JIAN FENG YAO

Méthodes bayésiennes en segmentation d'image et estimation par rabotage des modèles spatiaux

Thèses d'Orsay, 1990

<http://www.numdam.org/item?id=BJHTUP11_1990__0280__P0_0>

ORSAY

n° d'ordre:

UNIVERSITÉ DE PARIS–SUD
CENTRE D'ORSAY

THESE

Présentée pour obtenir

Le TITRE de DOCTEUR EN SCIENCE

PAR

Jian Feng YAO

... ∞ ...

SUJET: *Méthodes Bayésiennes en Segmentation d'Image et Estimation par Rabotage des Modèles Spatiaux*

soutenue le 11 Janvier 1990 devant un jury composé de

M. Didier DACUNHA–CASTELLE *Président*

Mme. Christiane COCOZZA–THIVENT

M. Robert AZENCOTT

M. Pierre A. DEVIJVER

M. Xavier GUYON

Code Matière A.M.S (1980): 60G35, 68T10, 62F12

A ma grand-mère,

A mes parents,

A Ying ...

Remerciements

Toute ma reconnaissance va, en premier lieu, au professeur Xavier GUYON dont la direction du travail présenté ici a toujours été efficace et d'une amicale compréhension. La confiance et la disponibilité qu'il a accordées à mon égard ont été particulièrement précieuses.

Le Laboratoire de Statistique Appliqué (C.N.R.S UA 473) du Centre Scientifique d'Orsay m'a réservé à la fois un acceuil plein de sympathie et un environnement scientifique hautement favorable qui m'ont permis de mener à bien ce travail. Que tous les membres de cette équipe trouvent ici l'expression de ma profonde gratitude.

Je remercie tout particulièrement le professeur Didier DACUNHA-CASTELLE pour sa bienveillance et son soutien qu'il m'a constamment manifestés pour mes activités de recherche. Sa participation au jury en tant que président est pour moi un grand honneur.

Madame Christiane COCOZZA-THIVENT, professeur à l'Université de Technologie de Compiège, et Monsieur Pierre A. DEVIJVER, directeur scientifique de l'Ecole Nationale Supérieur de Télécommunication de Bretagne, ont accepté la lourde charge de rapporteur de thèse. Leurs critiques fructueuses ont largement contribué à l'amélioration du document. Je leur en suis vivement reconnaissant.

Mes remerciement s'adressent également au professeur Robert AZENCOTT pour l'honneur qu'il m'a fait en acceptant de juger cette thèse.

Je ne saurai oublier de remercier le professeur Hans KÜNSCH de l'ETH de Zürich qui, par ses suggestions sur ce travail, m'a fait profiter de ses solides expériences scientifiques.

C'est pour moi un vif regret de ne pas pouvoir mentionner tous ceux qui m'ont aidé. Qu'ils soient ici assurés de l'expression de ma profonde reconnaissance.

Orsay, Janvier 1990

Jian Feng YAO

Bayesian image segmentation technics and estimation problem for strationary fields with tapered data

ABSTRACT

In the first part, we begin with an unified methodological analysis on image modelling problem. Then, following a factorization model approach and by an explicit calculation of the misclassification error rate, we quantify significant improvement on pixel classification made by addition of spatial and contextual knowledge on given images. We give then test experiments for a comparative study on *MAP, ICM, MPM* restoration methods together with some contextual classification procedures. We study also the choice problem for regularization parameters.

The second topic deals with parameter estimations for stationary process on two or higher dimensional lattice. Data tapering technics allows us to remove simultaneously edge effets (bias phenomenon) and non-positivity of unbiased estimators. We first extended already known results for univariate process to the multivariate case and give consistance and asymptotical normality results for spectrographic estimators with tapered data. Then as application, consistance and asymptotical normality of Whittle's estimator are proved. Other applications of Whittle's contrast function for gaussian Markov process and for hypopthesis testing are also given.

Partie I

Autour de la segmentation bayésienne d'image

Introduction

Le problème de la segmentation d'image constitue le thème de cette première partie. Segmenter une image donnée signifie en chercher une certaine *description.* Cette recherche est guidée par une *propriété* P : sur une zone de la segmentation la propriété P est constante. Par exemple en télédétection, P est un indice de texture qui varie suivant que la zone correspond aux forêts, aux terrains cultivés, à l'océan, aux zones urbaines...; en météorologie et examen du ciel, P est un indice d'allongement, d'orientation et de texture des images.

Référons nous ici à une définition de T. Pavlidis[1] : "Etant donnée une définition de *l'uniformité* , une *segmentation* est une partition de l'image en des sous-ensembles connexes dont chacun est uniforme mais de sorte qu'aucune union des sous-ensembles adjacents ne soit uniforme". Il s'agit donc d'une partition connexe de l'ensemble des *pixels* (picture elements) dont chacune des composantes correspond à une région de l'image où la propriété considérée P montre une valeur constante: un *objet* de l'image.

Notons $x = \{x_s, s \in S\}$ l'image où S est l'ensemble des pixels et $\lambda = \{\lambda_s, s \in S\}$ la carte descriptive de la propriété qu'on cherche à estimer en segmentant x. Généralement, la propriété P ne prend qu'un nombre fini de valeurs qualitatives, disons dans $\Omega = \{1, \cdots, K\}$. Les éléments de Ω se nomment *les labels.*

La définition de l'uniformité évoquée ci-dessus constitue l'étape d'une modélisation: avant la segmentation de l'image x, nous avons besoin de savoir comment se comporte l'image x si la carte des labels est λ. Les modèles stochastiques adoptés considèrent cette dépendance comme une loi de probabilité $\mathbf{P}(x|\lambda)$. Cette modélisation peut être soit locale, i.e. dans une fenêtre de taille réduite autour d'un pixel et transportée ensuite sur tout le réseau S par similarité , soit globale sur S tout entier.

Pour que cette modélisation soit efficace, on introduira une information dont dispose l'utilisateur. Cette information *a priori* $\mathbf{P}(\lambda)$ sert à exprimer des exigences que l'on impose à la carte des labels λ à reconstituer. Ces exigences concernent généralement l'interaction des labels $\{\lambda_s, s \in S\}$, leur géométrie relative ou leur dépendance mutuelle.

[1] texte original (en Anglais) dans "Structural Pattern Recognition". Springer: New York, 1977.

De ces modélisations $\mathbf{P}(x|\lambda)$ et $\mathbf{P}(\lambda)$ résultent des méthodes stochastiques bayésiennes de segmentation qui fournissent une estimation $\hat{\lambda} = \{\hat{\lambda}_s, s \in S\}$ de la carte des labels sur la base de *la distribution a posteriori* $\mathbf{P}(\lambda|x)$. Celle-ci s'obtient par la formule de Bayes:

$$\mathbf{P}(\lambda|x) = \mathbf{P}(x|\lambda)\mathbf{P}(\lambda)/\mathbf{P}(x).$$

Nous pouvons répartir l'ensemble des méthodes bayésiennes en les deux classes suivantes:

1. La classification contextuelle des pixels estime le label en chaque pixel $s \in S$, en fonction non de l'image tout entière x mais seulement de sa restriction à une fenêtre autour de s, c-à-d,

   $$\hat{\lambda}_s \quad \text{est fonction de} \quad x(V_s) = \{x_t, t \in V_s\},$$

   où V_s, la fenêtre concernée, possède une taille réduite;

2. La segmentation markovienne globale utilise quant à elle toute l'image x pour l'estimation $\hat{\lambda}_s$ du label en pixel s. Cette approche globale nécessite souvent des modélisations par des champs de Markov (d'où la terminologie) que ce soit pour $\mathbf{P}(x|\lambda)$ ou pour $\mathbf{P}(\lambda)$.

Cette première partie comprend quatre chapitres:

- Le premier chapitre tente une analyse *méthodologique* des différentes segmentations bayésiennes exploitant des modèles stochastiques. La littérature sur ce sujet étant très abondante, ce chapitre tente d'en analyser certaines idées fondamentales. Les deux classes de méthodes mentionnées ci-dessus, contextuelles locales ou globales ont apparu séparément dans les applications. Notre analyse, tout en les distinguant, fournit un point de vu commun de l'ensemble. Notamment, les classifications contextuelles peuvent s'interpréter comme des *variantes localisées* de la méthode globale *MPM*.

- Le second chapitre apporte un nouveau développement à la classification contextuelle. Nous développons une étude quantitative en évaluant la probabilité d'erreur de classification dans une situation où les textures à discriminer peuvent être considérées comme réalisations des processus gaussiens stationnaires et factorisants. Nous avons quantifié l'amélioration apportée par l'introduction du contexte spatial de l'image dans une procédure de classification: la probabilité d'erreur de classification décroît exponentiellement en fonction de la taille de la fenêtre utilisée.

- Le troisième chapitre fournit une étude expérimentale et comparative dans laquelle notamment sont évaluées, qualitativement et quantitativement, les performances respectives des classifications contextuelles (*simple* ou *géométrique*) et des restaurations markoviennes par le *MAP*, l'*ICM* et le *MPM*. Les classification-restaurations sont réalisées sur les mêmes images de test. Cette étude nous a permis d'une part de mieux comprendre, en ce qui concernent les trois segmentations markoviennes, leurs comportements différents vis à vis du choix des paramètres des champs markoviens. Nous avons, par ordre croissant de sensibilité, l'*ICM*, le *MPM* et le *MAP*. D'autre part, selon le critète du poucentage d'erreur de classification, les deux classifications contextuelles fournissent de résultats satisfaisants par rapport aux méthodes globales. De plus, ce deux méthodes, travaillant pixel par pixel sur des vecteurs de faible dimension et ne demandant pas d'algorithme itératif, sont économiques.

