre elles est l'algorithme EM.

## A.2 L'algorithme EM

L'algorithme EM, proposé par Dempster [Dempster et al., 1977], est basé sur la notion de données incomplètes ou bien manquantes. Il permet de trouver le MV des paramètres de modèles probabilistes lorsque le modèle dépend de variables (données) non observées.

L'algorithme EM est un algorithme itératif qui alterne deux étapes d'évaluation :

**Étape Espérance (E)** Cette étape est vue comme celle reconstruisant les données manquantes. Elle permet de calculer l'espérance de la vraisemblance en tenant compte des dernières variables observées.

**Étape Maximisation (M)** Cette étape sert à estimer les hyperparamètres de l'étape E. Plus précisément, elle estime le MV des paramètres en maximisant la vraisemblance trouvée à l'étape E. Les hyperparamètres estimés en M seront utilisés comme point de départ d'une nouvelle phase d'évaluation de l'étape E.

### A.2.1 Définition générale

Soient  $Y$  les données observées incomplètes et  $Z$  les données complètes non-observées. L'idée générale de l'algorithme EM est de choisir  $Z$  de manière à rendre plus aisée l'estimation au sens du MV sur les données complètes. Nous considérerons que la VA  $Z$  possède une PDF  $\mathcal{P}_Z(Z; \theta)$  fonction de  $Z$  et des hyperparamètres  $\theta$ . L'algorithme itère entre deux étapes comme suit :

**Algorithme 9** Algorithme EM**Entrées:** Une estimation initiale  $\theta_0$ ,  $t = 0$ 1: **répéter**2: **Étape E** : L'espérance conditionnelle suivante est calculée :

$$Q(\theta|\theta_t) = \mathbb{E}[\ell(\mathbf{z}; \theta) | Y, \theta_t] \quad (\text{A.3})$$

3: **Étape M** :

$$\theta_{t+1} = \arg \max_{\theta \in \Theta} Q(\theta|\theta_t) \quad (\text{A.4})$$

4:  $t = t + 1$ 5: **jusqu'à** Convergence.**A.2.2 Propriétés**

- L'algorithme EM jouit d'une excellente stabilité numérique. Il mène à une augmentation systématique de la vraisemblance des observations, ce qui lui confère une propriété importante du point de vue de l'optimisation. Il permet aussi d'incorporer élégamment des contraintes sur les hyperparamètres à estimer.
- L'inconvénient opposé à ces avantages est la lenteur de convergence au voisinage de l'optimum, ce qui est directement lié au taux de données manquantes. Sous peu de conditions, l'algorithme EM garantit la convergence vers un point stationnaire de la fonction de vraisemblance. Cependant, cet optimum n'est pas global mais local, la plupart du temps dépendant du point de départ.
- La qualité des hyperparamètres estimés au final dépend de l'initialisation (par la méthode des moments par exemple).
- La propriété importante de l'algorithme EM est que la maximisation de  $Q(\theta|\theta_t)$  entraîne un accroissement de la fonction de vraisemblance  $\ell(\mathbf{y}; \theta)$  des observations.

**A.2.3 Autres variantes de l'algorithme EM**

L'algorithme EM, bien que très performant, pose parfois quelques problèmes de convergence notamment qui ont donné lieu à d'autres variantes. Parmi ceux-ci, on trouve :

- **EM généralisé** : proposé par Dempster [Dempster et al., 1977] et permet de simplifier le problème de l'étape de maximisation pour assurer la convergence vers un maximum local de vraisemblance. Donc, il n'est pas nécessaire de maximiser  $Q$  à chaque étape car une simple amélioration de  $Q$  est suffisante.

- **EM stochastique** : proposé par Celeux & Diebolt [Celeux & Diebolt, 1985] et permet de réduire le risque de tomber dans un optimum local de la vraisemblance.

**A.3 Paradigme bayésien en images**

Le paradigme bayésien s'impose comme un cadre théorique de choix pour la prise en compte de l'*a priori* imposé sur les signaux/images pour diverses tâches : restauration, débruitage, segmentation, reconnaissance, classification, etc.

### A.3.1 Lois a priori et a posteriori

Pour l'exemple de débruitage avec un bruit additif gaussien blanc et pour estimer  $X$  ayant observé  $Y$ ,  $\hat{x}$  est une solution candidate si elle minimise la distance (au sens  $\ell_2$ ) entre  $Y$  et  $X$  ce qui correspond à la minimisation de la variance du bruit. La solution n'est pas unique et en plus, elle n'est pas stable en présence de bruit. Par conséquent, il faut imposer des contraintes supplémentaires pour réduire le nombre de candidats à la solution dans l'espace  $\mathcal{X}$ . En traitement d'images, nous parlons parfois de régularisation de la solution. Un exemple est de forcer la solution à appartenir à un espace de régularité donnée (norme  $L_p$  finie, norme  $L_p$  d'une mesure différentielle, variation totale bornée, etc.).

Au delà du problème de débruitage, ce raisonnement peut être étendu quelle que soit la tâche que l'on projette d'accomplir sur une image. Ainsi, et de manière plus générale,  $X$  dénotera le paramètre pertinent (considéré comme une VA) que l'on veut extraire de l'image ayant observé  $Y$ . Le paradigme bayésien offre un cadre théorique flexible pour imposer des contraintes sur  $X$  au travers d'une loi de probabilité sur l'espace  $\mathcal{X}$ . Cette loi est alors appelée **loi a priori**.

Le second ingrédient est la loi de l'observation  $Y$  sachant  $X$ . Celle-ci est essentiellement obtenue par la loi du bruit contaminant les mesures dans le modèle de formation de l'image. A partir de la loi a priori et de la conditionnelle a priori, on peut définir la PDF jointe définie sur l'espace produit  $\mathcal{X} \times \mathcal{Y}$  :

$$\mathcal{P}_{X,Y}(x,y) = \mathcal{P}_{Y|X}\mathcal{P}_X(x) \quad (\text{A.5})$$

Ayant les observations  $y$ , il est aisé de prouver que la PDF **conditionnelle a posteriori**, déduite par la règle de Bayes, est :

$$\mathcal{P}_{X|Y}(x|y) = \frac{\mathcal{P}_{Y|X}(y|x)\mathcal{P}_X(x)}{\int_{\mathcal{X}} \mathcal{P}_{Y|X}(y|x)\mathcal{P}_X(x)dx} \quad (\text{A.6})$$

### A.3.2 Estimation bayésienne

Soit  $Y$  l'image observée de laquelle nous voulons déduire l'image non-observée  $X$ . L'approche bayésienne permet d'ajuster la solution  $\hat{x}$  aux données tout en obéissant à des contraintes de "qualité". Il existe différentes manières de choisir l'estimée  $\hat{x}$  de  $x$  tout en assurant une certaine régularité.

En estimation bayésienne, un estimateur est défini comme une application :

$$\begin{aligned} \mathcal{D} : \mathcal{Y} &\longrightarrow \mathcal{X} \\ y &\longrightarrow \hat{x} = \mathcal{D}(y) \end{aligned}$$

pour laquelle l'estimée  $\hat{x}$  est la plus proche possible de  $x$ . Le terme proche suppose que l'on dispose d'une métrique pour quantifier l'écart entre  $x$  et son estimée. Définissons la fonction coût  $C(x, \hat{x})$  positive vérifiant  $C(x, \hat{x}) = 0$  (une pseudo-distance puisque l'inégalité triangulaire n'est pas forcément vérifiée). Le risque bayésien est le risque moyen relativement à la probabilité jointe de  $X$  et  $\hat{X}$  :

$$\begin{aligned} R(x, \mathcal{D}(y)) &= \mathbb{E}_{X, \hat{X}} [C(x, \mathcal{D}(y))] = \int_{\mathcal{X} \times \mathcal{Y}} C(x, \mathcal{D}(y)) \mathcal{P}_{X,Y}(x, y) dx dy \\ &= \int_{\mathcal{Y}} \int_{\mathcal{X}} C(x, \mathcal{D}(y)) \mathcal{P}_{Y|X}(y|x) \mathcal{P}_X(x) dx dy \quad (\text{A.7}) \end{aligned}$$

Un estimateur bayésien est celui minimisant ce risque. Ainsi, chaque estimateur bayésien différent correspond à une fonction de coût spécifique. D'ailleurs les performances d'un estimateur bayésien sont affectées non seulement par la qualité de l'a priori, mais aussi par le choix de la fonction de coût.

Nous allons dans la suite expliciter certains des estimateurs les plus utilisés dans la pratique en traitement du signal et des images.

