

Mémoire d'habilitation à diriger des recherches

Jérôme IDIER

Problèmes inverses en restauration de signaux et d'images

soutenu le 10 juillet 2000 devant le jury composé de

M.	Patrick	DUVAUT	Examineur
Mme	Line	GARNERO	Examinatrice
MM.	Alain	HILLION	Examineur
	Philippe	LOUBATON	Rapporteur
Mme	Odile	MACCHI	Examinatrice
M.	Kenneth	SAUER	Rapporteur
Mme	Josiane	ZERUBIA	Rapporteur

Table des matières

Organisation du document	5
I Notice individuelle et thèmes d'activité scientifique	7
1 Notice individuelle	9
1.1 État civil	9
1.2 Diplômes universitaires	9
1.3 Expérience professionnelle	9
1.4 Autres fonctions et affiliations	9
1.5 Encadrement de thèses	10
1.6 Participation à des jurys de thèses	10
1.7 Collaborations universitaires	11
1.8 Valorisation	11
1.9 Expérience pédagogique	12
1.10 Autres contributions	13
2 Thèmes d'activité scientifique	15
2.1 Introduction	15
Un exemple classique, l'analyse spectrale AR	16
Localisation de sources et « asymptomania »	17
2.2 Thème 1 – Déconvolution impulsionnelle	19
2.3 Thème 2 – Méthodes markoviennes en imagerie	22
Sous-thème 2a – Modèles « à énergie non convexe »	22
Sous-thème 2b – Inversion d'un problème non-linéaire : la tomographie d'impédance	22
Sous-thème 2c – Modèles « à énergie convexe »	23
Sous-thème 2d – Chaînes de Markov cachées et champs unilatéraux	25
2.4 Thème 3 – Estimation spectrale en temps court	28
Sous-thème 3a – Approche autorégressive	28
Sous-thème 3b – Synthèse de Fourier	29
2.5 Conclusion et perspectives	31
Approches semi-quadratiques	31
Problèmes aveugles	31
3 Notice bibliographique (octobre 1991-99)	35

Articles dans des revues internationales avec comité de lecture	35
Articles dans des comptes rendus de conférences internationales avec comité de lecture	35
Articles dans des comptes rendus de conférences nationales avec comité de lecture	37
Conférences sans actes	38
Chapitres de livres	39
Rapports internes	39
Documents pédagogiques	40
Rapports de contrats	40
 II Restauration de signaux et d'images : une mise en perspective	 41
 4 Approches pénalisées	 43
4.1 La régularisation au sens de Tikhonov	43
4.1.1 Généralités	43
4.1.2 Limitation	44
4.2 Détection-estimation	45
4.2.1 Principe	45
4.2.2 Inconvénients	47
4.3 Approche non quadratique	48
4.4 Critères augmentés semi-quadratiques	50
4.4.1 Dualité entre critères non quadratiques et critères augmentés SQ	50
4.4.2 Minimisation de critères SQ	52
4.5 Conclusion	53
 5 Modèles et outils stochastiques	 55
5.1 Introduction	55
5.2 Cadre statistique	55
5.3 Constructions statistiques des fonctions Ω et Φ	58
5.4 Réglage autodidacte des hyperparamètres	59
5.4.1 Outils « déterministes » : courbe en L, validation croisée	60
5.4.2 Outils probabilistes : maximisation de vraisemblance	60
5.5 Conclusion	66
 III Références bibliographiques	 67
 IV Publications et documents annexés	 77
 A Inversion d'opérateurs linéaires mal posés à support large	 79
 B Critères convexes semi-quadratiques et variables auxiliaires interactives	 81

C	Champs de Markov stationnaires sur un réseau rectangulaire fini	83
D	Stabilité structurelle de méthodes de prédiction par moindres carrés	85

Organisation du document

Ce document est divisé en quatre parties :

- La première partie, **Notice individuelle et thèmes d'activité scientifique**, réunit en trois chapitres les éléments d'un bilan factuel de mon activité scientifique.
 - Le premier chapitre, *Notice individuelle*, rassemble curriculum vitæ, encadrement de thèses et participation à des jurys, expérience pédagogique, etc.
 - Le deuxième chapitre, *Thèmes d'activité scientifique*, présente mon activité de recherche depuis mon recrutement au CNRS en octobre 1991, découpée en thèmes et sous-thèmes, avec des indications chronologiques et la distinction nette entre des thèmes aboutis et des thèmes émergents.
 - Le troisième chapitre, qui s'intitule *Notice bibliographique (octobre 1991-99)*, énumère les travaux dont je suis l'auteur ou le co-auteur depuis octobre 1991. Dans l'ensemble du document, les citations s'y rapportant sont faites exclusivement par des numéros entre crochets.
- La deuxième partie s'intitule **Restauration de signaux et d'images : une mise en perspective**. Il s'agit d'une présentation synthétique et personnelle, en deux chapitres qui s'enchaînent, de ma principale thématique de recherche : les problèmes inverses en restauration de signaux et d'images. Contrairement à la présentation éclatée du Chapitre 3, la cohérence d'ensemble et la progressivité du propos sont privilégiée dans cette partie du document.
 - Le quatrième chapitre, *Approches pénalisées*, aborde les problèmes inverses en traitement de signal et en imagerie sous un angle *déterministe*, à partir des travaux fondateurs de Tikhonov dans les années 1960. La présentation effectuée met l'accent sur la restauration de signaux ou d'images « réguliers par morceaux » : fonctions en escalier, images constituées de régions distinctes, défauts isolés dans un milieu homogène, etc.
 - Le cinquième chapitre, *Modèles et outils stochastiques*, donne l'interprétation *statistique* des outils introduits au chapitre précédent, puis développe de nouveaux outils propres au cadre probabiliste bayésien. Une grande partie de ce chapitre est consacrée à l'inversion « non supervisée », et l'accent est mis sur les aspects inférentiels plutôt que sur les aspects algorithmiques.
- La troisième partie, **Références bibliographiques**, rassemble la bibliographie pour l'ensemble du document, hormis le contenu de la notice bibliographique du Chapitre 3. Dans tous les chapitres, les citations se rapportant à cette bibliographie principale sont faites exclusivement par nom d'auteur et année, entre crochets.
- La quatrième partie, **Publications et documents annexés**, est constituée de trois articles parus et d'un article soumis, donnant un aperçu représentatif de mes travaux.

Dans une certaine mesure, les Parties I, II, et les quatre documents de la Partie IV peuvent être lues indépendamment.

Première partie

Notice individuelle et thèmes d'activité scientifique

Chapitre 1

Notice individuelle

1.1 État civil

Nom :	IDIER
Prénom :	Jérôme
Date et lieu de naissance :	le 30 octobre 1966, à VINCENNES (94)
Nationalité :	Française
Situation de famille :	Célibataire, sans enfant
Adresse personnelle :	3, rés. Les Champs Lasniers, 91940 LES ULIS

1.2 Diplômes universitaires

- Ingénieur de l'École supérieure d'électricité (1988)
- Docteur en Science (Automatique et traitement du signal) de l'Université PARIS-SUD (UPS, 1991)

1.3 Expérience professionnelle

De septembre 1991 à septembre 1995 : CR2 au CNRS, Laboratoire des signaux et systèmes (L2S, UMR 8506, CNRS/SUPÉLEC/UPS)

Depuis octobre 1995 : CR1 au CNRS, L2S

1.4 Autres fonctions et affiliations

- Depuis 1988 : Membre du GdR 134 « Traitement du signal et des images » puis du GdR-PRC ISIS
- Depuis 1992 : Membre du Conseil de laboratoire du L2S
- Depuis mars 1998 : Membre de la Commission de spécialistes de la 61^e section de l'Université PARIS-SUD
- Enseignant vacataire à l'Université PARIS-SUD, l'Université de Cergy-Pontoise, SUPÉLEC, l'ENSTA (voir la rubrique 1.9).

1.5 Encadrement de thèses

- Jean-François GIOVANNELLI, allocataire MENRT, thèse soutenue en février 1995 : « Estimation spectrale en temps court. Application à l'imagerie Doppler » [GIOVANNELLI 1995]. Encadrement à 50% (avec Guy DEMOMENT, Professeur des Universités à l'UPS). Fonction actuelle : Maître de conférence (UPS).
- Mila NIKOLOVA, boursière du gouvernement français, thèse soutenue en février 1995 : « Inversion markovienne de problèmes linéaires mal posés. application à l'imagerie tomographique » [NIKOLOVA 1995], Encadrement à 20% (avec Ali MOHAMMAD-DJAFARI, chercheur CNRS au L2S). Situation actuelle : Chargée de recherche CNRS (ENST).
- Frédéric DUBLANCHET, boursier DGA, thèse soutenue en octobre 1996 : « Contribution de la méthodologie bayésienne à l'analyse spectrale de raies pures et à la goniométrie haute résolution » [DUBLANCHET 1996]. Encadrement à 50% (avec Patrick DUVAUT, Professeur des Universités à l'ENSEA, Cergy-Pontoise). Situation actuelle : Ingénieur d'affaire (Canal Plus).
- Thierry MARTIN, allocataire MENRT, thèse soutenue en novembre 1997 : « Inversion bayésienne du problème non linéaire de tomographie d'impédance électrique modélisé par une méthode d'éléments finis » [MARTIN 1997]. Encadrement à 100%. Situation actuelle : ingénieur R&D (ATDI).
- Marc FAYOLLE, boursier DGA, thèse soutenue en octobre 1998 : « Modélisation unilatérale composite pour la restauration d'images » [FAYOLLE 1998]. Encadrement à 70% (avec Yves GOUSSARD, Professeur à l'ÉPM). Situation actuelle : ingénieur R&D (CGR-GE).
- Stéphane BRETTE, allocataire moniteur normalien en 1993-95 : « Méthodes d'imagerie markoviennes. Application au contrôle non destructif » ; thèse achevée mais non rédigée et non soutenue. Encadrement à 50% (avec Ali MOHAMMAD-DJAFARI, chercheur CNRS au L2S). Situation actuelle : PRAG (IUT de Ville d'Avray)¹.
- Philippe CIUCIU, allocataire MENRT, 3^e année de thèse en 1999-2000 : « Méthodes markovienne pour l'estimation spectrale et application en imagerie Doppler ». Encadrement à 80% (avec Jean-François GIOVANNELLI, Maître de conférence à l'UPS).
- Myriam FIANI, boursière EDF, 2^e année de thèse en 1999-2000 : « Reconstruction 3D appliquée au contrôle non destructif de soudures par radiographie ». Encadrement à 50% (avec Stéphane GAUTIER, ingénieur-chercheur à EdF).
- Amandine BLANC, boursière ONERA, 2^e année de thèse en 1999-2000 : « Identification de réponse impulsionnelle et restauration d'images: apports de la diversité de phase ». Encadrement à 50% (avec Laurent MUGNIER, ingénieur-chercheur à l'ONERA).
- Andrea RIDOLFI, allocataire MENRT, 2^e année de thèse en 1999-2000 : « Méthodes markoviennes pour la vélocimétrie. Applications à l'imagerie médicale ». Encadrement à 30% (avec Jean-François GIOVANNELLI, Maître de conférence à l'UPS).

1.6 Participation à des jurys de thèses

- Frédéric CHAMPAGNAT, décembre 1993, UPS, Orsay [CHAMPAGNAT 1993].
- Nasser CHEBIRA, février 1994, UPS, Orsay [CHÉBIRA 1994].
- Jean-François GIOVANNELLI, février 1995, UPS, Orsay [GIOVANNELLI 1995].
- Frédéric DUBLANCHET, octobre 1996, UPS, Orsay [DUBLANCHET 1996].

1. Stéphane BRETTE a pris ce poste en 1995 sans avoir rédigé son manuscrit de thèse, et n'a malheureusement pas trouvé le temps de le faire par la suite.

- Thierry MARTIN, novembre 1997, UPS, Orsay [MARTIN 1997].
- Marc FAYOLLE, octobre 1998, UPS, Orsay [FAYOLLE 1998].
- Aude CORVAISIER-RICHE, juin 1999, Lyon I [CORVAISIER-RICHE 1999].
- Lionel DELPHIN-POULAT, octobre 1999, Rennes I [DELPHIN-POULAT 1999].

1.7 Collaborations universitaires

- Hervé CARFANTAN, Maître de conférence, Université Paul-Sabatier, TOULOUSE [85].
- Patrick DUVAUT, Professeur des universités, ENSEA ETIS Groupe signal, Université de CERGY-PONTOISE [17, 43, 68].
- Stéphane GAUTIER, Ingénieur-chercheur, EdF-DRD, CHATOU [14, 38, 39].
- Yves GOUSSARD, Professeur agrégé (*Associate professor*), Institut de génie biomédical, École polytechnique de MONTRÉAL [5, 6, 20, 37, 44, 46, 58, 62, 63].
- Alain HERMENT, Directeur de recherche INSERM, « Imagerie médicale quantitative », C.H.U. Pitié-Salpêtrière [23, 24, 29, 61].
- Laurent MUGNIER, Ingénieur-chercheur, ONERA CHÂTILLON [33].
- Jean-Luc STARCK, Chercheur au CEA, DAPNIA/SEI-SAP, SACLAY [79].

1.8 Valorisation

- Je suis co-inventeur avec Yves GOUSSARD et Frédéric CHAMPAGNAT de deux logiciels de déconvolution dont les licences d'exploitation ont été acquises par le CEA en 1995.
- J'ai participé à dix contrats d'études, dont six comme responsable scientifique, pour le compte d'industriels (THOMSON, L'ORÉAL, SCHLUMBERGER) ou d'établissements publics (le CNES, le CEA) ; voir page 40.
- En liaison avec le Service du partenariat et de la valorisation du CNRS (Délégation Ile-de-France sud), j'ai suscité et négocié la mise en place de trois accords de collaboration entre les tutelles du L2S (CNRS/SUPÉLEC, puis CNRS/SUPÉLEC/UPS) et d'autres organismes scientifiques :
 - Une convention de recherche a été signée en octobre 1998 avec l'École polytechnique de MONTRÉAL (ÉPM). Son objet est : « Développement de méthodes pour la résolution de problèmes inverses en imagerie et applications ». Les questions concernant la propriété industrielle et l'exploitation des résultats ont été plus spécifiquement traitées pour couvrir les retombées d'un projet franco-qubécois de conception de prothèses personnalisées (voir page 24).
 - Une convention a été signée fin 1998 avec EdF, portant sur « Reconstruction 3D appliquée au contrôle non destructif de soudures par radiographies ». Je travaille sur ce thème avec Stéphane GAUTIER, du Groupe « Traitements avancés de l'information » de la Direction « Recherche et développement » d'EdF. La thèse de Myriam FIANI s'inscrit dans ce cadre.
 - Une convention a été signée en 1999 avec l'Office National d'Études et de Recherches Aérospatiales (ONERA), portant sur « Identification de réponse impulsionnelle et restauration d'images : apports de la diversité de phase ». Je travaille sur ce sujet, qui relève de l'imagerie optique à haute résolution, avec Laurent MUGNIER, du Département « Optique théorique et appliquée » de l'ONERA. La thèse d'Amandine BLANC s'inscrit dans ce cadre.

1.9 Expérience pédagogique

Cours magistraux	91-92	92-93	93-94	94-95	95-96	96-97	97-98	98-99
<i>Problèmes inverses, Traitement du signal, SUPÉLEC formation continue</i>	3h30	4h30	1h30	1h30	3h	5h	3h30	1h30
<i>Signal et communication, ENSTA 2^e année</i>			6h					
<i>Bases mathématiques et algorithmes pour le traitement du signal, ENSTA 2^e année</i>			6h	12h	8h	8h		
<i>Problèmes inverses, SUPÉLEC 3^e année</i>							18h	
<i>Problèmes inverses, DEA TIS, CERGY-PONTOISE²</i>							15h	15h
<i>Problèmes inverses et modèles markoviens, DEA ATS, UPS³</i>							21h	21h
Travaux dirigés	91-92	92-93	93-94	94-95	95-96	96-97	97-98	98-99
<i>Processus markoviens, ENSTA 1^{re} année</i>	10h	10h						
<i>Probabilités, statistiques, signal aléatoire, SUPÉLEC 1^{re} et 2^e année</i>	30h	30h	21h					
<i>Modélisation des phénomènes aléatoires, ENSTA 1^{re} année</i>					12h	24h	24h	12h
<i>Analyse numérique et optimisation, SUPÉLEC 2^e année</i>				18h	18h	18h	18h	18h
<i>Bases mathématiques et algorithmes pour le traitement du signal, ENSTA 2^e année</i>	20h	20h	24h	24h				
<i>Problèmes inverses, SUPÉLEC 3^e année</i>							6h	
<i>Problèmes inverses et modèles markoviens, DEA ATS, UPS³</i>				6h	6h	6h	3h	3h
Travaux pratiques	91-92	92-93	93-94	94-95	95-96	96-97	97-98	98-99
<i>Problèmes inverses, SUPÉLEC formation continue</i>	3h	3h						
<i>Traitement du signal, DESS et FIUPSO, UPS</i>	42h							
Total équivalent TD	95h	69h	65h	68h	52h	67h	137h	89h

2. DEA Traitement des images et du signal, Université de CERGY-PONTOISE

3. DEA Automatique et traitement du signal, Université de PARIS-SUD, ORSAY

1.10 Autres contributions

- Expert pour les revues internationales : *IEEE Transactions on Signal Processing*, *IEEE Transactions on Image Processing*, *Graphical models and Image Processing*, *Reliable Computing*, *Signal Processing*, et nationales : *Compte-rendus de l'Académie des sciences*, *Traitement du signal*.
- « Expert consulté » par le Groupe « Problèmes inverses » de l'Observatoire français des techniques avancées (1996-98).
- Membre du comité de programme de *SPIE 98 annual meeting, Bayesian Inference for Inverse Problems, San Diego, CA, USA, July 19-24, 1998*
- Conférencier invité à : *3rd International Conference on Inverse Problems in Engineering, Port Ludlow, U.S.A., 1999*.
- Coordinateur du volume « Approche bayésienne pour les problèmes inverses » du traité IC2 des Éditions HERMES (rédaction en cours ; sortie prévue : automne 2000).
- Intervenant lors de l'École d'été 2000 EdF-CEA « Traitement de l'image » dans la session « Problèmes inverses en imagerie » (préparation en cours pour juin 2000).

Chapitre 2

Thèmes d'activité scientifique

2.1 Introduction

Mon activité de recherche s'insère dans le groupe *Problèmes inverses* du Laboratoire des signaux et systèmes. Elle se concentre essentiellement sur des problèmes d'*imagerie*, au sens large du terme : déconvolution, segmentation, tomographie, problèmes le plus souvent multivariés en coordonnées d'espace et de temps. La plupart sont des *problèmes inverses*, au sens où ils sont associés à un *problème direct*, exprimable et résoluble dans le cadre des équations de la physique.

Les problèmes inverses ont connu un intérêt croissant depuis une vingtaine d'années. On les rencontre dans de très nombreuses applications : optique, radar, calorimétrie, spectroscopie, géophysique, acoustique, radioastronomie, contrôle non destructif, génie biomédical, instrumentation et imagerie en général. La résolution de ces problèmes présente la caractéristique désagréable d'être très sensible aux erreurs sur les données, ce sont souvent des *problèmes mal-posés*. Un moyen désormais classique de résoudre cette difficulté est fourni par la *régularisation*. De façon privilégiée dans mon activité, les outils méthodologiques mis en œuvre pour y parvenir sont les modèles probabilistes de type *markovien*, utilisés comme *a priori* en estimation bayésienne, ou encore les fonctions de pénalisation « gibbsiennes » qui en constituent le pendant dans un cadre déterministe.

Certaines applications du traitement de signal (comme les télécommunications) gèrent des flots de données *en ligne*. Au contraire, les problèmes inverses tels qu'on les rencontre en géophysique, en contrôle industriel non destructif, en astronomie, etc, peuvent être abordés *en temps différé*. D'autre part, les conditions d'acquisition des données dépendent plus de contraintes expérimentales que des *desiderata* du traiteur de signaux. Il en résulte que, selon mon expérience pratique de l'inversion, l'objectif est souvent de construire des solutions *raisonnables* et *robustes* dans un contexte difficile : données lacunaires, fortement bruitées, modèle d'observation approximatif, absence de modèle paramétrique satisfaisant pour le signal à reconstruire, etc. La théorie de l'estimation statistique fournit un cadre effectif pour construire ces solutions. Mais il est alors important de remarquer que les études asymptotiques (par exemple à rapport signal-à-bruit ou à nombre de données tendant vers l'infini) ne rendent pas compte du caractère éventuellement mal-posé des problèmes en situation réelle. En traitement de signal, le choix d'un estimateur en fonction de son comportement asymptotique est classique et légitime *lorsque ce comportement correspond au régime établi dans un système alimenté par un flot de données*. Dans le contexte spécifique des problèmes inverses, cette démarche devient risquée : ignorant la nécessité de régulariser, elle est susceptible de produire des solutions inexploitable car instables.

* *
*

Dans mon activité scientifique, j'aborde également des problèmes de traitement du signal pour lesquels la dénomination de « problèmes inverses » ne s'impose pas, avec l'idée selon laquelle la régularisation peut néanmoins fournir un cadre théorique et des solutions pratiques dignes d'intérêt. Considérons par exemple le problème de l'analyse spectrale autorégressive (AR) (figurant dans mon activité de recherche au Sous-thème 3a, page 28), puis celui de la localisation de sources (abordée au Sous-thème 3b, page 29).

Un exemple classique, l'analyse spectrale AR

Le problème de l'analyse spectrale AR désigne le problème suivant :

Problème 1 (analyse spectrale autorégressive) *Estimer le spectre d'un processus AR stationnaire au sens large (d'ordre P fixé), à partir de N données correspondant à des échantillons successifs d'une réalisation de ce processus.*

Il s'agit clairement d'un problème d'*estimation* des paramètres du modèle, qui peut être formulé au sens des moindres carrés [MULLIS et ROBERTS 1976], du maximum de vraisemblance [PHAM 1988], du maximum *a posteriori* [KITAGAWA et GERSCH 1985]. La référence [2], constituant l'Annexe D (page 85) du présent document, est une contribution personnelle récente à ce domaine. On trouvera aussi au Chapitre 4 du document pédagogique [84] une introduction à l'algorithme de Levinson comme méthode d'inversion de systèmes de Toeplitz.

La formulation « classique » [KAY et MARPLE 1981] mêle intimement le problème de l'analyse spectrale AR à une problématique voisine :

Problème 2 (extrapolation de la corrélation d'un AR) *Soit r_0, \dots, r_P les $P+1$ premiers coefficients de corrélation d'un processus AR (stable et causal) d'ordre P ; trouver les $P+1$ paramètres AR qui décrivent de façon équivalente la loi de ce processus.*

Contrairement au problème de l'analyse spectrale, ce second problème n'est pas un problème d'estimation, mais un « simple » problème de changement de paramétrisation (dont on obtient la solution par inversion d'un système linéaire de Toeplitz). Or, la solution du Problème 1 est classiquement « déduite » de celle du Problème 2, pour $P+1$ coefficients de corrélation *empiriques* estimés au préalable à partir des N données. Cette formulation présente plusieurs faiblesses.

- La question du *traitement statistique des données*, centrale dans le Problème 1, devient une question apparemment subalterne : le choix d'estimateurs empiriques pour les coefficients de corrélation ; tandis que la part belle est faite à la théorie de la prédiction linéaire, cadre adapté au Problème 2, mais pas au Problème 1.
- Les estimateurs empiriques choisis convergent vers les « vrais » coefficients de corrélation lorsque le nombre de données tend vers l'infini (si elles sont issues d'un « vrai » processus AR stable). On retrouve ici l'argument « asymptotique ». Mais cet argument ne justifie en aucune façon la supériorité avérée de l'approche AR à *faible nombre de données* [BURG 1967] sur des méthodes plus anciennes, tel que le calcul de périodogrammes moyennés, qui sont aussi convergents.

Remarquons finalement que l'approche AR de l'analyse spectrale repose sur le choix d'un modèle de description des données, et que ce choix est rarement fondé, en pratique, sur des éléments

« objectifs » (c'est-à-dire sur la nature AR d'un hypothétique vrai modèle), mais plutôt :

- sur la nécessité factuelle de régulariser le problème, régularisation obtenue ici par le choix d'un modèle paramétrique *parcimonieux* [GIOVANNELLI 1995] ;
- sur des informations qualitatives *a priori* sur le spectre attendu (par exemple, on attend un spectre plutôt « piqué ») ;
- sur les avantages pratiques du choix d'un modèle AR (en terme de facilité de calcul des estimateurs associés).

Localisation de sources et « asymptomania »¹

Pour certaines applications, telles que la localisation de sources en traitement d'antenne, mon centre d'intérêt se distingue sensiblement de celui de l'école majoritaire, au-delà d'une simple question de formulation. En effet, beaucoup de contributions accordent une importance capitale aux propriétés statistiques asymptotiques de solutions proposées. Parmi les arguments avancés pour justifier « l'asymptomania » en traitement de signal, figure en bonne place le fait que la statistique à petit nombre d'échantillons produit peu d'expressions analytiques, donc peu de théorèmes d'optimalité [STOICA 1998]. Ce point de vue privilégie une approche académique qui offre peu d'outils pour aborder certaines applications. Même en traitement d'antenne, la situation d'un grand nombre de *clichés* pour un même scénario n'est pas si usuelle. De plus, les propriétés de convergence sont obtenues sous des hypothèses statistiques dont la validité est rarement vérifiable. A toutes fins pratiques, l'adaptativité des méthodes de localisation aux évolutions de scénario liés à des cibles mobiles [JAUFFRET et LE CADRE 1999] ne devrait-elle pas constituer un enjeu aussi important que la question de la convergence en environnement stationnaire ?

Tandis que l'étude du comportement asymptotique des estimateurs relève de la « statistique classique », l'importance de la régularisation est plutôt mise en avant par d'autres traditions scientifiques, parmi lesquelles figure la « statistique bayésienne ». Chaque école cultive son propre champ d'investigation (avec les mêmes outils du calcul des probabilités) : faible nombre d'échantillons d'une part, situation asymptotique d'autre part. On peut finalement regretter la rareté des travaux fédérateurs. Ainsi, le comportement asymptotique des estimateurs bayésiens est mal connu ; les résultats sur ce sujet sont peu nombreux [IBRAGIMOV et HAS'MINSKII 1981, LEHMANN 1983] et mal diffusés dans la communauté du traitement de signal. Il en ressort pourtant que les « variantes bayésiennes » d'un estimateur de maximum de vraisemblance convergent héritent de cette propriété asymptotique. On peut même observer la situation contraire dans le cas des mélanges de population : pénaliser la vraisemblance de façon adéquate permet de se débarrasser de problèmes de dégénérescence qui sont bien connus dans le cas des mélanges statistiques [36]. En conclusion, le respect des propriétés asymptotiques n'interdit pas la régularisation, tandis que le comportement à petit nombre d'échantillons devrait y inciter.

Faute de théorèmes pour les petits nombres d'échantillons, le développement de méthodes d'échantillonnage stochastique [TIERNEY 1994, ROBERT 1997] rend de plus en plus souvent possibles des études statistiques empiriques appropriées, peut-être moins élégantes mais plus pragmatiques. Ainsi, j'ai entrepris en 1994-95 avec Frédéric CHAMPAGNAT (alors post-doctorant au L2S) une étude statistique portant sur l'identification de mélanges de population à petit nombre de données. Le cas étudié était représentatif du problème posé par l'estimation des hyperparamètres en déconvolution impulsionnelle, situation proche de l'analyse de raies spectrales. A l'aide d'expressions analytiques complétées par des simulations de type Monte-Carlo, nous avons obtenu quelques résultats significatifs ; en particulier, que le maximum de la *vraisemblance généralisée* (MVG), bien que non convergent, possède un bon comportement à faible nombre de données,

1. L'appellation « asymptomania » est utilisée par Petre STOICA, chercheur reconnu en localisation de sources et lui-même « asymptomane » auto-proclamé [STOICA 1998].

qui est même meilleur que celui du maximum de vraisemblance. Ces travaux ont été présentés lors d'une conférence internationale [26] et synthétisés sous la forme d'un rapport technique [82], mais nous avons vainement tenté de les publier dans *IEEE Transactions on Information Theory* : étudier un estimateur non convergent n'a pas été jugé intéressant.