- Le quatrième chapitre mène une étude plus approfondie sur le choix de ces paramètres, parfois appelés *paramètres de régularisation.* Le choix par validation croisée y est notamment expérimenté.

Le premier et le troisième chapitre se sont largement inspirés de [I–43], le deuxième de [I–21], le quatrième chapitre reprend sans modification le [I–15].

Chapitre 1

Les méthodes

On considère une image donnée $x = \{x_s, s \in S\}$ où l'ensemble des pixels S est un rectangle $[1,m] \times [1,n]$ de $\mathbb{Z}^2$. La carte des labels d'objet décrivant l'image x doit être estimée. Les labels, qualitatifs, sont en nombre fini, disons $\lambda_s \in \Omega = \{1,\ldots,K\}$.

La littérature sur ce sujet et ses applications étant très abondantes, ce chapitre tente d'en analyser certaines idées fondamentales.

1.1 La classification contextuelle (méthodes locales)

Historiquement, les premiers algorithmes de classification dérivaient directement de la discrimination classique. On assignait, à chaque pixel s, le label $\hat{\lambda}_s$ ne dépendant de l'image x que par sa valeur au pixel s seul:

$$\hat{\lambda}_s = \text{Arg}\max_{j\in\Omega} \mathbf{P}(\lambda_s = j|x_s). \tag{1.1}$$

Cette classification ne tient pas compte des informations spatiales contenues dans l'image x. Ces informations, baptisées sous le nom du *contexte spatial*, sont de diverses natures. Citons deux situations intervenant dans une segmentation de textures d'objet:

1. en général, la taille d'un objet de l'image dépasse considérablement celle représentée par un seul pixel, de sorte que les pixels voisins représentent probablement un même objet et par conséquent pourraient posséder des propriétés similaires;

2. l'emplacement relatif dans l'espace des objets de l'image, fournissent des contraintes géométriques sur la carte des textures $\{\lambda_s, s \in S\}$.

Nous appelons cette classification (1.1) *la classification aveugle*, car elle ne "voit" pas le contexte spatial de l'image x.

Le souci d'une plus grande efficacité invite naturellement à intégrer ce contexte spatial dans les algorithmes de classification. Les premiers germes de cette idée ont été introduits par WELSH & SALTER[I–47]. Leur algorithme, dit *contextuel*, fait intervenir, afin de classer un pixel s, l'ensemble des données de l'image x dans une fenêtre autour de s. Bien que cet élargissement n'atteigne pas encore l'image x tout entière, on peut raisonnablement espérer une classification meilleure que la classification aveugle.

Précisons maintenant cette notion de classification contextuelle.

Définition 1 *Une* classification contextuelle des pixels[1] *consiste à retenir l'estimation des labels* $\hat{\lambda} = \{\hat{\lambda}_s, s \in S\}$ *où chaque label* $\hat{\lambda}_s$ *retenu maximise, conditionnellement à une imagette* $x(V_s)$, *la* marginale a posteriori locale *suivante*

$$\hat{\lambda}_s = \operatorname{Arg}\max_{j \in \Omega} \mathbf{P}\left[\lambda_s = j | x(V_s)\right]. \tag{1.2}$$

V_s *représente une fenêtre autour de* s *et l'imagette* $x(V_s)$ *la restriction de l'image* x *à cette fenêtre.* ♣

Généralement, les fenêtres $\{V_s\}$ sont des translatées d'une fenêtre V_0 de l'origine:

$$V_s = \{s\} + V_0 \ , \tag{1.3}$$

avec l'ajustement nécessaire pour des pixels frontières du réseau S. La fenêtre originelle V_0 prend couramment la forme suivante

$$V_0 = \left\{ s \in \mathbb{Z}^2 : \|s\|_2 \leq q \right\} \ , \tag{1.4}$$

où la norme $\|s\|_2$ d'un pixel $s = (a, b)$ vaut

$$\|s\|_2 = \sqrt{a^2 + b^2}.$$

Avec q égal successivement à $1, \sqrt{2}, 2, \sqrt{5}, \ldots$, la fenêtre V_0, ainsi que les V_s, sont dites du premier, second, troisième, ... , ordre (cf. Figure 1.1).

Notons $V_s^* = V_s \backslash \{s\}$ la fenêtre privée du pixel central s et $d = |V_s^*|$ son cardinal. La marginale a posteriori locale apparaissant dans la règle contextuelle (1.2) peut encore s'écrire sous la forme qui suit, avec $\lambda(V_s^*) = (\lambda_t, t \in V_s^*)$,

$$\begin{aligned} \mathbf{P}\left[\lambda_s = j | x(V_s)\right] &= \sum_{\rho \in \Omega^d} \mathbf{P}\left[\lambda_s = j, \lambda(V_s^*) = \rho | x(V_s)\right] \\ &= \mathbf{P}[x(V_s)]^{-1} \sum_{\rho \in \Omega^d} \mathbf{P}\left[x(V_s) | \lambda(V_s) = (j, \rho)\right] \times \mathbf{P}\left[\lambda(V_s) = (j, \rho)\right]. \end{aligned} \tag{1.5}$$

[1] WELSH & SALTER[I–47] tiennent compte aussi d'une fonction de coût plus complexe. Notre formulation correspond à une fonction de coût additive ordinaire et qui vaut, en un pixel s, 1 si $\hat{\lambda}_s \neq \lambda_s$.

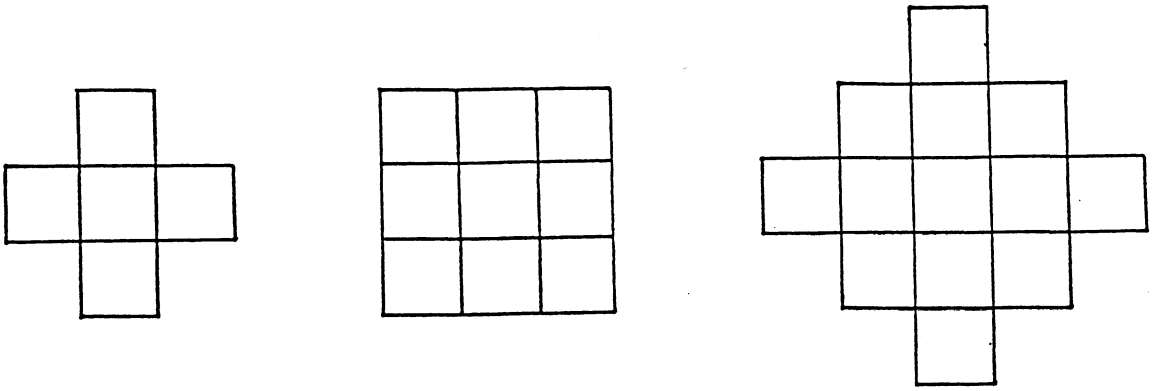


Figure 1.1: Exemple des fenêtres V_0: du premier au troisième ordre.

Ici, j et ρ sont respectivement des labels au pixel s et sur V_s^*.

Cette décomposition indique clairement qu'une classification contextuelle nécessite deux types d'informations locales sur les fenêtres $\{V_s\}$:

1. une modélisation de $\mathbf{P}[x(V_s)|\lambda(V_s)]$, le comportement de l'image par rapport aux labels dans des fenêtres;
2. une modélisation a priori $\mathbf{P}[\lambda(V_s)]$ sur l'interaction souhaitée des labels, dans ces mêmes fenêtres.

Nous allons séparément examiner ces deux types d'informations.

Sur $\mathbf{P}[x(V_s)|\lambda(V_s)]$ — *le comportement de l'image par rapport au label.*
La nature de l'image x et du label λ change d'une situation à l'autre, ce qui explique la multitude des modèles adoptés pour le comportement conditionnel image/label. Une approche répandue fait l'hypothèse de l'indépendance conditionnelle sur des fenêtres $\{V_s\}$, c'est à dire,

$$\mathbf{P}[x(V_s)|\lambda(V_s)] = \prod_{t \in V_s} \mathbf{P}(x_t|\lambda_t). \tag{1.6}$$

pour tout $s \in \Omega$. WELSH & SALTER[I–47], SWAIN & al.[I–40], HASLETT [I–23], KITTLER & FÖGLEIN[I–28] , HJORT[I–25], DEVIJVER & al.[I–12] entre autres retiennent cette hypothèse. Quant à $\mathbf{P}(x_t|\lambda_t)$, lorsque l'image x représente des niveaux de gris, il est souvent modélisé par une variable gaussienne $\mathcal{N}(\tilde{\mu}_j, \Sigma_j)$ si $\lambda_t = j$. Par rapport à ce choix, l'apport de SWITZER[I–41] est significatif: il propose une modélisation jointe, sur la fenêtre V_s, des

variables $(x(V_s)|\lambda(V_s))$: $(x(V_s)|\lambda(V_s))$ sera $p \times |V_s|$–multivarié si x_s est p–multivarié, de loi connue ou déterminée après un processus d'apprentissage. Nous reviendrons sur cette approche dans le Chapitre 2.

Sur $\mathbf{P}[\lambda(V_s)]$ — *le comportement des labels.*
L'idée est toujours de préserver un minimum de "régularité" spatiale dans la répartition des labels. On distingue deux types d'approche: l'approche dite *géométrique* et l'approche *markovienne*. Ce qui nous amène à une analyse séparée de ces deux courants d'idées dans les prochains paragraphes (1.1.1) et (1.1.2).

1.1.1 Modélisation géométrique de $\lambda(V_s)$

La formule (1.5) évaluant les marginales a posteriori locales invite à la constatation suivante: parmi les nombreuses $|\Omega|^{|V_s|}$ configurations locales possibles des labels sur la fenêtre V_s, beaucoup sont, pour des raisons d'homogénéité d'objets, très improbables, i.e. d'une probabilité négligeable. On ne retiendra alors qu'un nombre limité de configurations, disons probables, garantissant la régularité souhaitée des objets. Les *modèles géométriques* de HJORT[I–25](voir aussi OWEN[I–33]) retiennent comme seules configurations locales $\lambda(V_s)$ celles où:

- d'une part deux labels distincts au plus peuvent intervenir,
- d'autre part, un sous-ensemble de V_s de label donné est connexe.