### A.3.2.1 Maximum a posteriori (MAP)

Soit la fonction de coût suivante :

$$C(x, \hat{x}) = \begin{cases} 0 & \text{si } x = \hat{x} \\ 1 & \text{si } x \neq \hat{x} \end{cases} \quad (\text{A.8})$$

L'estimateur bayésien correspondant à la fonction de coût ci-dessus est le maximum a posteriori, i.e. :

$$\hat{x}_{\text{MAP}} = \arg \max_{x \in \mathcal{X}} \mathcal{P}_{X|Y}(x|y) \quad (\text{A.9})$$

### A.3.2.2 Espérance conditionnelle a posteriori (ECP)

Considérons le cas de l'erreur quadratique pour la fonction de coût :

$$C(x, \mathcal{D}(y)) = (x - \mathcal{D}(y))^2 \quad (\text{A.10})$$

L'estimateur bayésien correspondant à la fonction de coût ci-dessus est l'espérance conditionnelle a posteriori, i.e. :

$$\hat{x}_{\text{ECP}} = \mathbb{E}[X|Y] = \frac{\int_{\mathcal{X}} x \mathcal{P}_{Y|X}(y|x) \mathcal{P}_X(x) dx}{\int_{\mathcal{X}} \mathcal{P}_{Y|X}(y|x) \mathcal{P}_X(x) dx} \quad (\text{A.11})$$

### A.3.2.3 Médiane conditionnelle a posteriori (MCP)

Lorsque la fonction de coût devient l'erreur absolue :

$$C(x, \mathcal{D}(y)) = |x - \mathcal{D}(y)| \quad (\text{A.12})$$

L'estimateur bayésien correspondant à la fonction de coût ci-dessus est la médiane conditionnelle a posteriori, i.e. :

$$\hat{x}_{\text{MCP}} = \left\{ \hat{x} : \mathcal{P}(X > \hat{x} | Y = y) = \mathcal{P}(X \leq \hat{x} | Y = y) \right\} = \left\{ \hat{x} : \int_{\hat{x}}^{\infty} \mathcal{P}_{X|Y}(x|y) dx = \frac{1}{2} \right\} \quad (\text{A.13})$$



## Annexe B

# Estimation de McCulloch

---

### B.1 Tables

TAB. B.1 – Estimé du paramètre  $\alpha$  en fonction de  $\nu_\alpha$  et  $\nu_\beta$ .

$\nu_\alpha$	$\nu_\beta$						
	0.0	0.1	0.2	0.3	0.5	0.7	1.0
2.439	2.000	2.000	2.000	2.000	2.000	2.000	2.000
2.5	1.916	1.924	1.924	1.924	1.924	1.924	1.924
2.6	1.808	1.813	1.829	1.829	1.829	1.829	1.829
2.7	1.729	1.730	1.737	1.745	1.745	1.745	1.745
2.8	1.664	1.663	1.663	1.663	1.676	1.676	1.676
3.0	1.563	1.560	1.553	1.548	1.547	1.547	1.547
3.2	1.484	1.480	1.471	1.460	1.448	1.438	1.438
3.5	1.391	1.386	1.378	1.364	1.337	1.318	1.318
4.0	1.279	1.273	1.266	1.250	1.210	1.184	1.150
5.0	1.128	1.121	1.114	1.101	1.067	1.027	0.973
6.0	1.029	1.021	1.014	1.004	0.974	0.935	0.874
8.0	0.896	0.892	0.887	0.883	0.855	0.823	0.769
10.0	0.818	0.912	0.806	0.801	0.780	0.756	0.691
15.0	0.698	0.695	0.692	0.689	0.676	0.656	0.595
25.0	0.593	0.590	0.588	0.586	0.579	0.563	0.513

TAB. B.2 – Estimé du paramètre  $\beta$  en fonction de  $\nu_\alpha$  et  $\nu_\beta$ .

$\nu_\alpha$	$\nu_\beta$						
	0.0	0.1	0.2	0.3	0.5	0.7	1.0
2.439	0.000	2.160	1.000	1.000	1.000	1.000	1.000
2.5	0.000	1.592	3.390	1.000	1.000	1.000	1.000
2.6	0.000	0.759	1.800	1.000	1.000	1.000	1.000
2.7	0.000	0.482	1.048	1.694	1.000	1.000	1.000
2.8	0.000	0.360	0.760	1.232	2.229	1.000	1.000
3.0	0.000	0.253	0.518	0.823	1.575	1.000	1.000
3.2	0.000	0.203	0.410	0.632	1.244	1.906	1.000
3.5	0.000	0.165	0.332	0.499	0.943	1.560	1.000
4.0	0.000	0.136	0.271	0.404	0.689	1.230	2.195
5.0	0.000	0.109	0.216	0.323	0.539	0.827	1.917
6.0	0.000	0.096	0.190	0.284	0.472	0.693	1.759
8.0	0.000	0.082	0.163	0.243	0.412	0.601	1.596
10.0	0.000	0.074	0.147	0.220	0.377	0.546	1.362
25.0	0.000	0.056	0.112	0.167	0.285	0.428	1.274

TAB. B.3 – Les différentes valeurs de la fonction  $\vartheta_3(\alpha, \beta)$ .

$\alpha$	$\beta$				
	0.0	0.25	0.5	0.75	1.0
2.0	1.908	1.908	1.908	1.908	1.908
1.9	1.914	1.915	1.916	1.918	1.921
1.8	1.921	1.922	1.927	1.936	1.947
1.7	1.927	1.930	1.943	1.961	1.987
1.6	1.933	1.940	1.932	1.997	2.043
1.5	1.939	1.952	1.988	2.045	2.116
1.4	1.946	1.967	2.022	2.106	2.211
1.3	1.955	1.984	2.037	2.188	2.333
1.2	1.935	2.007	2.125	2.294	2.491
1.1	1.980	2.040	2.205	2.435	2.696
1.0	2.000	2.085	2.311	2.624	2.973
0.9	2.040	2.149	2.461	2.886	3.356
0.8	2.098	2.244	2.676	3.265	3.912
0.7	2.189	2.392	3.004	3.844	4.775
0.6	2.337	2.635	3.542	4.808	6.247
0.5	2.588	3.073	4.534	6.636	9.144

## Annexe C

# Modèle multivarié gaussien généralisé anisotrope (AMGGD)

---

Nous introduisons, dans cette annexe, le modèle multivarié gaussien généralisé anisotrope (AMGGD) pour la classe des PDFs marginales à queues lourdes (comportement leptokurtique), afin de modéliser les dépendances inter- et intra-échelles entre les coefficients des images dans l'espace des transformées (voir [Boubchir & Fadili, 2005c, Boubchir & Fadili, 2005b] pour plus de détails).

### C.1 Modèle multivarié analytique

**Définition C.2** Si un vecteur de VAs  $\mathbf{X} = (X_1, X_2, \dots, X_d)^T$  dans  $\mathbb{R}^d$  suit une loi multivariée gaussienne généralisée **anisotrope** centrée, alors sa PDF s'écrit :

$$f_{\mathbf{X}}(\mathbf{x}; \alpha, \Sigma) = \frac{\det \Sigma^{-\frac{1}{2}}}{(A(\alpha) B(\alpha))^d} \exp \left( - \left\| \frac{\Sigma^{-\frac{1}{2}} \mathbf{x}}{B(\alpha)} \right\|_{\alpha}^{\alpha} \right) \quad (\text{C.1})$$

où

$$A(\alpha) = \frac{2}{\alpha} \Gamma \left( \frac{1}{\alpha} \right), \alpha > 0 \quad (\text{C.2})$$

et

$$B(\alpha) = \sqrt{\frac{\Gamma \left( \frac{1}{\alpha} \right)}{\Gamma \left( \frac{3}{\alpha} \right)}} \quad (\text{C.3})$$

$\alpha$  est le paramètre de forme et  $\Sigma$  une matrice symétrique définie positive. Pour rappel,  $\|\mathbf{x}\|_{\alpha}^{\alpha} = \sum_{i=1}^d |x_i|^{\alpha}$  est la norme  $\ell_{\alpha}$  du vecteur  $\mathbf{x}$ .

Cette définition englobe le cas particulier  $d = 1$  (une VA) et le cas des distributions normales multivariées ( $\alpha = 2$ ). Ce modèle de PDF jouit de plusieurs propriétés intéressantes que nous avons établies.

## C.2 Estimation des hyperparamètres

Pour mettre en application notre modèle multivarié de PDF, nous avons proposé des estimateurs (moments et maximum de vraisemblance) des hyperparamètres associés :  $\alpha$  et  $\Sigma$ .