* *
*

Les applications à l'imagerie, historiquement plus récentes, ne présentent pas les mêmes rigidités. Je me trouve naturellement en phase avec les traiteurs d'image *bas-niveau*, dans le domaine de la vision par ordinateur en particulier. Dans ce domaine, la pénalisation de critères d'adéquation aux données ne suscite pas de controverse, qu'elle soit présentée dans un cadre déterministe fonctionnel, par exemple [BLAKE et ZISSERMAN 1987, COHEN 1996], ou probabiliste bayésien [MARROQUIN et coll. 1987]. Les outils des traiteurs d'image sont aussi les miens : énergies de Gibbs, modèles markoviens, maximum *a posteriori*, maximisation de vraisemblances marginales, méthodes d'optimisation déterministes ou stochastiques. En revanche, les images recherchées en imagerie ne sont pas des cartes d'attributs à extraire d'images « photographiques ». Les modèles *a priori* utilisés sont donc en général différents, et surtout, le modèle direct fait intervenir des transformations plus complexes, linéaires à large support (Sous-thème 2a) ou même non linéaires (Sous-thème 2b).

La crédibilité des méthodes développées en imagerie dépend souvent, *in fine*, de la résolution de problèmes calculatoires. Ces problèmes prennent une part importante dans mon activité : *analyse combinatoire* pour la déconvolution impulsionnelle (Thème 1), programmation dynamique (Sous-thème 2d), *optimisation convexe* (Sous-thème 2c) ou *non convexe* (Sous-thème 2a), *algorithmes stochastiques* pour la résolution de problèmes inverses sous forme *non supervisée* (Thème 1 et Sous-chapitre 2.5 : *Perspectives*).

* *
*

La suite de ce chapitre est un résumé de mon activité scientifique découpée en trois thèmes, dont deux sont subdivisés en sous-thèmes :

- Thème 1 – Déconvolution impulsionnelle (1991-95)
- Thème 2 – Méthodes markoviennes en imagerie (depuis 1992)
 - Sous-thème 2a – Modèles « à énergie non convexe » (1992-95)
 - Sous-thème 2b – Inversion d'un problème non-linéaire : la tomographie d'impédance (1993-97)
 - Sous-thème 2c – Modèles « à énergie convexe » (depuis 1994)
 - Sous-thème 2d – Chaînes de Markov cachées et champs unilatéraux (1991-98)
- Thème 3 – Estimation spectrale en temps court (depuis 1993)
 - Sous-thème 3a – Approche autorégressive (1992-94)
 - Sous-thème 3b – Synthèse de Fourier (depuis 1994)

L'ordre choisi est approximativement chronologique, et les périodes d'activité sont indiquées pour chaque thème et sous-thème. Le découpage adopté conserve une part d'arbitraire, par un effet classique de « vases communicants » entre thèmes de recherche : par exemple, une partie de l'analyse de raies spectrales figurant dans *Synthèse de Fourier* (Sous-thème 3b) partage les outils du Thème 1 : *Déconvolution impulsionnelle* ; une autre partie de ce même Sous-thème *Synthèse de Fourier* utilise des modèles markoviens qui pourraient être rangés dans *Modèles « à énergie convexe »* (Sous-thème 2c).

2.2 Thème 1 – Déconvolution impulsionnelle (1991-95)

Ce thème a été initié au L2S en 1986-87 par Yves GOUSSARD pour la caractérisation des milieux inhomogènes à l'aide de techniques échographiques. Ces techniques intéressent de nombreux domaines d'application comme le génie biomédical, le contrôle non destructif en milieu industriel et la sismique-réflexion en géophysique. Sous certaines hypothèses physiques, le problème direct peut être approché par une convolution mono-dimensionnelle (1D) entre une ondelette (*i.e.*, la forme de l'onde incidente) et une réflectivité caractéristique du milieu à la verticale du capteur. La caractérisation du milieu étudié devient alors un problème de déconvolution dont la résolution passe par la prise en compte de propriétés structurales, mono- ou multi-dimensionnelles.

MENDEL et ses collaborateurs (USC, Los Angeles) ont proposé dans les années 80 de modéliser les échogrammes « parfaits » comme des trains d'impulsions, dans un cadre bayésien, en utilisant un modèle de processus composé Bernoulli-gaussien (BG). Un vecteur BG est un vecteur aléatoire à deux composantes $\mathbf{X} = (\mathbf{Q}, \mathbf{R})$, où \mathbf{Q} est un vecteur binaire i.i.d. de *densité d'événements* $\lambda = P(Q_1 = 1)$ et \mathbf{R} est défini conditionnellement à \mathbf{Q} comme un vecteur gaussien centré indépendant dont la diagonale de la covariance est proportionnelle à \mathbf{Q} : $\mathbf{R} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \text{diag}\{\mathbf{Q}\})$.

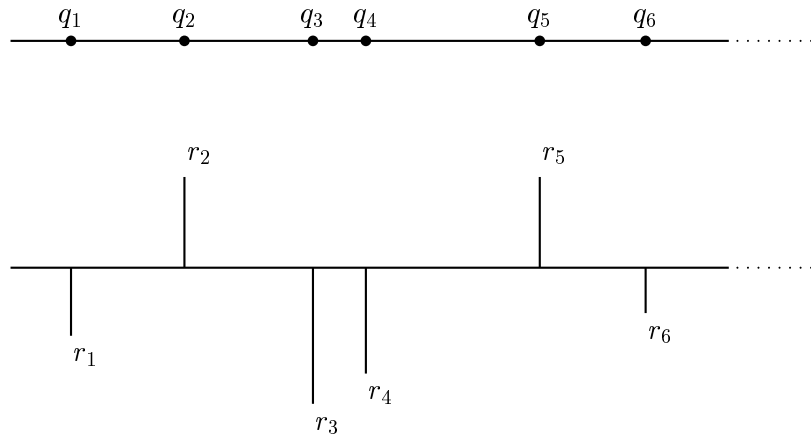


FIG. 2.1: *Processus composé (q_m, r_m) : q_m est un processus ponctuel qui commande l'apparition des événements, dont l'amplitude est distribuée suivant la loi du processus r_m .*

A la même époque, l'équipe de MENDEL a également proposé l'algorithme SMLR (pour *Single most likely replacement*) pour approcher des solutions de type maximum *a posteriori* [MENDEL 1983, GOUTSIAS et MENDEL 1986]. Leur contribution dans ce domaine a inspiré les premiers travaux du L2S en déconvolution impulsionnelle en 1986-90. Ils ont débouché en particulier sur une variante du SMLR, beaucoup plus simple que l'original, substituant un modèle de réponse impulsionnelle finie au modèle ARMA adopté par MENDEL [GOUSSARD et coll. 1990].

Ma thèse (1988-91) était consacrée à la prise en compte de la stratification sous la forme de modèles 2D markoviens unilatéraux [IDIER 1991]. L'objectif était d'aboutir à un traitement conjoint pour un ensemble d'échogrammes juxtaposés, plus efficace que le traitement séparé de chacun des échogrammes par SMLR. Les articles [7] et [8] sont les deux principales publications extraites de ce travail, que je n'aborderai pas dans le présent document. Il me paraît cependant utile de mentionner que mon travail de thèse a été une des premières contributions du L2S dans le domaine de l'utilisation des champs de Markov et des algorithmes stochastiques en imagerie, thèmes que je me suis efforcé de développer par la suite.

Yves GOUSSARD ayant quitté le L2S (en 1990), nous avons poursuivi nos travaux communs sur le thème de la déconvolution BG. A l'occasion de la thèse de Frédéric CHAMPAGNAT (1991-93), nos efforts ont porté sur le développement de variantes plus robustes [CHAMPAGNAT 1993] : prise en compte d'un modèle de bruit coloré et de rotations de la phase de l'onde réfléchie [31], de saturations dans les échogrammes mesurés [47]. En 1993-95, nous avons poursuivi ces travaux en développant des stratégies d'estimation des paramètres des modèles pour le signal BG et pour le bruit. Nous avons ainsi obtenu des méthodes de déconvolution BG *non supervisées* utilisant des algorithmes stochastiques de type SEM (pour *Stochastic Expectation-Maximisation*) [6, 81]. Ces travaux sont à rapprocher de contributions voisines dans la communauté française en traitement statistique du signal (par exemple, [LAVIELLE 1995]).

Arrivé à maturité, ce thème a engendré finalement d'autres utilisations fructueuses des modèles impulsionnels et des techniques de placement d'impulsions, hors du contexte initial. Le cas de la détection de raies spectrales, correspondant à la thèse de Frédéric DUBLANCHET, est l'exemple le plus caractéristique (voir Sous-thème 3b, page 29). D'une façon plus indirecte, des questions ou des perspectives initialement formulées dans le cadre de la déconvolution d'impulsions ou d'objets stratifiés ont finalement débouché sur des problématiques plus générales :

- Dans mon manuscrit de thèse, les méthodes de continuation de type GNC sont évoquées comme une alternative possible (mais plus coûteuse) aux algorithmes de type SMLR pour la déconvolution d'objets stratifiés [IDIER 1991, p.132]. Par la suite, j'ai souhaité concrétiser cette idée, ce qui a été fait de façon spécifique au cours du stage de DEA de Thierry MARTIN [MARTIN 1993], mais aussi au cours de la thèse de Mila NIKOLOVA [NIKOLOVA 1995], dans le cadre plus générique de la *minimisation de critères markoviens non convexes en imagerie* (Sous-thème 2a, §2.3, page 22). De fait, le cas du critère Bernoulli-gaussien

$$\mathcal{J}(\mathbf{r}, \mathbf{q}) = \sigma_b^{-2} \|\mathbf{z} - \mathbf{H}\mathbf{r}\|^2 + \sigma^{-2} \mathbf{r}^t \text{diag}\{\mathbf{q}\} \mathbf{r} - \alpha \mathbf{1}^t \mathbf{q}$$

référéncé dans [IDIER 1991, p.30] partage les caractéristiques de certains critères minimisés en restauration d'image (voir Sous-chapitre 4.2, *Détection-estimation*, page 45) : il s'agit d'un critère *semi-quadratique*, à variables cachées *binaires*. De plus la minimisation est explicite vis-à-vis des variables cachées :

$$\begin{aligned} \min_{\mathbf{q}} \mathcal{J}(\mathbf{r}, \mathbf{q}) &= \sigma_b^{-2} \|\mathbf{z} - \mathbf{H}\mathbf{r}\|^2 + \sum_m \min_{q_m} q_m (\sigma^{-2} r_m^2 - \alpha) \\ &= \sigma_b^{-2} \|\mathbf{z} - \mathbf{H}\mathbf{r}\|^2 + \sum_m \min \{0, \sigma^{-2} r_m^2 - \alpha\}, \end{aligned}$$

et le résultat fait intervenir la fonction quadratique tronquée de [BLAKE et ZISSERMAN 1987], pour laquelle le GNC a été d'abord introduit.

- Un élément central dans mon travail de thèse était la construction d'un champ de Markov binaire stationnaire sur un graphe spécifique. Ce travail était en partie fondé sur ceux de PICKARD, sur un réseau cartésien [PICKARD 1977, PICKARD 1980]. A cette occasion, s'est posée la question de la caractérisation des champs markoviens stationnaires sur un réseau cartésien fini, les champs de Pickard étant les seuls alors connus. Les développements ultérieurs forment le Sous-thème 2d (page 25).
- En déconvolution impulsionnelle, nous avons abordé le problème du réglage non supervisé des hyperparamètres par des techniques de maximisation de vraisemblance de type MVG, EM et SEM [6, 81]. Ultérieurement, j'ai mis en œuvre des approches similaires ou apparentées dans des contextes différents : pour l'analyse de raies spectrales [17, 66], la segmentation de signaux [18, 44, 75] et d'images [15, 46], la classification [36], l'estimation d'aberrations optiques [33].

Le réglage non supervisé des hyperparamètres est une question très importante pour la restauration de signaux et d'images, et qui demeure encore largement une question ouverte. Le Chapitre 5 (page 55) lui est largement consacré.

Globalement, on peut avancer que les travaux menés par Yves GOUSSARD, Frédéric CHAMPAGNAT et moi-même entre 1990 et 95 constituent une référence dans le domaine de la déconvolution impulsionnelle, et sont à ce titre régulièrement cités dans les contributions ultérieures, nationales [SALLARD et coll. 1997, DOUCET et DUVAUT 1997, BOUBAL 1998] et internationales [CHENG et coll. 1996, KAARESEN 1998, SANTAMARIA et coll. 1999].

2.3 Thème 2 – Méthodes markoviennes en imagerie (depuis 1992)

Une partie importante de mon activité est consacrée à l'étude et à la construction de modèles régularisants pour les problèmes d'imagerie tels que la déconvolution et la reconstruction tomographique, sans oublier les aspects algorithmiques liés aux problèmes d'inférence associés (maximisation de vraisemblance *a posteriori*, calcul d'espérance conditionnelle). Ces travaux concernent surtout les modèles markoviens, car ils permettent facilement d'incorporer des informations structurelles liées aux propriétés locales des objets. Comme en traitement d'image, la possibilité de décrire des zones homogènes séparées par des discontinuités est souvent cruciale en imagerie. Le caractère central de ce thème dans mon activité scientifique justifie qu'un chapitre spécifique lui soit consacré ; il s'agit du Chapitre 4, page 43.

Sous-thème 2a – Modèles « à énergie non convexe » (1992-95)

En imagerie, la plupart des techniques algorithmiques adaptées aux modèles markoviens à étiquettes ou « à énergie non convexe » (recuit simulé, ICM) se révèlent inexploitable dès lors que les équations d'observation — le modèle direct — ne sont pas restreintes à des opérateurs locaux ou bien conditionnés, ce qui est un cas fréquent (en tomographie, en déconvolution). La thèse de Mila NIKOLOVA [NIKOLOVA 1995], co-encadrée par Ali MOHAMMAD-DJAFARI et moi-même, a été l'occasion de développer une approche déterministe, concurrente du recuit simulé et plus réaliste en terme de coût de calcul. Pour cela, nous avons adapté à la résolution générale de problèmes inverses linéaires mal-posés un procédé d'optimisation par non-convexité graduelle (GNC, [BLAKE et ZISSERMAN 1987]). Il s'agit d'une méthode sous-optimale de relaxation déterministe initialement limitée à la segmentation d'image. Outre la suggestion de l'approche GNC en imagerie, ma contribution a concerné la résolution de difficultés spécifiques posées par le caractère mal posé et par le support large de l'opérateur d'observation [4, 65]. La référence [4], publiée dans les *IEEE Transactions on Image Processing*, est fournie en Annexe A, page 79. Par la suite, cette approche a été appliquée avec succès en tomographie à courants de Foucault [NIKOLOVA et MOHAMMAD-DJAFARI 1996] et en imagerie à ondes diffractées [CARFANTAN et MOHAMMAD-DJAFARI 1995].

Sous-thème 2b – Inversion d'un problème non-linéaire : la tomographie d'impédance (1993-97)

Compte tenu des avancées dans le domaine de l'inversion de problèmes linéaires (déconvolution, reconstruction à partir de projections, synthèse de Fourier), et en particulier de l'utilisation d'estimateurs non-linéaires dans ce domaine, il m'a semblé souhaitable d'aborder également le cas de modalités d'imagerie correspondant à une équation d'observation non-linéaire, dans le cas où la linéarisation est une approximation trop grossière.

A la suite d'une étude effectuée pour la Société SCHLUMBERGER en 1992 [32], la tomographie d'impédance en mode quasi-statique m'a paru constituer un sujet idéal pour entamer cette étude. En effet, la résolution du problème direct associé est bien maîtrisée et numériquement peu coûteuse à l'aide de méthodes d'approximation aux éléments finis. D'autre part, les applications possibles sont variées : imagerie médicale, exploration du sous-sol en géophysique, contrôle industriel non-destructif. Enfin, la bibliographie sur le sujet nous a montré que la plupart des méthodes d'inversion utilisées en tomographie d'impédance s'appuient sur des outils régularisants très frustes [MURAI et KAGAWA 1985], voire méconnaissent le caractère mal-posé du problème [TASTO et SCHOMBERG 1978, WEXLER et coll. 1985].

En tomographie d'impédance en mode quasi-statique, les dispositifs expérimentaux sont des

systèmes d'injection de courants (continus ou très basse fréquence) et de mesure de tensions à la surface d'un matériau de résistivité interne variable qu'il s'agit d'imager. L'objectif a été d'apporter une contribution d'ordre méthodologique permettant d'augmenter la fiabilité et la résolution des images reconstruites à partir de couples tension-courant. Ce thème a constitué le sujet de la thèse de Thierry MARTIN, que j'ai encadrée. En particulier, nous avons validé une approche par modélisation markovienne décrivant les interactions entre les valeurs de résistivité affectées à des éléments voisins dans le maillage. Il est intéressant de remarquer que les algorithmes de minimisation du critère d'adéquation aux données ainsi pénalisés bénéficient du caractère creux du système à inverser pour résoudre le problème direct [3]. Ces travaux ont contribué à valoir au L2S une dotation financière (*Stitching Fund Award*) de la Fondation SCHLUMBERGER au titre du mécénat, en 1996. Bien qu'ils n'aient pas de suite immédiate, je considère qu'il s'agit d'un sujet d'avenir, d'autant plus que d'autres domaines de l'ingénierie comme la thermique ou la mécanique rencontrent des problèmes inverses de structure voisine.

La poursuite de ces travaux supposerait l'évolution vers la résolution de problèmes 3D, ce qui ne pose pas *a priori* de problème de principe. Une perspective plus riche consisterait à envisager la question du « remaillage » du réseau en cours d'inversion. En effet, cette formulation permettrait une restauration à la fois multi-résolution et à degré de résolution adaptative spatialement. L'économie dans la description numérique de l'objet restauré que permettrait le remaillage me semble particulièrement attrayante dans le cas d'objets 3D. Cette perspective peut également concerner l'inversion des problèmes linéaires.

Sous-thème 2c – Modèles « à énergie convexe » (depuis 1994)

Les méthodes de minimisation de critères non convexes restant numériquement coûteuses, j'ai effectué un investissement important dans le domaine de la spécification et la minimisation d'énergies de Gibbs convexes pour l'imagerie à partir de 1994. Cet investissement a débuté pendant la thèse de Stéphane BRETTE, co-encadrée par Ali MOHAMMAD-DJAFARI et moi-même, et s'est poursuivi ensuite sous plusieurs formes.

Au cours des travaux de thèse de Stéphane BRETTE, nous avons adopté des approches numériques *semi-quadratiques* (SQ) introduites par [GEMAN et REYNOLDS 1992] et [GEMAN et YANG 1995], en les spécialisant au cas d'un critère convexe correspondant à un problème de déconvolution 2D à noyau séparable [22].

Entre 1996 et 1998, l'algorithme proposé dans [22] a été adapté au problème de la déconvolution d'images SPOT dans le cadre de contrats avec le CNES, par Frédéric CHAMPAGNAT puis Thierry MARTIN, successivement post-doctorants au L2S [87, 89]. J'ai supervisé ces travaux dans un environnement concurrentiel peu fréquent [53] : en 1996, par l'intermédiaire du GdR-PRC ISIS, le CNES a mis en « compétition » six laboratoires de compétence reconnue en déconvolution d'image, sans contrainte méthodologique, dans le but de sélectionner une méthode pour le traitement au sol des images fournies par les satellites SPOT de la prochaine génération (SPOT 5). L'évaluation a été effectuée « en double aveugle », le CNES effectuant lui-même les traitements à partir des éléments fournis par les concurrents et les faisant évaluer par quatre groupes d'experts photo-interprètes. La méthode markovienne proposée par le L2S l'a emporté sur les méthodes des cinq autres laboratoires extérieurs... mais pas sur la propre méthode du CNES.

Depuis 1997, j'ai également abordé pour le compte du CNES d'autres problèmes de traitement d'images SPOT : identification de réponse impulsionnelle [88], suppression de l'effet de colonnage [85, 88]. Ces travaux se poursuivent actuellement, en collaboration avec Hervé CARFANTAN, Maître de conférence à l'Université Paul-Sabatier (TOULOUSE). En particulier, une validation effectuée récemment au CNES sur données simulées et sur données réelles a montré que la

méthode de suppression de l'effet de colonnage atteint des résultats bien supérieurs à ceux d'une méthode concurrente du CNES.

A l'occasion d'un co-encadrement de stage de fin d'études, j'ai étudié avec Jean-Luc STARCK l'application de l'algorithme SQ proposé dans [22] à la déconvolution d'images astronomiques (voir [79]).

Enfin, en collaboration avec Yves GOUSSARD, dans le cadre d'un projet franco-qubécois soutenu par le Ministère français des affaires étrangères et le Ministère québécois des relations internationales (1994-96, puis 1997-99), nous avons généralisé ce même algorithme à la déconvolution 3D dans le but d'améliorer la qualité des images tomographiques produites par les scanners de la génération actuelle [20, 37]. Le projet lui-même, intitulé « Conception automatisée et fabrication par prototypage rapide de prothèses personnalisées », associe des chercheurs universitaires de compétences complémentaires : biomécanique, chirurgie orthopédique, prototypage rapide, imagerie. Il est pour moi l'occasion de valoriser des méthodes d'inversion dans un contexte pluri-disciplinaire et applicatif. L'état actuel du projet a dépassé le strict cadre universitaire : nous rédigeons actuellement des demandes de dépôt pour trois brevets, dont un relève de l'imagerie, et nous poursuivons des négociations avec un industriel pour aborder une phase plus coûteuse de validation, technologique puis médicale.

* *
*

Après ces expériences probantes d'utilisation de schémas de minimisation SQ, j'ai plus récemment consacré des travaux à la caractérisation des critères SQ dans le cas de fonctions de pénalisation qui s'écrivent sous la forme

$$\varphi(\mathbf{x}) = \sum_{k=1}^K \phi(\mathbf{v}_k^t \mathbf{x} - w_k), \quad (2.1)$$

où \mathbf{x} regroupe les variables inconnues (image ou signal à restaurer, ...), ϕ est une fonction scalaire convexe, et les \mathbf{v}_k et les w_k sont respectivement des vecteurs et des scalaires pré-spécifiées. Ces travaux s'appliquent en particulier aux modèles markoviens utilisés couramment en imagerie. Ils permettent d'étendre des résultats de convergence [CHARBONNIER et coll. 1997] à des structures d'algorithmes SQ beaucoup plus générales.

D'autre part, j'ai établi que les algorithmes SQ s'identifient à la famille des algorithmes de *moindres carrés repondérés* (*reweighted least squares*), préexistants dans le domaine de la *statistique robuste* [REY 1983].

Enfin, en étudiant la convexité des critères SQ de type « Geman-Reynolds » [GEMAN et REYNOLDS 1992] issus de l'expression (2.1), j'ai montré qu'il était possible de construire des énergies convexes SQ à variables de ligne *interactives*. Ce résultat offre des perspectives intéressantes, car il permet d'envisager la construction de critères *convexes* s'inspirant de critères préexistants à lignes binaires interactives difficiles et coûteux à minimiser. On peut espérer ainsi préserver au moins une partie des bonnes propriétés de ces critères (en termes de modèle), en se débarrassant des mauvaises (les problèmes calculatoires). L'exploration systématique de ce principe est une vaste entreprise et fera l'objet de travaux ultérieurs.

Ces travaux récents sont contenus dans la référence [10], soumise pour publication aux *IEEE Transactions on Image Processing*. Elle est fournie en Annexe B, page 81.

Sous-thème 2d – Chaînes de Markov cachées et champs unilatéraux (1991-98)

Aspects théoriques

Depuis la contribution fondatrice de [PICKARD 1977, PICKARD 1980], les champs unilatéraux sont restés très méconnus et sous-exploités en regard de leurs propriétés théoriques. Par définition, un champ aléatoire \mathbf{X} à S composantes (on peut avoir $S = \infty$) et à valeur dans un ensemble fini E est unilatéral si il existe une numérotation de ses composantes $\mathbf{X} = \{X_s, s = 1, \dots, S\}$ pour laquelle les conditionnelles « présent sachant passé » se simplifient sous la forme :

$$\forall s \in \{2, \dots, S\}, x_1, \dots, x_s \in E, \\ P(X_s = x_s | X_1 = x_1, \dots, X_{s-1} = x_{s-1}) = P(X_s = x_s | X_r = x_r, r \in \mathcal{P}_s),$$

où \mathcal{P}_s est un sous-ensemble de $\{1, \dots, s-1\}$ qui réunit les *prédécesseurs* du site s [IDIER 1991, p.90], [BRÉMAUD 1999, p.266]. Autrement dit, la loi d'un champ unilatéral s'écrit, d'après la règle de Bayes séquentielle :

$$P(X_1 = x_1, \dots, X_S = x_S) = P(X_1 = x_1) \prod_{s=2}^S P(X_s = x_s | X_r = x_r, r \in \mathcal{P}_s). \quad (2.2)$$

Il est facile de montrer qu'un champ unilatéral est un champ de Markov. En particulier, le cas $\mathcal{P}_s = s-1$ définit une chaîne de Markov. Dans le cas d'une répartition des sites sur une grille cartésienne 2D et d'un balayage lexicographique, on obtient plus généralement les « modèles markoviens en réseau » (*Markov mesh models*), parmi lesquels figurent les champs de Pickard.

Les champs de Pickard sont stationnaires sur les réseaux rectangulaires, c'est-à-dire que les lois marginales de sous-ensembles de pixels sont invariantes par décalage à l'intérieur du domaine. Ainsi, choisir un champ de Pickard comme modèle *a priori* en reconstruction d'image assure que tous les pixels ont la même loi (marginale) *a priori*, alors que ce résultat est présumé faux pour la plupart des champs de Markov 2D [GEMAN et GEMAN 1984].

A la suite de mes travaux de thèse, qui constituent à ma connaissance le seul exemple de caractérisation et d'utilisation d'un modèle unilatéral possédant des propriétés similaires à celles du *Curious Lattice Process* de Pickard [PICKARD 1977], nous avons maintenu avec Frédéric CHAMPAGNAT une activité théorique dont l'objectif était de caractériser les modèles markoviens possédant la propriété de stationnarité sur un réseau fini cartésien rectangulaire. Pour les graphes de voisinage d'ordre un, nous sommes finalement parvenus à caractériser ces modèles dans les principaux cas : il s'avère que la propriété de stationnarité sur un réseau fini est caractéristique des champs de Pickard dans le cas gaussien et dans le cas symétrique, mais pas dans le cas binaire. Dans ce dernier cas, nous avons mis en évidence et caractérisé le seul autre modèle markovien d'ordre un stationnaire sur un réseau fini. Ce modèle, bien que non unilatéral, est simulable en un nombre fini d'opérations. Ces travaux sont publiés dans *IEEE Transactions on Information Theory* [5], article fourni en Annexe C, page 83.

Il nous a ensuite semblé pénible et peu gratifiant d'étendre la caractérisation à des graphes de voisinage plus grands... sauf dans le cas gaussien. De façon étonnante, nous avons en effet pu montrer que les champs de Pickard gaussiens de voisinage quelconque s'identifient avec les processus AR gaussiens « quart-de-plan ». Ce résultat a plusieurs conséquences formelles originales. D'une part, il permet de montrer que ces modèles maximisent l'entropie sous des conditions de connaissance de corrélations (« problème de Burg 2D »). D'autre part, il fournit les « équations manquantes » aux équations de Yule-Walker 2D pour exprimer l'équivalence des deux représentations : coefficients de prédiction d'une part, coefficients de corrélation d'autre part. Ce nouvel ensemble de résultats est rassemblé dans un article à paraître dans *Advanced Applied Probability* [1].

Modèles unilatéraux à étiquettes

Du côté des applications au traitement d'image, [DEVIJVER et DEKESSEL 1988] fait partie des rares exemples de contribution exploitant les propriétés des champs de Pickard, ou plus généralement des champs unilatéraux (on peut citer également [DERIN et coll. 1984, LACROIX 1987, QIAN et TITTERINGTON 1989]). Les restrictions 1D des champs stationnaires sont des chaînes de Markov [5], ce qui fournit des algorithmes de programmation dynamique et des algorithmes *aller-retour* pour calculer différents types d'estimateurs dans des problèmes de débruitage ou de segmentation, ainsi que des algorithmes de réestimation pour estimer les paramètres des modèles par maximum de vraisemblance [QIAN et TITTERINGTON 1990]. Malheureusement, pour un modèle *a priori* markovien unilatéral, la loi *a posteriori* reste markovienne mais elle ne reste pas forcément unilatérale, sauf dans le cas d'une image simplement bruitée [FAYOLLE 1998]. C'est pourquoi les applications de ces modèles sont en grande partie restreintes aux problèmes de classification et de segmentation.