Ces restrictions conduisent à une réduction considérable du nombre des configurations $\{\lambda(V_s)\}$ possibles. L'évaluation de la probabilité marginale locale (1.5) s'en trouve nettement simplifiée.

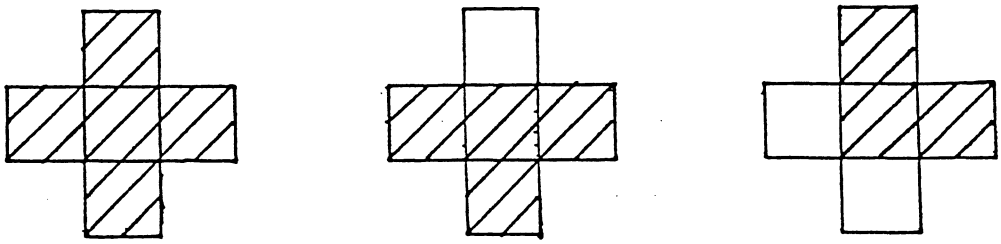
Décrivons plus précisément le modèle géométrique du premier ordre: le *p-q-r* modèle ([I–25],[I–33]). V_s est constitué du pixel s et de ses quatre plus proches voisins (Nord-Est-Sud-Ouest). Trois types de configurations locales $\mathcal{X}, \mathcal{T}, \mathcal{L}$ sont retenus (cf. Figure 1.2).

Le type $\mathcal{X}$ correspond à la situation où le pixel central est intérieur à une région homogène; les types $\mathcal{T}$ et $\mathcal{L}$ à celles où il est au bord d'une région.

Ces configurations locales seront notées comme suit

$$\begin{aligned}
&\text{Type } \mathcal{X} \quad : \quad \mathcal{X}(j), \quad j \in \Omega, \\
&\text{Type } \mathcal{T} \quad : \quad \mathcal{T}(j,k;\ell), \quad j,k \in \Omega, \quad j \neq k \ ,; \ \ell = 1,\ldots,4, \\
&\text{Type } \mathcal{L} \quad : \quad \mathcal{L}(j,k;\ell), \quad j,k \in \Omega, \quad j \neq k \ ,; \ \ell = 1,\ldots,4.
\end{aligned}$$

L'indice ℓ repère l'orientation des configurations du type $\mathcal{T}$ et $\mathcal{L}$ correspondant respectivement à une rotation de 0°, 90°, 180°, 270°. Ainsi, le nombre de termes



Type $\mathcal{X}$ Type $\mathcal{T}$ Type $\mathcal{L}$

Figure 1.2: Modèle p-q-r: les 3 types de configurations locales retenues $\mathcal{X}$, $\mathcal{T}$ et $\mathcal{L}$ (et pour $\mathcal{T}$, $\mathcal{L}$ les 3 autres configurations obtenues par rotation de 90°, 180°, 270°). Deux labels: vide ou hachuré.

sommés dans (1.5) passe de $|\Omega|^4$ à $(8|\Omega|-7)$. Notons p, $q/4$, $r/4$ respectivement la probabilité d'une configuration du type $\mathcal{X}$, $\mathcal{T}$, $\mathcal{L}$ avec $p+q+r=1$, $\{\pi_j\}$ la loi a priori sur Ω, c'est à dire pour tout $s \in \Omega$

$$\mathbf{P}(\lambda_s = j) = \pi_j\,,\ j \in \Omega. \tag{1.7}$$

Alors les probabilités marginales locales $\mathbf{P}(\lambda(V_s))$ retenues sont:

$$\begin{aligned} \mathbf{P}[\lambda(V_s) = \mathcal{X}(j)] &= \pi_j(p + \pi_j(q+r)), \\ \mathbf{P}[\lambda(V_s) = \mathcal{T}(j,k;\ell)] &= \frac{1}{4}\pi_j\pi_k q, \\ \mathbf{P}[\lambda(V_s) = \mathcal{L}(j,k;\ell)] &= \frac{1}{4}\pi_j\pi_k r. \end{aligned} \tag{1.8}$$

où $j,k \in \Omega$, $j \neq k$, et $\ell = 1,\ldots,4$.

Les marginales locales a posteriori (1.5) deviennent:

$$\mathbf{P}[\lambda_s = j | x(V_s)] = \mathbf{P}[x(V_s)]^{-1}\pi_j(f_{\mathcal{X}} + f_{\mathcal{T}} + f_{\mathcal{L}})\ , \tag{1.9}$$

avec

$$\begin{aligned} f_{\mathcal{X}} &= (p + \pi_j(q+r))\ \mathbf{P}[x(V_s)|\mathcal{X}(j)]\ , \\ f_{\mathcal{T}} &= \frac{q}{4}\sum_{k\neq j}\pi_k \sum_{\ell=1}^{4} \mathbf{P}[x(V_s)|\mathcal{T}(j,k;\ell)]\ , \\ f_{\mathcal{L}} &= \frac{r}{4}\sum_{k\neq j}\pi_k \sum_{\ell=1}^{4} \mathbf{P}[x(V_s)|\mathcal{L}(j,k;\ell)]\ . \end{aligned}$$

La dimension paramétrique de $\{\mathbf{P}(\lambda(V_s))\}$ est passée de $(|\Omega|^5 - 1)$ à $(|\Omega| + 1)$ (les π_j plus p, q, r).

Donnons la

Définition 2 *Une* classification contextuelle géométrique, *ou plus simplement une* classification géométrique, *est une classification contextuelle suivant des marginales locales a posteriori* $\mathbf{P}[\lambda_s = j|x(V_s)]$ *du type (1.9), i.e. obtenues par une réduction géométrique des configurations locales.* ♣

Notons que HJORT[I–25] propose aussi des procédures d'estimation des paramètres contextuels (p, q, r). Si l'on dispose d'un échantillon où les objets correspondants sont convenablement identifiés (données d'apprentissage), l'estimation est facile. Sinon, un procédé d'estimation directe à partir de l'image x, du type EM, basé sur des vraisemblances jointes des fenêtres éloignées, est construit.

Naturellement, cette méthode de réduction s'étend sur des fenêtres d'ordre supérieur. Dans le cas du second ordre, les configurations locales que l'on pourrait retenir sont illustrées dans la Figure 1.3. Il s'agit du *p-q-r-s-t-u* modèle de HJORT[I–25].

Introduisons maintenant une classification géométrique encore plus simple: la *classification contextuelle simple* CCS. Bien que le procédé soit antérieur, on le comprend mieux sous l'angle de la règle géométrique (1.9). SWITZER[I–41] considère une classification dans laquelle il existe une complète continuité spatiale locale qui correspond à ne retenir parmi les configurations locales $\lambda(V_s)$ que celles du type $\mathcal{X}$, c'est à dire celle où le label est constant sur V_s (cf. Figure 1.2). D'une manière équivalente, on a fait le choix $p = 1$.

Nous résumons cette situation par la

Définition 3 *Une* classification contextuelle simple *est la règle suivante:*

$$\hat{\lambda}_s = \operatorname{Arg}\max_{j\in\Omega} \pi_j \mathbf{P}[x(V_s)|\mathcal{X}(j)] \tag{1.10}$$

$$= \operatorname{Arg}\max_{j\in\Omega} \pi_j \mathbf{P}[x(V_s)|\lambda_t = j, t \in V_s] \tag{1.11}$$

♣

A la suite de [I–41], MARDIA[I–31] précise un modèle statistique de $\mathbf{P}[x(V_s)|\mathcal{X}(j)]$ qui lui permet de mettre en évidence, par des calculs analytiques explicites des erreurs de classification, la nette amélioration apportée par l'intégration du contexte spatial. Ces calculs sont étendus à d'autres modélisations ([I–21]) et le contrôle des erreurs de classification sera présenté au Chapitre 2.

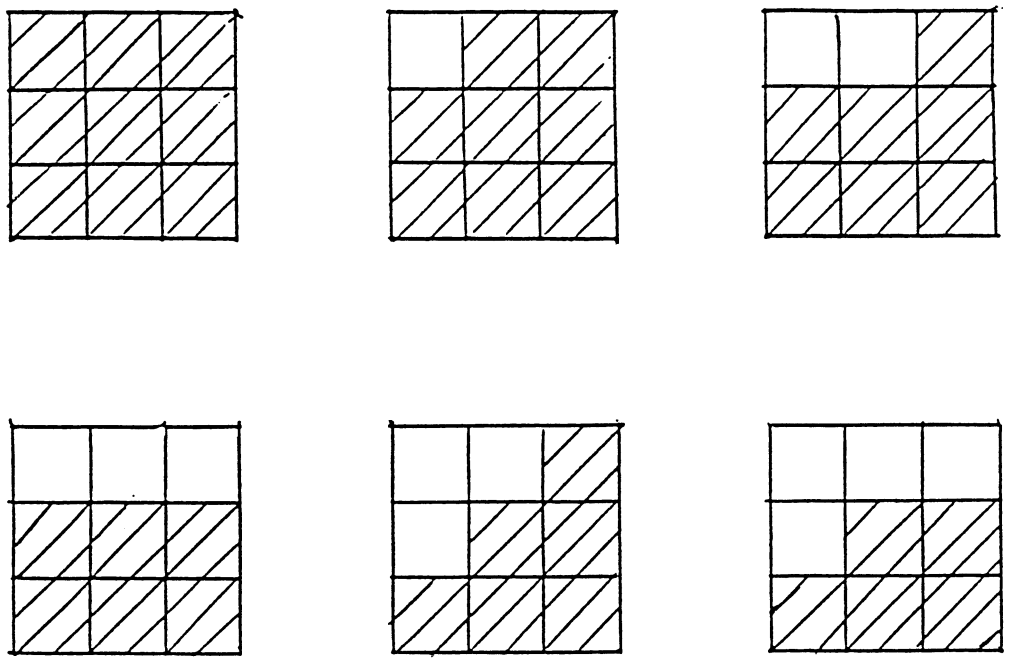


Figure 1.3: Les configurations locales retenues dans le modèle *p-q-r-s-t-u* (D'autres configurations locales sont obtenues par rotation). Deux labels: vide ou hachuré.