L'intérêt de la paramétrisation de l'Eq. C.1 est que

$$\text{Cov}[\mathbf{X}] = \Sigma \quad (\text{C.4})$$

Ainsi, sachant  $\mathbf{X}$  (estimé par la méthode des moments) comme indiqué ci-dessus, nous pouvons mettre en place l'estimateur du maximum de vraisemblance (MV) pour obtenir  $\alpha$ . L'estimé de  $\alpha$ , dans ce cas, est donné par la proposition suivante :

**Proposition C.1** *L'équation de vraisemblance correspondant à  $\alpha$  sachant  $\Sigma$  s'écrit comme suit*

$$\hat{\alpha}_{MV} = \arg \min_{\alpha > 0} \ell(\alpha) = \arg \min_{\alpha > 0} - \sum_{i=1}^n \log f_{\mathbf{X}}(\mathbf{x}_i; \alpha, \Sigma) \quad (\text{C.5})$$

où  $\alpha$  est la racine de l'équation suivante :

$$\begin{aligned} F(\alpha) &= \frac{\alpha \partial \ell(\alpha)}{\partial \alpha} \\ &= \sum_{i,j} \left| \frac{y_{i,j}}{B(\alpha)} \right|^\alpha \cdot \log \left| \frac{y_{i,j}}{B(\alpha)} \right|^\alpha - \frac{1}{2} \sum_{i,j} \left| \frac{y_{i,j}}{B(\alpha)} \right|^\alpha \cdot \left( \Psi \left( \frac{1}{\alpha} \right) - 3\Psi \left( \frac{3}{\alpha} \right) \right) - nd - \frac{3nd}{2\alpha} \left( \Psi \left( \frac{1}{\alpha} \right) - \Psi \left( \frac{3}{\alpha} \right) \right) \\ &= 0 \end{aligned} \quad (\text{C.6})$$

avec

$$y_{i,j} = \Sigma_j^{-\frac{1}{2}} \mathbf{x}_i \quad (\text{C.7})$$

$j$  représente la  $j^{\text{ème}}$  composante du vecteur  $\mathbf{y}_i$  où  $\mathbf{y}_i = \Sigma^{-1/2} \mathbf{x}_i$ ,  $\mathbf{x}_{i, \{i=1, \dots, n\}}$  sont les réalisations de  $\mathbf{X}$  et  $\Psi$  est la fonction digamma<sup>10</sup>.

**Preuve:**

Preuve de l'Eq.C.6

$$\begin{aligned} \hat{\alpha} &= \arg \min_{\alpha \in \mathbb{R}^+} \ell(\alpha) \\ &= \arg \min_{\alpha \in \mathbb{R}^+} - \sum_{i=1}^n \log f_{\mathbf{X}}(\mathbf{x}_i; \alpha, \Sigma) \\ &= \arg \min_{\alpha \in \mathbb{R}^+} - \sum_{i=1}^n \left( \log \frac{\det |\Sigma|^{-\frac{1}{2}}}{(A(\alpha)B(\alpha))^d} - \left\| \frac{\Sigma^{-\frac{1}{2}} \mathbf{x}_i}{B(\alpha)} \right\|^\alpha \right) \\ &= \arg \min_{\alpha \in \mathbb{R}^+} \left( \sum_{i,j} \left| \frac{\Sigma_j^{-\frac{1}{2}} \mathbf{x}_i}{B(\alpha)} \right|^\alpha + \frac{n}{2} \log \det |\Sigma| + nd \log \left( \frac{2}{\alpha} \Gamma \left( \frac{1}{\alpha} \right) \right) + nd \log \left( \Gamma^{\frac{1}{2}} \left( \frac{1}{\alpha} \right) \Gamma^{-\frac{1}{2}} \left( \frac{3}{\alpha} \right) \right) \right) \end{aligned}$$

<sup>10</sup>Pour rappel,  $\Psi(x) = \frac{d \log(\Gamma(x))}{dx}$ .

$$\begin{aligned}
F(\alpha) = \frac{\partial \ell(\alpha)}{\partial \alpha} = 0 \implies & \sum_{i,j} \left( \exp \left( \log \left| \Sigma_j^{-\frac{1}{2}} B^{-1}(\alpha) \mathbf{x}_i \right| \cdot \alpha \right) \right)' - \frac{nd}{\alpha} - \frac{nd}{\alpha^2} \Psi \left( \frac{1}{\alpha} \right) - \frac{nd}{2\alpha^2} \log \Gamma \left( \frac{1}{\alpha} \right) \\
& + \frac{3nd}{2\alpha^2} \log \Gamma \left( \frac{3}{\alpha} \right) = 0 \\
& \sum_{i,j} \left| \Sigma_j^{-\frac{1}{2}} B^{-1}(\alpha) \mathbf{x}_i \right|^\alpha \cdot \left[ \log \left| \Sigma_j^{-\frac{1}{2}} B^{-1}(\alpha) \mathbf{x}_i \right| - \alpha \left( \frac{1}{2\alpha^2} \left( \Psi \left( \frac{1}{\alpha} \right) - 3\Psi \left( \frac{3}{\alpha} \right) \right) \right) \right] \\
& - \frac{nd}{\alpha} - \frac{3nd}{2\alpha^2} \left[ \Psi \left( \frac{1}{\alpha} \right) - \Psi \left( \frac{3}{\alpha} \right) \right] = 0 \\
& \sum_{i,j} \left| \frac{\Sigma_j^{-\frac{1}{2}} \mathbf{x}_i}{B(\alpha)} \right|^\alpha \cdot \log \left| \frac{\Sigma_j^{-\frac{1}{2}} \mathbf{x}_i}{B(\alpha)} \right|^\alpha - \frac{1}{2} \sum_{i,j} \left| \frac{\Sigma_j^{-\frac{1}{2}} \mathbf{x}_i}{B(\alpha)} \right|^\alpha \left( \Psi \left( \frac{1}{\alpha} - 3\Psi \left( \frac{3}{\alpha} \right) \right) \right) - nd \\
& - \frac{3nd}{2\alpha} \left[ \Psi \left( \frac{1}{\alpha} \right) - \Psi \left( \frac{3}{\alpha} \right) \right] = 0 \tag{C.8}
\end{aligned}$$

■

Le résultat suivant établit l'existence de l'estimateur du MV de  $\alpha$ .

**Proposition C.2 Existence d'une solution**

- (i)  $F(\alpha)$  a au moins une racine sur  $\mathbb{R}^{+*}$ , et donc il existe au moins une solution (non nécessairement unique)  $\hat{\alpha}_{MV}$  à l'estimateur du MV sachant  $\Sigma$ .
- (ii) si  $M = \max_{i,j \in \mathbb{N}} |\mathbf{y}_{i,j}| \leq \sqrt{3}$  alors au minimum global  $\hat{\alpha}_G$  de  $\ell(\alpha)$  on aura toujours  $\ell(\hat{\alpha}_G) < \frac{nd \log(3)}{2}$ .

**Remarque C.1** La preuve de la proposition ci-dessus est longue et technique, et a été volontairement omise ici pour des raisons de synthèse.

L'intérêt du résultat Prop.C.2-(ii) réside dans le fait que nous pouvons rejeter adaptativement certains minima locaux de  $\ell(\alpha)$  (en l'occurrence ceux placés au dessus du  $\frac{nd \log(3)}{2}$ ).

D'après l'Eq. C.4 et la Proposition C.2, nous suggérons l'algorithme d'estimation des hyperparamètres suivant :

---

**Algorithme 10** Estimation des hyperparamètres

---

1: Estimer  $\Sigma$  :

$$\hat{\Sigma} = \frac{1}{n} \sum_i \mathbf{x}_i \mathbf{x}_i^T \tag{C.9}$$

2: Estimer  $\alpha$  :

obtenir  $\hat{\alpha}$  en résolvant numériquement l'Eq. C.6 par une descente de gradient. Cette descente tirera profit de la Prop.C.2-(ii) de façon à éviter d'être piégé dans un minimum local.

---



## Annexe D

# Transformée de Curvelet 2<sup>ème</sup> génération

---

Au cours des dix dernières années, les ondelettes ont eu un immense succès dans le domaine du traitement d'images, et ont été utilisées pour de nombreux problèmes tels que la compression et la restauration d'images [Mallat, 1999]. Ces problèmes ont souvent pour préalable la recherche d'une représentation de l'image qui soit la plus parcimonieuse possible, au sens où un petit nombre de paramètres permet d'obtenir une approximation précise de l'image.

Cependant, il apparaît aujourd'hui clairement que les ondelettes ne sont pas optimales pour l'analyse d'objets anisotropes dans l'image (les lignes, les contours...), mais restent efficaces pour la détection de structures isotropes à différentes échelles. Depuis quelques années, de nouvelles transformées multi-échelles ont été développées -comme les curvelets, contourlets et bandlets- qui intègrent de notion de directionnalité et qui permettent de rechercher des objets de manière optimale dont l'efficacité en traitement d'image a déjà été prouvée.

### D.1 Transformée multi-échelle orientée discrète : FDCT

Les curvelets ont été proposées par E. Candès et D. Donoho [Candès & Donoho, 1999], elles constituent une nouvelle famille de frames d'ondelettes géométriques plus efficaces que les transformées traditionnelles. Elles sont conçues pour représenter de façon parcimonieuse les contours. Par exemple, sur la Fig.D.1(a), les ondelettes prendraient beaucoup de coefficients pour représenter précisément un tel contour. Comparées aux ondelettes, les curvelets peuvent représenter un contour lisse avec moins de coefficients pour la même précision (Fig.D.1(b)).

La transformée de curvelets est une transformée multi-échelle multi-directionnelle avec des atomes indexés par un paramètre de position, d'échelle et de direction [Candès & Donoho, 1999, Candès & Donoho, 2002, Candès et al., 2006]. La Fig.D.2 montre le partitionnement spectral engendré par la FDCT [Candès et al., 2006]. Une curvelet ainsi que son spectre sont montrés dans la Fig.D.3.