Avec Yves GOUSSARD, nous avons repris les travaux de DEVIJVER sur la segmentation non supervisée de signaux et d'images par chaînes de Markov et champs de Pickard cachés [DEVIJVER et DEKESSEL 1988]. Nous avons proposé une diminution du nombre de paramètres du modèle, passant de l'ordre de N^2 à N pour une segmentation sur N niveaux, corrigeant ainsi certains défauts de la méthode proposée. Cette réduction a été obtenue grâce à l'utilisation d'un modèle de chaîne de Markov télégraphique cachée pour les lignes ou les colonnes de l'image et à une exploitation plus poussée de l'hypothèse de stationnarité [15, 44, 46]. Nous avons également proposé de maximiser une version pénalisée de la vraisemblance [36], ce qui permet d'éviter les dégénérescences classiques pour ce type de problème d'estimation des paramètres de mélange statistique [NÁDAS 1983]. Au total, la méthode de segmentation proposée est non supervisée (sauf le nombre de classe) et d'un coût remarquablement faible pour une méthode markovienne non supervisée [71].

Modèles unilatéraux à variables de contour

Dans le travail de thèse de Marc FAYOLLE [FAYOLLE 1998], un modèle voisin a été étudié avec le même objectif, la segmentation non supervisée de signaux et d'images. Ce modèle est fondé sur des variables cachées modélisant cette fois des ruptures (en 1D) ou des contours (en 2D) plutôt que des étiquettes.

En 1D, ce type de modèle est baptisé *chaîne faible* dans le contexte de la vision par ordinateur ; il est très proche d'un modèle AR d'ordre un avec des ruptures. Des travaux préexistants avaient abordé la segmentation 1D supervisée par programmation dynamique [BLAKE 1989, DE CESARE 1996]. Nous avons complété ces travaux en développant un algorithme EM permettant un réglage automatique des paramètres [18, 75].

En 2D, le problème de l'estimation des hyperparamètres est beaucoup plus délicat. Cependant, les modèles unilatéraux possédant une fonction de partition égale à un, une approche stochastique de type Monte-Carlo par chaîne de Markov [ROBERT 1997] permet d'échantillonner la nappe de vraisemblance des paramètres d'un modèle markovien beaucoup plus simplement si ce dernier est unilatéral. De plus, cette simplicité structurelle n'est pas réservée au cas de la segmentation : dans le cas d'un modèle gaussien ou gaussien par morceaux, elle reste encore valable pour un modèle d'observation linéaire à support arbitraire. Cette idée a été concrétisée à la fin de la thèse de Marc FAYOLLE, dans le cas d'un modèle unilatéral gaussien par morceaux intégrant des variables de contours binaires interactives [78]. Indépendamment, [DUNMUR et TITTERINGTON 1997] développe une idée très proche. Nous partageons les mêmes conclusions positives dans le cas du débruitage. Dans le cas de la déconvolution d'images

« photographiques », cas non traité dans [DUNMUR et TITTERINGTON 1997], cette approche reste numériquement très efficace, mais elle est susceptible de fournir de « mauvais » réglages de paramètres. C'est la validité du modèle comme *a priori* pour de telles images qui semble devoir être mise en cause, car il est connu que l'unilatéralité induit des corrélations préférentielles indésirables. Mais le caractère markovien est lui-même discutable [GRAY et coll. 1994] : les champs de Markov sont-ils de « bons » modèles en imagerie ? Une partie importante du Chapitre 5 est consacré au problème de l'estimation d'hyperparamètres et des difficultés méthodologiques.

2.4 Thème 3 – Estimation spectrale en temps court (depuis 1993)

Dans ce thème s'inscrivent plusieurs thèses co-encadrées ; chronologiquement :

- celle de Jean-François GIOVANNELLI [GIOVANNELLI 1995] (1992-95) : « Estimation spectrale en temps court. Application à l'imagerie Doppler »,
- celle de Frédéric DUBLANCHET [DUBLANCHET 1996] (1994-96) : « Contribution de la méthodologie bayésienne à l'analyse spectrale de raies pures et à la goniométrie haute résolution »,
- celle de Philippe CIUCIU (commencée en 1996) : « Méthodes markovienne pour l'estimation spectrale et application en imagerie Doppler »,
- celle de Andrea RIDOLFI (commencée en 1998) : « Méthodes markoviennes pour la vélocimétrie. Applications à l'imagerie médicale ».

Du point de vue applicatif, ce thème a pour origine les problèmes de vélocimétrie Doppler ultrasonore biomédicale, étudiée de longue date au L2S en collaboration avec Alain HERMENT, Directeur de recherche INSERM dans l'équipe « Imagerie médicale quantitative » (Pitié-Salpêtrière). L'objectif de la vélocimétrie Doppler biomédicale est typiquement d'évaluer la vitesse du flux sanguin dans une artère, avec de multiples variantes possibles : estimation d'un paramètre scalaire représentant la vitesse moyenne dans une section d'artère, variant au cours du temps, estimation de spectres complets juxtaposés spatialement,... [29], [GIOVANNELLI 1995]. Le domaine de la vélocimétrie par résonance magnétique, plus récent, fait également partie de nos préoccupations [23, 24]. Pour des raisons expérimentales, un très faible nombre de données sont disponibles par spectre à reconstruire, de l'ordre de dix. Autant dire qu'une vision « asymptotomaniac » [STOICA 1998] de ces problèmes laisserait peu d'espoir. Heureusement, la continuité spatiale, temporelle et même spectrale de la carte de vitesse peuvent être prises en compte comme éléments régularisants.

Entre 1994 et 1996, la cartographie aérienne par radar Doppler a également servi de support au développement de nos estimateurs spectraux régularisés. Avec Jean-François GIOVANNELLI, en fin de thèse puis en post-doctorat, et avec Philippe CIUCIU en début de thèse, nous avons travaillé sur ce sujet pour THOMSON-AIRSYS [80, 90, 91, 92].

Sous-thème 3a – Approche autorégressive (1992-94)

Jusqu'en 1994, nos méthodes se sont surtout appuyées sur les travaux fondateurs de KITAGAWA et GERSCH, consacrés aux modèles AR « longs » (c'est-à-dire d'ordres élevés : de l'ordre du nombre de données) [KITAGAWA et GERSCH 1985] et à l'analyse spectrale AR adaptative [KITAGAWA et GERSCH 1985].

D'une part, [KITAGAWA et GERSCH 1985] aborde la prise en compte de la régularité d'un spectre en fonction de la fréquence. Les paramètres d'un modèle AR d'ordre élevé sont estimés, mais le résultat est stabilisé par une pénalisation introduisant l'hypothèse de « douceur » du spectre de façon habile. Le critère minimisé reste un critère quadratique :

$$J(\mathbf{a}) = \|\mathbf{Z}\mathbf{a}\|^2 + \lambda \sum_{p=1}^P p^{2k} a_p^2, \quad (2.3)$$

où $\mathbf{a} = [1, -\mathbf{a}^t]^t$; \mathbf{Z} est une matrice de Toeplitz contenant les données, dont la structure précise dépend de l'hypothèse de fenêtrage adoptée [KAY et MARPLE 1981] ; λ est un paramètre de régularisation (positif) et k détermine un *ordre de la douceur*, conformément à l'identité de

Parseval

$$\int_0^1 \left| \frac{d^k A(\nu)}{d\nu^k} \right|^2 d\nu = (2\pi)^{2k} \sum_{p=1}^P p^{2k} a_p^2, \quad (2.4)$$

où $A(\nu) = 1 - \sum_{p=1}^P a_p e^{-2i\pi\nu}$ est l'expression dont le module carré forme le dénominateur du spectre AR.

D'autre part, [KITAGAWA et GERSCH 1985] aborde la prise en compte de la régularité d'une nappe temps-fréquence en fonction du temps. Dans le cadre du lissage de Kalman, cette hypothèse est introduite sous la forme d'une équation d'état portant sur la suite des vecteurs AR :

$$\mathbf{a}_{t+1} = \mathbf{a}_t + \mathbf{e}_t, \quad (2.5)$$

où \mathbf{e} est un vecteur gaussien indépendant. Bien que cet argument ne soit pas donné dans [KITAGAWA et GERSCH 1985], on peut à nouveau justifier ce choix par une identité de Parseval :

$$\int_0^1 \left| \frac{d^k (A_{t+1}(\nu) - A_t(\nu))}{d\nu^k} \right|^2 d\nu = (2\pi)^{2k} \sum_{p=1}^P p^{2k} (a_{t+1,p} - a_{t,p})^2. \quad (2.6)$$

Notre contribution a consisté à introduire simultanément les deux types de régularité pour estimer des nappes spectrales « espace-fréquence » en vélocimétrie radar Doppler, en conservant une mise en œuvre par lissage de Kalman [92, 80]. Ce travail s'appuie sur le fait qu'un lisseur de Kalman minimise implicitement un critère des moindres carrés pénalisés [JAZWINSKI 1970] : de même que l'identité (2.6) justifie l'équation d'état (2.5), l'identité (2.4) définit la loi stationnaire d'un vecteur d'état \mathbf{a} . Dans [80], l'équation d'état proposée correspond à la fois à une identité de type (2.6) et à une loi stationnaire associée à l'identité (2.4).

Une question importante portant sur les multiples méthodes d'estimation AR consiste à savoir lesquelles assurent la stabilité du filtre tout pôle défini par le régresseur. La pratique des estimateurs régularisés introduits par KITAGAWA et GERSCH nous laissaient penser qu'ils définissent des filtres structurellement stables. Par la suite, une étude plus générale nous a permis de montrer ce résultat, en mettant en évidence le rôle joué par la matrice de déplacement de la matrice normale [40]. Ce travail a été publié récemment dans *IEEE Transactions on Signal Processing* sous forme d'un article court [2], fourni en Annexe D, page 85. Il semble avoir un certain impact dans le domaine de la modélisation des systèmes [LOPEZ-VALCARCE et DASGUPTA 1999].

Sous-thème 3b – Synthèse de Fourier (depuis 1994)

Dans le domaine des radars Doppler utilisés en surveillance aérienne, les spectres à estimer comportent non seulement des composantes étalées, dues par exemple à des phénomènes météorologiques, pour lesquelles les modèles AR longs sont satisfaisants, mais aussi des composantes quasi ponctuelles, dont la fréquence correspond à la vitesse instantanée d'un avion présent dans la zone surveillée. La restitution de ces composantes piquées est évidemment cruciale. Or la modélisation AR n'est pas d'une grande souplesse pour décrire des spectres composites. C'est pourquoi nous avons commencé à explorer une autre approche à partir de 1995. Cette approche est de type *non paramétrique*. Notons $\mathbf{z}_K = [z_1, \dots, z_K]^t$ les observations, $X(\nu)$ le spectre complexe cherché (en fait le but de l'estimation est plutôt $|X(\nu)|^2$),

$$x_k = \int_0^1 X(\nu) e^{2i\pi\nu k} d\nu, \text{ et } \mathbf{x}_K = [x_1, \dots, x_K]^t.$$

On formule alors le problème de l'analyse spectrale de la façon suivante :

$$\text{Trouver } X(\nu), \nu \in [0,1], \text{ tel que } \mathbf{z}_K = \mathbf{x}_K + \text{bruit}. \quad (2.7)$$

Cette formulation diffère de la formulation statistique habituelle, qui consiste à trouver la densité spectrale de puissance d'un processus aléatoire à partir de \mathbf{z}_K considéré comme un fragment de trajectoire. Elle se rapproche plutôt des problèmes de synthèse de Fourier en astronomie [GULL et DANIELL 1978]. Il existe évidemment une infinité de solutions, si bien que ce problème est *mal posé* au sens de Hadamard. On se trouve alors dans un contexte très proche de celui du traitement d'image bas-niveau, et il est assez naturel de s'intéresser à la construction puis à la minimisation de critères pénalisés de type :

$$\mathcal{J}(X) = \|\mathbf{z}_K - \mathbf{x}_K\|^2 + \lambda\varphi(X). \quad (2.8)$$

Plusieurs cas ont été étudiés :

- Si on choisit pour ϕ une norme ou une pseudo-norme *quadratique*, on peut montrer que le minimiseur de (2.8) a pour module carré un périodogramme, simple dans le cas $\phi(X) = \int_0^1 |X(\nu)|^2 d\nu$, fenêtré sinon [41, 9]. On retrouve donc l'analyse spectrale *non paramétrique* standard comme un cas particulier.
- Des estimateurs plus résolus peuvent être obtenus en choisissant pour φ une fonctionnelle non quadratique. On obtient un cas particulier important avec $\phi(X) = \int_0^1 \phi(|X(\nu)|) d\nu$.
 - Si ϕ est une fonction convexe croissante sur \mathbb{R}_+ , alors ϕ et \mathcal{J} sont des fonctionnelles convexes, ce qui permet d'aborder l'analyse de raies spectrales sous forme robuste, non paramétrique mais résolutive [41] (voir aussi [SACCHI et coll. 1998]).
 - Pour détecter et localiser des raies spectrales, on est plutôt amené à considérer pour ϕ une fonction à seuil, non convexe, telle que la fonction quadratique tronquée. Sous une forme mathématiquement équivalente ou proche, ce problème a été formulé sous l'angle de la détection-estimation dans la thèse de Frédéric DUBLANCHET [DUBLANCHET 1996], en collaboration avec l'équipe ETIS de l'ENSEA. Notre objectif était alors d'adapter les algorithmes de type SMLR (développés précédemment pour la déconvolution de processus Bernoulli-gaussiens – voir le Thème 1, page 19). La principale évolution a été de rechercher les « événements » (ici, des raies spectrales) dans un continuum au lieu d'un ensemble discret de positions préspecifiées, c'est-à-dire de remplacer le modèle Bernoulli-gaussien par un modèle Poisson-gaussien. Cet objectif a été atteint en adjoignant au SMLR un algorithme exploitant le gradient du critère en fonction de la position des événements détectés [17]. Par rapport à des approches comparables en traitement d'antenne (maximum de vraisemblance déterministe ou stochastique), notre contribution a plusieurs qualités : elle ne nécessite pas la connaissance préalable du nombre d'« événements », ni une solution initiale proche de l'optimum, et ses performances sont bien supérieures sur des cas d'école.
- La thèse de Philippe CIUCIU est plus particulièrement consacrée à l'introduction de termes markoviens permettant la restitution de spectres réguliers par morceaux. De même que l'introduction de modèles locaux permet de préserver les contours des zones homogènes d'une image, elle est susceptible de favoriser la régularité d'un spectre tout en respectant les ruptures. Si la régularité porte sur les modules spectraux (plutôt que sur les amplitudes complexes), l'obtention d'un critère convexe en X passe par la question suivante : à quelle condition la fonction $\phi(|X_1| - |X_2|)$ est-elle convexe en $(X_1, X_2) \in \mathbb{C}^2$? Il s'avère que la convexité de ϕ ne suffit pas. La référence [12] présente des conditions suffisantes et des algorithmes de minimisation semi-quadratiques adaptés. Dans la suite de ces travaux, nous nous sommes intéressés à la restitution de raies spectrales dans du bruit coloré, en combinant deux modèles « à énergie convexe » [34].

2.5 Conclusion et perspectives

Le mode de présentation adopté dans les sections précédentes permet de distinguer nettement les thèmes aboutis et les thèmes en développement, et, pour ces derniers, mes préoccupations « prioritaires ». Les deux thèmes qui devraient rester les plus actifs à court et moyen terme sont la modélisation markovienne « à énergie convexe » d'une part, et les problèmes d'estimation spectrale d'autre part.

Approches semi-quadratiques

Plus spécifiquement, les formulations semi-quadratiques constituent actuellement une grande partie de mon activité de recherche. Il s'agit d'un investissement à caractère algorithmique : comment obtenir toujours plus efficacement le minimiseur de l'énergie *a posteriori* en déconvolution, en reconstruction, ... ? Mais il permet également d'aborder des problèmes plus fondamentaux sous un angle original.

Je pense d'une part à la spécification de nouvelles familles de modèles gibbsiens ; par exemple, les modèles 2D « à énergie convexe » et à variables de ligne interactives proposés en Annexe B, mais aussi à des extensions semi-quadratiques originales du terme de douceur (2.3) – page 28 pour les modèles « AR longs » (travail en cours).

Je pense également à la résolution de problèmes non supervisés pour ce qui concerne les hyperparamètres. D'une façon plus générale, la mise au point de méthodes pénalisées exemptes de réglage empirique représente à la fois un enjeu pratique évident pour certaines applications, mais aussi l'occasion d'asseoir la crédibilité des approches pénalisées dans le domaine académique du traitement du signal. Des résultats spécifiques ont déjà été obtenus : en segmentation non supervisée par champs de Markov unilatéraux [15], en estimation impulsionnelle [6, 17]. La poursuite de ce type de travaux est une des priorités de mon activité, en particulier dans le domaine de l'imagerie par modélisation markovienne. Pour cela, le support probabiliste semble être un outil formel très adapté, car il permet l'écriture de vraisemblances marginales, mais de nombreuses difficultés subsistent. Cette perspective est abordée plus en détail au Chapitre 5.

Problèmes aveugles

Les problèmes inverses « aveugles », tel que la déconvolution à noyau inconnu, constituent des problèmes voisins des problèmes « non supervisés » abordés au Chapitre 5. En dehors de certains cas particuliers (par exemple, la déconvolution *impulsionnelle* [CHENG et coll. 1996] – voir Figure 2.2), la déconvolution « aveugle » comporte, outre les difficultés d'ordre calculatoire, un obstacle essentiel, d'ordre informationnel (voir Figure 2.3). En l'absence de toute information *a priori* sur l'objet \mathbf{x} et sur la réponse impulsionnelle \mathbf{h} , il est évidemment impossible de discriminer le couple (\mathbf{x}, \mathbf{h}) de tout autre couple $(\mathbf{h}_1 \star \mathbf{x}, \mathbf{h}_2)$ tel que $\mathbf{h}_1 \star \mathbf{h}_2 = \mathbf{h}$. À l'extrême, la solution dégénérée $(\mathbf{h} \star \mathbf{x}, I)$ (où I est l'opérateur identité) joue le rôle d'une candidate indésirable mais « naturelle », dont il n'est pas forcément facile de se débarrasser. Les Figures 2.2 et 2.3 illustrent ce problème, sous la forme d'une indétermination *sur le contenu spectral*. Il s'y ajoute, en général, une indétermination *sur la phase* : fréquence par fréquence, entrée et filtre peuvent être déphasés d'une quantité arbitraire opposée sans affecter le résultat de la convolution.

Une revue de la littérature actuelle [LAGENDIJK et coll. 1990, YOU et KAVEH 1996, KUNDUR et HATZINAKOS 1996, CHAN et CHIU-KWONG 1998] peut pourtant laisser penser que des solutions systématiques, « sans informations *a priori* », existent.

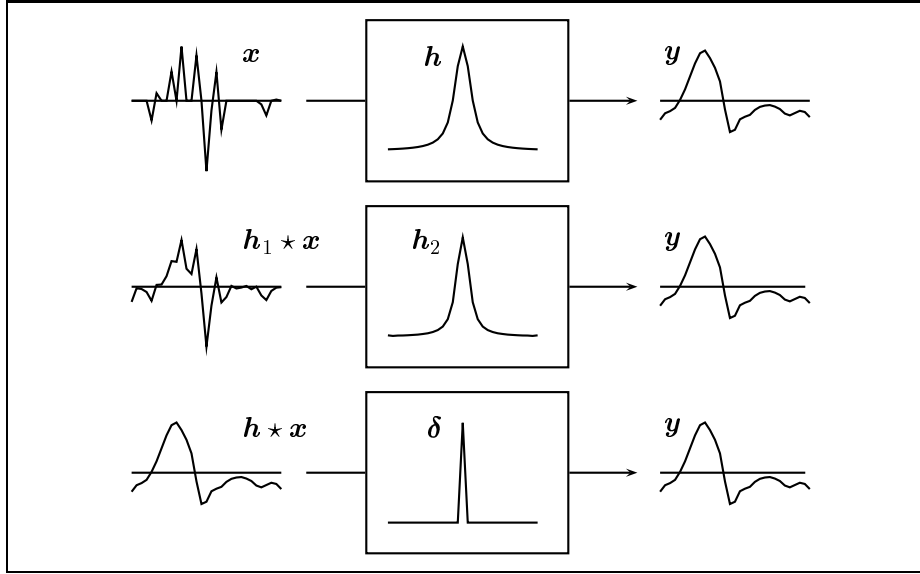


FIG. 2.2: Exemple d'un problème de déconvolution myope 1D pour lequel le signal d'entrée x inconnu est blanc. Le signal y désigne la sortie non bruitée. Dans ce cas, le couple solution (x, h) est en concurrence avec d'autres couples $(h_1 \star x, h_2)$ tels que $h_1 \star h_2 = h$, mais qui ne satisfont pas l'hypothèse de blancheur de l'entrée.

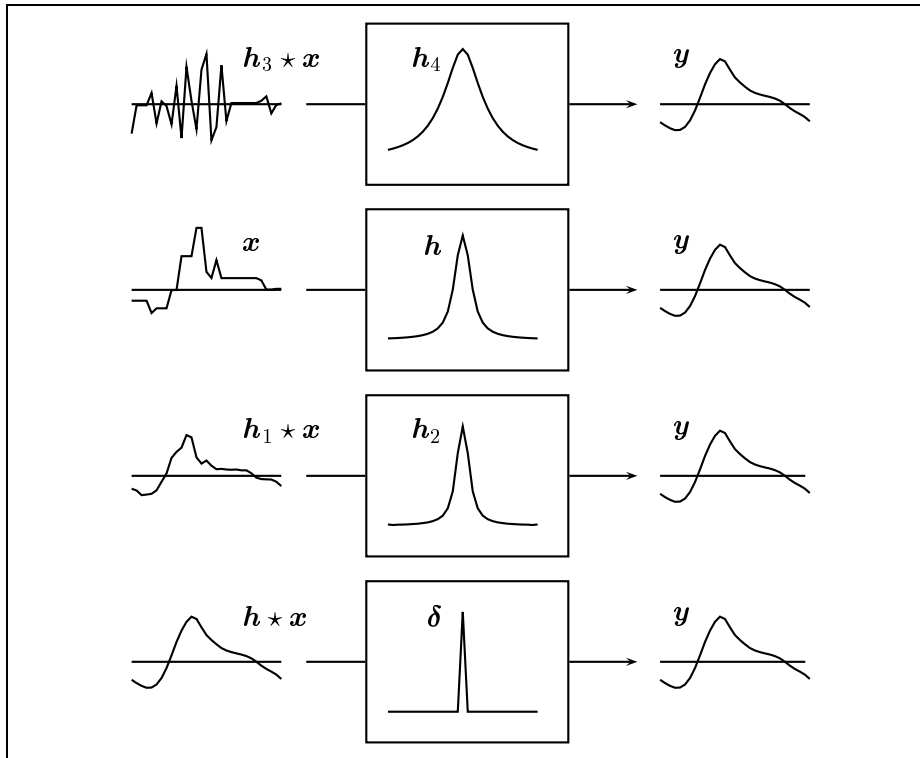


FIG. 2.3: Exemple d'un problème de déconvolution aveugle 1D pour lequel le signal d'entrée x inconnu est corrélé. Le signal y désigne la sortie non bruitée. Dans ces conditions, le couple solution (x, h) est en concurrence avec des couples $(h_1 \star x, h_2)$ tels que $h_1 \star h_2 = h$, dont l'entrée est moins corrélée que x (en haut de la figure) ou plus corrélée (en bas de la figure). Un simple critère de corrélation ou d'énergie ne suffit plus pour discriminer la solution correcte.

Parmi les solutions souvent référencées, [LAGENDIJK et coll. 1990] propose un estimateur de maximum de vraisemblance joint (noyau, hyperparamètre) sous hypothèse d'objet \mathbf{x} gaussien AR 2D et de noyau RIF 2D. Cette méthode repose sur l'identification des coefficients d'un processus ARMA, modèle résultant pour l'image observée. Mais le fonctionnement pratique d'une telle méthode semble très aléatoire, car il est manifeste que ce sont les hypothèses structurelles adoptées (l'objet est un AR 2D, le noyau est un filtre RIF 2D) qui permettent ici de lever l'indétermination illustrée par la Figure 2.3. Or ces hypothèses correspondent à des modèles très grossiers. D'ailleurs de nouvelles informations structurantes apparaissent en cours de route dans [LAGENDIJK et coll. 1990, Section IV], pour pallier des difficultés qui sont attribuées à des aspects numériques ou algorithmiques. La source de ces difficultés me semble malheureusement beaucoup plus essentielle : l'indétermination fondamentale subsiste en l'absence d'informations suffisamment structurantes. Si de telles informations sont effectivement accessibles, il est excessif de parler de déconvolution *aveugle* ; on pourra parler, par exemple, de déconvolution *myope*.

On retrouve la même ambiguïté dans des contributions plus récentes. Ainsi, dans [CHAN et CHIU-KWONG 1998], un estimateur de maximum *a posteriori* joint $(\mathbf{x}, \mathbf{h})^{\text{MAP}}$ est proposé, et il est affirmé en introduction : « *our algorithm can recover both the images and psf's without any a priori information on the psf* ». Un peu plus loin, différents couples solutions $(\hat{\mathbf{x}}, \hat{\mathbf{h}})$ sont obtenus pour différentes valeurs d'un paramètre de régularisation α_2 , illustrant implicitement l'indétermination fondamentale de la Figure 2.3, avec le commentaire : « *... α_2 can be chosen proportionately according to the amount of desired deblurring* ». Finalement, des informations *a priori* sont nécessaires pour choisir α_2 , choix laissé à l'opérateur.

Selon ma propre expérience en déconvolution myope pour l'imagerie satellitaire [88], les résultats pratiques ne peuvent être que parcellaires, et les méthodes statistiques — maximum de vraisemblance $\hat{\mathbf{h}}^{\text{MV}}$, maximum *a posteriori* joint $(\mathbf{x}, \mathbf{h})^{\text{MAP}}$ — sont trop fragiles car dépendantes de modèles *a priori* grossiers et d'hyperparamètres à régler. Dans [88], j'ai proposé une méthode partiellement supervisée. Elle s'appuie sur la présence supposée de quelques zones polygonales à peu près uniformes dans l'image nette (champs, toits d'immeubles, ...), dont l'opérateur peut approximativement localiser les sommets dans l'image floue. La restauration de ces zones uniformes à bord francs (dont on pourrait envisager d'automatiser la localisation) constitue une information *a priori* suffisamment structurante pour lever l'indétermination fondamentale en déconvolution aveugle.

Un autre axe de recherche réaliste consiste à enrichir les données acquises, avec l'objectif explicite de lever l'indétermination. C'est le cas de l'imagerie optique par diversité de phase en présence d'aberrations (turbulence atmosphérique, défauts optiques), qui correspond au sujet de la thèse d'Amandine BLANC et à une collaboration avec Laurent MUGNIER, du Département « Optique théorique et appliquée » de l'ONERA. Les aberrations de phase inconnues ϕ se traduisent dans le plan focal par une convolution $\mathbf{z} = \mathbf{h}(\phi) \star \mathbf{x} + \text{bruit}$, avec $\mathbf{h}(\phi) = |TF(Pe^{i\phi})|^2$ (P est la fonction caractéristique de la pupille). Le principe de la diversité de phase est d'acquérir une seconde image hors du plan focal $\mathbf{z}' = \mathbf{h}(\phi + \delta\phi) \star \mathbf{x} + \text{bruit}'$, à déphasage $\delta\phi$ connu. Il faut souligner ici que la relation entre $\mathbf{h}(\phi)$ et $\mathbf{h}(\phi + \delta\phi)$ n'est pas linéaire. Les données complémentaires \mathbf{z}' peuvent donc permettre de lever l'indétermination fondamentale, y compris par les méthodes statistiques mentionnées plus haut [33].

Chapitre 3

Notice bibliographique (octobre 1991-99)

Les références annexées en Partie IV de ce mémoire sont [4], en Annexe A, page 79 ; [10], en Annexe B, page 81 ; [5], en Annexe C, page 83 ; [2], en Annexe D, page 85.