1.1.2 Modélisation markovienne de $\lambda(V_s)$

Les *champs markoviens*[2] fournissent un outil naturel pour modéliser l'interaction spatiale et la géométrie régulière des labels d'objet. Le choix des paramètres d'un champ markovien permet notamment de contrôler la taille, l'orientation ... des régions homogènes présentes dans l'image.

Chronologiquement CHOW[I–7] semble être parmi les premiers à avoir exploité l'interaction des labels se trouvant à proximité dans un algorithme de classification. Pour sa procédure de reconnaissance de caractères manuscrits, il utilise implicitement un champ de Markov unilatéral où chaque site est conditionnellement dépendant de ses deux voisins Nord et Ouest. Ce type de champs markoviens unilatéraux, *champs à germes markoviens (Markov mesh fields)*, sont indépendamment introduits par ABEND & al.[I–1] d'une manière précise. Par la suite apparaît l'utilisation des MRF généraux non causaux en analyse d'image:

[2]Ces champs sont aussi appelés *champs de Gibbs* et seront désignés dans la suite par le sigle MRF (Markov random field).

initialement destinés à modéliser des textures dans HASSNER & SKLANSKY[I–24] et CROSS & JAIN[I–9] , le premier algorithme de segmentation markovienne globale, s'appuyant sur le *recuit simulé* est donné par S. & D. GEMAN[I–16] .

La segmentation markovienne globale fera l'objet du prochain paragraphe (1.2). Ici nous allons analyser des approches markoviennes apparues dans des classifications contextuelles, essentiellement celles issues des champs à germes markoviens.

La définition suivante vient de [I–1].

Définition 4 *Un champ* $\{\lambda_s, s \in S\}$ *est* un champ à germes markoviens *si pour tout site* $s = (a, b)$, λ *possède la propriété de Markov unilatérale suivante:*

$$\mathbf{P}[\lambda_{a,b} | \lambda_{x,y} : x < a \text{ ou } y < b] = \mathbf{P}[\lambda_{a,b} | U_{a,b}], \tag{1.12}$$

où $U_{a,b}$ *est un petit ensemble des voisins "Nord-Ouest" (cf. Figure 1.4):*

$$U_{a,b} \subset \{(x, y) \in S : x \leq a \text{ et } y \leq b\} \ . \tag{1.13}$$

qui se trouvent soit au dessus soit à gauche de s . ♣

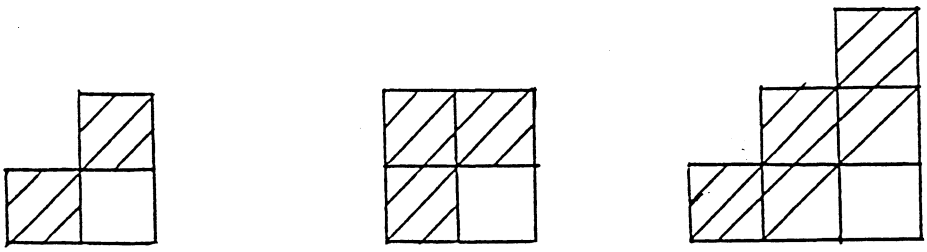


Figure 1.4: Exemple de $U_{a,b}$: voisins unilatéraux (hachurés) du pixel (a, b).

Par exemple, un champ à germes markoviens admettant comme relation de voisinage les 3 plus proches voisins unilatéraux est défini par:

$$\mathbf{P}[\lambda_{a,b} | \lambda_{x,y} : x < a \text{ ou } y < b] = \mathbf{P}\left[\lambda_{a,b} \,\middle|\, \begin{matrix} \lambda_{a-1,b-1} & \lambda_{a-1,b} \\ \lambda_{a,b-1} & \end{matrix}\right], \tag{1.14}$$

avec les ajustements nécessaires au bord.

L'absence d'un ordre naturel dans le plan $\mathbb{Z}^2$ fait que les propriétés des chaînes de Markov ne se généralisent pas aux champs markoviens généraux. Les champs à germes markoviens constituent de ce point de vue une généralisation directe des chaînes de Markov: la notion de causalité pour l'ordre lexicographique est conservée et permet un calcul direct récursif pour la probabilité jointe sur un

réseau S. Pour le champ (1.14) et avec $S = [1,m] \times [1,n]$, nous avons pour la probabilité jointe:

$$\mathbf{P}[\lambda_{a,b} \ : \ 1 \leq a \leq m,\ 1 \leq b \leq n] = \prod_{a=1}^{m} \prod_{b=1}^{n} \ \mathbf{P}\left[\lambda_{a,b} \left| \begin{array}{ll} \lambda_{a-1,b-1} & \lambda_{a-1,b} \\ \lambda_{a,b-1} & \end{array} \right.\right], \quad (1.15)$$

avec des ajustements convenables pour des sites-bords de S: $\{(a,1)$ et $(1,b) : 1 \leq a \leq n, 1 \leq b \leq m\}$. Ce développement permet d'entrevoir l'intérêt particulier de ces modèles pour une classification contextuelle: les marginales locales $\{\mathbf{P}[\lambda(V_s)]\}$ sur des fenêtres s'expriment explicitement à partir des paramètres de base. Par exemple, les probabilités de transitions suivantes:

$$p^{x}_{jk} = \mathbf{P}[\lambda_{2,1} = k | \lambda_{1,1} = j],\ \ p^{y}_{jk} = \mathbf{P}[\lambda_{1,2} = k | \lambda_{1,1} = j],$$

$$p^{xy}_{j\mu\nu k} \ = \ \mathbf{P}\left[\lambda_{2,2} = k \left| \begin{array}{ll} \lambda_{1,1} = j & \lambda_{1,2} = \mu \\ \lambda_{2,1} = \nu & \end{array} \right.\right],$$

combinées avec une loi à priori $\{\pi_j, j \in \Omega\}$ suffisent à déterminer la distribution jointe (1.15), en particulier sur des fenêtres $\{V_s\}$.

Un type particulier des champs à germes markoviens, les *champs de Pickard*, a attiré l'attention de bien des spécialistes. Il s'agit de champs stationnaires générés à partir d'une distribution quadruple $\mathbf{P}_c$ sur un carré (PICKARD[I–34],[I–35]):

$$\begin{pmatrix} (1,1) & (1,2) \\ (2,1) & (2,2) \end{pmatrix},$$

qui est supposée invariante par toute symétrie du carré (homogénéïté) et vérifie l'indépendance conditionnelle suivante des deux directions:

$$\mathbf{P}_c[\lambda_{2,1}, \lambda_{1,2} \mid \lambda_{1,1}] = \mathbf{P}_c[\lambda_{2,1} | \lambda_{1,1}] \ \times \ \mathbf{P}_c[\lambda_{1,2} | \lambda_{1,1}] \ .$$

La distribution quadruple $\mathbf{P}_c$ est alors spécifiée par les 3 distributions:

$$\mathbf{P}_c\left(\lambda_{2,2} \left| \begin{array}{ll} \lambda_{1,1} & \lambda_{1,2} \\ \lambda_{2,1} & \end{array} \right.\right), \ \ \mathbf{P}_c(\lambda_{1,2} | \lambda_{1,1}), \ \ \mathbf{P}_c(\lambda_{1,1}).$$

Dans le cas binaire, $\lambda_{i,j} \in \{-1,1\}$, PICKARD[I–34] remarque que $\mathbf{P}_c$ est déterminée par les 3 paramètres suivants:

$$\theta = \mathbf{P}_c(\lambda_{1,1} = 1),\ \ a = \mathbf{P}_c(\lambda_{1,2} = 1 | \lambda_{1,1} = 1) \ , \quad (1.16)$$

$$d = \mathbf{P}_c\left(\lambda_{2,2} = 1 \left| \begin{array}{ll} \lambda_{1,1} = 1 & \lambda_{1,2} = 1 \\ \lambda_{2,1} = 1 & \end{array} \right.\right) \ .$$

Explicitons maintenant la construction des champs de Pickard ([I–34]): étant donnée la loi jointe de deux variables discrètes Z_1, Z_2, on calcule les marginales de

Z_1, Z_2 et la loi conditionnelle $Z_2|Z_1$ pourvu que Z_1, Z_2 possèdent les mêmes atomes. On pose $Z_{i+1}|Z_1, \ldots, Z_i \stackrel{\mathcal{D}}{=} Z_{i+1}|Z_i$, pour $i = 1, 2, \cdots$. Alors $\{Z_i\}_1^\infty$ est une chaîne de Markov homogène. Si en plus $Z_1, Z_2 \stackrel{\mathcal{D}}{=} Z_2, Z_1$, la chaîne est stationnaire et la chaîne inverse lui est identique.