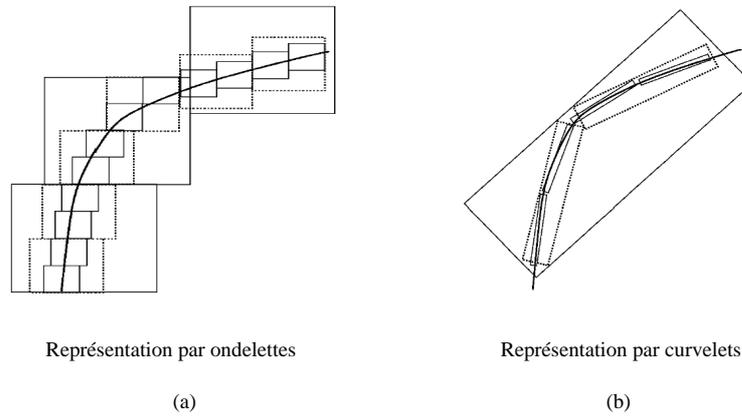


FIG. D.1 – Comparaison de l'approximation non-linéaire des ondelettes et des curvelets.

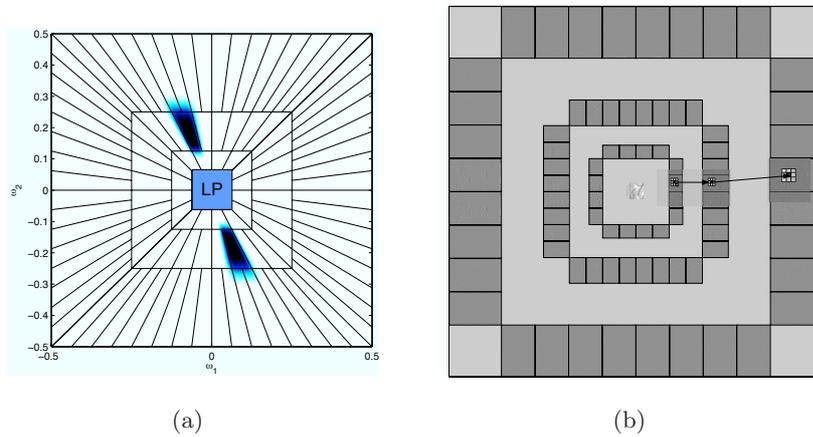


FIG. D.2 – Décomposition pyramidale orientée par la FDCT : (a) Partition spectrale de la FDCT. (b) FDCT de l'image de Lena. Chaque rectangle représente une portion angulaire.

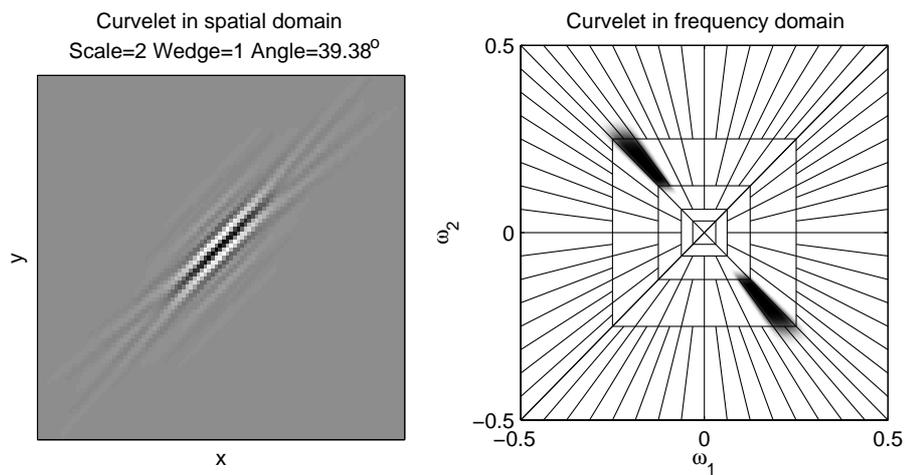


FIG. D.3 – Une curvelet.

## D.2 Propriétés

- La transformée de curvelets est une transformée multi-échelle multi-directionnelle avec des atomes indexés par un paramètre de position  $r$ , d'échelle  $j$  et de direction  $o$ .
- Les atomes sont anisotropes répondant à une loi d'échelle paraboliques où les rectangles d'Heisenberg satisfont  $(2^{-j}) \approx \text{largeur} \approx \text{longueur}^2 (2^{-j/2})$
- La transformée de Fourier  $\hat{\varphi}_{j,o}(r)$  de ces atomes est définie en coordonnées polaires par :

$$\hat{\varphi}_{j,o}(r) = 2^{3j/4} W(2^j r) V\left(\frac{[2^{2j/2}] \theta}{2\pi}\right) \quad (\text{D.1})$$

où  $W$  et  $V$  sont les fenêtres radiale et angulaire.

- Soit  $f_M$  l'approximation non-linéaire d'une fonction  $f \in C^2$  par morceaux au-delà de contours  $C^2$ .  $f_M$  est obtenue en ne gardant que les  $M$  coefficients de la FDCT de  $f$  les plus grands en valeur absolue. Alors l'erreur d'approximation non-linéaire est telle que :

$$\|f - f_M\|_{\ell_2}^2 \leq C (\log M)^3 M^{-2} \quad (\text{D.2})$$

où  $C$  est une constante qui ne dépend que de  $f$ .

- La FDCT implémente une trame ajustée (Tight frame). En effet, on peut écrire que :

$$f = \frac{1}{A} \sum_{\gamma} \langle f, \varphi_{\gamma} \rangle \varphi_{\gamma} \quad (\text{D.3})$$

où de façon équivalente, la relation de conservation d'énergie suivante :

$$\|f\|_{\ell_2}^2 = \frac{1}{A} \sum_{\gamma} |\langle f, \varphi_{\gamma} \rangle|^2 \quad (\text{D.4})$$

où  $\gamma = \{j, o, t\}$  (échelle, orientation, position) et  $A$  est la borne de la trame ajustée (i.e. le paramètre de redondance).

- La FDCT est une transformée redondante :  $A \leq 7.8$ .
- Complexité algorithmique : la FDCT possède une implémentation rapide par la FFT (algorithme wrapping [Candès et al., 2006]) qui est donc en  $O(N^2 \log_2 N)$  pour une image de taille  $N \times N$ .



## Annexe E

# Publications & diffusion des résultats

---

Dans cette annexe, on trouve la liste de publications et séminaires effectuées durant la thèse.

### Journaux internationaux avec comité de lecture

- L. Boubchir, J. Fadili, "*A Closed-form Nonparametric Bayesian Estimator in the Wavelet-domain of Images Using an Approximate  $\alpha$ -stable Prior*", Pattern Recognition Letters, Vol. 27, No. 12, pp. 1370–1382, 2006.
- J. Fadili, L. Boubchir, "*Analytical Form for a Bayesian Wavelet Estimator of Images Using the Bessel  $K$  Formes Densities*", IEEE Transactions on Image Processing, Vol. 14, No. 2, pp. 231–240, 2005.

### Conférences internationales

- L. Boubchir, J. Fadili, D. Bloyet, "*Bayesian Denoising in the Wavelet-domain Using an Analytical Approximate  $\alpha$ -stable prior*", in *Proc. of ICPR'2004*; the 17<sup>th</sup> International Conference on Pattern Recognition, Vol 4, pp. 889–892, August 23-26, Cambridge, United Kingdom, 2004.
- L. Boubchir, J. Fadili, "*Multivariate Statistical Modeling of Images with the Curvelet Transform*", in *Proc. of IEEE ISSPA'2005*; the IEEE International Conference on Signal Processing and Its Applications, pp. 747–750, August 28-31, Sydney, Australia, 2005.
- L. Boubchir, J. Fadili, "*Bayesian Denoising Based on the MAP Estimator in Wavelet-domain Using Bessel  $K$  Form Prior*", in *Proc. of IEEE ICIP'2005*; the IEEE International Conference on Image Processing, Vol I, pp. 113–116, September 11-14, Genoa, Italy, 2005.
- J. Fadili, L. Boubchir, "*Sparse representations and Bayesian Denoising*", *SIAM IS'2006*; the SIAM Conference on Imaging Science, May 15-18, Minneapolis, USA, 2006.
- J. Fadili, J. -L. Starck, L. Boubchir, "*Morphological Diversity and Sparse Image Denoising*", *IEEE ICASSP'2007*; the International Conference on Acoustics, Speech and Signal Processing, Vol I, pp. 589–592, April 15-20, Honolulu, Hawaii, USA, 2007.

## Conférences nationales

- L. Boubchir, J. Fadili, D. Bloyet, "*Le processus  $\alpha$ -stable pour l'estimation bayésienne non-paramétrique des images dans le domaine des ondelettes*", In *Proc. of JEMOSICO'2003*; Journées d'Etudes sur les Méthodes pour les Signaux Complexes en Traitement d'Images, pp. 105–118, December 9-10, INRIA Rocquencourt, Paris, France, 2003.
- L. Boubchir, J. Fadili, "*Modélisation statistique multivariée des images dans le domaine de la transformée de Curvelet*", In *Proc. of GRETSI'2005*; 20<sup>th</sup> GRETSI Symposium on Signal and Image Processing, pp. 233-236, September 6-9, Louvain-la-Neuve, Belgium, 2005.
- L. Boubchir, "*Algorithme EM pour l'estimation des hyperparamètres du débruiteur bayésien d'images basé sur l'a priori des Formes K de Bessel*", in *Proc. of JETIM'2006*; Journées d'Etudes Algéro-Française en Imagerie Médicale, pp. 47–54, Novembre 21-22, Alger, Algérie, 2006.