Articles dans des revues internationales avec comité de lecture

- [1] F. Champagnat et **J. Idier**, « On the correlation structure of unilateral AR processes », à paraître dans *Adv. Appl. Prob.*, vol. 32, n° 2, pp. 1–21, juin 2000.
- [2] **J. Idier** et J.-F. Giovannelli, « Structural stability of least squares prediction methods », *IEEE Trans. Signal Processing*, vol. 46, n° 11, pp. 3109–3111, novembre 1998.
- [3] T. Martin et **J. Idier**, « Estimating a conductivity distribution via a FEM-based nonlinear Bayesian method », *European Physical Journal, Applied Physics*, vol. 1, pp. 87–91, janvier 1998.
- [4] M. Nikolova, **J. Idier** et A. Mohammad-Djafari, « Inversion of large-support ill-posed linear operators using a piecewise Gaussian MRF », *IEEE Trans. Image Processing*, vol. 7, n° 4, pp. 571–585, avril 1998.
- [5] F. Champagnat, **J. Idier** et Y. Goussard, « Stationary Markov random fields on a rectangular finite lattice », *IEEE Trans. Inf. Theory*, vol. 44, n° 7, pp. 2901–2916, novembre 1998.
- [6] F. Champagnat, Y. Goussard et **J. Idier**, « Unsupervised deconvolution of sparse spike trains using stochastic approximation », *IEEE Trans. Signal Processing*, vol. 44, n° 12, pp. 2988–2998, décembre 1996.
- [7] **J. Idier** et Y. Goussard, « Markov modeling for Bayesian restoration of two-dimensional layered structures », *IEEE Trans. Inf. Theory*, vol. IT-39, n° 4, pp. 1356–1373, juillet 1993.
- [8] **J. Idier** et Y. Goussard, « Multichannel seismic deconvolution », *IEEE Trans. Geosci. Remote Sensing*, vol. GE-31, n° 5, pp. 961–979, septembre 1993.

Articles soumis

- [9] J.-F. Giovannelli et **J. Idier**, « Bayesian interpretation of periodograms », rapport technique soumis à *IEEE Trans. Signal Processing*, GPI-LSS, 1999.
- [10] **J. Idier**, « Convex half-quadratic criteria and interacting auxiliary variables for image restoration », rapport technique soumis à *IEEE Trans. Image Processing*, GPI-LSS, 1999.

Articles dans des comptes rendus de conférences internationales avec comité de lecture

- [11] **J. Idier**, « Regularization tools and models for image and signal reconstruction », in *3rd Intern. Conf. Inverse Problems in Engng.*, Port Ludlow, U.S.A., conférence invitée, juin 1999, pp. 1–6.
- [12] P. Ciuciu, **J. Idier** et J.-F. Giovannelli, « Markovian high resolution spectral analysis », in *Proc. IEEE ICASSP*, Phoenix, U.S.A., mars 1999, pp. 1601–1604.
- [13] L. Delphin-Poulat, C. Mokbel et **J. Idier**, « Frame-synchronous stochastic matching based on the Kullback-Leibler information », in *Proc. IEEE ICASSP*, Seattle, U.S.A., mai 1998, pp. 89–92.
- [14] S. Gautier, **J. Idier**, A. Mohammad-Djafari et B. Lavayssière, « X-ray and ultrasound data fusion », in *Proc. IEEE ICIP*, Chicago, U.S.A., octobre 1998, pp. 366–369.
- [15] Y. Goussard, **J. Idier** et A. De Cesare, « Unsupervised image segmentation using a telegraph parameterization of Pickard random fields », in *Proc. IEEE ICASSP*, Munich, Germany, avril 1997, pp. 2777–2780.
- [16] H. Carfantan, A. Mohammad-Djafari et **J. Idier**, « A single site update algorithm for nonlinear diffraction tomography », in *Proc. IEEE ICASSP*, Munich, Germany, avril 1997, pp. 2837–2840.
- [17] F. Dublanchet, **J. Idier** et P. Duvaut, « Direction-of-arrival and frequency estimation using Poisson-Gaussian modeling », in *Proc. IEEE ICASSP*, Munich, Germany, avril 1997, pp. 3501–3504.
- [18] M. Fayolle et **J. Idier**, « EM algorithm for unsupervised segmentation of a piecewise ar process », in *Proc. IEEE ICASSP*, Munich, Germany, avril 1997, pp. 3545–3548.
- [19] T. Martin et **J. Idier**, « A FEM-based nonlinear MAP estimator in electrical impedance tomography », in *Proc. IEEE ICIP*, Santa Barbara, USA, octobre 1997, pp. 684–687.
- [20] N. Villain, Y. Goussard, S. Brette et **J. Idier**, « Computed tomography image restoration using convex-potential 3-D Markov random fields », in *Proc. IEEE EMBS*, Chicago, U.S.A., 1997, pp. 561–564.
- [21] **J. Idier**, A. Mohammad-Djafari et G. Demoment, « Regularization methods and inverse problems: an information theory standpoint », in *2nd Intern. Conf. Inverse Problems in Engng.*, Le Croisic, France, juin 1996, pp. 321–328.
- [22] S. Brette et **J. Idier**, « Optimized single site update algorithms for image deblurring », in *Proc. IEEE ICIP*, Lausanne, Switzerland, septembre 1996, pp. 65–68.
- [23] A. Herment, J.-F. Giovannelli, E. Mousseaux, **J. Idier**, A. Decesare et J. Bittoun, « Regularized estimation of flow patterns in MR velocimetry », in *Proc. IEEE ICIP*, Lausanne, Switzerland, septembre 1996, pp. 291–294.
- [24] A. Herment, E. Mousseaux, J.-F. Giovannelli, **J. Idier**, J. Bittoun et O. Jolivet, « MR velocity mapping: Improvement of noise robustness by using a regularized estimation of flow patterns », in *Computer Assisted Radiology*, Paris, juin 1996, vol. 1124, pp. 116–120.
- [25] T. Martin et **J. Idier**, « A Bayesian non-linear inverse approach for electrical impedance tomography », in *2nd Intern. Conf. Inverse Problems in Engng.*, Le Croisic, France, juin 1996, pp. 473–480.
- [26] F. Champagnat et **J. Idier**, « An alternative to standard maximum likelihood for Gaussian mixtures », in *Proc. IEEE ICASSP*, Detroit, U.S.A., mai 1995, pp. 2020–2023.
- [27] **J. Idier**, J.-F. Giovannelli et B. Querleux, « Bayesian time-varying AR spectral estimation for ultrasound attenuation measurement in biological tissues », in *Proceedings of the Section on Bayesian Statistical Science*, Alicante, Spain, 1994, pp. 256–261, American Statistical Association.

- [28] S. Brette, **J. Idier** et A. Mohammad-Djafari, « Scale invariant Markov models for linear inverse problems », in *Proc. of the Section on Bayesian Statistical Sciences*, Alicante, Spain, 1994, pp. 266–270, American Statistical Association.
- [29] J.-F. Giovannelli, **J. Idier**, B. Querleux, A. Herment et G. Demoment, « Maximum likelihood and maximum a posteriori estimation of Gaussian spectra. Application to attenuation measurement and color Doppler velocimetry », in *Proc. Int. Ultrasonics Symp.*, Cannes, France, novembre 1994, vol. 3, pp. 1721–1724.
- [30] M. Nikolova, A. Mohammad-Djafari et **J. Idier**, « Inversion of large-support ill-conditioned linear operators using a Markov model with a line process », in *Proc. IEEE ICASSP*, Adelaide, Australia, 1994, vol. V, pp. 357–360.
- [31] F. Champagnat, **J. Idier** et G. Demoment, « Deconvolution of sparse spike trains accounting for wavelet phase shifts and colored noise », in *Proc. IEEE ICASSP*, Minneapolis, U.S.A., 1993, pp. 452–455.
- [32] **J. Idier** et Y. Goussard, « Modélisation markovienne pour l'imagerie des milieux stratifiés », in *24^e Journées de statistiques*, Bruxelles, Belgique, 1992, pp. 266–268, ASU.

Articles dans des comptes rendus de conférences nationales avec comité de lecture

- [33] A. Blanc, L. Mugnier et **J. Idier**, « Identifiabilité et estimation myope d'aberrations optiques par diversité de phase », in *Actes 17^e coll. GRETSI*, Vannes, France, septembre 1999, pp. 63–66.
- [34] P. Ciuciu, **J. Idier** et J.-F. Giovannelli, « Analyse spectrale non paramétrique haute résolution », in *Actes 17^e coll. GRETSI*, Vannes, France, septembre 1999, pp. 721–724.
- [35] **J. Idier**, « Critères convexes et processus auxiliaires interactifs pour la restauration d'images », in *Actes 17^e coll. GRETSI*, Vannes, France, septembre 1999, pp. 399–402.
- [36] A. Ridolfi et **J. Idier**, « Penalized maximum likelihood estimation for univariate normal mixture distributions », in *Actes 17^e coll. GRETSI*, Vannes, France, septembre 1999, pp. 259–262.
- [37] N. Villain, Y. Goussard, **J. Idier** et S. Brette, « Restauration par champ de Markov 3D à potentiels convexes appliquée aux images tomographiques », in *Actes 17^e coll. GRETSI*, Vannes, France, septembre 1999, pp. 599–602.
- [38] S. Gautier, **J. Idier**, A. Mohammad-Djafari et B. Lavayssière, « Fusion de données gammagraphiques et ultrasonores », in *Actes 16^e coll. GRETSI*, Grenoble, France, septembre 1997, pp. 781–784.
- [39] S. Gautier, **J. Idier**, A. Mohammad-Djafari et B. Lavayssière, « Traitement d'échogrammes ultrasonores par déconvolution aveugle », in *Actes 16^e coll. GRETSI*, Grenoble, France, septembre 1997, pp. 1431–1434.
- [40] **J. Idier** et J.-F. Giovannelli, « Stabilité structurelle des méthodes de prédiction linéaire », in *Actes 16^e coll. GRETSI*, Grenoble, France, septembre 1997, pp. 543–546.
- [41] **J. Idier**, J.-F. Giovannelli et P. Ciuciu, « Interprétation régularisée des périodogrammes et extensions non quadratiques », in *Actes 16^e coll. GRETSI*, Grenoble, France, septembre 1997, pp. 695–698.
- [42] T. Martin et **J. Idier**, « Reconstruction bayésienne en tomographie d'impédance électrique », in *Numélec 97: seconde conférence européenne sur les méthodes numériques en électromagnétisme*, Lyon, France, mars 1997, pp. 30–31.

- [43] F. Dublanchet, P. Duvaut, **J. Idier** et G. Demoment, « EXPULSE COMPLEXE, estimation bayésienne de sinusoides par déconvolution de la transformée de Fourier discrète du signal », in *Actes 15^e coll. GRETSI*, Juan-les-Pins, France, septembre 1995, pp. 37–40.
- [44] **J. Idier** et Y. Goussard, « Formules de réestimation pour un modèle de chaîne de Markov cachée stationnaire réversible », in *Actes 15^e coll. GRETSI*, Juan-les-Pins, France, septembre 1995, pp. 177–180.
- [45] T. Martin et **J. Idier**, « Problème mal posé et reconstruction bayésienne en tomographie d'impédance électrique », in *Actes 15^e coll. GRETSI*, Juan-les-Pins, France, septembre 1995, pp. 905–908.
- [46] G. Perrot, Y. Goussard et **J. Idier**, « Paramétrisation « télégraphique » de champs de Pickard pour la segmentation non-supervisée d'images », in *Actes 15^e coll. GRETSI*, Juan-les-Pins, France, septembre 1995, pp. 669–672.
- [47] F. Champagnat et **J. Idier**, « Un nouvel algorithme de déconvolution impulsienne avec prise en compte de saturation », in *Actes 14^e coll. GRETSI*, Juan-les-Pins, 1993, pp. 559–562.
- [48] G. Demoment et **J. Idier**, « Problèmes inverses et déconvolution », in *2^e Congrès français d'acoustique*, Arcachon, avril 1992, pp. 929–936, Journal de physique IV.
- [49] Y. Goussard et **J. Idier**, « Modélisation markovienne pour l'imagerie des milieux stratifiés », in *Journées thématiques « Approches markoviennes en signal et images »*, Paris, 1992, pp. 105–119, GDR TDSI.

Conférences sans actes

- [50] **J. Idier** et S. Gautier, « La détection de défauts en contrôle non destructif par ultrasons : utilisation de la déconvolution impulsienne », Paris, France, octobre 1999, Société de calcul mathématique, séminaire invité.
- [51] **J. Idier**, « Comparaison de méthodes pour l'estimation en imagerie », Paris, France, janvier 1998, GDR-PRC ISIS, GT1.
- [52] **J. Idier**, P. Ciuciu et J.-F. Giovannelli, « Analyse spectrale à temps court et périodogrammes non quadratiques », Palaiseau, France, janvier 1998, CMAPX, École Polytechnique, séminaire invité.
- [53] B. Rougé, G. Demoment et **J. Idier**, « Méthodes de restauration : adéquation aux besoins du CNES en imagerie optique (opération recherche-industrie) », Paris, France, mars 1998, 7^{ème} séminaire des partenaires du GDR-PRC ISIS.
- [54] **J. Idier**, « Estimation des hyperparamètres en estimation d'impulsions », Paris, France, juillet 1997, GDR-PRC ISIS, GT7.
- [55] **J. Idier**, « Interprétation et outils probabilistes pour l'inversion en imagerie », Paris, France, mars 1997, OFTA, séminaire invité.
- [56] **J. Idier**, « Interprétation régularisée des périodogrammes et extensions non quadratiques », Cergy, France, novembre 1997, ENSEA, Équipe de traitement des images et du signal, séminaire invité.
- [57] **J. Idier**, « Image processing using Markov random fields », Milan, Italie, juillet 1997, Politecnico di Milano, séminaire invité.
- [58] N. Villain, Y. Goussard et **J. Idier**, « Restauration d'images tomographiques par champ de Markov 3D », in *Proc. ACFAS Conf.*, Trois-Rivières, Canada, 1997.
- [59] S. Brette et **J. Idier**, « Déconvolution d'images à noyau séparable », Paris, France, novembre 1996, GDR-PRC ISIS, GT7.

- [60] F. Dublanchet et **J. Idier**, « Détection et estimation de directions d'arrivées par une approche bayésienne avec modélisation impulsionnelle », Paris, France, mai 1996, GDR-PRC ISIS, GT4.
- [61] A. Herment, E. Mousseaux, J.-F. Giovannelli, **J. Idier**, O. Jolivet et J. Bittoun, « Improved robustness of MR velocity mapping by using a spatial regularized estimation of flow patterns », in *Fourth scientific meeting of the International Society for Magnetic Resonance in Medicine*, New York, avril 1996, vol. 2, p. 1288.
- [62] **J. Idier** et Y. Goussard, « Algorithme itératif pour identification d'un modèle markovien d'images », Montréal (Québec), Canada, 1995, Journées d'optimisation.
- [63] **J. Idier**, Y. Goussard et S. Doré, « Problèmes d'imagerie pour la conception de prothèses médicales », Paris, France, novembre 1995, GDR-PRC ISIS, GT7.
- [64] **J. Idier**, « Bayesian restoration of sparse spikes. Application to spectral analysis », Cambridge, U.S.A., décembre 1995, Stochastic Systems Group, MIT, séminaire invité.
- [65] M. Nikolova, **J. Idier** et A. Mohammad-Djafari, « Relaxation déterministe en imagerie markovienne », Paris, France, 1993, GDR TDSI, GT2.

Chapitres de livres

- [66] G. Demoment et **J. Idier**, *Chapitre III: Approche bayésienne pour la résolution des problèmes inverses en imagerie*, pp. 59–77, ARAGO, vol. 22. OFTA, Paris, TEC & DOC edition, 1999.
- [67] A. Mohammad-Djafari et **J. Idier**, *A scale invariant Bayesian method to solve linear inverse problems*, pp. 121–134, Kluwer Academic Publ., G. Heidbreder edition, 1996.
- [68] F. Dublanchet, P. Duvaut et **J. Idier**, *Complex sinusoid analysis by Bayesian deconvolution of the discrete Fourier transform*, pp. 323–328, Maximum entropy and Bayesian methods. Kluwer Academic Publ., Santa Fe, U.S.A., K. Hanson edition, 1995.
- [69] S. Brette, **J. Idier** et A. Mohammad-Djafari, *Scale invariant Markov models for Bayesian inversion of linear inverse problems*, pp. 199–212, Maximum entropy and Bayesian methods. Kluwer Academic Publ., Cambridge, England, J. Skilling & S. Sibusu edition, 1994.
- [70] A. Mohammad-Djafari et **J. Idier**, *Maximum Likelihood Estimation of the Lagrange Parameters of the Maximum Entropy Distributions*, pp. 131–140, Kluwer Academic Publ., Seattle, U.S.A., C.R. Smith, G.J. Erikson and P.O. Neudorfer edition, 1991.

Rapports internes

- [71] **J. Idier**, Y. Goussard et A. Ridolfi, « Unsupervised image segmentation using a telegraph parameterization of Pickard random fields », Rapport interne, IGB/GPI-LSS, 1999.
- [72] N. Villain, Y. Goussard, **J. Idier** et S. Brette, « 3D edge-preserving image enhancement for computed tomography », rapport technique, IGB/GPI-LSS, 1999.
- [73] P. Ciuciu, **J. Idier** et J.-F. Giovannelli, « Markovian high resolution nonparametric spectral analysis », rapport technique, GPI-LSS, 1999.
- [74] T. Martin et **J. Idier**, « Stability and accuracy of finite element direct solvers for electrical impedance tomography », rapport technique, GPI-LSS, 1998.
- [75] M. Fayolle et **J. Idier**, « Unsupervised estimation for a hidden stationary piecewise AR(1) », rapport technique, GPI-LSS, 1998.

- [76] M. Fayolle, Y. Goussard et **J. Idier**, « Sequential edge detection algorithm for images modeled by compound Markov random fields », rapport technique, IGB/GPI-LSS, 1998.
- [77] M. Fayolle, Y. Goussard et **J. Idier**, « Utilisation des symétries pour l'estimation de la fonction de partition », rapport technique, IGB/GPI-LSS, 1998.
- [78] M. Fayolle, **J. Idier** et F. Champagnat, « Fully Bayesian estimation of unilateral fields », rapport technique, IGB/GPI-LSS, 1998.
- [79] O. Dubois, **J. Idier** et J.-L. Starck, « Deconvolution of astronomical images using a Markov random field model », rapport technique, DAPNIA/SEI-SAP, CEA/Saclay, Gif-sur-Yvette, France, 1998.
- [80] J.-F. Giovannelli, **J. Idier**, G. Desodt et D. Muller, « Analyse spectrale AR adaptative régularisée », rapport technique, GPI-LSS, 1996.
- [81] Y. Goussard et **J. Idier**, « Déconvolution non-supervisée de trains d'impulsions par optimisation stochastique », Rapport interne, IGB/GPI-LSS, 1995.
- [82] F. Champagnat et **J. Idier**, « Generalized marginal likelihood for Gaussian mixtures », rapport technique, GPI-LSS, 1995.

Documents pédagogiques

- [83] H. Piet-Lahanier et **J. Idier**, *Bases mathématiques pour le traitement numérique du signal*, Service d'Édition de l'ENSTA, 1993.
- [84] H. Piet-Lahanier et **J. Idier**, *Algorithmes pour le traitement du signal*, Service d'Édition de l'ENSTA, 1993.

Rapports de contrats

- [85] H. Carfantan et **J. Idier**, « Estimation de paramètres instrumentaux à partir d'images SPOT », Rapport de contrat CNRS-CNES, OMP/GPI-LSS, 1999.
- [86] J.-F. Giovannelli et **J. Idier**, « Méthodes et algorithmes d'inversion de données en spectrométrie de neutrons : analyse bibliographique prospective », Rapport de contrat SUPÉLEC-CEA, GPI-LSS, 1999.
- [87] T. Martin et **J. Idier**, « Prise en compte de la géométrie des capteurs et de l'optique en restauration d'images satellitaires », Rapport de contrat CNES, GPI-LSS, 1998.
- [88] **J. Idier** et C. Heinrich, « Restitution de paramètres instrumentaux », Rapport de contrat CNES, GPI-LSS, 1997.
- [89] F. Champagnat, **J. Idier** et G. Demoment, « Restauration convexe markovienne », Rapport de contrat CNES, GPI-LSS, 1996.
- [90] P. Ciuciu, J.-F. Giovannelli et **J. Idier**, « Analyse spectrale post-moderne. application aux signaux radars », Rapport de contrat CNRS-Société THOMSON, GPI-LSS, 1996.
- [91] J.-F. Giovannelli et **J. Idier**, « Une nouvelle approche non-paramétrique de l'imagerie radar Doppler », Rapport de contrat CNRS-Société THOMSON, GPI-LSS, 1995.
- [92] J.-F. Giovannelli et **J. Idier**, « Caractérisation spectrale du fouillis de radar Doppler. Méthodes autorégressives adaptatives régularisées », Rapport de contrat CNRS-Société THOMSON, GPI-LSS, 1994.
- [93] J.-F. Giovannelli et **J. Idier**, « Mesure de l'atténuation acoustique de la peau. étude de faisabilité », Rapport de contrat CNRS-Société L'ORÉAL, GPI-LSS, 1993.
- [94] **J. Idier**, A. Kartmann et J. Oksman, « Étude d'un procédé d'optimisation d'instrument », Rapport interne, GPI-LSS, 1992.

Deuxième partie

Restauration de signaux et d'images : une mise en perspective

Chapitre 4

Approches pénalisées

4.1 La régularisation au sens de Tikhonov

4.1.1 Généralités

En général, on construit des solutions admissibles à un problème mal posé par restriction à une classe restreinte, étant donné une connaissance *a priori*. Suivant l'approche développée dans [TIKHONOV et ARSÉNINE 1976], cette restriction est introduite en pénalisant les candidats suivant une mesure de régularité supposée correspondre à cette connaissance. Supposons qu'une fonction monovariée x^* doive être estimée à partir de données imparfaites

$$z = Ax^* + \text{bruit}, \quad (4.1)$$

où A est un opérateur linéaire. L'approche introduite par Tikhonov requiert la sélection d'une fonctionnelle régularisante $\|Dx\|^2$, où D est aussi un opérateur linéaire, par exemple un opérateur différentiel. La solution estimée \hat{x} est alors définie comme celle qui minimise la fonctionnelle

$$\mathcal{J}(x) = \|z - Ax\|^2 + \lambda \|Dx\|^2, \quad (4.2)$$

où λ est un paramètre de régularisation, qui contrôle le compromis entre régularité de la solution et fidélité aux données. En pratique, on peut imposer des conditions peu restrictives permettant d'assurer que \mathcal{J} est strictement convexe et possède un unique minimiseur.

Le premier terme de $\mathcal{J}(x)$ dans (4.2) est une norme quadratique qui pénalise l'écart entre les données z et la « sortie modèle » pour une fonction admissible x . Dans la suite, on s'intéresse plutôt au second terme de $\mathcal{J}(x)$, *i.e.*, à la construction de la fonctionnelle régularisante. Dans la contribution originale de Tikhonov, la fonctionnelle régularisante est donnée par

$$\|Dx\|^2 = \sum_{r=0}^R \int c_r(t) |x^{(r)}(t)|^2 dt,$$

où les poids c_r sont des fonctions strictement positives et $x^{(r)}$ désigne la r^{e} dérivée de x . Qualitativement, il est clair qu'un tel choix correspond à une hypothèse *a priori* portant sur la « douceur » du signal à estimer x^* .

Des extensions multivariées ont été proposées. Plus spécifiquement, le cas bidimensionnel (2D) a été développé pour le traitement d'image. Les travaux novateurs de Hunt [HUNT 1977] ont introduit la déconvolution d'image fondée sur l'équivalent discrétisé (aux différences finies) de (4.2) et le calcul rapide de la solution par passage dans le domaine de Fourier.

4.1.2 Limitation

Un grand nombre de problèmes inverses concernent des objets *homogènes par morceaux*, tels que des fonctions en escalier, des images constituées de régions distinctes, des défauts isolés dans un milieu homogène. Dans la plupart de ces situations, les méthodes directement reliées à l'approche de Tikhonov ne permettent pas de détecter les discontinuités, ni même d'en restituer approximativement la position.

A titre d'illustration, considérons l'expérience simulée suivante : soit $\mathbf{z} = [z_0, \dots, z_N]^t$ un vecteur de données bruitées, régulièrement échantillonnées à partir d'une fonction x^* sur $[0,1]$, univariée et douce par morceaux : $z_n = x^*(n/N) + b_n$. La fonction x^* et le vecteur \mathbf{z} sont représentés à la Figure 4.1.

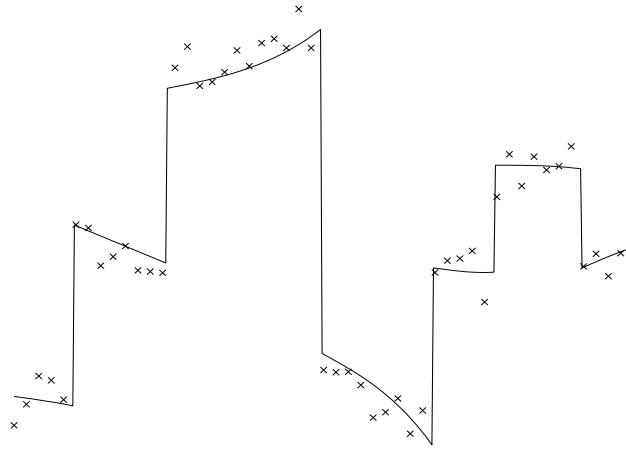


FIG. 4.1: Une fonction x^* monovariée, douce par morceaux, et 51 données bruitées (z_n) , $n = 0, \dots, N = 50$.

Avec $R = 2$, $c_0 = c_1 = 0$, $c_2 = 1$ et $\|\mathbf{z} - A\mathbf{x}\|^2 = \sum_{n=0}^N |z_n - x(n/N)|^2$, (4.2) s'écrit

$$\mathcal{J}(x) = \|\mathbf{z} - A\mathbf{x}\|^2 + \lambda \int_0^1 |x''(t)|^2 dt. \quad (4.3)$$

Introduisons une approximation discrète pour (4.3), obtenue par différences finies:

$$J_M(\mathbf{x}) = \|\mathbf{z} - A\mathbf{x}\|^2 + \lambda M^3 \sum_{m=1}^{M-1} |2x_m - x_{m-1} - x_{m+1}|^2, \quad (4.4)$$

où $\|\mathbf{z} - A\mathbf{x}\|^2 = \sum_{n=0}^N |z_n - x_{nM/N}|^2$, $\mathbf{x} = [x_0, \dots, x_M]^t$, et $M+1$ est un multiple de $N+1$. Définissons alors le vecteur estimé $\hat{\mathbf{x}}^\lambda$ comme le minimiseur de J_M . La convergence point à point de $\hat{\mathbf{x}}^\lambda$ vers le minimiseur unique de (4.3) quand $M \rightarrow \infty$ est un résultat classique [NASHED 1981]. La Figure 4.2 représente $\hat{\mathbf{x}}^\lambda$ pour $M = 400$ et pour la « meilleure » valeur de λ au sens L_1 , c'est-à-dire celle qui minimise $C(\hat{\mathbf{x}}^\lambda, x^*)$, avec

$$C(\mathbf{x}, x^*) = \sum_{m=0}^M |x_m - x^*(m/M)|. \quad (4.5)$$

Cette procédure de choix d'hyperparamètre est artificielle, puisqu'elle requiert la connaissance de x^* , mais elle autorise des comparaisons « équitables ».

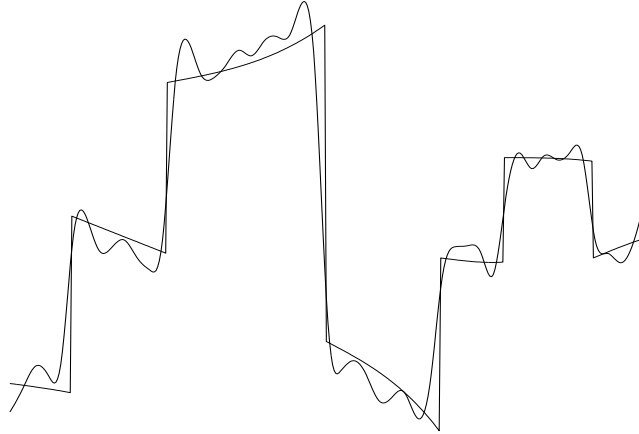


FIG. 4.2: Estimée linéaire douce $\hat{\mathbf{x}}^\lambda$ obtenue comme minimiseur de (4.4), pour $M = 400$ et pour la valeur optimale de λ au sens L_1 : $C(\hat{\mathbf{x}}^\lambda, \mathbf{x}^*) = 18,16\%$.

La solution $\hat{\mathbf{x}}^\lambda$ n'est pas très satisfaisante, car uniformément douce. En comparaison, une simple interpolation linéaire par morceaux des données successives fournit une erreur de norme L_1 égale à 17,04%, donc sensiblement inférieure.