Soit MC_* la chaîne de Markov homogène et stationnaire horizontalement générée selon ce procédé avec $Z_1 = \begin{pmatrix} \lambda_{1,1} \\ \lambda_{2,1} \end{pmatrix}$, $Z_2 = \begin{pmatrix} \lambda_{1,2} \\ \lambda_{2,2} \end{pmatrix}$ et leur distribution jointe $\mathbf{P}_c$. Un segment $Z_{j+1}, \ldots, Z_{j+n}$ de longueur n de cette chaîne spécifie une distribution jointe sur un lattice $2 \times n$ qui ne dépend pas de j. Soient alors $W_k = (\lambda_{k,j+1}, \cdots, \lambda_{k,j+n})$, $k = 1, 2$ les deux lignes de ce segment. Une nouvelle chaîne de Markov MC_n homogène et stationnaire générée selon le même procédé ci-dessus mais avec W_1, W_2 comme générateurs constitue un champ homogène sur $\mathbb{Z} \times [1, n]$. Un champ de Pickard $\mathbf{P}_{mn}$ sur un lattice fini $[1, m] \times [1, n]$ est défini comme un segment de longueur m de la chaîne MC_n. Pickard[I–35] démontre alors que $\mathbf{P}_{mn}$ est un champ stationnaire à germes markoviens dont la dépendance conditionnelle unilatérale relève des 3 plus proches voisins comme indiqué par l'équation (1.14).

Les champs de Pickard possèdent plusieurs propriétés intéressantes comme modèles de labels pour une classification contextuelle ([I–23],[I–12]). Donnons un exemple: si on note V_s^* l'ensemble des 4 plus proches voisins de s, conditionnellement à λ_s, les $\{\lambda_t,\ t \in V_s^*\}$ sont indépendants:

$$\mathbf{P}[\lambda(V_s^*) \mid \lambda_s] = \prod_{t \in V_s^*} \mathbf{P}[\lambda_t | \lambda_s]. \tag{1.17}$$

Remarquons que le membre de droite est complètement déterminé par la distribution basique $\mathbf{P}_c$. L'adoption d'un tel modèle de $\lambda(V_s)$, jointe à l'indépendance conditionnelle (1.6) $x(V_s)/\lambda(V_s)$ de l'image par rapport au label dans la fenêtre V_s conduit à la classification contextuelle "voisinage" d' HASLETT[I–23]:

$$\hat{\lambda}_s = \operatorname{Arg} \max_{j \in \Omega} \mathbf{P}\left[\lambda_s = j | x(V_s)\right]. \tag{1.18}$$

où:

$$\mathbf{P}\left[\lambda_s = j | x(V_s)\right] \ \propto \ \mathbf{P}[x_s | \lambda_s = j] \prod_{t \in V_s^*} H_t(j) \ ,$$

avec pour $t \in V_s^*$,

$$H_t(j) \ = \ \sum_{k \in \Omega} \mathbf{P}[x_t | \lambda_t = k]\, \mathbf{P}[\lambda_t = k | \lambda_s = j] \ .$$

Dans cette approche, par rapport à la classification aveugle, les $\{H_t(j)\}$ apparraissent comme des facteurs de corrections dûs à l'introduction du contexte. Une extension de cette classification se trouve également proposée dans [I–23] où les quatres voisins sont remplacés par la ligne et la colonne qui contiennent le pixel central s.

Les modèles à germes markoviens sont largement exploités par DEVIJVER & ses collaborateurs. Citons les travaux récents de [I–12] et [I–13]. Dans [I–13], DEVIJVER propose une méthode semi-localisée (avec $s = (a, b)$):

$$\hat{\lambda}_{a,b} = \text{Arg} \max_{j \in \Omega} \mathcal{L}_{a,b}(j) \quad , \tag{1.19}$$

avec

$$\mathcal{L}_{a,b}(j) = \mathbf{P}\left(\lambda_{a,b} = j | x_{ik} : 1 \leq i \leq a+1,\ 1 \leq k \leq b+1\right) \quad .$$

Utilisant un champ "d'ordre 3" (1.14), la démarche proposée possède la particularité que les marginales locales a posteriori $\{\mathcal{L}_{a,b}(j)\}$ peuvent être récursivement calculées suivant un balayage ligne par ligne du réseau S. Le problème important d'apprentissage, i.e. l'estimation des paramètres du modèle adopté, y est aussi examiné.

1.2 La segmentation markovienne (méthodes globales)

Les champs markoviens ont déjà été abordés dans le paragraphe précédent (1.1.2) lorsqu'ils interviennent dans une classification contextuelle. Cependant, la presque totalité des champs exploités dans ce contexte sont unilatéraux à germes markoviens ou bénéficient de certaines conditions d'indépendance conditionnelle.

En segmentation d'image et plus généralement en vision par ordinateur, l'intérêt croissant pour la modélisation par champs markoviens fait suite au travail de S. & D. GEMAN[I–16]. La modélisation markovienne est une façon de traduire l'information a priori que l'on a sur l'objet et le problème d'optimisation sur un espace de configurations de très grand cardinal (pour une image binaire 10×10, il y a 2^{100} configurations) peut être résolu via un schéma de recuit simulé (le recuit sera précisé dans le paragraphe 1.2.1). Concernant le recuit et ses applications, on pourra consulter [I–27], [I–16], [I–22], [I–42]. Différentes approches utilisant des modélisations markoviennes en segmentation de textures ou en détection de contours, ... sont proposées par exemple dans [I–19], [I–18], [I–10], [I–20], [I–8], [I–4], [I–5] [I–14].

Reprenons les notations du début de ce chapitre: $\{x_s, s \in S\}$ et $\{\lambda_s, s \in S\}$ sont respectivement l'image observée et sa carte de labels à reconstituer. L'approche markovienne de segmentation modélise le processus de label λ par un champ de Markov (ou de Gibbs):

$$\mathbf{P}(\lambda) = \frac{1}{Z} \exp\left(-U(\lambda)/T\right) \quad , \tag{1.20}$$

où par analogie au langage de la mécanique statistique, $U(\lambda)$ et T sont respectivement *l'énergie* et *la température* du système. Z est *la fonction de partition*

qui normalise la distribution. Le modèle d'Ising figure parmi les champs de Gibbs historiques: l'ensemble des états $\Omega = \{+1, -1\}$ correspond aux deux orientations magnétiques des spins en chaque site et l'énergie du système est donnée par:

$$U(\lambda) = - \sum_{\langle s,t \rangle} \lambda_s \lambda_t \ , \tag{1.21}$$

où $\langle s,t \rangle$ signifient que les sites s et t sont voisins. Un état stable du système, i.e. d'énergie minimale, est un état dans lequel les spins voisins s'orientent plutôt d'une manière similaire. On imagine sans peine que le modèle d'Ising peut être aussi un modèle décrivant des plages de label constant géométriquement régulières.

Passons maintenant à la modélisation de l'image x conditionnellement au label λ. Ces structures $\mathbf{P}(x|\lambda)$ restent aussi variées que de différentes applications existantes. Soit $D(x|\lambda)$ définie par:

$$\mathbf{P}(x|\lambda) = \exp\left(-D(x|\lambda)\right) \ , \tag{1.22}$$

$D(x|\lambda)$ correspond à une certaine mesure de la "distorsion" de l'image x par rapport au label λ. En guise d'exemple, nous allons décrire une modélisation par champ markovien gaussien, employée par COHEN & COOPER[I–8] dans leur algorithme de segmentation.

Exemple de modélisation de $\mathbf{P}(x|\lambda)$ ([I–8]):

- x est une image de niveau de gris. Conditionnellement à λ donné , l'ensemble des pixels S admet une répartition en régions homogènes:

  $$S = S(1) \cup S(2) \cup \ldots \cup S(K) \ ,$$

  avec

  $$S(j) = \{s \in S : \ \lambda_s = j\} \ , \ j \in \Omega \ .$$

  Le modèle conditionnel est un champ de Gibbs:

  $$\mathbf{P}(x|\lambda) = \exp\left(-D(x|\lambda)\right) \ ,$$

  Sur chaque région homogène $S(j)$, la contribution à cette énergie, $D_j(x|\lambda)$ suit une forme gaussienne:

  $$\begin{aligned} D_j(x|\lambda) &= D_j\left(x(S_j)|\lambda(S_j)\right) \\ &= \frac{1}{2}\left({}^tW(j)\, Q(j)\, W(j) - \log\det Q(j)\right) \ , \end{aligned} \tag{1.23}$$

  où $W(j) = [\ (x_s - \mu_j),\ s \in S(j)\ ]$, $Q(j) = [\ Q_{st}(j),\ s,t \in S(j)\ ]$ avec

  $$Q_{st}(j) = \begin{cases} 1/\sigma_j^2 & \text{si } s = t \\ -\alpha_j(\tau)/\sigma_j^2 & \text{si } \langle s,t \rangle_\tau \\ 0 & \text{sinon} \end{cases} \ .$$

$\langle s,t\rangle_\tau$ signifie que les pixels s,t sont des voisins à distance $\tau = s-t$. Les $\alpha_j(\tau)$ sont nuls si $||\tau||$ dépasse un certain seuil q. Ainsi, chaque type de texture j est paramétré par

$$\left\{\mu_j, \sigma_j^2, \alpha_j(\tau),\ ||\tau|| \leq q\right\} \quad ,$$

qui représentent respectivement le niveau moyen de la texture, sa variance conditionnelle résiduelle et les coefficients de dépendance conditionnelle avec les pixels voisins (cf. [I–43] et Partie II pour la généralisation au cas vectoriel). L'énergie totale $D(x|\lambda)$ est la résultante des contributions, supposées indépendantes, de différentes régions:

$$D(x|\lambda) = \sum_{j=1}^{K} D_j\left(x(S_j)|\lambda(S_j)\right) \ . \qquad \spadesuit$$

La segmentation bayésienne globale s'obtiendra à partir des deux modélisations $\mathbf{P}(\lambda)$ et $\mathbf{P}(x|\lambda)$: la distribution a posteriori $\mathbf{P}(\lambda|x)$ est encore un champ de Gibbs:

$$\mathbf{P}(\lambda|x) = \frac{1}{Z_x} \exp\left(-U(\lambda|x)/T\right) \quad , \tag{1.24}$$

avec l'énergie a posteriori

$$U(\lambda|x) = D(x|\lambda) + U(\lambda)/T \ . \tag{1.25}$$

Par exemple, lorsque deux textures seulement interviennent et si on adopte simultanément le modèle d'Ising (1.21) et le modèle gaussien (1.23), cette énergie a posteriori prend la forme

$$U(\lambda|x) = \frac{1}{2} \sum_{j=1,-1} \left({}^t W(j)\, Q(j)\, W(j) - \log\det Q(j)\right) \ - \ \frac{1}{T} \sum_{\langle s,t\rangle} \lambda_s \lambda_t \ .$$

Ces distributions a posteriori servent de base à trois techniques de segmentations markoviennes que nous allons décrire maintenant.