## Conférences nationales sans actes et sans comité de lecture

- L. Boubchir, J. Fadili, D. Bloyet, "*Le processus  $\alpha$ -stable pour l'estimation bayésienne non-paramétrique des images dans la domaine des ondelettes*", *Conf. IFR 47*; Institut Fédératif de Recherche en Neuro-Imagerie Fonctionnelle, Cyceron, Septembre 13, Caen, France, 2003.

## Séminaires

- L. Boubchir, J. Fadili, "*Approches bayésiennes pour le débruitage des images dans la domaine des transformées multi-échelles*", *Séminaire Ecole Doctorale SIMEM*; Avril 11, Université de Caen, 2005.
- L. Boubchir, J. Fadili, "*Paradigme bayésien et transformées multi-échelles pour la restauration d'images*", *Séminaire GREYC*; Novembre 11, CREYC, Université de Caen, 2005.

# Bibliographie

- [Abramovich & Benjamini, 1995] Abramovich, F. & Benjamini, Y. (1995). Thresholding of wavelet coefficients as multiple hypotheses testing procedure. *In Antoniadis, A. and Oppenheim, G., editors, Wavelets and Statistics, volume 103 of Lecture Notes in Statistics*, (pp. 5–14).
- [Abramovich & Benjamini, 1996] Abramovich, F. & Benjamini, Y. (1996). Adaptive thresholding of wavelet coefficients. *Computational Statistics & Data Analysis*, 22, 351–361.
- [Abramovich et al., 2000] Abramovich, F., Benjamini, Y., Donoho, D., & Johnstone, I. (2000). Adapting to unknown sparsity by controlling the false discovery rate.
- [Abramovich & Sapatinas, 1999] Abramovich, F. & Sapatinas, T. (1999). Bayesian approach to wavelet decomposition and shrinkage.
- [Abramovich et al., 1998] Abramovich, F., Sapatinas, T., & Silverman, B. (1998). Wavelet thresholding via a bayesian approach. *J. R. Statist. Soc. B*, 60, 725–749.
- [Abramowitz & Stegun, 1972] Abramowitz, M. & Stegun, I. A. (1972). *Handbook of Mathematical Functions*. New York : Dover Publications.
- [Achim et al., 2001] Achim, A., Bezerianos, A., & Tsakalides, P. (2001). Novel bayesian multiscale method for speckle removal in medical ultrasound images. *IEEE Trans. Med. Imag.*, 20, 772–783.
- [Achim & Kuruoglu, 2004] Achim, A. & Kuruoglu, E. E. (2004). Image denoising bivariate  $\alpha$ -stable distributions in the complex wavelet domain. *IEEE Signal Processing letters.*, 12(1), 17–20.
- [Akaike et al., 1973] Akaike, H., Petrox, B. N., & Caske, F. (1973). Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory. Akademiai Kiado, Boudapest.*, (pp. 267–281).
- [Andrews & Mallows, 1974] Andrews, D. F. & Mallows, C. L. (1974). Scale mixtures of normality. *Journal of the Royal Statistical Society, Series B* 36, (pp. 99–102).
- [Antoniadis et al., 2001] Antoniadis, A., Bigot, J., & Sapatinas, T. (2001). Wavelet estimators in nonparametric regression : A comparative simulation study. *Journal of Statistical Software*, 6(6), 1–83.
- [Bates & McLaughlin, 2000] Bates, S. & McLaughlin, S. (2000). The estimation of stable distribution parameters from teletraffic data. *IEEE Transactions on Signal Processing*, 48, 865–870.
- [Benjamini & Hochberg, 1995] Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate : a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, 57, 289–300.
- [Boubchir, 2006] Boubchir, L. (2006). Algorithme EM pour l'estimation des hyperparamètres du débruiteur bayésien d'images basé sur l'a priori des formes k de Bessel. *In Proc. of JETIM'2006 ; Journées d'Etudes Algéro-Française en Imagerie Médicale*, (pp. 47–54).

- [Boubchir & Fadili, 2005a] Boubchir, L. & Fadili, J. (2005a). Bayesian denoising based on the MAP estimator in wavelet-domain using Bessel  $K$  Form prior. *In Proc. of IEEE ICIP'2005; the IEEE International Conference on Image Processing*, 1, 113–116.
- [Boubchir & Fadili, 2005b] Boubchir, L. & Fadili, J. (2005b). Modélisation statistique multivariée des images dans le domaine du curvelet. *In Proc. of GRETSI'2005; the 20th GRETSI Symposium on Signal and Image Processing*, (pp. 233–236).
- [Boubchir & Fadili, 2005c] Boubchir, L. & Fadili, J. (2005c). Multivariate statistical modeling of images with the curvelet transform. *In Proc. of IEEE ISSPA '2005; the IEEE International Conference on Signal Processing and Its Applications*, (pp. 747–750).
- [Boubchir & Fadili, 2006] Boubchir, L. & Fadili, J. (2006). A closed-form nonparametric bayesian estimator in the wavelet-domain of images using an approximate  $\alpha$ -stable prior. *Pattern Recognition Letters*, 27(12), 1370–1382.
- [Boubchir et al., 2003] Boubchir, L., Fadili, J., & Bloyet, D. (2003). Le processus  $\alpha$ -stable pour l'estimation bayésienne non-paramétrique des images dans la domaine des ondelettes. *In Proc. of JEMOSICO'2003; Journées d'Etudes sur les Méthodes pour les Signaux Complexes en Traitement d'Images*, (pp. 105–118).
- [Boubchir et al., 2004] Boubchir, L., Fadili, J., & Bloyet, D. (2004). Bayesian denoising in the wavelet-domain using an analytical approximate  $\alpha$ -stable prior. *In Proc. of ICPR'2004; the 17<sup>th</sup> International Conference on Pattern Recognition*, 4, 889–892.
- [Bruce & Gao, 1996] Bruce, A. G. & Gao, H. Y. (1996). Understanding waveshrink : variance and bias estimation. *Biometrika*, 83, 727–745.
- [Buadès et al., 2005] Buadès, A., Coll, B., & Morel, J. M. (2005). A non-local algorithm for image denoising. *In Proc. IEEE CVPR*, 2, 60–65.
- [Cai, 1999] Cai, T. T. (1999). Adaptive wavelet estimation : a block thresholding and oracle inequality approach. *Annals of Statistics*, 27, 898–924.
- [Cai & Silverman, 2000] Cai, T. T. & Silverman, B. W. (2000). *Incorporating Information on Neighboring Coefficients into Wavelet Estimation*. Technical report, Departement of Statistics, Purdu University, USA.
- [Candès et al., 2006] Candès, E. J., Demanet, L., Donoho, D. L., & L.ying (2006). Fast discrete curvelet transforms. *SIAM Multiscale Model. Simul.*, 5-3, 861–899.
- [Candès & Donoho, 1999] Candès, E. J. & Donoho, D. L. (1999). Curvelets : A surprisingly effective nonadaptive representation of objects with edges. *Curves and Surface, Vanderbilt University Press, Nashville, TN*, (pp. 123–143).
- [Candès & Donoho, 2004] Candès, E. J. & Donoho, D. L. (2004). New tight frames of curvelets and the problem of approximating piecewise  $c^2$  images with piecewise  $c^2$  edges. *Comm. Pure Appl. Math.*, 57, 219–266.
- [Candès & Donoho, 2002] Candès, E. J. & Donoho, D. L. (June 2002). Recovering edges in ill-posed inverse problems : optimality of curvelet frames. *Annals of Statistics*, 30(3), 784–842.
- [Celeux & Diebolt, 1985] Celeux, G. & Diebolt, D. (1985). The EM algorithm : a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly*, 2(1), 73–82.
- [Chang et al., 2000a] Chang, S., Yu, B., & Vetterli, M. (2000a). Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing*, 9(9), 1522–1531.