Si le nombre P et les positions $\mathbf{t} = [t_1, \dots, t_P]$ des discontinuités étaient connues, une approche adaptée serait de remplacer le critère \mathcal{J} par

$$\mathcal{J}_{\mathbf{t}}(x) = \|\mathbf{z} - A\mathbf{x}\|^2 + \lambda \sum_{p=0}^P \int_{t_p}^{t_{p+1}} |x(t)''|^2 dt, \quad (4.6)$$

avec $t_0 = 0$ et $t_{P+1} = 1$, critère à minimiser en x sans hypothèse de régularité en t_1, \dots, t_P (c'est pourquoi la somme d'intégrales ne s'écrit pas $\int_0^1 |x(t)''|^2 dt$). Quant à l'équivalent discret, on peut l'exprimer sous la forme

$$J_{\mathbf{b}}(\mathbf{x}) = \|\mathbf{z} - A\mathbf{x}\|^2 + \lambda M^3 \sum_{m=1}^{M-1} b_m |2x_m - x_{m-1} - x_{m+1}|^2, \quad (4.7)$$

où $\mathbf{b} = [b_1, \dots, b_{M-1}]$ est un vecteur binaire de **variables de bord** : $b_m = 0$ correspond à la présence d'une discontinuité à la m^{e} position. En pratique, cette approche est très limitée, car l'ignorance de la position des discontinuités fait partie intégrante du problème. Les sections qui suivent sont consacrées aux principales idées et outils introduits dans la communauté du traitement des signaux et des images pour aborder la question de la restauration de fonctions régulières par morceaux.

4.2 Détection-estimation

4.2.1 Principe

Dans le milieu des années 80, la gestion maîtrisée des discontinuités a marqué une avancée importante, à la fois sur les plans théoriques et pratiques [MUMFORD et SHAH 1985, TERZOPOULOS 1986, BLAKE et ZISSERMAN 1987]. Elle consiste à considérer *conjointement* le problème de l'estimation de x et celui de la détection d'un vecteur inconnu positionnant les discontinuités (*i.e.*, \mathbf{t} en continu ou \mathbf{b} en discret). Pour ce faire, la minimisation conjointe de $\mathcal{J}_{\mathbf{t}}(x)$ en (x, \mathbf{t}) ou

de $J_M(\mathbf{x}, \mathbf{b})$ en (\mathbf{x}, \mathbf{b}) n'est pas adéquate. En effet, il n'est pas difficile de voir que cette stratégie conduit à un nombre maximum de discontinuités (*i.e.*, $P = \infty$ en continu et $P = M$ en discret). Tel n'est pas le cas si un « prix à payer » $\alpha > 0$ est imposé pour chaque discontinuité. Il en résulte un **critère augmenté** qui s'écrit

$$\mathcal{K}(x, \mathbf{t}) = \mathcal{J}_{\mathbf{t}}(x) + \alpha P,$$

ou bien

$$K(\mathbf{x}, \mathbf{b}) = J_{\mathbf{b}}(\mathbf{x}) - \alpha \sum_{m=1}^{M-1} b_m \quad (4.8)$$

pour l'équivalent discret.

Dans l'expression (4.8), la pénalité ne dépend que du nombre de discontinuités, pas de leur position relative. Des variantes sont facilement envisageables ; par exemple, on peut vouloir sur-pénaliser des discontinuités proches, en introduisant un prix à payer dégressif avec la distance entre discontinuités voisines. Quand la pénalité dépend de la position relative des discontinuités, les variables de bord sont dites **interactives**. En segmentation d'image, des modèles spécifiques de variables de bord interactives ont été introduites pour favoriser la fermeture des contours entre régions [GEMAN et GEMAN 1984, MARROQUIN et coll. 1987, JENG et WOODS 1991], ou encore pour pénaliser la formation de lignes parallèles adjacentes [BEDINI et coll. 1994].

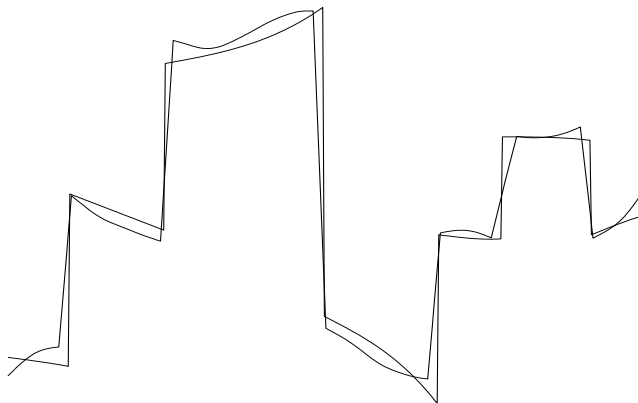


FIG. 4.3: *Estimée douce par morceaux $\hat{\mathbf{x}}^{\lambda, \alpha}$ obtenue comme minimiseur conjoint de (4.8) pour des valeurs optimales de λ, α au sens L_1 : $C(\hat{\mathbf{x}}^{\lambda, \alpha}, \mathbf{x}^*) = 16,22\%$.*

Plusieurs traits communs se dégagent des travaux effectués en détection-estimation pour le traitement d'image et la vision par ordinateur (tel que [MUMFORD et SHAH 1985, TERZOPOULOS 1986, BLAKE et ZISSERMAN 1987]) :

- la présence de variables de bord *binaires* dans un critère augmenté. De telles variables peuvent être qualifiées de **variables cachées** (sous-entendu, vis-à-vis du procédé d'observation), dans la mesure où elles sont exclues de l'équation d'observation (4.1). Conceptuellement et pratiquement, la manipulation de variables cachées fournit une grande variété d'outils pour prendre en compte des connaissances *a priori* sous forme pénalisée.
- Dans la plupart des cas, le critère augmenté est **semi-quadratique** (SQ) : une fonction K est dite SQ si elle dépend de deux ensembles de variables, disons \mathbf{x} et \mathbf{b} , de telle façon que K est une fonction quadratique de \mathbf{x} mais pas de (\mathbf{x}, \mathbf{b}) . L'exploitation de cette caractéristique a fait récemment l'objet de nombreux travaux en restauration d'image. Le Sous-chapitre 4.4 lui est consacré.

4.2.2 Inconvénients

Le coût de calcul est le principal inconvénient pratique de l'approche « détection-estimation ». En effet, des méthodes numériquement lourdes sont souvent nécessaires pour gérer correctement le problème combinatoire induit par l'introduction des variables binaires. La plupart de ces méthodes sont fondées sur le principe de la *relaxation*, dans un cadre stochastique (*e.g.*, le *recuit simulé* [GEMAN et GEMAN 1984]) ou déterministe (*e.g.*, les méthodes de continuation telles que la *non convexité graduelle* — GNC pour *Graduated Non Convexity* [BLAKE et ZISSERMAN 1987]; voir aussi [4]).

Plusieurs auteurs [BOUMAN et SAUER 1993, LI et coll. 1995] ont mis en avant le manque de stabilité de l'estimée $\hat{\mathbf{x}}^{\lambda, \alpha}$ comme une autre faiblesse de la détection-estimation. Mathématiquement, ce manque de stabilité correspond au fait que $\hat{\mathbf{x}}^{\lambda, \alpha}$ n'est pas une fonction *continue* des données. En d'autres termes, la troisième condition de Hadamard [TIKHONOV et ARSÉNINE 1976] n'est pas satisfaite par $\hat{\mathbf{x}}^{\lambda, \alpha}$, donc le problème est encore mal posé malgré la pénalisation. La Figure 4.4 illustre ce problème, suivant un exemple donné dans [LI et coll. 1995].

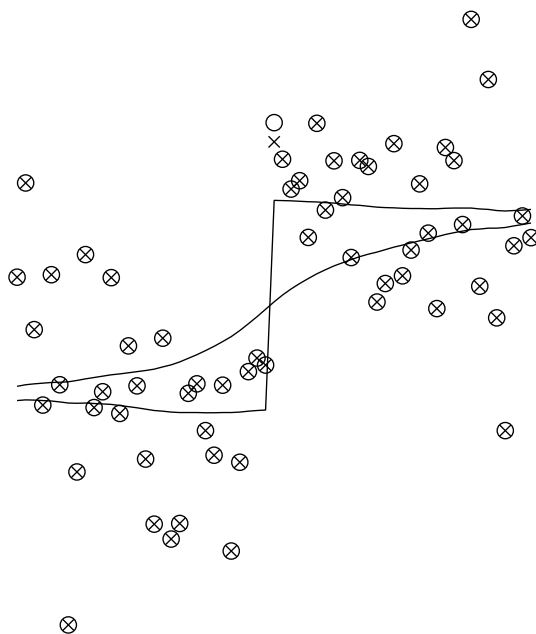


FIG. 4.4: *Instabilité de $\hat{\mathbf{x}}^{\lambda, \alpha}$ comme fonction des données : en trait continu, deux estimées $\hat{\mathbf{x}}^{\lambda, \alpha}$ issues du même jeu de données à l'exception d'une valeur. Les données sont représentées par des cercles pour un jeu et par des croix pour l'autre.*

En fait, $\hat{\mathbf{x}}^{\lambda, \alpha}$ est quand même « stable par morceaux », et ce comportement est intrinsèque à la capacité de détection des discontinuités d'une telle approche. En effet, il est bien connu que la résolution d'un problème de décision à hypothèses multiples revient à partitionner de l'espace des données (ici, \mathbb{R}^N) en autant de régions E_k que d'hypothèses concurrentes [BRÉMAUD 1984, Chapitre XII]. On définit E_k comme le sous-ensemble des valeurs de \mathbf{z} pour laquelle l'hypothèse adoptée est la k^{e} selon la règle de décision envisagée. Ici, on a $k = 2^M$, et il est facile de montrer que $\hat{\mathbf{x}}^{\lambda, \alpha}$ est continu quand \mathbf{z} est à l'intérieur de n'importe lequel des E_k . Autrement dit, les discontinuités se produisent seulement à la frontière entre sous-ensembles, c'est-à-dire quand la variation des données provoque un changement de décision. Dans certains cas, loin d'être un inconvénient, cette caractéristique est souhaitable puisqu'elle témoigne d'une prise de décision automatique. En revanche, si la restauration du signal ne constitue qu'une aide au diagnostic et à la décision d'un expert humain, l'approche « détection-estimation » est sans doute peu

recommandable, car à la fois risquée et numériquement coûteuse.

4.3 Approche non quadratique

Les approches pénalisées non quadratiques ont pris une importance croissante en restauration de signaux et d'images peu avant le milieu des années 90 [BOUMAN et SAUER 1993, KÜNSCH 1994, LI et coll. 1995]. Le principe consiste à remplacer la pénalisation quadratique de Tikhonov par une autre fonction paire, mieux adaptée à la préservation des discontinuités. Par exemple,

$$\mathcal{J}_\phi(x) = \|z - Ax\|^2 + \lambda \int_0^1 \phi(x''(t)) dt \quad (4.9)$$

est une généralisation de (4.3) pour le cas continu, de même que

$$J_\phi(x) = \|z - Ax\|^2 + \frac{\lambda}{M} \sum_{m=1}^{M-1} \phi\left(\frac{2x_m - x_{m-1} - x_{m+1}}{1/M^2}\right), \quad (4.10)$$

pour le cas discret. De façon à préserver les bords entre régions homogènes, donc à autoriser ponctuellement des variations importantes de l'estimée, il est assez clair que ϕ doit croître moins vite qu'une parabole. Le choix de la fonction ϕ et les arguments attachés sont finalement très voisins de l'utilisation des **normes robustes** en statistique [HUBER 1981, REY 1983]. Deux groupes de fonctions ont été principalement proposés dans la littérature :

- les fonctions « L_{21} », *i.e.*, paires, convexes, C^1 , asymptotiquement linéaire, C^2 autour de 0. Un exemple typique et très utilisé est la branche d'hyperbole (Figure 4.6(b))

$$\phi(u) = \sqrt{s^2 + u^2}, \quad s > 0.$$

Le minimiseur de (4.10) qui correspond à ce choix est représenté à la Figure 4.5. Il fournit un assez bon compromis entre restitution des *plateaux* et préservation des frontières.

D'autres familles de fonctions convexes fournissent un résultat qualitativement proche ; par exemple, les normes L_p [BOUMAN et SAUER 1993],

$$\phi(u) = |u|^p, \quad 1 \leq p < 2$$

(qui ne sont pas C^2 autour de 0, et même non différentiable pour $p = 1$, et croissent plus vite que le régime linéaire) ; ou encore, la fonction « équitable » (pour *fair* [REY 1983]), qui, bien que restant convexe, croît moins vite que le régime linéaire :

$$\phi(u) = |u/s| - \ln(1 + |u/s|), \quad s > 0.$$

- les fonctions « L_{20} », *i.e.*, paires, asymptotiquement constantes, C^2 autour de 0. Un exemple typique donné par [GEMAN et MCCLURE 1987] et représenté à la Figure 4.6(c) est

$$\phi(u) = \frac{u^2}{r^2 + u^2}, \quad r > 0.$$

Les fonctions « L_{20} » sont non convexes (sauf les fonctions constantes). En conséquence, la minimisation globale de la fonctionnelle (4.9) n'a pas forcément un sens mathématique, alors que l'équivalent discret, c'est-à-dire la minimisation globale de la fonction (4.10), est correctement défini car posé dans un espace de dimension fini. Le minimiseur de (4.10)

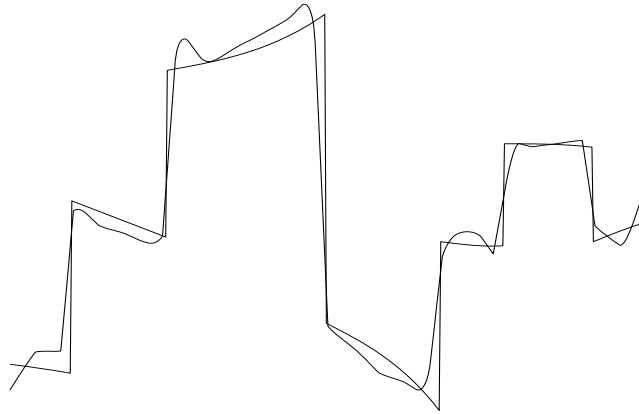


FIG. 4.5: *Estimée douce non linéaire $\hat{\mathbf{x}}^{\lambda,s}$ obtenue comme minimiseur de (4.10), avec $\phi(u) = \sqrt{s^2 + u^2}$ pour des valeurs optimales de λ, s au sens L_1 : $C(\hat{\mathbf{x}}^{\lambda,\alpha}, x^*) = 16,57\%$.*

n'est pas représenté ici ; son comportement est très similaire à celui de la solution obtenue par détection-estimation (Figure 4.3).

D'autres familles de fonctions non convexes sont parfois recommandées dans la littérature ; par exemple, les fonctions paires, concaves et croissantes sur \mathbb{R}_+ , et qui sont donc non dérivables en zéro [GEMAN et REYNOLDS 1992] telles que

$$\phi(u) = \frac{|u|}{r + |u|}, \quad r > 0.$$

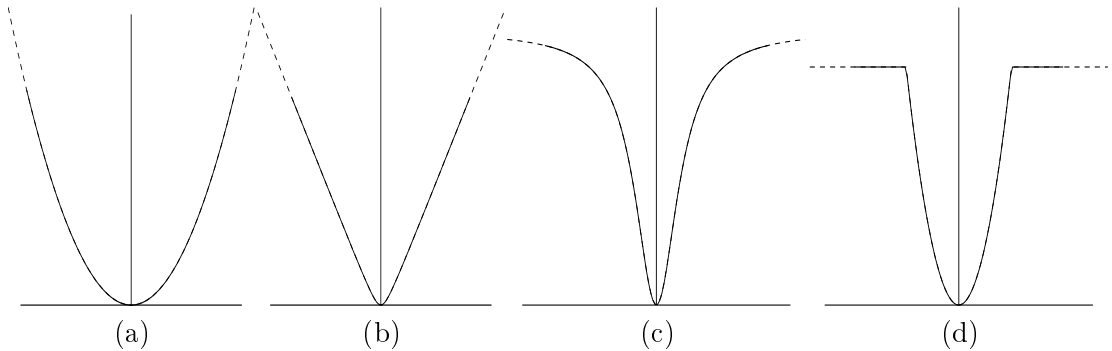


FIG. 4.6: $\phi(u) =$ (a) u^2 , (b) $\sqrt{s^2 + u^2}$, (c) $\frac{u^2}{r^2 + u^2}$, (d) $\min\{u^2, r^2\}$

En pratique, on constate que les deux classes de fonctions ϕ « L_{20} » et « L_{21} » fournissent des solutions nettement différentes en termes de comportement et de coût de calcul.

D'une part, l'approche « L_{21} » assure la convexité des critères \mathcal{J}_ϕ et J_ϕ . Sous des conditions techniques simples, cette propriété permet de définir un minimum global unique pour (4.10) ainsi que pour (4.9). De plus, les critères convexes n'admettent pas de minima locaux, donc la convergence vers un minimiseur global de (4.10) est démontrée pour des algorithmes de minimisation standards (de type gradient ou descente coordonnée par coordonnée par exemple [BERTSEKAS 1995]). Une autre propriété intéressante de la solution ainsi obtenue est sa « stabilité » [BOUMAN et SAUER 1993, KÜNSCH 1994, LI et coll. 1995], c'est-à-dire qu'elle satisfait la troisième condition de Hadamard, contrairement aux solutions obtenues par détection-estimation.

D'autre part, l'approche « L_{20} » partage les caractéristiques de l'approche « détection-estimation » : les frontières restituées sont très franches, la solution est seulement stable par morceaux ; et le coût de calcul d'algorithmes évitant les éventuels minima locaux est important. La similarité des caractéristiques des approches « L_{20} » et « détection-estimation » n'est pas fortuite, ce que met en évidence la section suivante.

4.4 Critères augmentés semi-quadratiques

La manipulation de critères SQ s'est imposée récemment comme un outil numérique puissant pour la restauration d'image avec préservation des contours [CHARBONNIER et coll. 1994, COHEN 1996, CHARBONNIER et coll. 1997, VOGEL et OMAN 1998] ; voir aussi [22]. Comme indiqué ci-dessus dans le cadre de l'approche « détection-estimation », les critères SQ ont d'abord incorporé des variables de bords *exclusivement binaires*, soit interactives soit découplées. D'après [GEMAN et REYNOLDS 1992], la présente section montre que les critères SQ couvrent un domaine beaucoup plus vaste. Plus précisément, de nombreuses approches pénalisées non quadratiques admettent une formulation équivalente dans le cadre SQ. La conséquence est double. Du point de vue formel, cette équivalence permet une meilleure compréhension des véritables choix de modèles de signaux et d'images. Du point de vue pratique, la formulation SQ apporte de nouveaux outils algorithmiques pour la minimisation de critères non quadratiques.

4.4.1 Dualité entre critères non quadratiques et critères augmentés SQ

[BLAKE et ZISSERMAN 1987] ont remarqué qu'un critère SQ à variables de bord découplées (tel que le critère $K(\mathbf{x}, \mathbf{b})$ défini par (4.8)) pouvait être considéré comme un *équivalent augmenté* d'un certain critère non quadratique $J(\mathbf{x})$, au sens où

$$\min_{\mathbf{b} \in \mathcal{B}} K(\cdot, \mathbf{b}) = J,$$

avec $\mathcal{B} = \{0,1\}^{M-1}$. En qualifiant le critère K d'*augmenté*, on souligne qu'il dépend non seulement de \mathbf{x} , mais aussi des variables *auxiliaires* \mathbf{b} , tandis que J ne dépend que de \mathbf{x} .

Plus précisément, d'après (4.7) et (4.8), on a

$$\begin{aligned} \min_{\mathbf{b} \in \{0,1\}^{M-1}} K(\mathbf{x}, \mathbf{b}) &= \|\mathbf{z} - \mathbf{A}\mathbf{x}\|^2 + \sum_{m=1}^{M-1} \min_{b_m \in \{0,1\}} \left(\lambda M^3 b_m |2x_m - x_{m-1} - x_{m+1}|^2 - \alpha b_m \right) \\ &= \|\mathbf{z} - \mathbf{A}\mathbf{x}\|^2 + \sum_{m=1}^{M-1} \min \left\{ 0, \lambda M^3 |2x_m - x_{m-1} - x_{m+1}|^2 - \alpha \right\} \\ &= J_{\phi_r}(\mathbf{x}) + \text{const}, \end{aligned}$$

où J_{ϕ_r} est défini par (4.10) pour $\phi_r(u) = \min\{u^2, r^2\}$, c'est-à-dire pour la fonction *quadratique tronquée* représentée Figure 4.6(d), avec $r = \sqrt{\alpha/\lambda M}$.

En conséquence, $J = J_{\phi_r}$ et K ont même minimum, et surtout même *minimiseur* $\hat{\mathbf{x}}$ sur tout fermé X de \mathbb{R}^M , au sens où

$$\hat{\mathbf{x}} \text{ minimise } J \text{ sur } X \Leftrightarrow \exists \hat{\mathbf{b}} / (\hat{\mathbf{x}}, \hat{\mathbf{b}}) \text{ minimise } K \text{ sur } X \times \{0,1\}^{M-1}.$$

Pour trouver $\hat{\mathbf{x}}$ en pratique, on a donc le choix de s'intéresser à la minimisation de J , ou bien à celle de K . En pratique, Blake & Zisserman définissent K comme critère dans le cadre de l'approche « détection-estimation », déduisent l'expression de J de celle de K , et abordent le

problème de la minimisation du critère non convexe J par *non convexité graduelle*, plutôt que le problème combinatoire équivalent posé par la minimisation de K .

Plus tard, D. Geman et ses collaborateurs [GEMAN et REYNOLDS 1992, GEMAN et YANG 1995] ont généralisé la construction de Blake & Zisserman à des variables auxiliaires découplées non discrètes. Ce faisant, ils renversent l'ordre de construction : ils montrent qu'il existe des critères augmentés semi-quadratiques K pour de larges classes de critères J non quadratiques, au sens où

$$\inf_{\mathbf{b} \in \mathcal{B}} K(\cdot, \mathbf{b}) = J \quad (4.11)$$

pour un ensemble \mathcal{B} à préciser, en général différent de $\{0, 1\}$. Sous une forme étendue ultérieurement ([CHARBONNIER et coll. 1997], [10]), cette construction s'applique aux fonctions ϕ qui satisfont les hypothèses suivantes :

$$\phi \text{ est paire,} \quad (4.12)$$

$$\phi(\sqrt{\cdot}) \text{ est concave sur } \mathbb{R}_+, \quad (4.13)$$

$$\phi \text{ est continue en zéro et } C^1 \text{ sur } \mathbb{R}^*. \quad (4.14)$$

Sous ces conditions, la relation de dualité suivante peut être obtenue à partir de résultats classiques en analyse convexe [ROCKAFELLAR 1970] :

$$\phi(u) = \inf_{b \in \mathbb{R}_+} (bu^2 + \psi(b)), \quad (4.15)$$

où

$$\psi(b) = \sup_{u \in \mathbb{R}} (\phi(u) - bu^2). \quad (4.16)$$

D'après (4.10) et (4.15), il est facile de déduire (4.11) pour le critère augmenté SQ

$$\begin{aligned} K(\mathbf{x}, \mathbf{b}) &= J_{\mathbf{b}}(\mathbf{x}) + \frac{\lambda}{M} \sum_{m=1}^{M-1} \psi(b_m) \\ &= \|\mathbf{z} - \mathbf{A}\mathbf{x}\|^2 + \lambda M^3 \sum_{m=1}^{M-1} b_m |2x_m - x_{m-1} - x_{m+1}|^2 + \frac{\lambda}{M} \sum_{m=1}^{M-1} \psi(b_m). \end{aligned} \quad (4.17)$$

L'équivalence (4.11) possède une extension fonctionnelle dans le cas où ϕ est convexe [AUBERT et VESE 1997].

Qu'elles soient convexes ou non, la plupart des fonctions ϕ proposées dans la littérature satisfont (4.12)-(4.14). C'est le cas des deux fonctions des Figures 4.6(b) et 4.6(c).

- Pour la branche d'hyperbole $\phi(u) = \sqrt{s^2 + u^2}$, $s > 0$, (4.16) donne

$$\psi(b) = \begin{cases} s^2 b + 1/4b & \text{si } b \in]0, b_{\infty} = 1/2s] \\ 1 & \text{si } b \geq b_{\infty} \end{cases}$$

- Pour $\phi(u) = \frac{u^2}{r^2 + u^2}$, $r > 0$, (4.16) donne

$$\psi(b) = \begin{cases} (1 - r\sqrt{b})^2 & \text{si } b \in]0, b_{\infty} = 1/r^2] \\ 1 & \text{si } b \geq b_{\infty} \end{cases}$$

Dans les deux cas, les variables auxiliaires b_m ne sont plus binaires mais continues. Cependant, en pratique, leurs valeurs sont proches de 0 ou de b_{∞} , particulièrement dans le cas non convexe.

4.4.2 Minimisation de critères SQ

Principe de la relaxation

[GEMAN et REYNOLDS 1992] mettent en avant les avantages structurels liés à la semi-quadraticité pour minimiser le critère augmenté K plutôt que J (voir aussi [GEMAN et YANG 1995]). Plus précisément, il s'agit de procéder par *relaxation*, c'est-à-dire en considérant le sous-problème à \mathbf{x} fixé, puis à \mathbf{b} fixé, alternativement jusqu'à convergence. Ces deux sous-problèmes sont simples, car l'un est quadratique, et l'autre est séparable. Les références [GEMAN et REYNOLDS 1992] et [GEMAN et YANG 1995] s'intéressent spécifiquement au cas où ϕ est non convexe, et développent des versions SQ du recuit simulé fondées sur le principe de l'échantillonnage alterné.

Plusieurs autres auteurs [CHARBONNIER et coll. 1994, COHEN 1996, CHARBONNIER et coll. 1997] (voir aussi [22]) s'intéressent au pendant déterministe, bien adapté au cas où ϕ est convexe. Partant d'un couple initial (\mathbf{x}, \mathbf{b}) , il s'agit d'entrelacer minimisation en \mathbf{x} à \mathbf{b} fixé et en \mathbf{b} à \mathbf{x} fixé :

- ① Le critère augmenté $K(\mathbf{x}, \mathbf{b})$ est une fonction quadratique de \mathbf{x} . Calculer le minimiseur en \mathbf{x} à \mathbf{b} fixé est donc un simple problème d'inversion de système linéaire. On peut remarquer que cette étape correspond à une version adaptative de la régularisation de Tikhonov, dans la mesure où un facteur b_m multiplie le paramètre de régularisation devant chaque terme de pénalisation quadratique dans l'expression (4.17).
- ② Le critère $K(\mathbf{x}, \mathbf{b})$ étant une fonction séparable des variables b_m , il peut être minimisé comme une fonction de \mathbf{b} sous forme parallèle. De plus, l'équation de remise à jour pour chaque b_m est explicite : on peut montrer que l'infimum de (4.15) est atteint pour

$$\hat{b}(u) = \begin{cases} b_\infty & \text{si } u = 0, \\ \phi'(u)/2u & \text{sinon,} \end{cases} \quad (4.18)$$

et l'expression de ψ n'est donc pas requise pour exprimer $\hat{b}(u)$.

Cas d'une fonction ϕ convexe

Quand ϕ est convexe, il a été démontré récemment que cette procédure de descente *blocs de coordonnées par blocs de coordonnées* converge vers un unique minimum sous des conditions larges [CHARBONNIER et coll. 1997], [10]. Dans une formulation peu différente, propre à l'estimation robuste, une technique de descente similaire (identifiée comme une méthode de *moindres carrés repondérés*, ou *Reweighted Least Squares* — RLS) et le résultat de convergence associé sont connus depuis plus longtemps [YARLAGADDA et coll. 1985].

Bien que d'autres techniques de descente soient envisageables pour minimiser K sous les mêmes conditions de régularité (par exemple, de type gradient), l'approche RLS a un avantage remarquable : si le problème inverse considéré a été précédemment résolu par pénalisation quadratique de type Tikhonov, les extensions non quadratiques s'obtiennent par adaptation triviale :

- introduire la régularisation adaptative dans la méthode Tikhonov, qui devient ainsi l'étape ① de la méthode RLS (on peut noter ici que cette étape ne dépend pas de l'expression de la fonction non quadratique ϕ) ;
- ajouter l'étape ② correspondant à la remise à jour des variables auxiliaires (4.18) (seule cette équation de remise à jour fait intervenir l'expression de ϕ) ;
- reboucler.