1.2.1 Segmentation par le MAP

Cette méthode du *Maximum A Posteriori* consiste à choisir $\hat{\lambda}$ qui maximise la probabilité a posteriori globale ([I–16]):

$$\hat{\lambda}_{\text{map}} = \text{Arg}\max_{\lambda\in\Omega^S} \mathbf{P}(\lambda|x) = \text{Arg}\min_{\lambda\in\Omega^S} U(\lambda|x). \tag{1.26}$$

Elle correspond à la règle bayésienne associée à la fonction de coût global:

$$C(\lambda,\hat{\lambda}) = \begin{cases} 1 & \text{s'il existe } s \text{ , tel que } \lambda_s \neq \hat{\lambda}_s \\ 0 & \text{sinon.} \end{cases} \tag{1.27}$$

Nous avons déjà signalé que la difficulté de la méthode réside dans la complexité numérique de la minimisation de l'énergie a posteriori $U(\lambda|x)$. L'ensemble des configurations Ω^S est fini et de grande cardinalité. L'algorithme du recuit simulé apporte une solution à ce problème. Celui-ci exploite un procédé d'échantillonnage pour une distribution de Gibbs, appelé *l'échantillonneur de Gibbs.* Nous allons donner un bref rappel de ces deux mécanismes.

La description suit S. & D. GEMAN[I–16] . Notons dans ce paragraphe une distribution de Gibbs quelconque sur Ω^S par

$$\Pi(\lambda) = \frac{1}{Z} \exp\{-V(\lambda)/T\}. \tag{1.28}$$

Echantillonneur de Gibbs. Le procédé est itératif. On commence par fixer un ordre de parcours infini du réseau S. Supposons qu'à des instants $1, 2, \ldots, t, \ldots$, on visite des sites (pixels) $n_1, n_2, \ldots, n_t, \ldots$. On va construire une chaîne de configurations $\{\lambda_t\}$. L'échantillonneur démarre avec une configuration initiale η. A chaque instant, la transition se réalise par le relaxation de la configuration courante en un seul site, le site visité n_t. C'est-à-dire,

$$\lambda_s(t+1) = \lambda_s(t) \ , \quad \text{si} \ \ s \neq n_t,$$

et en site n_t, on affecte à $\lambda_{n_t}(t+1)$ une réalisation $\sigma \in \Omega$ tirée à partir de *la caractéristique locale conditionnelle* du site

$$\Pi(\lambda_{n_t} = j | \lambda_s = \lambda_s(t), \ s \neq n_t), \tag{1.29}$$

qui est par définition la loi conditionnelle du site par rapport au reste du réseau.

La convergence de l'échantillonneur est assurée par le théorème suivant ([I–16]):

Théorème A (Relaxation stochastique). *Supposons que la suite $\{n_t, t \geq 1\}$ repasse infiniment souvent en chaque site $s \in S$. Alors, quel que soit la configuration initiale $\eta \in \Omega^S$ et pour tout $\omega \in \Omega^S$,*

$$\lim_{t \to \infty} \mathbf{P}(\lambda(t) = \omega | \lambda(0) = \eta)) = \Pi(\omega). \tag{1.30}$$

♣

De plus, l'échantillon est ergodique: si $Y(\omega)$ est une fonction sur Ω^S, nous avons ([I–16]):

Théorème B (Ergodicité). *Supposons qu'il existe un τ de sorte que $S \subset \{n_{t+1}, n_{t+2}, \ldots, n_{t+\tau}\}$ quel que soit t. Alors, pour toute fonction Y sur*

Ω^S et quelle que soit la configuration initiale η, la convergence

$$\lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} Y(\lambda(t)) = \int_{\Omega^S} Y(\omega)\, d\Pi(\omega) \ , \tag{1.31}$$

a lieu avec une probabilité 1. ♣

Le recuit simulé. Notons par Π_T la dépendance en la température T de la distribution de Gibbs (1.28). Soit Ω_0 l'ensemble des configurations minimales:

$$\Omega_0 = \{\omega \in \Omega^S : V(\omega) = \min_{\lambda} V(\lambda)\}.$$

Notre propos est de trouver une configuration minimale. Intuitivement, on observe que si T tend vers 0, la distribution Π_T converge vers Π_0, la distribution uniforme sur Ω_0. D'où l'idée du recuit: on se donne un *schéma de refroidissement*, i.e. une suite décroissante de températures $\{T_t, t \geq 1\}$ tendant vers 0, et on construit encore une chaîne de configurations $\{\lambda(t), t \geq 1\}$. Cette construction est "presque" la même que dans l'échantillonneur de Gibbs, sauf que cette fois-ci, les caractéristiques locales (1.29) utilisées dans la relaxation au site n_t sont celles de la distribution $\Pi_{T(t)}$ associée à la température courante $T(t)$. Alors, si celle-ci ne décroît pas plus vite qu'avec une vitesse logarithmique inverse $C/\log(t)$, nous avons ([I–16]):

$$\lim_{t\to\infty} \mathbf{P}(\lambda(t) = \omega | \lambda(0) = \eta)) = \Pi_0(\omega), \tag{1.32}$$

quelle que soit la configuration initiale η et pour tout $\omega \in \Omega^S$ dès que C est assez grande (pour le choix optimal de C, cf. HAJEK[I–22]). La minimisation se trouve ici "asymptotiquement" résolue: $\lambda(t)$ du recuit fournit une approximation d'une configuration d'énergie minimale. Signalons cependant que le schéma théorique optimal de refroidissement $\{T(t)\}$ (cf. [I–16], [I–22]) est en général trop lent. Dans les applications, divers schémas pragmatiques et praticables ont été adoptés, voir par exemple [I–16], [I–20]. ♠

Nous venons de voir que le recuit simulé , sur la base de l'énergie a posteriori $U(\lambda|x)$, donne une solution approchée de l'estimateur $\hat{\lambda}_{\text{map}}$. Lorsque les labels sont binaires, i.e. $\Omega = \{0, 1\}$, il existe un algorithme de PORTEOUS & al.[I–36] qui donne la solution *exacte* de l'estimateur $\hat{\lambda}_{\text{map}}$. L'algorithme est déduit d'une adaptation de l'algorithme de Ford-Fulkerson (minimisation d'un flot dans un graphe non orienté). En particulier l'existence de cet algorithme permet de comparer le MAP exact et le MAP via un schéma approximatif de refroidissement (cf. [I–36] et Chapitre 3).

1.2.2 Segmentation par le MPM

Cette méthode du *Maximum des Marginales a Posteriori*, proposée par MARROQUIN & al.[I–32] , consiste à choisir

$$\hat{\lambda}_{\text{mpm}} = \{\hat{\lambda}_s\}, \tag{1.33}$$

telle qu'en chaque pixel s, $\hat{\lambda}_s$ maximise la marginale a posteriori en s :

$$\hat{\lambda}_s = \text{Arg}\max_{j\in\Omega} \mathbf{P}(\lambda_s = j|x). \tag{1.34}$$

Elle correspond à la règle bayésienne associée à la fonction de coût additif:

$$C(\lambda,\hat{\lambda}) = \sum_{s\in S}[1-\delta(\lambda_s,\hat{\lambda}_s)], \tag{1.35}$$

où δ est le symbole de Kronecker. ($\delta(x,y) = 1$ si $x = y$, 0 sinon). Par conséquent, le MPM minimise l'erreur moyenne de classification. Si la marginale a posteriori en x est localisée à une fenêtre $x(V_s)$, on retombe sur les méthodes contextuelles: en ce sens les méthodes contextuelles peuvent être vues comme des méthodes M.P.M. localisées.

Une difficulté liée au M.P.M. tient au fait que les marginales du champ de Gibbs a posteriori { $\mathbf{P}(\lambda|x)$ } ne s'explicitent pas. Ses auteurs proposent de les estimer par une méthode de Monte-Carlo via l'échantillonneur de Gibbs. Celui-ci génère une chaîne de Markov $\{\lambda(t), t \geq 1\}$ ergodique (cf. Théorème B, équation (1.31)). Les marginales a posteriori sont alors estimées par

$$\mathbf{P}(\lambda_s = j|x) \approx \frac{1}{n_1 - n_0}\sum_{n_0+1}^{n_1} \delta(\lambda_s(t), j). \tag{1.36}$$

où n_1 est la longueur de la chaîne simulée et n_0 le nombre de visites préliminaires permettant au système d'entrer en "équilibre" ([I–32]).

1.2.3 Segmentation par l'ICM

Cette méthode a été proposée par BESAG[I–3]. Il s'agit d'une minimisation itérative de l'énergie a posteriori. On fixe, comme pour le recuit simulé, un parcours infini du réseau S: $\{n_t, t \geq 1\}$. La suite des configurations minimisantes $\{\lambda(t), t \geq 1\}$ est construite de la manière suivante: partant d'une configuration initiale $\lambda(0) = \eta$, à chaque instant t une relaxation de la configuration courante a lieu au site visité n_t. Mais contrairement à l'échantillonneur de Gibbs et au recuit simulé , la nouvelle configuration au site n_t, $\lambda_{n_t}(t+1)$ est affectée cette fois-ci du mode de la caractéristique locale conditionnelle (cf. l'équation (1.29)) avec l'énergie a posteriori $U(\lambda|x)$. D'où le nom de *"Iterated Conditional Modes"*. Il

s'agit donc d'un algorithme déterministe le long duquel l'énergie globale décroît vers un minimum local.