- 
- [Chang et al., 2000b] Chang, S., Yu, B., & Vetterli, M. (2000b). Spatially adaptive wavelet thresholding with context modeling for image denoising. *IEEE Trans. Image Processing*, 9(9), 1532–1546.
- [Chipman et al., 1997] Chipman, H., Kolaczyk, E., & McCulloch, R. (1997). Adaptive bayesian wavelet shrinkage. *J. Am. Statist. Ass.*, 92, 1413–1421.
- [Claypoole & Baraniuk, 2000] Claypoole, R. & Baraniuk, R. (2000). A multiresolution wedgelet transform for image processing.
- [Clyde & George, 2000] Clyde, M. & George, E. (2000). Flexible empirical bayes estimation for wavelets. *J. R. Statist. Soc. B*, 62, 681–698.
- [Clyde et al., 1998] Clyde, M., Parmigiani, G., & Vidakovic, B. (1998). Multiple shrinkage and subset selection in wavelets. *Biometrika*, 85(2), 391–401.
- [Clyde & George, 1999] Clyde, M. A. & George, E. I. (1999). Empirical bayes estimation in wavelet nonparametric regression. In P. Muller & B. Vidakovic (Eds.), *Bayesian Inference in Wavelet Based Models* (pp. 309–322). New York : Springer-Verlag.
- [Coifman & Donoho, 1995] Coifman, R. R. & Donoho, D. L. (1995). Translation-invariant denoising. in *Lecture Notes in Statistics : Wavelets and Statistics*, vol. New York : Springer-Verlag, (pp. 125–150).
- [Cover & Thomas, 1991] Cover, T. M. & Thomas, J. A. (1991). *Elements of Information Theory*. New York : John Wiley and Sons, Inc.
- [Crouse & Baraniuk, 1997] Crouse, M. & Baraniuk, R. (1997). Contextual hidden markov models for wavelet-domain signal processing. in *Proc. 31st Asilomar Conf. on Signals, Systems and Computers*.
- [Crouse et al., 1998] Crouse, M., Nowak, R., & Baraniuk, R. (1998). Wavelet-based statistical signal processing using hidden markov models. *IEEE Transactions on Signal Processing*, 46(4), 886–902.
- [Dabov et al., 2006] Dabov, K., Foi, A., Katkovnik, V., & Egiazarian, K. (2006). Image denoising with block-matching and 3D filtering. In *Proc. SPIE Electronic Imaging : Algorithms and Systems V*, volume 6064 San Jose, CA, USA.
- [Delyon & Juditsky, 1996] Delyon, B. & Juditsky, A. (1996). On minimax wavelet estimators. *Applied Computational Harmonic Analysis*, 3, 215–228.
- [Dempster et al., 1977] Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B*, 39(1), 1–38.
- [Do & Vitterli, 2003a] Do, M. N. & Vitterli, M. (2003a). *Contourlets, in beyond Wavelets*. New York : G. V. Welland ed, Academic press.
- [Do & Vitterli, 2003b] Do, M. N. & Vitterli, M. (2003b). Framing pyramids. *IEEE Transactions on Signal Processing*, 51, 2329–2342.
- [Donoho, 1998] Donoho, D. L. (1998). Widgelets : Nearly minimax estimation of edges. *Annals of Statistics*, 27(3), 859–997.
- [Donoho & Johnstone, 1994] Donoho, D. L. & Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3), 425–455.
- [Donoho & Johnstone, 1995] Donoho, D. L. & Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432), 1200–1224.

- [Donoho & Johnstone, 1998a] Donoho, D. L. & Johnstone, I. M. (1998a). Minimax estimation via wavelet shrinkage. *Annals of Statistics*, 26(3), 879–921.
- [Donoho & Johnstone, 1998b] Donoho, D. L. & Johnstone, I. M. (1998b). Minimax estimation via wavelet shrinkage. *Annals of Statistics*, 26(3), 879–921.
- [Donoho et al., 1995] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., & picard, D. (1995). Wavelet shrinkage : Asymptopia ? *J. R. Statist. Soc. B.*, 57(2), 301–337.
- [Efromovich, 1999] Efromovich, S. (1999). Quasi-linear wavelet estimation. *JASA*, 94, 189–204.
- [Efromovich, 2000] Efromovich, S. (2000). Sharp linear and block shrinkage wavelet estimation. *Statist. Proba. Lett.*, 49, 323–329.
- [Eslami & Radha, 2006] Eslami, R. & Radha, H. (2006). Translation-invariant contourlet transform and its application to image denoising. *IEEE Transaction on Image Processing*, 15(11), 3362–3374.
- [Eubank, 1999] Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing*. New York : Marcel dekker, 2nd edition.
- [Fadili & Boubchir, 2005] Fadili, J. M. & Boubchir, L. (2005). Analytical form for a bayesian wavelet estimator of images using the Bessel  $K$  Form densities. *IEEE Transactions on Image Processing*, 14(2), 231–240.
- [Fadili et al., 2007] Fadili, J. M., Starck, J. L., & Boubchir, L. (2007). Morphological diversity and sparse image denoising. *In Proc. of IEEE ICASSP'2007*, 1, 589–592.
- [Fama & Roll, 1971] Fama, E. F. & Roll, R. (1971). Parameters estimates for symmetric stable distribution. *Journal of the American Statistical Association*, 66(334), 331–338.
- [Fan & Xia, 2001] Fan, G. & Xia, X.-G. (2001). Image denoising using local contextual markov model in the wavelet domain. *IEEE Signal Processing Letters*, 8(5), 125–128.
- [Fan & Gijbels, 1996] Fan, J. & Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. London : Chapman & Hall.
- [Figueiredo et al., 1999] Figueiredo, M. A. T., Leitao, J. M. N., & Jain, A. K. (1999). On fitting mixture models. *In Energy Minimization Methods in Computer Vision and Pattern Recognition* (pp. 54–69).
- [Fisher, 1928] Fisher, R. A. (1928). Moments and product moments of sampling distributions. *Proc. London Math. Soc.*, 30, 199–238.
- [Foi et al., 2007] Foi, A., Katkovnik, V., & Egiazarian, K. (2007). Pointwise shape-adaptive dct for high-quality denoising and deblocking of grayscale and color images. *IEEE Trans. Image Processing*, 16(5), 1395–1411.
- [Freeman & Adelson, 1991] Freeman, W. T. & Adelson, E. H. (1991). The design and use of steerable filters. *IEEE Pat. Anal. Mach. Intell.*, 13(9), 891–906.
- [Gao, 1998] Gao, H. Y. (1998). Wavelet shrinkage denoising using the non-negative garrote. *Journal of Computational and Graphical Statistics*, 7(4), 469–488.
- [Gao & Bruce, 1997] Gao, H. Y. & Bruce, A. G. (1997). Waveshrink with firm shrinkage. *Statistica Sinica*, 7, 855–874.
- [Geman & Geman, 1984] Geman, D. & Geman, S. (1984). Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images. *IEEE Trans. PAMI*, 6, 721–741.
- [Gradshteyn & Ryzhik, 1980] Gradshteyn, I. & Ryzhik, I. (1980). *Table of Integrals, Series and Products*. Academic Press, A. Jeffrey edition.

- 
- [Green & Silverman, 1994] Green, E. I. & Silverman, B. W. (1994). *Nonparametric Regression and Generalised Linear Models*. London : Chapman & Hall.
- [Grenander, 1993] Grenander, U. (1993). *General Pattern Theory*. Oxford University Press.
- [Grenander & Srivastava, 2001] Grenander, U. & Srivastava, A. (2001). Probability models for clutter in natural images. *IEEE Trans. PAMI*, 23(4), 424–429.
- [Hall et al., 1997] Hall, P., Penev, S., Kerkyacharian, G., & Picard, D. (1997). Numerical performance of block thresholded wavelet estimators. *Statist. Comput.*, 7, 115–124.
- [Hsung et al., 1999] Hsung, T. C., Lun, D. P. K., & Siu, W. C. (1999). Denoising by singularity detection. *IEEE Trans. on Signal Proc.*, 47, 3139–3144.
- [Huang & Cressie, 2000] Huang, S. & Cressie, N. (2000). Deterministic/stochastique wavelet decomposition for recovery of signal from noisy data. *Technometrics*, 42, 262–276.
- [Huang & Lu, 2000] Huang, S. & Lu, H. (2000). Bayesian wavelet shrinkage for nonparametric mixed effects models. *Statist. Sinica*, 10, 1021–1040.
- [Härdle, 1990] Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge : Cambridge University Press.
- [Härdle et al., 1998] Härdle, W., Kerkyacharian, G., Pickard, D., & Tsybakov, A. (1998). *Wavelets, approximation and statistical applications*. Lecture Note en Statistics 129.
- [Jalobeanu, 2001] Jalobeanu, A. (2001). Modèle, estimation bayésienne et algorithmes pour la déconvolution d’images satellitaires et aériennes. Master’s thesis, Université de Nice-Sophia Antipolis, Nice.
- [Jansen & Bultheel, 1999] Jansen, M. & Bultheel, A. (1999). Geometrical priors for noise-free wavelet coefficient configurations in image denoising. in *Bayesian inference in wavelet based models*, P. Muller and B. Vidakovic, Eds. New York : Springer-Verlag, (pp. 223–242).
- [Jansen & Bultheel, 2001] Jansen, M. & Bultheel, A. (2001). Empirical bayes approach to improve wavelet thresholding for image noise reduction. *Journal of the Amer. Statist. Assoc.*, 96(454), 629–639.
- [Jansen et al., 1997] Jansen, M., Malfait, M., & Bultheel, A. (1997). Generalized cross-validation for wavelet thresholding. *Signal Processing*, 56, 33–44.
- [Johnstone & Silverman, 1998] Johnstone, I. & Silverman, B. (1998). *Empirical Bayes approaches to mixture problems and wavelet regression*. Technical report, Department of Mathematics, University of Bristol, UK.
- [Kervrann & Boulanger, 2006] Kervrann, C. & Boulanger, J. (2006). Optimal spatial adaptation for patch-based image denoising. *IEEE Trans. on Image Processing*, 15(10), 2866–2878.
- [Kingsbury, 1999] Kingsbury, N. (1999). Images processing with complexe wavelets. In *Phil. Trans. Roy. Soc. London A, Special issue for the discussion meeting on 'Wavelets : the key to intermittent information ?'*, 357, 2543–2560.
- [Kotz, 2004] Kotz, S. (2004). *Continous Multivariate Distributions, Models and Applications*. John Wiley & Sons.
- [Koutrouvelis, 1980] Koutrouvelis, I. A. (1980). Regression-type estimation of the parameters of stable laws. *Journal of the American Statistical Association*, 75(372), 918–928.
- [Koutrouvelis, 1981] Koutrouvelis, I. A. (1981). An iterative procedure for the estimation of the parameters of stable laws. *Communications in Statistics-Simulation and Computation*, 10(1), 17–28.