Cas d'une fonction ϕ non convexe

Quand la fonction ϕ n'est pas convexe, la technique RLS converge vers un des minima de J , local ou global [DELANEY et BRESLER 1998], comme d'autres algorithmes simples de descente déterministe. Le recours à des méthodes numériquement plus lourdes comme le recuit simulé ou le GNC est alors nécessaire.

De même qu'il existe des versions du recuit simulé qui tirent parti de la semi-quadraticité du critère augmenté [GEMAN et REYNOLDS 1992, GEMAN et YANG 1995], on peut mentionner ici l'intérêt d'une version RLS du GNC (utilisée dans [88] mais non publiée). Soit deux fonctions ϕ_0 et ϕ_1 , qui sont respectivement « L_{20} » et « L_{21} ». Posons $\phi_\theta = (1 - \theta)\phi_0 + \theta\phi_1$, et considérons que θ passe par petits incréments de un à zéro. On obtient alors une propriété intéressante : dans l'étape ② du RLS, l'expression de la remise à jour (4.18) fondée sur ϕ_θ est une simple combinaison linéaire des expressions valables pour ϕ_0 et ϕ_1 respectivement (alors que la fonction duale ψ_θ n'est pas une combinaison linéaire de ψ_0 et de ψ_1). Finalement, l'étape ① ne dépendant pas de ϕ_θ , on obtient ainsi une version RLS du GNC de façon immédiate.

4.5 Conclusion

Dans le cadre des méthodes pénalisées pour la restauration d'objets homogènes par morceaux, trois types d'approches ont été mis en évidence. Par ordre de complexité croissante :

- L'approche originale de Tikhonov correspond à la pénalisation *quadratique*. Elle est à la fois la plus ancienne et la moins coûteuse en temps de calcul, puisqu'elle fournit une solution par simple inversion d'un système linéaire. Sauf en situation favorable (données peu bruitées et très redondantes), cette solution n'est pas satisfaisante pour restaurer des objets homogènes par morceaux. Dans l'exemple illustrant ce chapitre, cette méthode atteint au mieux une norme d'erreur peu compétitive, égale à 18,16% (Figure 4.2).
- La pénalisation *convexe non quadratique* fournit des estimateurs robustes qui manifestent une bien meilleure capacité de préservation des frontières entre régions. La norme d'erreur atteinte dans l'exemple devient égale à 16,57% (Figure 4.5). De plus ces estimateurs peuvent être calculés itérativement par des méthodes à point fixe de type gradient. Dans ce chapitre, on a mis l'accent sur une autre méthode de descente à point fixe, de type *moindres carrés repondérés*, dont chaque itération requiert l'inversion d'un système linéaire de type Tikhonov. Formellement, cette méthode s'identifie à une méthode de descente *bloc de coordonnées par blocs de coordonnées* sur une forme augmentée semi-quadratique du critère de départ.
- L'approche *détection-estimation* constitue la formulation semi-quadratique équivalente pour une approche pénalisée non convexe (plus précisément, utilisant des fonctions de pénalisation quadratiques tronquées). Elle est la seule qui possède une réelle capacité de détection des frontières entre régions distinctes. Dans l'exemple illustrant ce chapitre, cette capacité a permis de diminuer légèrement la norme d'erreur : 16,22% (Figure 4.3) au lieu de 16,57% pour la pénalisation convexe. Cependant, cette capacité est obtenue pour un coût de calcul accru. De plus, le réglage des hyperparamètres, qu'il soit empirique ou automatique, est rendu plus difficile par l'instabilité de la solution (de même qu'elle est une fonction discontinue des données, elle n'est en général pas continue en fonction des hyperparamètres).

Il est intéressant de terminer par un constat « historique » : depuis les travaux pionniers des années 60 et 70 [PHILLIPS 1962, TWOMEY 1962, TIKHONOV et ARSÉNINE 1976, HUNT 1977], l'évolution des méthodes couramment étudiées et appliquées en restauration de signaux et d'images n'a pas toujours suivi un ordre de complexité croissant. En effet, les méthodes quadratiques

(filtrage de Wiener avec mise en œuvre dans le domaine de Fourier [HUNT 1977], filtrage de Kalman 1D [DEMOMENT et REYNAUD 1985] et 2D [SAINT-FÉLIX 1987]) ont cédé le pas, après les travaux des frères Geman [GEMAN et GEMAN 1984], à des méthodes numériquement beaucoup plus coûteuses, fondées sur des variables cachées discrètes et adoptant le recuit simulé comme technique d'optimisation. Le recours à la minimisation de critères *convexes non quadratiques*, beaucoup plus économique et déjà présent sous des formes spécifiques dans les années 80 (par exemple dans le cadre du maximum d'entropie [MOHAMMAD-DJAFARI et DEMOMENT 1988]), devient prépondérant dans le courant des années 90 seulement, à la suite de travaux tels que [SAUER et BOUMAN 1993].

Chapitre 5

Modèles et outils stochastiques

5.1 Introduction

Dans le chapitre précédent, j'ai présenté une classe importante d'outils pour l'inversion de problèmes inverse mal posés : celle des solutions construites comme minimiseurs de critères pénalisés. Pour un modèle direct $\mathbf{z} = A(x^*) + \text{bruit}$, l'adoption de cette approche conduit au choix préalable

- d'une fonction mesurant l'adéquation aux données $\Omega(\mathbf{z} - A(x))$,
- d'une fonction régularisante $\Phi(x)$,
- d'une valeur d'hyperparamètre λ gérant le compromis entre régularité et adéquation aux données.

Sous réserve d'existence et d'unicité, la solution est ensuite définie comme le minimiseur d'un critère

$$\mathcal{J}(x) = \Omega(\mathbf{z} - A(x)) + \lambda\Phi(x). \quad (5.1)$$

Le choix des fonctions Ω et Φ et de la valeur de λ sont des questions d'autant plus délicates que les arguments objectifs suffisent rarement pour trancher cette question. Seul le choix de Φ a été discuté dans le chapitre précédent, essentiellement sur la base d'arguments qualitatifs : tel signal à estimer est « doux », tel autre est « homogène par morceaux », etc. Or l'un des objectifs de la modélisation statistique est précisément l'appréhension d'informations vagues ou incertaines. De fait, de nombreuses contributions dans le domaine de l'inversion en imagerie s'inscrivent explicitement dans un cadre statistique [GEMAN et GEMAN 1984, BESAG 1986, DEMOMENT 1989, BOUMAN et SAUER 1993]. Il est donc légitime de présenter les grandes lignes de cette approche, et de voir plus particulièrement dans quelle mesure elle apporte des réponses à des questions difficiles à trancher dans le cadre déterministe du chapitre précédent, tel que le choix de Ω , Φ et λ . La construction structurelle des fonctions Ω et Φ dans un cadre statistique est envisagée au Sous-chapitre 5.3. Dans ce même cadre, le réglage non supervisé (ou encore, *autodidacte*, c'est-à-dire à partir du seul jeu de données disponible \mathbf{z}) du paramètre λ et d'autres hyperparamètres est abordé au Sous-chapitre 5.4.

5.2 Cadre statistique

Proposer un modèle probabiliste pour un signal consiste à supposer que ce signal est la réalisation d'un processus aléatoire, dans l'espoir de mieux structurer les informations disponibles en les traduisant sous forme de caractéristiques de ce processus. A proprement parler, il ne s'agit donc pas, comme on le dit souvent, de supposer que le signal lui-même est aléatoire.

A certaines conditions, le cadre statistique *bayésien* est approprié pour interpréter un critère tel que (5.1). Adopter ce cadre consiste à supposer simultanément que les incertitudes (le *bruit*) et le signal x^* sont issus de quantités aléatoires¹. Supposons qu'observation, bruit et signal sont des vecteurs réels², respectivement $\mathbf{z} \in \mathbb{R}^N$, $\mathbf{n} \in \mathbb{R}^N$ et $x^* = \mathbf{x}^* \in \mathbb{R}^M$:

$$\mathbf{z} = A(\mathbf{x}^*) + \mathbf{n}. \quad (5.2)$$

Supposons également que les fonctions Φ et Ω vérifient la condition de normalisation

$$\exists T_x \in \mathbb{R}_+^*, \quad Z_x(T_x) = \int_{\mathbb{R}^M} e^{-\Phi(\mathbf{x})/T_x} d\mathbf{x} < +\infty, \quad (5.3)$$

$$\exists T_n \in \mathbb{R}_+^*, \quad Z_b(T_n) = \int_{\mathbb{R}^N} e^{-\Omega(\mathbf{n})/T_n} d\mathbf{n} < +\infty \quad (5.4)$$

(T_x et T_n sont des paramètres dits de *température*, dont le rôle apparaîtra dans la suite). On peut alors définir des fonctions qui ont les caractéristiques de densités de probabilité (mesurables, à valeurs réelles positives ou nulles, intégrables de somme unité) :

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}; T_x) = \frac{1}{Z_x(T_x)} e^{-\Phi(\mathbf{x})/T_x}, \quad (5.5)$$

$$f_{\mathbf{N}}(\mathbf{n}) = f_{\mathbf{N}}(\mathbf{n}; T_n) = \frac{1}{Z_b(T_n)} e^{-\Omega(\mathbf{n})/T_n}. \quad (5.6)$$

Supposer que signal et bruit sont des réalisations de vecteurs aléatoires \mathbf{X} et \mathbf{N} , indépendants de densités respectives $f_{\mathbf{X}}$ et $f_{\mathbf{N}}$, permet de poser la construction de solutions approchant \mathbf{x}^* dans le cadre de l'inférence statistique. On appelle alors $f_{\mathbf{X}}$ la loi *a priori*,

$$f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) = f_{\mathbf{N}}(\mathbf{z} - A(\mathbf{x})) \quad (5.7)$$

la vraisemblance des données connaissant \mathbf{x} , et

$$f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}) = \frac{f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{Z}}(\mathbf{z})} = \frac{f_{\mathbf{X},\mathbf{Z}}(\mathbf{x},\mathbf{z})}{\int_{\mathbb{R}^M} f_{\mathbf{X},\mathbf{Z}}(\mathbf{x},\mathbf{z}) d\mathbf{x}}, \quad (5.8)$$

la vraisemblance *a posteriori*. La loi jointe du couple (\mathbf{X}, \mathbf{Z}) :

$$f_{\mathbf{X},\mathbf{Z}}(\mathbf{x},\mathbf{z}) = f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})f_{\mathbf{X}}(\mathbf{x}) \quad (5.9)$$

joue un rôle central pour deux raisons : d'une part, elle réalise formellement la fusion des deux sources d'information, connaissance *a priori* et information attachée aux données. D'autre part, on peut mathématiquement déduire de la loi jointe $f_{\mathbf{X},\mathbf{Z}}$ toutes les autres lois, marginales telles que $f_{\mathbf{X}}$, $f_{\mathbf{N}}$ et $f_{\mathbf{Z}}$ et conditionnelles telles que $f_{\mathbf{Z}|\mathbf{X}}$ et $f_{\mathbf{X}|\mathbf{Z}}$; en ce sens, cette loi résume toute

1. Dans le cadre statistique *non bayésien*, (5.1) est interprétable comme une anti-log-vraisemblance pénalisée; seul le premier terme est alors supposé correspondre à un modèle probabiliste (pour le bruit), tandis que le second est introduit de façon *ad hoc*.

2. Les cas pour lesquels signal ou observation appartiennent à un espace de dimension infinie ne se prêtent que partiellement à une formulation statistique. En particulier, l'approche de Tikhonov est interprétable dans le cadre statistique bayésien, même sans discrétisation [FRANKLIN 1970]; voir aussi [9]. En revanche, dans le cas de fonctionnelles Ω ou Φ non quadratiques, la minimisation du critère (5.1) n'a pas d'interprétation statistique systématique. Cette restriction limite la portée des outils introduits dans ce chapitre : par exemple, dans le cas d'un problème inverse dans un espace de dimension infinie, discrétisé sur un nombre de points fini variable, la signification et le comportement du maximum de vraisemblance $\hat{\theta}^{\text{MV}}$ du Sous-chapitre 5.4.2 est inconnu quand le nombre de points tend vers l'infini pour approcher le problème initial.

Cadre énergétique		Cadre probabiliste
Énergie $\Phi(\mathbf{x})$	\leftrightarrow	Densité de probabilité $f_{\mathbf{X}}(\mathbf{x}) = \frac{\exp\{-\Phi(\mathbf{x})/T\}}{\int_{\mathbf{x}} \exp\{-\Phi(\mathbf{x})/T\} d\mathbf{x}}$
Énergie quadratique $\Phi(\mathbf{x}) = \ \mathbf{x} - \mathbf{m}\ _{\mathbf{W}}^2$	\leftrightarrow	Loi gaussienne : $\mathcal{N}(\mathbf{m}, \frac{T}{2}\mathbf{W}^{-1})$
Énergie séparable : $\Phi(\mathbf{x}) = \sum_{s \in \mathcal{S}} G_s(x_s)$	\leftrightarrow	Bruit blanc : $f_{\mathbf{X}}(\mathbf{x}) = \prod_{s \in \mathcal{S}} f_s(x_s)$
Énergie de Gibbs : $\Phi(\mathbf{x}) = \sum_{c \in \mathcal{C}} G_c(\mathbf{x})$ $c \in \mathcal{C} \subset \mathcal{P}(\mathcal{S})$ clique, G_c potentiel associé (Théorème de Hammersley-Clifford)	\leftrightarrow	Modèle markovien : $f(x_s x_r, r \in \mathcal{S}) = f(x_s x_r, r \in \mathcal{V}_s),$ $\mathcal{V}_s = \{r \in \mathcal{C} \ni s, r \neq s\}$
$\mathcal{J}(\mathbf{x}) = \Omega(\mathbf{z} - \mathbf{A}\mathbf{x}) + \lambda\Phi(\mathbf{x})$	\leftrightarrow	$f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}, \mathbf{z}) = f_{\mathbf{Z} \mathbf{X}}(\mathbf{z} \mathbf{x}) f_{\mathbf{X}}(\mathbf{x})$ (règle de Bayes)
Solution pénalisée $\arg \min_{\mathbf{x}} \mathcal{J}(\mathbf{x})$	\equiv	Maximum <i>a posteriori</i> $\hat{\mathbf{x}}^{\text{MAP}} = \arg \max_{\mathbf{x}} f_{\mathbf{X} \mathbf{Z}}(\mathbf{x} \mathbf{z})$

TAB. 5.1: La minimisation de critères pénalisés possède une interprétation probabiliste bayésienne. Ce même mécanisme permet de définir des distributions de probabilité à partir de fonctions d'énergie, et inversement.

l'information statistique. Le maximiseur de la vraisemblance *a posteriori* est l'estimateur du maximum *a posteriori*, estimateur bayésien le plus couramment utilisé en imagerie :

$$\begin{aligned}
 \hat{\mathbf{x}}^{\text{MAP}} &\stackrel{\Delta}{=} \arg \max_{\mathbf{x}} f_{\mathbf{X} | \mathbf{Z}}(\mathbf{x} | \mathbf{z}) \\
 &= \arg \max_{\mathbf{x}} f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}, \mathbf{z}) && \text{d'après (5.8)} \\
 &= \arg \max_{\mathbf{x}} f_{\mathbf{N}}(\mathbf{z} - \mathbf{A}(\mathbf{x})) f_{\mathbf{X}}(\mathbf{x}) && \text{d'après (5.9) et (5.7)} \\
 &= \arg \min_{\mathbf{x}} \Omega(\mathbf{z} - \mathbf{A}(\mathbf{x}))/T_n + \Phi(\mathbf{x})/T_x && \text{d'après (5.5) et (5.6).}
 \end{aligned} \tag{5.10}$$

En conséquence, l'estimateur du maximum *a posteriori* se confond avec le minimiseur du critère pénalisé (5.1) pour $\lambda = T_n/T_x$. Sous les conditions techniques permettant ce développement (c'est-à-dire, principalement, que le problème fasse intervenir des variables en nombre fini), il est donc manifeste que le cadre bayésien donne un sens statistique à la minimisation de critères pénalisés. Inversement, on pourrait dire que le résultat (5.10) donne une lecture déterministe d'un estimateur probabiliste maximum *a posteriori*³. Une fois défini, l'estimateur ne dépend plus du

3. D'autres estimateurs probabilistes classiques, tel que le *maximum a posteriori marginal*, ou encore la *moyenne*

cadre formel qui l'a engendré, pas plus que les moyens numériques à mettre en œuvre pour le calculer.

5.3 Constructions statistiques des fonctions Ω et Φ

Dans certaines situations pratiques, des modèles statistiques du bruit ou de l'objet inconnu sont déjà présents dans la mise en équation des phénomènes physiques observés. Par exemple, pour des mesures de comptage en imagerie corpusculaire (imageur CCD, tomographie à rayons X, tomographie à émission de positons), le choix de distributions de Poisson d'intensité liée à l'objet examiné est physiquement justifiée. En dehors de ce cadre restreint, les modèles statistiques ne préexistent pas à la formulation du problème d'inférence (trouver \mathbf{x} connaissant \mathbf{z}) : il s'agit de modèles « exogènes », construits spécifiquement pour traduire les informations disponibles à la mise en place de la procédure de traitement des données.

De nombreux travaux, principalement inspirés par ceux de E.T. JAYNES tel que [JAYNES 1968], ont pour idée-force que le cadre bayésien permet de construire ces modèles sous forme de lois de probabilités $f_{\mathbf{X}}$ et $f_{\mathbf{N}}$ de façon raisonnée, voire objective. Les actes publiés du congrès annuel *Maximum Entropy and Bayesian Methods* ont développé ce thème en traitement de signal depuis 1979. Plusieurs contributions du L2S, principalement de Ali MOHAMMAD-DJAFARI, se rangent sous cette bannière [MOHAMMAD-DJAFARI et DEMOMENT 1988, MOHAMMAD-DJAFARI et IDIER 1990, MOHAMMAD-DJAFARI 1992, MOHAMMAD-DJAFARI 1994] ; voir aussi [67, 69].

Il existe trop de variantes dans l'utilisation des principes informationnels qui sous-tendent ces travaux pour en faire une présentation exhaustive. Il s'agit principalement de méthodes dites « à maximum d'entropie » dans lesquelles on recherche une distribution de probabilité qui soit la plus proche d'une distribution de référence (au sens d'une distance de Kullback) tout en satisfaisant des contraintes qui traduisent une information incomplète connue *a priori*.

Dans certains cas, ces contraintes concernent l'invariance de la loi ou de la famille de lois par des transformations « naturelles » dans le contexte, telle qu'un changement d'échelle [67, 69]. Mes travaux de thèse [7] ou des travaux postérieurs tels que [1, 5] ont été faits dans cet esprit : réduire les degrés de liberté dans le choix de lois *a priori* par la prise en compte de propriétés probabilistes : stationnarité, markoviannité et lois marginales imposées dans le premier cas, stationnarité et markoviannité dans le second. Ces constructions sont formellement séduisantes mais de portée pratique limitée, car peu de problèmes de traitement de signal ou d'imagerie se prêtent à la formulation de telles contraintes.

Dans [MOHAMMAD-DJAFARI et DEMOMENT 1988, MOHAMMAD-DJAFARI et IDIER 1990, MOHAMMAD-DJAFARI 1992, MOHAMMAD-DJAFARI 1994], on s'appuie plutôt sur la connaissance d'espérances sur la loi cherchée, en exploitant la propriété constructive suivante [JAYNES 1968].

Propriété 1 La densité $f_{\mathbf{X}}$ qui maximise l'entropie

$$S(f_{\mathbf{X}}) = - \int_{\mathbf{x}} f_{\mathbf{X}}(\mathbf{x}) \ln \frac{f_{\mathbf{X}}(\mathbf{x})}{\mu(\mathbf{x})} d\mathbf{x}, \quad (5.11)$$

par rapport à la « mesure de référence » μ , étant donné les contraintes

$$E[\Phi_k(\mathbf{X})] = \varphi_k, \quad k = 1, \dots, K, \quad (5.12)$$

correspond à une loi exponentielle

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{Z} \mu(\mathbf{x}) e^{-\Phi(\mathbf{x})}, \quad \text{avec} \quad \Phi(\mathbf{x}) = \sum_{k=1}^K \lambda_k \Phi_k(\mathbf{x}) \quad \text{et} \quad Z = \int_{\mathbf{x}} \mu(\mathbf{x}) e^{-\Phi(\mathbf{x})} d\mathbf{x}.$$

a posteriori, ne possèdent pas d'interprétation déterministe « naturelle » ; le recours à ces estimateurs est envisagé

Les paramètres λ_k sont déterminés de façon unique par le respect des contraintes (5.12), qui peuvent encore s'écrire :

$$\frac{\partial Z}{\partial \lambda_k} = \varphi_k, \quad k = 1, \dots, K.$$

Beaucoup de lois (5.5) utilisées comme *a priori* pour la reconstruction de signaux et d'images sont extraites de familles exponentielles. Il est donc tentant d'« objectiviser » ce choix en considérant que la Propriété 1 est une justification apportée par la théorie de l'information. Malheureusement, alors que les contraintes de type (5.12) jouent un rôle privilégié en physique statistique en tant qu'*observables*, elles ne correspondent que rarement à des informations disponibles en imagerie. En l'absence de ces contraintes, l'appellation « méthode à maximum d'entropie » est abusive pour une méthode calculant (5.10). A la suite d'articles fondateurs tels que [GULL et DANIELL 1978], l'imagerie astronomique a constitué un domaine de prédilection pour les méthodes dites « à maximum d'entropie ». Sur les vertus supposées des principes informationnels dans ce domaine, je partage plutôt le point de vue pragmatique de [NITYANANDA et NARAYAN 1982].

Faute de connaître les quantités φ_k , on peut vouloir les remplacer par des estimateurs empiriques⁴ ; mais quelle que soit l'efficacité pratique de ce palliatif, il retire tout aspect constructif ou informationnel à la spécification de Φ , car il ne fait que reporter le problème du choix : pour quelles fonctions Φ_k va-t-on estimer empiriquement $E[\Phi_k(\mathbf{X})]$?

En conclusion, les constructions axiomatiques de type informationnel sont séduisantes mais trop spécifiques pour se substituer aux constructions faites « à la main ». Dans ces dernières, l'interprétation probabiliste ne joue pas un rôle crucial. Plus particulièrement, la plupart des modèles de Gibbs-Markov utilisés en imagerie sont construits par addition de potentiels formant une énergie de Gibbs. Pour cette construction, le cadre déterministe suffit. Le cadre probabiliste permet en outre de caractériser les modèles de Gibbs-Markov par une propriété d'indépendance conditionnelle, mais cette propriété est *non constructive*, c'est-à-dire qu'elle ne permet pas la spécification effective de modèles⁵ [BESAG 1974].

Ce constat n'est pas surprenant, dans la mesure où les informations *a priori* rencontrées en pratique sont surtout qualitatives : leur prise en compte ne rentre pas dans les théorèmes d'optimalité de la théorie de l'information ; elle relève plutôt du savoir-faire. L'approche empirique que je défends renonce d'avance à la quête (utopique ?) de « la méthode objective ». Dans une science appliquée comme le traitement de signal, la comparaison impartiale des résultats produits par des méthodes concurrentes, au sens d'un protocole prédéfini, me paraît constituer une meilleure garantie de pertinence par rapport au problème posé.

5.4 Réglage autodidacte des hyperparamètres

Dans cette section, θ désigne l'ensemble des hyperparamètres, c'est-à-dire des paramètres dont dépend l'estimateur $\hat{\mathbf{x}}^{\text{MAP}}$ défini par (5.10), ou, plus généralement, dont dépendent les estimateurs $\hat{\mathbf{x}}$ construits à partir de la loi jointe (5.9) : paramètres de température T_x et T_n déterminant le coefficient de régularisation $\lambda = T_n/T_x$, ainsi que tout autre degré de liberté paramétrant Φ ou Ω . Dans la suite, on fait apparaître explicitement cette dépendance par la notation $\hat{\mathbf{x}}^\theta$.

De même qu'il est souvent possible de calibrer les systèmes de télécommunication sur une séquence d'apprentissage, les hyperparamètres des méthodes d'inversion peuvent être réglés sur des signaux ou des images tests dans un grand nombre de situations. Mais ce réglage « par

4. C'est, par exemple, le schéma suivant lequel l'analyse spectrale autorégressive est dite « à maximum d'entropie » ; voir la page 16 à ce propos.

5. A l'exception notable des champs unilatéraux ; voir pages 25–27.

apprentissage » n'est pas toujours possible, par exemple si les caractéristiques de l'opérateur d'observation varient au cours du temps, ou bien si les caractéristiques de l'objet à restaurer sont trop mal connues. Dans ces cas difficiles, on peut envisager le « réglage autodidacte » de θ , c'est-à-dire le calcul d'un estimateur $\hat{\theta}$ (fonction des seules observations disponibles \mathbf{z}), qui permette un fonctionnement automatique de la procédure d'estimation par le calcul de $\hat{\mathbf{x}}_{\hat{\theta}}$.

A la différence d'un article de synthèse tel que [TITTERINGTON 1985], la présente section ne cherche pas à dresser un panorama exhaustif des méthodes envisageables, mais plutôt de présenter la problématique d'une façon cohérente, tout en insistant sur les aspects qui me semblent importants et qui sont parfois négligés dans la littérature existante.

5.4.1 Outils « déterministes » : courbe en L, validation croisée

Les méthodes « déterministes » d'estimation d'hyperparamètres sont peu nombreuses. On peut mentionner la méthode de la « courbe en L » de [HANSEN 1992], qui cherche la zone de courbure maximale sur la courbe paramétrée par λ ($x(\lambda) = \Omega(\mathbf{z} - A(\hat{\mathbf{x}}_\lambda))$, $y(\lambda) = \Phi(\hat{\mathbf{x}}_\lambda)$), tracée en échelle logarithmique.

Par un certain degré de généralité formelle et surtout par leurs résultats pratiques en imagerie dans le cas linéaire quadratique (A linéaire, Ω et Φ quadratiques) [THOMPSON et coll. 1991, FORTIER et coll. 1993], les méthodes de *validation croisée* se distinguent. Dans le principe de la validation croisée *ordinaire*, on s'intéresse aux estimateurs $\hat{\mathbf{x}}_\theta^{-k}$ obtenus avec toutes les données sauf la k^e , et plus spécifiquement à leur capacité de prédire la donnée manquante. Un critère sur la norme du résidu définit mathématiquement l'estimateur :

$$\hat{\theta}^{\text{vco}} \text{ minimise } \mathcal{V}(\theta) = \sum_k \left(z_k - [A(\hat{\mathbf{x}}_\theta^{-k})]_k \right)^2.$$

Malheureusement, en dehors des cas pour lesquels on possède une expression explicite pour les $\hat{\mathbf{x}}_\theta^{-k}$, le calcul de $\hat{\theta}^{\text{vco}}$ est trop coûteux pour être mis en pratique. Seul le cas linéaire quadratique a été exploré dans la littérature ; dans ce cas, on possède une expression explicite pour \mathcal{V} , et il est apparu préférable de remplacer $\hat{\theta}^{\text{vco}}$ par une variante appelée validation croisée *généralisée*, sur la base d'arguments statistiques non bayésiens consistant à voir $\mathcal{V}(\theta)$ comme une approximation de $\mathbb{E}_{\mathbf{Z}} [A(\mathbf{x}^*) - A(\hat{\mathbf{x}}^\theta)]$ [GOLUB et coll. 1979].

A ma connaissance, le comportement des estimateurs de validation croisée ordinaire n'est pas connu dans un contexte non quadratique tel que celui de la section (4.3) du chapitre précédent, pour les raisons pratiques évoquées ci-dessus. Quant aux estimateurs de validation croisée généralisée, ils ne sont pas définis en dehors du contexte linéaire quadratique.

5.4.2 Outils probabilistes : maximisation de vraisemblance

Dans le cadre bayésien, l'outil de prédilection pour inférer une valeur plausible de θ est la vraisemblance des données

$$f_{\mathbf{Z}}(\mathbf{z} ; \theta) = \int_{\mathbb{R}^M} f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}, \mathbf{z}) d\mathbf{x} = \int_{\mathbb{R}^M} f_{\mathbf{N}}(\mathbf{z} - A(\mathbf{x})) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}. \quad (5.13)$$

Plus précisément, on définit l'estimateur du **maximum de vraisemblance** $\hat{\theta}^{\text{mv}}$ comme le maximiseur de $f_{\mathbf{Z}}(\mathbf{z} ; \theta)$. Il existe d'autres formes voisines mais distinctes de maximisation de vraisemblance. Deux sont introduites brièvement ci-dessous ; le reste de ce chapitre est consacré au maximum de vraisemblance lui-même.