L'avantage de cet algorithme est qu'il est très rapide, arrêté au bout de 5 à 10 balayages et donnant de résultats expérimentaux satisfaisants ([I–3]). Cependant du fait de la convergence vers un minimum local, une bonne initialisation de l'algorithme ICM est importante. Lorsque dans une application l'algorithme du MAP ou du MPM s'avère trop coûteuse en temps de calculs, l'ICM est souvent employé.

Chapitre 2

Erreur d'une classification contextuelle simple

Ce chapitre examine, dans un cadre particulier explicité plus loin, l'amélioration apportée par la classification contextuelle simple (cf. Définition 3, paragraphe 1.1.1) par rapport à la classification classique, dite aveugle (équation (1.1)). Nous mesurons cette amélioration par une évaluation des erreurs théoriques de ces deux procédures de classification.

Rappelons quelques notations du chapitre précédent. $\{x_s, s \in S\}$ et $\{\lambda_s, s \in S\}$ sont respectivement l'image à segmenter et la carte des labels à reconstituer. Les labels $\{\lambda_s\}$ prennent K valeurs dans $\Omega = \{1, 2, \cdots, K\}$. Dans tout ce chapitre, la loi a priori sur Ω, $\{\pi_j, j \in \Omega\}$ sera fixée égale à celle la moins informative $\pi_j \equiv 1/K$, pour tout $j \in \Omega$. La classification contextuelle simple (CCS en abréviation) choisit en chaque pixel $s \in S$:

$$\text{Règle CCS :} \qquad \hat{\lambda}_s = \operatorname{Arg}\max_{j \in \Omega} \mathbf{P}\left[x(V_s) | \mathcal{X}(j)\right], \tag{2.1}$$

où $\mathcal{X}(j)$ est la configuration locale uniforme sur la fenêtre V_s, i.e. $\lambda_t \equiv j$, pour tout $t \in V_s$. Rappelons aussi la classification aveugle (Règle CA):

$$\text{Règle CA :} \qquad \hat{\lambda}_s = \operatorname{Arg}\max_{j \in \Omega} \mathbf{P}\left[x_s | \lambda_s = j\right]. \tag{2.2}$$

Dans l'imagerie de télédétection, les données des images sont composées des signaux reçus dans plusieurs bandes de fréquences distinctes. Pour une satellite du type SPOT par exemple, en chaque pixel on reçoit trois signaux dont les longueurs d'onde se situent respectivement entre 0.50–0.59 μm, 0.61–0.68 μm et 0.79–0.89 μm. Ce petit détour nous montre que dans des applications on est souvent confronté à des images multi-spectrales (ou vectorielles).

Le modèle statistique adopté ici est le suivant: nous supposerons d'abord que les textures multispectrales sont du type gaussien r-vectoriel:

$$\mathbf{P}[x_s | \lambda_s = j] \ \sim \mathcal{N}_r(\mu_j, \Lambda_j), \ \ j \in \Omega \ , \tag{2.3}$$

pour tout $s \in S$.[1]

Nous notons ces K textures gaussiennes par $\{\Pi_j, j \in \Omega\}$. Avec la condition d'uniformité sur la fenêtre V_s, si $\theta = |V_s|$, $y_s = x(V_s)$ est gaussien $r\theta$-vectoriel:

$$\mathbf{P}[y_s = x(V_s) | \lambda_t = j, t \in V_s] \sim \mathcal{N}_{r\theta}(\tilde{\mu}_j, \Sigma_j), \quad j \in \Omega \ , \tag{2.4}$$

où $\tilde{\mu}_j = 1_\theta \otimes \mu_j$, 1_θ est le vecteur unitaire de $\mathbb{R}^\theta$, $\otimes$ le produit de Kronecker et Σ_j la matrice de covariance de y_s sous Π_j. Nous résumons cette modélisation en disant simplement que

$$\text{sous } \Pi_j : \qquad x_s \sim \mathcal{N}_r(\mu_j, \Lambda_j) \quad \text{et} \quad y_s \sim \mathcal{N}_{r\theta}(\tilde{\mu}_j, \Sigma_j). \tag{2.5}$$

Notons $\ell_j, \tilde{\ell}_j$ les log-vraisemblances respectives de x_s et y_s sous Π_j (à une constante additive près). Les classifications CA et CCS s'explicitent en:

$$\text{Règle CA :} \quad \hat{\lambda}_s = \operatorname{Arg}\max_{j \in \Omega} \ell_j \tag{2.6}$$

$$\text{avec} \quad \ell_j = -\frac{1}{2}\left[{}^t(x_s - \mu_j)\, \Lambda_j^{-1}\, (x_s - \mu_j) + \log\det \Lambda_j\right],$$

et

$$\text{Règle CCS :} \quad \hat{\lambda}_s = \operatorname{Arg}\max_{j \in \Omega} \tilde{\ell}_j \tag{2.7}$$

$$\text{avec} \quad \tilde{\ell}_j = -\frac{1}{2}\left[{}^t(y_s - \tilde{\mu}_j)\, \Sigma_j^{-1}\, (y_s - \tilde{\mu}_j) + \log\det \Sigma_j\right].$$

Revenons à notre objectif principal: la quantification des erreurs théoriques de classification , soient

$$e_{jk} = \mathbf{P}[\hat{\lambda}_s = k | \lambda_s = j], \quad j, k \in \Omega, \quad j \neq k \ , \tag{2.8}$$

pour la règle aveugle CA et des expressions similaires $\{\tilde{e}_{jk}\}$ pour la règle CCS. Il se trouve que l'évaluation de $\{\tilde{e}_{jk}\}$ nécessite une connaissance supplémentaire sur la structure de la matrice de covariance de y_s, $\{\ \Sigma_j\}$ sous les différentes distributions $\{\ \Pi_j\}$. Afin de réaliser la comparaison souhaitée, nous supposerons que les K textures gaussiennes sont "factorisantes" (cette notion sera précisée dans le paragraphe suivant). Cette factorisation rendra possible l'évaluation des erreurs de classification.

Dans le paragraphe 2.1, nous définissons les processus factorisants. Le paragraphe 2.2 rappelle le résultat de la comparaison théorique obtenue par MARDIA[I–31] , utilisant ces modèles de factorisation dans la situation où les K textures se différencient au premier ordre ($\mu_j \neq \mu_k$ si $j \neq k$) tout en gardant la même structure du second ordre (Λ_j et Σ_j sont uniformes sur Ω). Dans le paragraphe 2.3, nous développons cette comparaison dans une situation où les K textures possèdent la même moyenne ($\mu_j \equiv \mu$, pour tout $j \in \Omega$) et se distinguent uniquement au second ordre. Le paragraphe 2.4 conclut enfin en examinant la situation mixte: à la fois la moyenne et la structure du second ordre sont distinctes parmi les K textures gaussiennes. Ces deux derniers paragraphes 2.3 et 2.4 sont issus de [I–21].

[1] Le symbole "$\sim$" est un alias de " suit une loi de ".

2.1 Les processus factorisants

Soit $\{x_s, s \in \mathbb{Z}^d\}$ un processus réel r-vectoriel sur $\mathbb{Z}^d$. Notons $\{R(s,t),\ s,t \in \mathbb{Z}^d\}$ les auto-covariances (matricielles) du processus, i.e.

$$R(s,t) = \mathbf{E}(x_s - \mu_s)^t(x_t - \mu_t)\ , \tag{2.9}$$

où $\{\mu_s, s \in \mathbb{Z}^d\}$ sont des vecteurs moyens du processus. Notons $\Lambda = R(0,0)$.

Définition 5 (MARDIA[I–31])[2] *Un processus $\{x_s, s \in \mathbb{Z}^d\}$ r-vectoriel est dit* factorisant spatial $\otimes$ multispectral[3] *s'il existe une fonction $\rho : \mathbb{Z}^d \times \mathbb{Z}^d \longrightarrow \mathbb{R}$ symétrique et du type positif telle que les auto-covariances $\{R(s,t)\}$ se factorisent en*

$$R(s,t) = \rho(s,t)\ \Lambda\ , \qquad s,t \in \mathbb{Z}^d\ . \tag{2.10}$$

ρ est appelée la fonction de corrélation spatiale *vérifiant $\rho(0,0) = 1$.* ♣

Il va de soi que cette notion de factorisation n'a d'intérêt que pour des processus au moins bi-variés: $r \geq 2$. Ainsi pour un processus factorisant, la dépendance des sites se factorise en deux facteurs: l'un *spatial* représenté par ρ qui est uniforme à travers les spectres $a = 1, \cdots, r$; l'autre *spectral*, $\Lambda = R(0,0)$.

La propriété suivante est immédiate partant de la définition ci-dessus. Rappelons que si $B = [B_{ij}]$, $D = [D_{kl}]$ sont des matrices quelconques de taille respective $m \times n$ et $p \times q$, leur produit de Kronecker $B \otimes D$ est la matrice de taille $mp \times nq$ définie par

$$B \otimes D = [B_{ij}\ D]\ \ ,1 \leq i \leq m,\ \ 1 \leq j \leq n\ .$$

Lemme 1 *Supposons que $\{x_s, s \in \mathbb{Z}^d\}$ est factorisant. Soient $A = \{s_1, s_2, \cdots, s_\theta\}$ θ sites distincts de $\mathbb{Z}^d$ et $y = x(A) = {}^t[{}^t x_{s_1}, \cdots, {}^t x_{s_\theta}]$ la variable concaténée correspondant à la restriction de x à l'ensemble A. Alors, la matrice de covariance de y, Σ se factorise en un produit de Kronecker:*

$$\Sigma = \mathbf{Var}(y) = C\ \otimes\ \Lambda\ , \tag{2.11}$$

où C est une matrice $\theta \times \theta$ déterminée par la fonction de corrélation spatiale ρ:

$$C = [\rho(s_i, s_j)\ ,\ 1 \leq i,j \leq \theta\ ]\ . \tag{2.12}$$

♣

[2] Mardia considérait les corrélations ρ isotropiques mais la généralisation est directe.