- [Kuruoglu, 1998] Kuruoglu, E. E. (1998). Signal processing in alpha-stable noise environments : A least  $l_p$ -norme approach. Master's thesis, Signal Processing and Communications Laboratory - Department of Engineering, University of Cambridge.
- [LePennec & Mallat, 2005] LePennec, E. & Mallat, S. G. (2005). Bandelet image approximation and compression. *SIAM journal of Multiscale Modeling and Simulation*, 4(3), 992–1039.
- [Lévy, 1954] Lévy, P. (1954). *Théorie de l'addition des variables aléatoires*. Paris : Gautier-Villars.
- [Li, 1995] Li, S. Z. (1995). *Markov Random Field Modeling in Computer Vision*. New York : Springer-Verlag.
- [Liu & Moulin, 2000] Liu, J. & Moulin, P. (2000). Analysis of interscale and intrascale dependencies between wavelet coefficients. in *Proc. Int. Conf. on Image Proc., (ICIP), Vancouver, Canada*, 1, 669–672.
- [Liu & Moulin, 2001] Liu, J. & Moulin, P. (2001). Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients. *IEEE Transactions on Image Processing*, 10(10), 1647–1658.
- [Ma & Nikias, 1995] Ma, X. & Nikias, C. L. (1995). Parameter estimation and blind channel identification in impulsive signal environments. *IEEE Transaction on Signal Processing*, 43(12), 2884–2897.
- [Magnus & Neudecker, 1998] Magnus, J. R. & Neudecker, H. (1998). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Son Ltd.
- [Malfait & Roose, 1997] Malfait, M. & Roose, D. (1997). Wavelet-based image denoising using a markov random field a priori model. *IEEE Trans. Image Processing*, 6, 549–565.
- [Mallat, 1989] Mallat, S. G. (1989). A theory for multiresolution signal decomposition : The wavelet representation. *IEEE trans. PAMI*, 11(7), 674–693.
- [Mallat, 1999] Mallat, S. G. (1999). *A Wavelet Tour of Signal Processing*. 2nd Edition, San Diego : Academic Press.
- [Mallat, 2006] Mallat, S. G. (2006). Geometrical grouplets. *submitted to ACHA*.
- [Mallat & LePennec, 2005] Mallat, S. G. & LePennec, E. (2005). Sparse geometric image representation with bandelets. *IEEE Transaction on Image Processing*, 14(4), 423–438.
- [Mandelbrot & van Ness, 1968] Mandelbrot, B. B. & van Ness, J. W. (1968). Fractional brownian motion, fractional noises and applications. *SIAM Rev.*, 10(4), 422–437.
- [Mardia, 1970] Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519–530.
- [Marron et al., 1998] Marron, J. S., Adak, S., Johnstone, I. M., Neumann, M. H., & Patil, P. (1998). Exact risk analysis of wavelet regression. *Journal of Computational and Graphical Statistics*, 7(3), 278–309.
- [Matalon et al., 2005] Matalon, B., Elad, B., & Zibulevsky, M. (2005). Improved denoising of images using modeling of the redundant contourlet transform. *Proceeding of the SPIE conference wavelets*, 5914.
- [Mathieu, 2002] Mathieu, J. (2002). Transformée en ondelettes et régression non-paramétrique dans un contexte bayésien. Master's thesis, Ecole Nationale Supérieure d'Ingénieur, Caen.
- [McCulloch, 1986] McCulloch, J. H. (1986). Simple consistent estimators of stable distribution parameters. *Communication on Statistics-Simulation*, 15(4), 1109–1136.

- 
- [McLachlan & Peel, 2000] McLachlan, G. J. & Peel, D. (2000). *Finite Mixture Models*. New York : Wiley.
- [Meng & Rubin, 1993] Meng, X. L. & Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm : a general framework. *Biometrika*, 2(80), 267–278.
- [Meyer, 1992] Meyer, Y. (1992). *Wavelets and Operators*. Cambridge : Cambridge University Press.
- [Moulin & Liu, 1999] Moulin, P. & Liu, J. (1999). Analysis of multiresolution image denoising schemes using generalized gaussian and complexity priors. *IEEE Transactions on Information Theory*, 45(3), 909–919.
- [Mumford & Gidas, 2001] Mumford, D. & Gidas, B. (2001). Stochastic models for generic images. *Quarterly of Applied Mathematics*, 4(1), 85–111.
- [Nason, 1994] Nason, G. P. (1994). *Wavelet regression by cross-validation*. Technical report, Dep. of Stat., Stanford University.
- [Nason, 1996] Nason, G. P. (1996). Wavelet shrinkage by cross-validation. *Journal of the Royal Statistical Society B*, 58, 463–479.
- [Nikias & Shao, 1995] Nikias, C. L. & Shao, M. (1995). *Signal Processing with Alpha-Stable Distributions and Applications*. Wiley-Interscience.
- [Nolan, 1997] Nolan, J. P. (1997). Numerical calculation of stable densities and distribution functions. *Statist. -Stochastic Models*, 13, 759–774.
- [Ogden & Parzen, 1996a] Ogden, R. T. & Parzen, E. (1996a). Change-point approach to data analytic wavelet thresholding. *Statist. Comput.*, 6, 93–99.
- [Ogden & Parzen, 1996b] Ogden, R. T. & Parzen, E. (1996b). Data dependent wavelet thresholding in nonparametric regression with change-point applications. *Computational Statistics & Data Analysis*, 22, 53–70.
- [Olshausen & Field, 1996] Olshausen, B. A. & Field, D. J. (1996). Emergence of simple-cell receptive-field properties by learning a sparse code of naturel images. *Nature*, 381(6583), 607–609.
- [Percival & Walden, 2000] Percival, D. B. & Walden, A. T. (2000). *Wavelet Methods for Time Series Analysis*. Cambridge press.
- [Pizurica et al., 2002] Pizurica, A., Philips, W., Lemahieu, I., & Acheroy, M. (2002). A joint inter- and intrascale statistical model for bayesian wavelet based image denoising. *IEEE Transaction on Image Processing*, 11(5), 545–557.
- [Po & Do, 2003] Po, D. D. Y. & Do, M. N. (2003). Directional multiscale statistical modeling of images x. *Proc. of SPIE conference on Wavelet Applications in Signal and Image Processing*, 5270, 69–79.
- [Portilla et al., 2003] Portilla, J., Strela, V., Wainwright, M. J., & Simoncelli, E. P. (2003). Image denoising using scale mixture of gaussians in the wavelet domain. *IEEE Transaction on Image Processing*, 12(11), 1338–1351.
- [Press, 1972] Press, S. J. (1972). *Applied Multivariate Analysis*. New York : Holt, Rinehart and Winston.
- [Press et al., 1992] Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical Recipes in C, The Art of Scientific Computing*. Cambridge : Cambridge University Press, 2nd edition.