Maximum de vraisemblance généralisé

Le *maximum de vraisemblance généralisé* (MVG) est défini comme le maximiseur $(\hat{\mathbf{x}}, \hat{\theta})^{\text{MVG}}$ de la loi jointe $f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}, \mathbf{z}; \theta)$ en fonction de (\mathbf{x}, θ) . La recherche de $(\hat{\mathbf{x}}, \hat{\theta})^{\text{MVG}}$ est souvent mise en œuvre par relaxation : la maximisation en \mathbf{x} à θ fixé est une étape de calcul de $\hat{\mathbf{x}}^{\text{MAP}}$ supervisé, et la maximisation en θ à \mathbf{x} fixé est une étape de maximisation de vraisemblance pour un modèle observé directement. Mais le MVG a deux défauts. Le premier et le plus souvent mis en avant concerne ses mauvaises propriétés statistiques asymptotiques [LITTLE et RUBIN 1983] (à ce propos, voir page 17). Le second est plus concret et plus spécifique : s'il existe (\mathbf{x}_0, θ_0) tel que

$$\lim_{\theta^x \rightarrow \theta_0} f_{\mathbf{X}}(\mathbf{x}; \theta) = \delta(\mathbf{x} - \mathbf{x}_0) \text{ et } f_{\mathbf{N}}(\mathbf{z} - A(\mathbf{x}_0); \theta_0) > 0,$$

la vraisemblance généralisée n'est pas bornée supérieurement et tend vers l'infini pour un couple (\mathbf{x}_0, θ_0) ne dépendant pas de \mathbf{z} . Ce cas de dégénérescence se produit en particulier pour (5.5) quand $T_x \rightarrow 0$, avec \mathbf{x}_0 qui maximise la loi *a priori* $f_{\mathbf{X}}(\mathbf{x})$. On rencontre cette situation dans plusieurs articles, à l'insu de leurs auteurs [LAKSHMANAN et DERIN 1989] ou non [GASSIAT et coll. 1992]. Dans d'autres cas, par exemple dans [26], la dégénérescence ne se produit pas. Le MVG est alors un estimateur dont le coût de calcul est faible, et dont l'utilisation est surtout à envisager à faible nombre de données.

Approche bayésienne hiérarchique

Muni d'un *a priori* f_{θ} pour θ , on peut intégrer θ hors du problème pour obtenir une vraisemblance *a posteriori* ne dépendant plus d'aucun hyperparamètre [MOLINA 1994]. Dans ce contexte « complètement bayésien », la loi jointe (5.9) s'interprète comme la loi conditionnelle $f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}, \mathbf{z} | \theta)$, et on obtient par marginalisation une nouvelle loi jointe *sans hyperparamètre*.

$$f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}, \mathbf{z}) = \int_{\theta} f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}, \mathbf{z} | \theta) f_{\theta}(\theta) d\theta. \quad (5.14)$$

En l'absence d'information qualitative ou quantitative sur θ , le choix de f_{θ} est une question sensible, d'autant qu'il n'est pas question de faire dépendre f_{θ} de nouveaux hyperparamètres. Les choix privilégiés sont des lois « non informatives », c'est-à-dire les plus uniformes possibles, éventuellement impropres, tout en générant une loi propre $f_{\mathbf{X}, \mathbf{Z}}$ par (5.14).

L'apport pratique de ce procédé me semble en fait assez modeste, dans la mesure où la loi jointe (5.14) finalement obtenue est celle qui aurait été fournie directement sous la forme (5.9), après intégration préalable des hyperparamètres hors de la loi *a priori* et de la loi du bruit, ou bien, plus directement, en choisissant directement des lois $f_{\mathbf{X}}$ et $f_{\mathbf{N}}$ sans hyperparamètres. Autrement dit, cette approche ne produit que des résultats qui peuvent être obtenus en interdisant d'emblée la présence de paramètres dans la construction des lois $f_{\mathbf{X}}$ et $f_{\mathbf{N}}$.

Maximum de vraisemblance

Il est communément admis que $\hat{\theta}^{\text{MV}}$ fournit une solution pertinente au problème du réglage autodidacte. D'autre part, il est également avéré que le calcul de cet estimateur est coûteux en imagerie quand \mathbf{X} est un champ de Markov (sauf quand la loi $f_{\mathbf{Z}}$ est gaussienne). Les progrès dans le domaine des échantillonneurs stochastiques et l'augmentation de la puissance de calcul ont permis récemment la diffusion de premiers résultats pratiques [MORRIS et coll. 1996, HIGDON et coll. 1997, SAQUIB et coll. 1998, JALOBÉANU et coll. 1998, DESCOMBES et coll. 1999], mais le comportement du maximum de vraisemblance reste en fait mal connu. Plutôt que d'aborder les aspects algorithmiques liés au calcul de $\hat{\theta}^{\text{MV}}$, mon objectif est précisément de discuter de

la pertinence de l'estimateur lui-même, problème essentiel, souvent masqué par les difficultés calculatoires. Dans cet esprit, introduisons d'abord la notion de *meilleur réglage*.

Définition 1 (meilleur réglage pour les hyperparamètres) *Étant donné un vecteur à estimer \mathbf{x}^* à partir d'observations \mathbf{z} , et un estimateur paramétré $\hat{\mathbf{x}}^\theta(\mathbf{z})$, on appelle meilleur réglage pour θ au sens d'une fonction de coût $C(\mathbf{x}, \mathbf{x}^*)$ toute valeur de θ minimisant $C(\hat{\mathbf{x}}^\theta, \mathbf{x}^*)$.*

En admettant implicitement que le meilleur réglage est unique, on le note θ^{MR} dans la suite. Évidemment, θ^{MR} dépend non seulement de \mathbf{z} à travers $\hat{\mathbf{x}}^\theta = \hat{\mathbf{x}}^\theta(\mathbf{z})$, mais aussi de \mathbf{x}^* : θ^{MR} est calculable seulement dans la situation d'un problème simulé, pour lequel \mathbf{x}^* est connu. Ainsi, pour comparer plusieurs estimateurs concurrents dans le chapitre précédent, on a choisi $C(\mathbf{x}, \mathbf{x}^*) = \|\mathbf{x} - \mathbf{x}^*\|_1$, et chaque estimateur a été ajusté à son meilleur réglage étant donné \mathbf{x}^* .

Peut-on espérer que $\hat{\theta}^{\text{MV}}$ approche θ^{MR} pour une certaine fonction de coût ? Dans un contexte restreint bien précis, on peut répondre par l'affirmative.

Définition 2 (problème inverse synthétique) *Appelons problèmes inverses synthétiques les problèmes d'estimation de \mathbf{x}^* à partir d'observations générées par (5.2), dans lesquels bruit \mathbf{n} et « vrai » objet \mathbf{x}^* ont été simulés indépendamment suivant leurs lois respectives $f_{\mathbf{N}}$ et $f_{\mathbf{X}}$, pour une certaine valeur θ^* de θ .*

Dans le cas d'un problème inverse synthétique, on peut prévoir (et vérifier expérimentalement) la double identité

$$\hat{\theta}^{\text{MV}} \simeq \theta^{\text{MR}} \simeq \theta^*, \quad (5.15)$$

aux trois conditions suivantes.

Condition 1 *Parmi tous les estimateurs de \mathbf{X}^* , $\hat{\mathbf{x}}^{\theta^*}(\cdot)$ est celui qui minimise le coût moyen⁶ $\mathbb{E}_{\theta^*}[C(\hat{\mathbf{x}}(\mathbf{Z}), \mathbf{X}^*)]$.*

Condition 2 *On dispose d'un théorème asymptotique affirmant la convergence du maximum de vraisemblance pour un nombre de données tendant vers l'infini :*

$$\hat{\theta}^{\text{MV}}(\mathbf{Z}) \xrightarrow{N \rightarrow \infty} \theta^*. \quad (5.16)$$

Condition 3 *On dispose d'un théorème ergodique affirmant, pour θ quelconque,*

$$\frac{1}{M} \left| C(\hat{\mathbf{x}}^\theta(\mathbf{Z}), \mathbf{X}^*) - \mathbb{E}_{\theta^*}[C(\hat{\mathbf{x}}^\theta(\mathbf{Z}), \mathbf{X}^*)] \right| \xrightarrow{M \rightarrow \infty} 0. \quad (5.17)$$

En particulier, dans le cas d'un coût homogène séparable $C(\mathbf{x}, \mathbf{x}^) = \sum_{m=1}^M c(x_m - x_m^*)$ et d'un problème stationnaire en $(\mathbf{X}^*, \mathbf{N}, \mathbf{Z})$ pour lequel $M = N$, (5.17) est une loi des grands nombres qui s'écrit*

$$\frac{1}{N} \sum_{n=1}^N c(\hat{x}_n^\theta(\mathbf{Z}), X_n^*) \xrightarrow{N \rightarrow \infty} \mathbb{E}_{\theta^*}[c(\hat{x}_1^\theta(\mathbf{Z}), X_1^*)].$$

Les conditions (5.16) et (5.17) expriment la convergence de suites de vecteurs ou de variables aléatoires, en un sens mathématique laissé volontairement imprécis : les arguments donnés ici sont indicatifs et n'ont évidemment pas valeur de démonstration. Des résultats théoriques de convergence du maximum de vraisemblance sont donnés dans [YOUNÈS 1989] et dans [COMETS

6. Il s'agit de la théorie bayésienne classique, introduisant chaque estimateur de Bayes comme le minimiseur d'un coût moyen, encore appelé *risque de Bayes*. Voir, par exemple, [ROBERT 1992]. La notation « $\mathbb{E}_{\theta^*}[\cdot]$ » indique que l'espérance est calculée sous le « vrai » modèle, pour la « vraie » valeur θ^* de θ .

et GIDAS 1992] dans le cas markovien pour un modèle d'observation de type (5.2). Quant aux conditions techniques de la seconde, elles doivent être suffisantes pour assurer

$$(5.17) \implies \theta^{\text{MR}} \xrightarrow{N \rightarrow \infty} \theta^*.$$

Ainsi $\hat{\theta}^{\text{MV}}$ et θ^{MR} se rapprochent l'un de l'autre (en s'approchant de θ^*), pour un nombre de données tendant vers l'infini.

Une illustration concrète de cette propriété peut être trouvée dans [FORTIER et coll. 1993]. Il s'agit d'un problème de déconvolution d'image sous hypothèse de bruit et d'image gaussiens⁷. L'estimateur calculé est $\hat{\mathbf{x}}^\theta = \hat{\mathbf{x}}^{\text{MAP}}$, qui, s'identifiant avec la moyenne *a posteriori* $\hat{\mathbf{x}}^{\text{EAP}} \triangleq \mathbb{E}[\mathbf{X} | \mathbf{z}]$, minimise l'espérance du coût quadratique $C(\mathbf{x}, \mathbf{x}^*) = \|\mathbf{x} - \mathbf{x}^*\|_2^2$. Lorsque \mathbf{x}^* est une réalisation pseudo-aléatoire gaussienne ayant la « bonne » structure de corrélation, la courbe de vraisemblance $f_{\mathbf{Z}}(\mathbf{z} ; \theta)$ et la fonction de coût quadratique $C(\hat{\mathbf{x}}^\theta, \mathbf{x}^*)$, tracées en fonction de θ , ont approximativement le même minimiseur. Un autre exemple en déconvolution d'image est donné dans la thèse de Marc FAYOLLE, dans le cas d'un champ de Pickard gaussien [FAYOLLE 1998].

Dans [FORTIER et coll. 1993] et dans [FAYOLLE 1998], le passage à une image réelle photographique, toutes choses égales par ailleurs, suffit à écarter notablement $\hat{\theta}^{\text{MV}}$ de la zone de meilleur réglage. Ce résultat est évidemment très négatif, puisqu'il contredit la conjecture (5.15), implicite dans tous les travaux sur le réglage autodidacte par maximisation de vraisemblance.

On peut sans doute diagnostiquer qu'un champ de Markov *gaussien homogène* est un mauvais modèle pour les images photographiques \mathbf{x}^* . De la même façon, l'emploi d'un modèle unilatéral gaussien par morceaux a été mis en cause pour la déconvolution d'image à la fin du Sous-thème 2d, page 26. On peut, en revanche, fonder quelque espoir dans les énergies non quadratiques ou semi-quadratiques du chapitre précédent.

Dans le cas mono-dimensionnel, il existe suffisamment de travaux concordant pour valider le réglage autodidacte par maximisation de vraisemblance. Ainsi, [6] se situe dans le contexte de la déconvolution impulsionnelle. Dans [18], une courbe tracée à la main est segmentée par une méthode de type détection-estimation (Sous-chapitre 4.2, chapitre précédent). Dans [66], un problème d'analyse de raies spectrales est résolu par une approche semi-quadratique. Dans tous ces cas, on effectue le réglage des hyperparamètres par maximum de vraisemblance, et la méthode autodidacte est validée sur un exemple qui n'est pas une réalisation pseudo-aléatoire du modèle.

Dans le cas de l'imagerie (déconvolution d'image, reconstruction à partir de projection, etc.), il est difficile de se prononcer car les contributions sont moins nombreuses et plus récentes sur le sujet, du fait des difficultés algorithmiques. Sur la base de travaux tels que ceux de [SAQUIB et coll. 1998] en tomographie, mon intuition est que la conjecture (5.15) ne peut pas être considérée comme systématiquement valide pour les modèles markoviens les plus utilisés, dans le cas où \mathbf{x}^* est une image réelle. Ainsi, dans [SAQUIB et coll. 1998], l'énergie de Gibbs *a priori* s'écrit

$$\Phi(\mathbf{x}) = \sum_{r \sim s} c_{rs} |x_r - x_s|^p,$$

où $r \sim s$ décrit un système à huit voisins et c_{rs} pondère sélectivement les potentiels diagonaux. Page 1036, les auteurs remarquent : «... *for most natural images except for a few texture images, the maximum likelihood estimate of p was less than 1.0, and for many images containing man made objects p was less than 0.4. ... Very small values of p may not lead to the best MAP reconstructions.* » Dans le commentaire qui suit dans le même article, la perte de qualité du MAP pour $p < 1$ est plutôt attribuée aux problèmes algorithmiques posés par la non convexité.

7. Dans un tel problème, il convient de remarquer que M tend vers l'infini comme N dans (5.16) et (5.17).

En fait on peut avancer plusieurs autres sources potentielles de difficulté, en amont des aspects algorithmiques : j'en envisage cinq ci-dessous.

Nombre fini de données

Les résultats (5.16) et (5.17) sont asymptotiques sur le nombre de données. Le rapport entre nombre de données et nombre d'hyperparamètres étant souvent très favorable, surtout lorsque les observations sont des images, il est peu crédible qu'une distance trop grande à l'asymptotique soit la cause principale de l'inadéquation entre $\hat{\theta}^{\text{MV}}$ et θ^{MR} . D'autant qu'à nombre de données égal, le problème se manifesterait de la même façon pour un problème synthétique.

Fonctions de coût

La Condition 1 associe à $\hat{\mathbf{x}}^{\text{MAP}}$ une fonction de coût très artificielle : on trouve $\hat{\mathbf{x}}^{\text{MAP}}$, par exemple, comme la limite pour $\varepsilon \rightarrow 0$ de l'estimateur minimisant

$$P_{\theta^*}(\|\hat{\mathbf{x}}(\mathbf{Z}) - \mathbf{X}^*\|_2 > \varepsilon) = E_{\theta^*}[\mathbb{1}_{\mathcal{D}(\mathbf{X}^*, \varepsilon)}(\hat{\mathbf{x}}(\mathbf{Z}))],$$

avec $\mathcal{D}(\mathbf{x}^*, \varepsilon) = \{\mathbf{x}, \|\mathbf{x} - \mathbf{x}^*\|_2 > \varepsilon\}$. Dans le cas linéaire gaussien, $\hat{\mathbf{x}}^{\text{MAP}} = \hat{\mathbf{x}}^{\text{EAP}}$ minimise aussi en moyenne la norme quadratique $C(\mathbf{x}, \mathbf{x}^*) = \|\mathbf{x} - \mathbf{x}^*\|_2$, plus naturelle. Tel n'est pas le cas dans les contextes non gaussiens étudiés dans [ZHOU et coll. 1997, SAQUIB et coll. 1998, JALOBÉANU et coll. 1998].

Des simulations simples permettraient de savoir si, pour des problèmes inverses synthétiques représentatifs en imagerie, un écart significatif sépare θ^{MR} de θ^* pour $\hat{\mathbf{x}}^{\text{MAP}}$ et $C(\mathbf{x}, \mathbf{x}^*) = \|\mathbf{x} - \mathbf{x}^*\|_2$. Si tel est le cas, il est intéressant de définir le « meilleur réglage moyen » :

Définition 3 (meilleur réglage moyen pour les hyperparamètres) *Étant donné un estimateur paramétré $\hat{\mathbf{x}}^\theta(\cdot)$, on appelle meilleur réglage moyen pour θ au sens d'une fonction de coût $C(\mathbf{x}, \mathbf{x}^*)$ toute valeur de θ minimisant $E_{\theta^*}[C(\hat{\mathbf{x}}^\theta(\mathbf{Z}), \mathbf{X}^*)]$.*

Le meilleur réglage moyen est seulement fonction de θ^* . On le note $\theta^{\text{MRM}}(\theta^*)$, en admettant qu'il est unique. Si la Condition 1 est vraie, on a $\theta^{\text{MRM}}(\theta^*) = \theta^*$. Mais, même si la Condition 1 n'est pas vraie (par exemple, pour $\hat{\mathbf{x}}^{\text{MAP}}$ et $C(\mathbf{x}, \mathbf{x}^*) = \|\mathbf{x} - \mathbf{x}^*\|_2$), $\theta^{\text{MRM}}(\hat{\theta}^{\text{MV}})$ s'approche du meilleur réglage θ^{MR} quand $\hat{\theta}^{\text{MV}}$ s'approche de θ^* , pour un problème inverse synthétique vérifiant les Conditions 2 et 3.

Pour une valeur donnée de θ , on peut approcher $\theta^{\text{MRM}}(\theta)$ en générant pseudo-aléatoirement un problème inverse synthétique pour $\theta^* = \theta$ et en sélectionnant le meilleur réglage pour ce problème. Ce principe permettrait d'obtenir simplement $\theta^{\text{MRM}}(\hat{\theta}^{\text{MV}})$ pour un problème inverse réel, une fois calculé $\hat{\theta}^{\text{MV}}$.

Un autre point de vue consisterait à respecter la Condition 1 pour des coûts réalistes, donc à abandonner $\hat{\mathbf{x}}^{\text{MAP}}$. Mais on connaît mal le comportement qualitatif d'autres estimateurs tel que l'espérance *a posteriori* $\hat{\mathbf{x}}^{\text{EAP}}$, ou la médiane *a posteriori*, (associée au coût $C(\mathbf{x}, \mathbf{x}^*) = \|\mathbf{x} - \mathbf{x}^*\|_1$), bien que leur calcul reste abordable par rééchantillonnage markovien.

A priori impropres

Beaucoup d'énergies de Gibbs utilisées comme *a priori* en traitement d'image s'écrivent

$$\Phi(\mathbf{x}) = \sum_{r \sim s} \phi_{rs}(x_r - x_s),$$

où les ϕ_{rs} prennent des valeurs finies sur \mathbb{R} et $r \sim s$ décrit un ensemble de couples de sites voisins (r, s) . Or ce cas ne vérifie pas la condition (5.3). En effet, dans le calcul de l'intégrale $Z_x(T_x)$, considérons le changement de variable linéaire inversible

$$\begin{cases} t_1 &= x_1, \\ t_2 &= x_2 - x_1, \\ t_3 &= x_3 - x_1, \\ &\vdots \\ t_M &= x_M - x_1, \end{cases}$$

où la numérotation des x_s est arbitraire. D'une part, le déterminant jacobien $\det \{\partial x_m / \partial t_n\}$ est constant car la transformation est linéaire. D'autre part, on vérifie facilement que $\Phi(\mathbf{x})$ s'exprime en fonction de t_2, \dots, t_M à l'exclusion de t_1 : $\Phi(\mathbf{x}) = \tilde{\Phi}(t_2, \dots, t_M)$. En utilisant le théorème de Fubini pour les fonctions mesurables positives, on trouve

$$Z_x(T_x) = |\partial x_m / \partial t_n| \int_{\mathbb{R}} dt_1 \int_{\mathbb{R}^{M-1}} e^{-\tilde{\Phi}(t_2, \dots, t_M)/T_x} dt_2 \dots t_M,$$

qui est infini à cause de la première intégrale. Dans ce cas, on dit que la loi *a priori* non normalisable $\exp -\Phi(\mathbf{x})/T_x$ est *impropre*. Dans la littérature, certaines références semblent méconnaître le problème (par exemple, [HIGDON et coll. 1997]). D'autres limitent le domaine de variation des pixels à un intervalle fini [ZHOU et coll. 1997] ou à un ensemble discret [JALOBÉANU et coll. 1998], ce qui a pour effet de rendre $Z_x(T_x)$ fini.

Certaines opérations mathématiques gardent un sens bien que la loi *a priori* soit impropre ; en particulier, la vraisemblance *a posteriori* $f_{\mathbf{X}} | \mathbf{Z}$ et la vraisemblance $f_{\mathbf{Z}}$, définies formellement par (5.8) et (5.13), peuvent rester des densité de probabilité. Cependant, comme il est impossible de simuler pseudo-aléatoirement la loi impropre $f_{\mathbf{X}}$, la notion de problème inverse synthétique dans lequel on peut envisager (5.15) devient sans objet. En revanche, comme la loi impropre $\exp -\Phi(\mathbf{x})/T_x$ peut être construite comme la limite d'une suite de lois propres, il serait intéressant d'étudier dans quelles conditions mathématiques on peut donner un sens à (5.15) *par continuité* (voir, à ce propos, la note de bas de page dans [SAQUIB et coll. 1998, page 1031]).

Changement de modèle

Il est raisonnable de penser que c'est le manque de réalisme du modèle qui reste la principale cause de difficulté. Faut-il pour autant remettre en cause radicalement les modèles markoviens (même non gaussiens) en imagerie ? Sous forme supervisée, il est avéré que ces modèles produisent des estimateurs satisfaisants, y compris dans un contexte pratique tel que l'imagerie satellitaire [85, 87, 88, 89]. Tout se passe donc comme si les modèles markoviens étaient suffisamment bons pour l'estimation de \mathbf{x} , mais pas pour celle de θ . Je partage totalement le diagnostic de [GRAY et coll. 1994] (que rejoignent ceux de [MORRIS et coll. 1996, DESCOMBES et coll. 1999]) : « *We should like to re-emphasize that the local properties of spatial Markov models are undoubtedly plausible descriptors of the local associations typical of many images, which is the way in which the models are often used. Nevertheless, it would be reassuring if models used as priors did in fact provide a realistic representation of our prior assumptions and if their (empirical) properties were more widely known.* ». Étant donné le coût numérique modeste associé au calcul de $\hat{\mathbf{x}}^{\text{MAP}}$ (du moins quand l'énergie *a posteriori* est convexe), une remise en cause totale des modèles markoviens utilisés serait irréaliste. En revanche, une meilleure connaissance de leur (in)adéquation à représenter certaines caractéristiques d'images réelles permettrait sans doute de progresser sur des questions difficiles comme le réglage autodidacte des hyperparamètres.

Énergies semi-quadratiques

Des travaux récents ([JALOBÉANU et coll. 1998], [66]) tirent parti de la structure semi-quadratique (SQ) des énergies augmentées introduites au Sous-chapitre 4.4 du Chapitre précédent (page 50) pour développer des échantillonneurs pseudo-aléatoires permettant d'approcher $\hat{\theta}^{\text{MV}}$. Il faut remarquer que la loi *a priori* utilisée n'est plus (5.5), mais plutôt la loi jointe faisant intervenir les variables auxiliaires

$$f_{\mathbf{X}, \mathbf{B}}(\mathbf{x}, \mathbf{b}) = \frac{e^{-\Psi(\mathbf{x}, \mathbf{b})/T_x}}{\int_{\mathbb{R}^M \times \mathcal{B}} e^{-\Psi(\mathbf{x}, \mathbf{b})/T_x} d\mathbf{x} d\mathbf{b}}, \quad (5.18)$$

où $\Psi(\mathbf{x}, \mathbf{b})$ est une fonction d'énergie SQ vérifiant

$$\forall \mathbf{x}, \quad \min_{\mathbf{b} \in \mathcal{B}} \Psi(\mathbf{x}, \mathbf{b}) = \Phi(\mathbf{x}).$$

Dans ce cadre, il devient naturel de définir la vraisemblance des données sous la forme

$$f_{\mathbf{Z}}^{\text{SQ}}(\mathbf{z}; \theta) = \int_{\mathbb{R}^M \times \mathcal{B}} f_{\mathbf{X}, \mathbf{B}, \mathbf{Z}}(\mathbf{x}, \mathbf{b}, \mathbf{z}) d\mathbf{x} d\mathbf{b} = \int_{\mathbb{R}^M} f_{\mathbf{N}}(\mathbf{z} - A(\mathbf{x})) \int_{\mathcal{B}} f_{\mathbf{X}, \mathbf{B}}(\mathbf{x}, \mathbf{b}) d\mathbf{b} d\mathbf{x}. \quad (5.19)$$

Partant des mêmes spécifications *a priori* Ω et Φ , les formulations SQ fournissent, par construction, le même estimateur $\hat{\mathbf{x}} = \hat{\mathbf{x}}^{\text{MAP}}$ que la formulation « directe » (donc, pour une même fonction de coût, elles partagent aussi le même réglage optimal θ^{MR}), mais pas la même fonction de vraisemblance pour les hyperparamètres : la fonction $f_{\mathbf{Z}}^{\text{SQ}}$ définie par (5.19) diffère de la fonction $f_{\mathbf{Z}}$ définie par (5.13), sauf si $\int_{\mathcal{B}} e^{-\Psi(\mathbf{x}, \mathbf{b})/T_x} d\mathbf{b}$ est proportionnel à $e^{-\Phi(\mathbf{x})/T_x}$, condition qui ne semble jamais satisfaite.

A quel maximum de vraisemblance peut-on se fier ? Au $\hat{\theta}^{\text{MV}}$ de la formulation directe ? Au $\hat{\theta}^{\text{MVGR}}$ déduit de la construction SQ de type « Geman-Reynolds » [GEMAN et REYNOLDS 1992] ? Au $\hat{\theta}^{\text{MVGy}}$ déduit de la construction SQ de type « Geman-Yang » [GEMAN et YANG 1995] ? Le problème est ouvert. En reprenant l'analyse de [GRAY et coll. 1994], il est probable que ces estimateurs approchent plus ou moins bien θ^{MR} suivant le réalisme statistique du modèle bayésien associé.

5.5 Conclusion

Dans ce chapitre, des outils statistiques ont été introduits, puis leur rôle dans la construction de modèles régularisants a été discuté au Sous-chapitre 5.3. Le problème de l'*inversion autodidacte* a été abordé ensuite, au Sous-chapitre 5.4. C'est précisément la capacité du cadre statistique bayésien à formaliser des « sur-problèmes » tels que la déconvolution aveugle ou l'*inversion autodidacte* qui me semble avant tout justifier son recours. Même si ce recours n'est pas « miraculeux », des opérations telles que marginalisation et conditionnement permettent d'introduire concrètement des estimateurs qui n'ont pas d'équivalents naturels dans le cadre déterministe.

Concernant les aspects calculatoires, il est incontestable que l'émergence des méthodes numériques de Monte-Carlo par chaînes de Markov [ROBERT 1996] en traitement de signaux et d'images constitue une avancée pratique importante pour de nombreux problèmes d'inférence dans ces domaines. Or, en dehors de tout argument « idéologique », l'interprétation bayésienne d'un problème d'inférence est un préalable indispensable à sa résolution par échantillonnage pseudo-aléatoire.

Troisième partie

Références bibliographiques

- [AUBERT et VESE 1997] G. AUBERT et L. VESE (1997), « A variational method in image recovery », *SIAM J. Num. Anal.*, **34**, n° 5, pages 1948–1979.
- [BEDINI et coll. 1994] L. BEDINI, I. GERACE et A. TONAZZINI (1994), « A Deterministic Algorithm for Reconstructing Images with Interacting Discontinuities », *Comput. Vision Graphics Image Process.*, **56**, n° 2, pages 109–123.
- [BERTSEKAS 1995] D. P. BERTSEKAS (1995), *Nonlinear programming*, Athena Scientific, Belmont, MA, USA.
- [BESAG 1974] J. E. BESAG (1974), « Spatial interaction and the statistical analysis of lattice systems (With discussion) », *J. R. Statist. Soc. B*, **36**, n° 2, pages 192–236.
- [BESAG 1986] J. E. BESAG (1986), « On the statistical analysis of dirty pictures (With discussion) », *J. R. Statist. Soc. B*, **48**, n° 3, pages 259–302.
- [BLAKE 1989] A. BLAKE (1989), « Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction », *IEEE Trans. Pattern Anal. Mach. Intell.*, **PAMI-11**, n° 1, pages 2–12.
- [BLAKE et ZISSERMAN 1987] A. BLAKE et A. ZISSERMAN (1987), *Visual reconstruction*, The MIT Press, Cambridge, MA, USA.
- [BOUBAL 1998] O. BOUBAL (1998), *Estimation du signal de cliquetis par inversion paramétrique des phénomènes de propagation acoustique et optimisation d'instrument*, thèse de doctorat, Université de Paris-sud, Orsay.
- [BOUMAN et SAUER 1993] C. A. BOUMAN et K. D. SAUER (1993), « A generalized Gaussian image model for edge-preserving MAP estimation », *IEEE Trans. Image Processing*, **2**, n° 3, pages 296–310.
- [BRÉMAUD 1984] P. BRÉMAUD (1984), *Introduction aux probabilités*, Springer Verlag, Berlin, Allemagne.
- [BRÉMAUD 1999] P. BRÉMAUD (1999), *Markov Chains. Gibbs fields, Monte Carlo Simulation, and Queues*, Texts in Applied Mathematics 31, Spinger, New York, NY, USA.
- [BURG 1967] J. P. BURG (1967), « Maximum Entropy Spectral Analysis », dans *Proc. of the 37th Meeting of the Society of Exploration Geophysicists*, pages 34–41, Oklahoma City, OK, USA.
- [CARFANTAN et MOHAMMAD-DJAFARI 1995] H. CARFANTAN et A. MOHAMMAD-DJAFARI (1995), « A Bayesian Approach for Nonlinear Inverse Scattering Tomographic Imaging », dans *Proc. IEEE ICASSP*, volume IV, pages 2311–2314, Detroit, U.S.A.
- [CHAMPAGNAT 1993] F. CHAMPAGNAT (1993), *Déconvolution impulsionnelle et extensions pour la caractérisation des milieux inhomogènes en échographie*, Thèse de doctorat, Université de Paris-Sud, Orsay.
- [CHAN et CHIU-KWONG 1998] T. F. CHAN et W. CHIU-KWONG (1998), « Total variation blind deconvolution », *IEEE Trans. Image Processing*, **7**, n° 3, pages 370–375.
- [CHARBONNIER et coll. 1994] P. CHARBONNIER, L. BLANC-FÉRAUD, G. AUBERT et M. BARLAUD (1994), « Two Deterministic Half-Quadratic Regularization Algorithms for Computed Imaging », dans *Proc. IEEE ICIP*, volume 2, pages 168–172, Austin, TX, USA.
- [CHARBONNIER et coll. 1997] P. CHARBONNIER, L. BLANC-FÉRAUD, G. AUBERT et M. BARLAUD (1997), « Deterministic edge-preserving regularization in computed imaging », *IEEE Trans. Image Processing*, **6**, n° 2, pages 298–311.
- [CHÉBIRA 1994] N. CHÉBIRA (1994), *Détection d'obstacles en temps réel par reconstruction de l'environnement 2D à partir de capteurs infrarouges 1D*, thèse de doctorat, Université de Paris-Sud, Orsay.
- [CHENG et coll. 1996] Q. CHENG, R. CHEN et T.-H. LI (1996), « Simultaneous Wavelet Estimation and Deconvolution of Reflection Seismic Signals », *IEEE Trans. Geosci. Remote Sensing*, **34**, pages 377–384.

- [COHEN 1996] L. D. COHEN (1996), « Auxiliary variables and two-step iterative algorithms in computer vision problems », *J. Math. Imag. Vision*, **6**, pages 59–83.
- [COHEN-BACRIE 1996] C. COHEN-BACRIE (1996), « Rapport de fin de stage CSN », Rapport DER-DDM-TAI, Électricité de France.
- [COMETS et GIDAS 1992] F. COMETS et B. GIDAS (1992), « Parameter estimation for Gibbs distributions from partially observed data », *Adv. Appl. Prob.*, **2**, pages 142–170.
- [CORVAISIER-RICHE 1999] A. CORVAISIER-RICHE (1999), *Problèmes inverses : analyse combinatoire. Application au contrôle non destructif par courants de Foucault*, thèse de doctorat, Université Claude Bernard, Lyon.
- [DE CESARE 1996] A. DE CESARE (1996), *Algorithmes rapides de restauration des signaux : Application à l'imagerie médicale*, thèse de doctorat, Université de Paris XI, Orsay.
- [DELANEY et BRESLER 1998] A. H. DELANEY et Y. BRESLER (1998), « Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography », *IEEE Trans. Image Processing*, **7**, n° 2, pages 204–221.
- [DELPHIN-POULAT 1999] L. DELPHIN-POULAT (1999), *Utilisation de modèles de Markov cachés pour une compensation synchrone à la trame, dans un contexte de reconnaissance de la parole*, thèse de doctorat, Université de Rennes I, CNET, Lannion.
- [DEMOMENT 1989] G. DEMOMENT (1989), « Image Reconstruction and Restoration: Overview of Common Estimation Structure and Problems », *IEEE Trans. Acoust. Speech, Signal Processing*, **ASSP-37**, n° 12, pages 2024–2036.
- [DEMOMENT et REYNAUD 1985] G. DEMOMENT et R. REYNAUD (1985), « Fast Minimum-Variance Deconvolution », *IEEE Trans. Acoust. Speech, Signal Processing*, **ASSP-33**, pages 1324–1326.
- [DERIN et coll. 1984] H. DERIN, H. ELLIOTT, R. CRISTI et D. GEMAN (1984), « Bayes smoothing algorithms for segmentation of binary images modeled by Markov random fields », *IEEE Trans. Pattern Anal. Mach. Intell.*, **PAMI-6**, pages 4.
- [DESCOMBES et coll. 1999] X. DESCOMBES, R. MORRIS, J. ZERUBIA et M. BERTHOD (1999), « Estimation of Markov Random Field prior parameters using Markov chain Monte Carlo Maximum Likelihood », *IEEE Trans. Image Processing*, **8**, n° 7, pages 954–963.
- [DEVIJVER et DEKESSEL 1988] P. A. DEVIJVER et M. DEKESSEL (1988), « Champs aléatoires de Pickard et modélisation d'images digitales », *Traitement du Signal*, **5**, n° 5, pages 131–150.
- [DOUCET et DUVAUT 1997] A. DOUCET et P. DUVAUT (1997), « Bayesian Estimation of State Space Models Applied to Deconvolution of Bernoulli-Gaussian Processes », *Signal Processing*, **57**, pages 147–161.
- [DUBLANCHET 1996] F. DUBLANCHET (1996), *Contribution de la méthodologie bayésienne à l'analyse spectrale de raies pures et à la goniométrie haute résolution*, Thèse de doctorat, Université de Paris-Sud, Orsay.
- [DUNMUR et TITTERINGTON 1997] A. P. DUNMUR et D. M. TITTERINGTON (1997), « Computational Bayesian Analysis of Hidden Markov Mesh Models », *IEEE Trans. Pattern Anal. Mach. Intell.*, **PAMI-19**, n° 11, pages 1296–1300.
- [FAYOLLE 1998] M. FAYOLLE (1998), *Modélisation unilatérale composite pour la restauration d'images*, Thèse de doctorat, Université de Paris-Sud, Orsay.
- [FORTIER et coll. 1993] N. FORTIER, G. DEMOMENT et Y. GOUSSARD (1993), « Comparison of GCV and ML methods of determining parameters in image restoration by regularisation », *J. Visual Comm. Image Repres.*, **4**, pages 157–170.
- [FRANKLIN 1970] J. N. FRANKLIN (1970), « Well-posed stochastic extensions of ill-posed linear problems », *J. Math. Anal. Applic.*, **31**, pages 682–716.

- [GASSIAT et coll. 1992] E. GASSIAT, F. MONFRONT et Y. GOUSSARD (1992), « On simultaneous signal estimation and parameter identification using a generalized likelihood approach », *IEEE Trans. Inf. Theory*, **IT-38**, pages 157-162.
- [GEMAN et REYNOLDS 1992] D. GEMAN et G. REYNOLDS (1992), « Constrained Restoration and the Recovery of Discontinuities », *IEEE Trans. Pattern Anal. Mach. Intell.*, **14**, n° 3, pages 367-383.
- [GEMAN et YANG 1995] D. GEMAN et C. YANG (1995), « Nonlinear Image Recovery with Half-Quadratic Regularization », *IEEE Trans. Image Processing*, **4**, n° 7, pages 932-946.
- [GEMAN et GEMAN 1984] S. GEMAN et D. GEMAN (1984), « Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images », *IEEE Trans. Pattern Anal. Mach. Intell.*, **PAMI-6**, n° 6, pages 721-741.
- [GEMAN et MCCLURE 1987] S. GEMAN et D. MCCLURE (1987), « Statistical Methods for Tomographic Image Reconstruction », dans *Proceedings of the 46th Session of the ICI, Bulletin of the ICI*, volume 52, pages 5-21, 1987.
- [GIOVANNELLI 1995] J.-F. GIOVANNELLI (1995), *Estimation de caractéristiques spectrales en temps court. Application à l'imagerie Doppler*, Thèse de doctorat, Université de Paris-Sud, Orsay.
- [GOLUB et coll. 1979] G. H. GOLUB, M. HEATH et G. WAHBA (1979), « Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter », *Technometrics*, **21**, n° 2, pages 215-223.
- [GOUSSARD et coll. 1990] Y. GOUSSARD, G. DEMOMENT et J. IDIER (1990), « A new algorithm for iterative deconvolution of sparse spike trains », dans *Proc. IEEE ICASSP*, pages 1547-1550, Albuquerque, U.S.A.
- [GOUTSIAS et MENDEL 1986] J. K. GOUTSIAS et J. M. MENDEL (1986), « Maximum-likelihood deconvolution: An optimization theory perspective », *Geophysics*, **51**, pages 1206-1220.
- [GRAY et coll. 1994] A. J. GRAY, J. W. KAY et D. M. TITTERINGTON (1994), « An empirical study of the simulation of various models used for images », *IEEE Trans. Pattern Anal. Mach. Intell.*, **16**, pages 507-513.
- [GULL et DANIELL 1978] S. F. GULL et G. J. DANIELL (1978), « Image reconstruction from incomplete and noisy data », *Nature*, **272**, pages 686-690.
- [HANSEN 1992] P. HANSEN (1992), « Analysis of discrete ill-posed problems by means of the L-curve », *SIAM Rev.*, **34**, pages 561-580.
- [HIGDON et coll. 1997] D. M. HIGDON, J. E. BOWSHER, V. E. JOHNSON, T. G. TURKINGTON, D. R. GILLAND et R. J. JASZCZAK (1997), « Fully Bayesian Estimation of Gibbs Hyperparameters for Emission Computed Tomography Data », *IEEE Trans. Medical Imaging*, **16**, n° 5, pages 516-526.
- [HUBER 1981] P. J. HUBER (1981), *Robust Statistics*, John Wiley, New York, NY, USA.
- [HUNT 1977] B. R. HUNT (1977), « Bayesian Methods in Nonlinear Digital Image Restoration », *IEEE Trans. Communications*, **C-26**, pages 219-229.
- [IBRAGIMOV et HAS'MINSKII 1981] I. IBRAGIMOV et R. HAS'MINSKII (1981), *Statistical estimation. Asymptotic theory*, Springer Verlag, New York, NY, USA.
- [IDIER 1991] J. IDIER (1991), *Imagerie des milieux stratifiés : modélisation markovienne, application à la déconvolution sismique*, Thèse de doctorat, Université de Paris-Sud, Orsay.
- [JALOBEANU et coll. 1998] A. JALOBEANU, L. BLANC-FÉRAUD et J. ZERUBIA (1998), « Estimation d'hyperparamètres pour la restauration d'images satellitaires par une méthode MCMCML », rapport de recherche 3469, INRIA, Sophia Antipolis.
- [JAUFFRET et LE CADRE 1999] C. JAUFFRET et J.-P. LE CADRE (1999), « trajectographie passive d'un objet en mouvement uniformément accéléré en dimension 3 », dans *Actes 17^e coll. GRETSI*, pages 673-676, Vannes.

- [JAYNES 1968] E. T. JAYNES (1968), « Prior Probabilities », *IEEE Trans. Systems Science and Cybern.*, **SSC-4**, n° 3, pages 227–241.
- [JAZWINSKI 1970] A. H. JAZWINSKI (1970), *Stochastic process and filtering theory*, Academic Press, New York, NY, USA.
- [JENG et WOODS 1991] F. C. JENG et J. W. WOODS (1991), « Compound Gauss-Markov random fields for image estimation », *IEEE Trans. Signal Processing*, **39**, n° 3, pages 683–697.
- [KAARESEN 1998] K. F. KAARESEN (1998), « Evaluation and applications of the iterated window maximization method for sparse deconvolution », *IEEE Trans. Signal Processing*, **46**, n° 3, pages 609–624.
- [KAY et MARPLE 1981] S. M. KAY et S. L. MARPLE (1981), « Spectrum Analysis – A Modern Perspective », *Proc. IEEE*, **69**, n° 11, pages 1380–1419.
- [KITAGAWA et GERSCH 1985] G. KITAGAWA et W. GERSCH (1985), « A Smoothness Priors Long AR Model Method for Spectral Estimation », *IEEE Trans. Automat. Contr.*, **AC-30**, n° 1, pages 57–65.
- [KITAGAWA et GERSCH 1985] G. KITAGAWA et W. GERSCH (1985), « A Smoothness Priors Time-Varying AR Coefficient Modeling of Nonstationary Covariance Time Series », *IEEE Trans. Automat. Contr.*, **AC-30**, n° 1, pages 48–56.
- [KUNDUR et HATZINAKOS 1996] D. KUNDUR et D. HATZINAKOS (1996), « Blind image deconvolution », *IEEE Signal processing magazine*, pages 43–64.
- [KÜNSCH 1994] H. R. KÜNSCH (1994), « Robust priors for smoothing and image restoration », *Ann. Inst. Stat. Math.*, **46**, n° 1, pages 1–19.
- [LACROIX 1987] V. LACROIX (1987), « Pixel labeling in a second-order Markov mesh », *Signal Processing*, **12**, pages 59–82.
- [LAGENDIJK et coll. 1990] R. LAGENDIJK, J. BIEMOND et D. E. BOEKEE (1990), « Identification and Restoration of Noisy Blurred Images Using The Expectation Minimization Algorithm », *IEEE Trans. Acoust. Speech, Signal Processing*, **ASSP-38**, pages 1180–1191.
- [LAKSHMANAN et DERIN 1989] S. LAKSHMANAN et H. DERIN (1989), « Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing », *IEEE Trans. Pattern Anal. Mach. Intell.*, **PAMI-11**, n° 8, pages 799–813.
- [LAVIELLE 1995] M. LAVIELLE (1995), « A stochastic algorithm for parametric and non-parametric estimation in the case of incomplete data », *Signal Processing*, **42**, pages 3–17.
- [LEHMANN 1983] E. LEHMANN (1983), *Theory of point estimation*, John Wiley, New York, NY, USA.
- [LI et coll. 1995] S. Z. LI, Y. H. HUANG et J. S. FU (1995), « Convex MRF potential functions », dans *Proc. IEEE ICIP*, volume 2, pages 296–299, Washington DC, USA.
- [LITTLE et RUBIN 1983] R. J. A. LITTLE et D. B. RUBIN (1983), « On Jointly Estimating Parameters and Missing Data by Maximizing the Complete-Data Likelihood », *Amer. Statist.*, **37**, pages 218–220.
- [LOPEZ-VALCARCE et DASGUPTA 1999] R. LOPEZ-VALCARCE et S. DASGUPTA (1999), « A New Proof for the Stability of Equation-Error Models », *IEEE Signal Processing Letters*, **6**, n° 6, pages 148–150.
- [MARROQUIN et coll. 1987] J. L. MARROQUIN, S. K. MITTER et T. A. POGGIO (1987), « Probabilistic solution of ill-posed problems in computational vision », *J. Amer. Stat. Assoc.*, **82**, pages 76–89.
- [MARTIN 1993] T. MARTIN (1993), « Relaxation déterministe et champs markoviens en problèmes inverses », rapport technique (stage de DEA), GPI-LSS, Gif-sur-Yvette, France.

- [MARTIN 1997] T. MARTIN (1997), *Inversion bayésienne du problème non linéaire de tomographie d'impédance électrique modélisé par une méthode d'éléments finis*, Thèse de doctorat, Université de Paris-Sud, Orsay.
- [MENDEL 1983] J. M. MENDEL (1983), *Optimal Seismic Deconvolution*, Academic Press, New York, NY, USA.
- [MOHAMMAD-DJAFARI 1992] A. MOHAMMAD-DJAFARI (1992), *Maximum Entropy and Linear Inverse Problems; A Short Review*, pages 253–264, Kluwer Academic Publ., Paris, France, ali mohammad-djafari and guy demoment edition.
- [MOHAMMAD-DJAFARI 1994] A. MOHAMMAD-DJAFARI (1994), « Maximum d'entropie et problèmes inverses en imagerie », *Traitement du Signal*, pages 87–116.
- [MOHAMMAD-DJAFARI et DEMOMENT 1988] A. MOHAMMAD-DJAFARI et G. DEMOMENT (1988), « Utilisation de l'entropie dans les problèmes de restauration et de reconstruction d'images », *Traitement du Signal*, **5**, n° 4, pages 235–248.
- [MOHAMMAD-DJAFARI et IDIER 1990] A. MOHAMMAD-DJAFARI et J. IDIER (1990), *Maximum entropy prior laws of images and estimation of their parameters*, pages 285–293, Kluwer Academic Publ., Laramie, T.W. Grandy edition.
- [MOLINA 1994] R. MOLINA (1994), « On the Hierarchical Bayesian Approach to Image Restoration: Application to Astronomical Images », *IEEE Trans. Pattern Anal. Mach. Intell.*, **PAMI-16**, n° 11, pages 1122–1128.
- [MORRIS et coll. 1996] R. MORRIS, X. DESCOMBES et J. ZERUBIA (1996), « An analysis of some models used in image segmentation », rapport de recherche 3016, INRIA, Sophia Antipolis.
- [MULLIS et ROBERTS 1976] C. T. MULLIS et R. A. ROBERTS (1976), « The use of second-order information in the approximation of discrete-time linear systems », *IEEE Trans. Acoust. Speech, Signal Processing*, **ASSP-24**, n° 3, pages 226–238.
- [MUMFORD et SHAH 1985] D. MUMFORD et J. SHAH (1985), « Boundary Detection by Minimizing Functionals », dans *IEEE Conf. Comp. Vision Pattern Recogn.*, pages 22–26, San Francisco, CA, USA.
- [MURAI et KAGAWA 1985] T. MURAI et Y. KAGAWA (1985), « Electrical Impedance Computed Tomography Based on a Finite Element Model », *IEEE Trans. Biomed. Eng.*, **32**, n° 3, pages 177–184.
- [NÁDAS 1983] A. NÁDAS (1983), « Hidden Markov chains, the forward-backward algorithm, and initial statistics », *IEEE Trans. Acoust. Speech, Signal Processing*, **ASSP-31**, n° 2, pages 504–506.
- [NASHED 1981] M. Z. NASHED (1981), « Operator-theoretic and computational approaches to ill-posed problems with applications to antenna theory », *IEEE Trans. Ant. Propag.*, **29**, pages 220–231.
- [NIKOLOVA 1995] M. NIKOLOVA (1995), *Inversion markovienne de problèmes linéaires mal posés. application à l'imagerie tomographique.*, Thèse de doctorat, Université de Paris-Sud, Orsay.
- [NIKOLOVA et MOHAMMAD-DJAFARI 1996] M. NIKOLOVA et A. MOHAMMAD-DJAFARI (1996), « Eddy current tomography using a binary Markov model », *Signal Processing*, **49**, pages 119–132.
- [NITYANANDA et NARAYAN 1982] R. NITYANANDA et R. NARAYAN (1982), « Maximum entropy image reconstruction – A practical non-information-theoretic approach », *J. Astrophys. Astr.*, **3**, pages 419–450.
- [PHAM 1988] D. T. PHAM (1988), « Maximum likelihood estimation of the autoregressive model by relaxation on the reflection coefficients », *IEEE Trans. Signal Processing*, **36**, n° 8, pages 1363–1367.

- [PHILLIPS 1962] D. L. PHILLIPS (1962), « A technique for the numerical solution of certain integral equation of the first kind », *J. Ass. Comput. Mach.*, **9**, pages 84–97.
- [PICKARD 1977] D. K. PICKARD (1977), « A curious binary lattice process », *J. Appl. Prob.*, **14**, pages 717–731.
- [PICKARD 1980] D. K. PICKARD (1980), « Unilateral Markov fields », *Adv. Appl. Prob.*, **12**, pages 655–671.
- [QIAN et TITTERINGTON 1989] W. QIAN et D. M. TITTERINGTON (1989), « On the use of Gibbs Markov chain models in the analysis of images based on second-order pairwise interactive distributions », *J. Appl. Statistics*, **16**, pages 267–281.
- [QIAN et TITTERINGTON 1990] W. QIAN et D. M. TITTERINGTON (1990), « Parameter estimation for hidden Gibbs chains », *Statistics & Probability Letters*, **10**, pages 49–58.
- [REY 1983] W. J. REY (1983), *Introduction to robust and quasi-robust statistical methods*, Springer-Verlag, Berlin.
- [ROBERT 1992] C. ROBERT (1992), *L'analyse statistique bayésienne*, Economica, Paris.
- [ROBERT 1996] C. ROBERT (1996), *Méthodes de Monte-Carlo par chaînes de Markov*, Economica, Paris.
- [ROBERT 1997] C. ROBERT (1997), *Simulations par la méthode MCMC*, Economica, Paris.
- [ROCKAFELLAR 1970] R. T. ROCKAFELLAR (1970), *Convex Analysis*, Princeton Univ. Press.
- [SACCHI et coll. 1998] M. D. SACCHI, T. J. ULRYCH et C. J. WALKER (1998), « Interpolation and extrapolation using a high-resolution discrete Fourier transform », *IEEE Trans. Signal Processing*, **46**, n° 1, pages 31–38.
- [SAINT-FÉLIX 1987] D. SAINT-FÉLIX (1987), *Restauration d'image : régularisation d'un problème mal-posé et algorithmique associée*, Thèse de doctorat, Thèse de doctorat d'État, Université de Paris-Sud.
- [SALLARD et coll. 1997] J. SALLARD, L. PARADIS et J.-L. LACOUME (1997), « Détection et localisation de défauts en échographie ultrasonore », dans *Actes 16^e coll. GRETSI*, pages 315–318, septembre 1997.
- [SANTAMARIA et coll. 1999] I. SANTAMARIA, C. J. PANTALEON, J. IBANEZ et A. ARTES (1999), « Deconvolution of seismic data using adaptive Gaussian mixtures », *IEEE Trans. Geosci. Remote Sensing*, **GE-37**, n° 2, pages 855–859.
- [SAQUIB et coll. 1998] S. S. SAQUIB, C. A. BOUMAN et K. D. SAUER (1998), « ML Parameter Estimation for Markov Random Fields with Applications to Bayesian Tomography », *IEEE Trans. Image Processing*, **7**, n° 7, pages 1029–1044.
- [SAUER et BOUMAN 1993] K. D. SAUER et C. A. BOUMAN (1993), « A Local Update Strategy for Iterative Reconstruction from Projections », *IEEE Trans. Signal Processing*, **41**, n° 2, pages 534–548.
- [STOICA 1998] P. STOICA (1998), « Asymptomania? », *Signal Processing Magazine*, **15**, n° 1, pages 16.
- [TASTO et SCHOMBERG 1978] M. TASTO et H. SCHOMBERG (1978), « Object reconstruction from projections and some non-linear extensions », *Pattern Recognition and Signal Processing*, pages 485–503.
- [TERZOPOULOS 1986] D. TERZOPOULOS (1986), « Regularization of Inverse Visual Problems Involving Discontinuities », *IEEE Trans. Pattern Anal. Mach. Intell.*, **PAMI-8**, n° 4, pages 413–424.
- [THOMPSON et coll. 1991] A. THOMPSON, J. C. BROWN, J. W. KAY et D. M. TITTERINGTON (1991), « A Study of Methods of Choosing the Smoothing Parameter in Image Restoration by Regularization », *IEEE Trans. Pattern Anal. Mach. Intell.*, **PAMI-13**, n° 4, pages 326–339.

- [TIERNEY 1994] L. TIERNEY (1994), « Markov chain for exploring posterior distribution », *Annals Statist.*, **22**, n° 4, pages 1701–1762.
- [TIKHONOV et ARSÉNINE 1976] A. TIKHONOV et V. ARSÉNINE (1976), *Méthodes de résolution de problèmes mal posés*, Éditions MIR, Moscou, Russie.
- [TITTERINGTON 1985] D. M. TITTERINGTON (1985), « Common Structure of Smoothing Techniques in Statistics », *Int. Statist. Rev.*, **53**, n° 2, pages 141–170.
- [TWOMEY 1962] S. TWOMEY (1962), « On the Numerical Solution of Fredholm Integral Equations of the First Kind by the Inversion of the Linear System Produced by Quadrature », *J. Assoc. Comp. Mach.*, **10**, pages 97–101.
- [VOGEL et OMAN 1998] R. V. VOGEL et M. E. OMAN (1998), « Fast, robust total variation-based reconstruction of noisy, blurred images », *IEEE Trans. Image Processing*, **7**, n° 6, pages 813–823.
- [WEXLER et coll. 1985] A. WEXLER, B. FRY et M. R. NEUMAN (1985), « Impedance-computed tomography algorithm and system », *Applied Optics*, **24**, n° 23, pages 3985–3992.
- [YARLAGADDA et coll. 1985] R. YARLAGADDA, J. B. BEDNAR et T. L. WATT (1985), « Fast algorithms for l_p deconvolution », *IEEE Trans. Acoust. Speech, Signal Processing*, **ASSP-33**, n° 1, pages 174–182.
- [YOU et KAVEH 1996] Y.-L. YOU et M. KAVEH (1996), « A regularization approach to joint blur identification and image restoration », *IEEE Trans. Image Processing*, **5**, n° 3, pages 416–428.
- [YOUNÈS 1989] L. YOUNÈS (1989), « Parametric inference for imperfectly observed Gibbsian fields », *Prob. Th. Rel. Fields*, **82**, pages 625–645.
- [ZHOU et coll. 1997] Z. ZHOU, R. LEAHY et Q. JINYI (1997), « Approximate maximum likelihood hyperparameter estimation for Gibbs priors », *IEEE Trans. Image Processing*, **6**, n° 6, pages 844–861.

Quatrième partie

Publications et documents annexés

Annexe A

Inversion d'opérateurs linéaires mal posés à support large à l'aide d'un champ de Markov gaussien par morceaux

- [4] M. Nikolova, J. Idier, et A. Mohammad-Djafari, « Inversion of large-support ill-posed linear operators using a piecewise Gaussian MRF », *IEEE Trans. Image Processing*, vol. 7, n° 4, pp. 571–585, avril 1998.

Annexe B

Critères convexes semi-quadratiques et variables auxiliaires interactives pour la restauration d'image

- [10] J. Idier, « Convex half-quadratic criteria and interacting auxiliary variables for image restoration », rapport technique soumis à IEEE Trans. Image Processing, GPI-LSS, 1999.

Annexe C

Champs de Markov stationnaires sur un réseau rectangulaire fini

- [5] F. Champagnat, J. Idier, et Y. Goussard, «Stationary Markov random fields on a rectangular finite lattice», *IEEE Trans. Inf. Theory*, vol. 44, n° 7, pp. 2901–2916, novembre 1998.

Annexe D

Stabilité structurelle de méthodes de prédiction par moindres carrés

- [2] J. Idier et J.-F. Giovannelli, « Structural stability of least squares prediction methods », *IEEE Trans. Signal Processing*, vol. 46, n° 11, pp. 3109–3111, novembre 1998.