- [Rissanen, 1996] Rissanen, J. (1996). Fisher information and stochastic complexity. *IEEE Transaction on Information Theory*, 42, 40–47.
- [Romberg et al., 1999] Romberg, J. K., Choi, H., & Baraniuk, R. (1999). Bayesian tree-structured image modeling using wavelet domain hidden markov model. *in Proc. of SPIE*, 3816, 31–44.
- [Rose & Smith, 2002] Rose, C. & Smith, M. (2002).  $k$ -statistics : Unbiased estimators of cumulants. In *Mathematical Statistics with Mathematica* chapter 7.2, (pp. 256–259). New York : Springer-Verlag.
- [Samorodnitsky & Taqqu, 1994] Samorodnitsky, G. & Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes : Stochastic Models with Infinite Variance*. New York : Chapman & Hall.
- [Schwarz, 1978] Schwarz, G. (1978). Estimation of the dimension of a model. *The Annals of Statistics*, 6, 461–464.
- [Sendur & Selesnick, 2002] Sendur, L. & Selesnick, I. W. (2002). Bivariate shrinkage function for wavelet-based denoising exploiting interscale dependency. *IEEE Transaction on Signal Processing*, 50(11), 2744–2756.
- [Simoncelli, 1999] Simoncelli, E. P. (1999). Bayesian denoising of visual images in the wavelet domain. In P. Muller & B. Vidakovic (Eds.), *Bayesian Inference in Wavelet Based Models* (pp. 291–308). New York : Springer-Verlag.
- [Simoncelli & Adelson, 1996] Simoncelli, E. P. & Adelson, E. H. (1996). Noise removal via bayesian wavelet coring. In *Third Int'l Conf on Image Proc*, volume 1 (pp. 379–382). Lausanne : IEEE Sig Proc Society.
- [Simoncelli et al., 1992] Simoncelli, E. P., Freeman, W. T., Adelson, E. H., & Heeger, D. H. (1992). Shiftable multiscale transforms. *IEEE Transaction Information Theory*, 38(2), 587–607.
- [Simoncelli & Portilla, 1998] Simoncelli, E. P. & Portilla, J. (1998). Texture characterization via joint statistics of wavelet coefficient magnitudes. In *5th IEEE Int'l Conf on Image Processing*, volume 1, Chicago.
- [Srivastava et al., 2002] Srivastava, A., Liu, X., & Grenander, U. (2002). Universal analytical forms for modeling image probabilities. *IEEE PAMI*, 24(9), 1200–1214.
- [Starck et al., 2002] Starck, J.-L., Candès, E. J., & Donoho, D. L. (2002). The curvelet transform for image denoising. *IEEE Transaction on Image Processing*, 11(6), 670–684.
- [Starck et al., 2003] Starck, J.-L., Nguyen, M. K., & Murtagh, F. (2003). Wavelets and curvelets for image deconvolution : a combined approach. *Signal Processing*, 83, 2279–2283.
- [Stein, 1982] Stein, C. (1982). Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, 9, 1135–1151.
- [Tsihrintzis & Nikias, 1996] Tsihrintzis, G. A. & Nikias, C. L. (1996). Fast estimation of the parameters of  $\alpha$ -stable impulsive interference. *IEEE Transactions on Signal Processing*, 44(6), 1492–1503.
- [url : base d'images, ] url : base d'images. <http://sipi.usc.edu/services/database/database.html>.
- [Vannucci & Corradi, 1999] Vannucci, M. & Corradi, F. (1999). Covariance structure of wavelet coefficients : theory and models in a bayesian perspective. *J. R. Statist. Soc. B*, 61, 971–986.
- [Vidakovic, 1998] Vidakovic, B. (1998). Nonlinear wavelet shrinkage with Bayes rules and Bayes factors. *Journal of the American Statistical Association*, 93(441), 173–179.

- 
- [Vidakovic, 1999] Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. New York : John Wiley & Sons.
- [Vidakovic & Ruggeri, 2000] Vidakovic, B. & Ruggeri, F. (2000). *BAMS Method : Theory And Simulations*. Technical report, Institute of Statistics and Decision Sciences, Duke University.
- [Wainwright et al., 2000] Wainwright, M., Simoncelli, E. P., & Willsky, A. (2000). Random cascades of gaussian scale mixtures for natural images. In *7th IEEE Int'l Conf on Image Processing*, volume 1, Vancouver.
- [Wand & Jones, 1995] Wand, M. P. & Jones, M. C. (1995). *Kernel Smoothing*. London : Chapman & Hall.
- [Weron, 1995] Weron, R. (1995). *Performance Of The Estimators Of Stable Law Parameters*. Technical report, HSC/95/1, Wroclaw University of Technology.
- [Weyrich & Warhola, 1995] Weyrich, N. & Warhola, G. T. (1995). Denoising using wavelets and cross-validation. *NATO Adv. Study Inst. C*, 454, 523–532.
- [Winkler, 1995] Winkler, G. (1995). *Image analysis, random fields and dynamic Monte Carlo methods : a mathematical introduction*. Springer-Verlag.
- [Wojtaszczyk, 1997] Wojtaszczyk, P. (1997). *A Mathematical Introduction to Wavelets*. Cambridge University Press.
- [Zhang & Jin, 1996] Zhang, S. & Jin, J. (1996). *Computation of Special Functions*. New York : Wiley & sons.
- [Zhu et al., 2000] Zhu, S. C., Liu, X. W., & Wu, Y. N. (2000). Exploring texture ensembles by efficient markov chain monte carlo-toward a 'trichromacy' theory of texture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(6), 554–569.
- [Zolotarev, 1986] Zolotarev, V. M. (1986). *One-dimensional Stable Distribution*. Amer. Math. Soc. Transl. of Math. Monographs, Vol. 65. Amer. Math. Soc., Providence, RI. (Transl. of the original 1983 Russian).





## Résumé

Les images issues d'une chaîne d'acquisition sont généralement dégradées par le bruit du capteur. La tâche qui consiste à restaurer une image de bonne qualité à partir de sa version bruitée est communément appelée débruitage. Celui-ci a engendré une importante littérature en pré-traitement des images. Lors de ce travail de thèse, et après avoir posé le problème du débruitage en présence d'un bruit additif gaussien, nous avons effectué un état de l'art méthodique sur ce sujet. Les méthodes présentées cherchent pour la plupart à reconstruire une solution qui présente une certaine régularité. En s'appuyant sur un cadre bayésien, la régularité de la solution, qui peut être imposée de différentes manières, a été formellement mise en place en passant dans le domaine des transformées multi-échelles. Ainsi, afin d'établir un modèle d'a priori, nous avons mené une modélisation des statistiques marginales et jointes des coefficients d'images dans le domaine des transformées multi-échelles orientées (e.g. curvelets) et non-orientées (e.g. ondelettes). Ensuite, nous avons proposé de nouveaux estimateurs bayésiens pour le débruitage. La mise en oeuvre de ces estimateurs est effectuée en deux étapes, la première consistant à estimer les hyperparamètres du modèle de l'a priori en présence du bruit et la seconde à trouver une forme analytique pour l'estimateur bayésien correspondant. Dans un premier temps, nous avons mis en place des estimateurs bayésiens univariés en mettant à profit les statistiques marginales des coefficients des images dans des représentations multi-échelles comme les ondelettes. Ces lois marginales ont été analytiquement modélisées par le biais des distributions :  $\alpha$ -stable et les Formes K de Bessel. Dans un second temps, nous avons amélioré les performances de nos estimateurs univariés en introduisant l'information géométrique dans le voisinage des coefficients. Plus précisément, nous avons proposé un cadre statistique bayésien multivarié permettant de prendre en compte les dépendances inter- et intra-échelles des coefficients, en mettant à profit les statistiques jointes de ces derniers dans le domaine des curvelets et des ondelettes non-décimées. Ensuite, nous avons mis en place l'estimateur bayésien multivarié correspondant basé sur une extension multivariée de la distribution des Formes K de Bessel. Une large étude comparative a finalement été menée afin de confronter nos algorithmes de débruitage à d'autres débruiteurs de l'état de l'art.

**Mots-clés:** débruitage, analyse harmonique appliquée, estimation bayésienne, modélisation statistique, transformées multi-échelles, ondelettes, curvelets, parcimonie.

## Abstract

Image data observed at the output of an image acquisition device are generally degraded by the sensor noise. The task which aims at recovering a good quality image from its noisy observations is widely known as denoising. Denoising has been at the heart of a flurry of research activity in the image processing literature. In this work, after defining the denoising problem when data are corrupted by an additive gaussian noise, we provide an extensive and methodical review of the literature. Most of image denoising methods try to narrow down the class of candidate solutions by imposing some prior regularity constraints on the recovered solution. We have chosen to formulate our prior in a bayesian framework, through multiscale transform coefficients of the image. Towards this end, and by appropriately exploiting the sparsity of these multiscale representations, we designed prior models to capture the marginal and joint statistics of such coefficients in oriented (e.g. curvelets) and non-oriented (e.g. wavelets) multiscale pyramids. These priors were then utilized in newly proposed bayesian denoisers. The implementation of these bayesian estimators relies on two key steps for which we suggested efficient solutions : (i) estimate the hyperparameters of the prior model in presence of noise, and (ii) find an analytical form for the corresponding bayesian estimator. In the first part of this thesis, we designed term-by-term univariate bayesian estimators by taking advantage of the marginal statistics of coefficients of images in sparse multiscale representations, e.g. wavelets. These marginal statistics were modelled analytically using  $\alpha$ -stable and Bessel K Form distributions. In the second part, we improved upon the performance of univariate estimators by introducing the geometrical information contained in the neighborhood of each representation coefficients. More precisely, we proposed a multivariate statistical bayesian framework which takes into account the intra- and inter-scale dependencies of coefficients and models the joint statistics of groups of coefficients in the curvelet and the undecimated wavelet domains. The associated multivariate bayesian estimator was also provided based on a multivariate extension of the Bessel K Form distribution. A comprehensive comparative study has been carried out to compare our denoising algorithms to state-of-the-art competitors.

**Keywords:** denoising, applied harmonic analysis, bayesian estimation, statistical modeling, multiscale transforms, wavelets, curvelets, sparsity.