[3] A ne pas confondre avec la factorisation dans les diverses directions du lattice $\mathbb{Z}^d$: $R(s) = \prod_{1 \leq k \leq d} R_k(s_k)$ si $s = (s_1, \ldots, s_d) \in \mathbb{Z}^d$.

Un résultat que nous allons donner maintenant concerne les processus markoviens gaussiens (GMRF) stationnaires. Il sera utilisé dans la seconde partie de la thèse consacrée à l'estimation des paramètres des modélisations.

Soit donc $\{x_s, s \in \mathbb{Z}^d\}$ un GMRF r-vectoriel et stationnaire sur $\mathbb{Z}^d$ vérifiant

$$\begin{cases} \mathbf{E}\,(x_s|x_t, t \neq s) &= \mu + \sum_{t \in V_0^*} B_t\,(x_{s+t} - \mu)\,, \\ \mathbf{Var}\,(x_s|x_t, t \neq s) &= \Gamma\,. \end{cases} \tag{2.13}$$

où V_0^* est une fenêtre symétrique autour et privée de l'origine, μ le vecteur moyen et $\{B_t\}, \Gamma$ des paramètres matriciels du processus. La proposition suivante donne une caractérisation d'un GMRF factorisant.

Proposition 1 *Le GMRF (2.13) est factorisant si et seulement si les coefficients matriciels $\{B_t\}$ sont proportionnels à la matrice identité I_r, i.e. il existe $\{b_t\}$ scalaires de sorte que*

$$B_t = b_t\, I_r\,, \tag{2.14}$$

pour tout $t \in V_0^*$. ♣

Preuve. Le fait que les proportionnalités à la matrice identité des $\{B_t\}$ impliquent la factorisation du processus s'établit directement en explicitant la densité spectrale du processus.

Prouvons la nécessité . Supposons donc le processus factorisant avec la fonction de corrélation spatiale $\{\rho(s),\ s \in \mathbb{Z}^d\}$. Nous allons calculer explicitement l'espérance conditionnelle

$$\mathbf{E}\,(x_s|x_t, t \neq s) = \mathbf{E}\,(x_s|x_t, t \in V_s^*)\,,$$

avec $V_s^* = \{s\} + V_0^*$.

Soient $\theta = |V_s^*|$ et $z = x(V_s^*) = {}^t[\,{}^t x_t,\ t \in V_s^*]$. Notons aussi $y = {}^t[\,{}^t x_s, {}^t z]$. D'après le Lemme 1,

$$\mathbf{Cov}\,(z) = C \otimes \Lambda\,,$$

avec la $\theta \times \theta$ matrice C égale à

$$C = [\rho(t_i - t_j);\ 1 \leq i, j \leq \theta]\,.$$

Posons le θ-vecteur $v = {}^t[\rho(s - t_j);\ 1 \leq i \leq \theta]$. Alors,

$$\mathbf{Cov}\,(y) = \begin{bmatrix} 1 & {}^t v \\ v & C \end{bmatrix} \otimes \Lambda = \begin{bmatrix} \Lambda & {}^t v \otimes \Lambda \\ v \otimes \Lambda & C \otimes \Lambda \end{bmatrix}.$$

Par un calcul classique du conditionnement gaussien,

$$\begin{aligned} \mathbf{E}\,(x_s|x_t, t \neq s) &= \mathbf{E}\,(x_s|z) = -({}^t v \otimes \Lambda)(C \otimes \Lambda)^{-1}\, z \\ &= -({}^t v C^{-1} \otimes I_r)\, z = -\sum_{i=1}^{\theta} a_i x_{t_i}\,, \end{aligned}$$

avec

$$[a_1, a_2, \cdots, a_\theta] = -{}^t v C^{-1} \ .$$

L'identification avec la définition (2.13) montre que

$$B_{t_i} = -a_i \, I_r \ ,$$

pour $1 \leq i \leq \theta$, $t_i \in V_0^*$. ♠

2.2 Classification suivant les niveaux moyens

Revenons maintenant au problème de classification. Nous modélisons les K textures gaussiennes $\{\Pi_j, \ j \in \Omega\}$ par K processus plans (sur $\mathbb{Z}^2$), gaussiens stationnaires et factorisants avec les K fonctions de corrélations spatiales $\{\rho_j\}$. Ce qui signifie que

$$\mathbf{Cov}\ (x_s, x_t | \lambda_s = \lambda_t = j) \ = \ \rho_j(s-t)\ \Lambda_j \quad . \tag{2.15}$$

En particulier pour la matrice de covariance de la variable concaténée $y_s = x(V_s)$ intervenant dans la classification contextuelle, nous avons

$$\text{sous} \quad \Pi_j \ : \quad \Sigma_j = \mathbf{Cov}\ (y_s) = C_j \otimes \Lambda_j \ , \tag{2.16}$$

où C_j est la $\theta \times \theta$ matrice $(\theta = |V_s|)$ de corrélation spatiale :

$$C_j = [\rho_j(t_k - t_l) \ ; \ t_k, t_l \in V_s] \quad . \tag{2.17}$$

Comme nous l'avons signalé , MARDIA[I–31] a examiné ces classifications lorsque les K textures se distinguent au niveau moyen

$$\mu_j \neq \mu_k \ , \ \text{si} \ j \neq k \ ,$$

tout en conservant la même structure du second ordre : $\Lambda_j \equiv \Lambda$ et $\rho_j \equiv \rho$ sur Ω. Notons alors $C_j \equiv C$ et $\Sigma_j \equiv C \otimes \Lambda$. Nous rappelons maintenant les résultats de MARDIA[I–31].

Dans ce contexte, les règles CA, CCS s'écrivent comme suit, après soustraction de termes constants ((2.6)-(2.7)):

$$\text{Règle CA :} \quad \hat{\lambda}_s = \operatorname{Arg}\max_{j \in \Omega} \ \ell_j \tag{2.18}$$

$$\text{avec} \quad \ell_j = {}^t\mu_j \, \Lambda^{-1} \, (x_s - \frac{1}{2}\mu_j),$$

et

$$\text{Règle CCS :} \quad \hat{\lambda}_s = \operatorname{Arg}\max_{j\in\Omega} \tilde{\ell}_j \tag{2.19}$$

$$\text{avec} \quad \tilde{\ell}_j = {}^t\tilde{\mu}_j (C \otimes \Lambda)^{-1} (y_s - \frac{1}{2}\tilde{\mu}_j).$$

La règle CCS sera explicitée comme suit: soient $\theta = |V_s|$ et $y_s = {}^t[\, {}^t x_t,\ t \in V_s]$. Alors,

$$\begin{aligned} {}^t\tilde{\mu}_j (C \otimes \Lambda)^{-1} &= {}^t(1_\theta \otimes \mu_j)\ (C \otimes \Lambda)^{-1} \\ &= ({}^t 1_\theta C^{-1}) \otimes ({}^t\mu_j\, \Lambda^{-1})\ , \end{aligned} \tag{2.20}$$

et si $c = {}^t[c_1, \cdots, c_\theta] = C^{-1} 1_\theta$,

$$\begin{aligned} {}^t\tilde{\mu}_j (C \otimes \Lambda)^{-1} y_s &= \left({}^t c \otimes ({}^t\mu_j\, \Lambda^{-1})\right)\ y_s \\ &= {}^t\mu_j\, \Lambda^{-1} \left(\sum_1^\theta c_i\, x_{t_i} \right) \quad . \end{aligned} \tag{2.21}$$

Notons $g_s = \sum_1^\theta c_i\, x_{t_i}$ et:

$$\nu^2 \overset{\text{déf}}{=} \sum_{i=1}^{\theta} c_i = {}^t 1_\theta\, C^{-1}\, 1_\theta \ . \tag{2.22}$$

Comme:

$$\begin{aligned} {}^t\tilde{\mu}_j (C \otimes \Lambda)^{-1} \tilde{\mu}_j &= \left({}^t c \otimes ({}^t\mu_j\, \Lambda^{-1})\right)\ (1_\theta \otimes \mu_j) \\ &= \left(\sum_1^\theta c_i \right)\ {}^t\mu_j\, \mu_j \\ &= \nu^2\, {}^t\mu_j\, \mu_j \ . \end{aligned} \tag{2.23}$$

La règle CCS équivaut à la discrimination basée sur le r-vecteur g_s:[4]

$$\text{Règle CCS :} \quad \hat{\lambda}_s = \operatorname{Arg}\max_{j\in\Omega} \tilde{\ell}_j \tag{2.24}$$

$$\text{avec} \quad \tilde{\ell}_j = {}^t\mu_j \Lambda^{-1} (g_s - \nu^2 \mu_j). \tag{2.25}$$

g_s est un vecteur (gaussien) de moyenne $\nu^2\mu_j$ et de covariance

$$\text{Cov}\ (g_s) \ = \ \text{Cov}\ (\sum_{i=1}^{\theta} c_i x_{t_i}) \tag{2.26}$$

[4]Ce fait a été relevé pour la première fois par SWITZER[I–41].

$$
\begin{aligned}
&= \sum_{i,k=1}^{\theta} c_i c_k \, \mathrm{Cov}(x_{t_i}, x_{t_k}) \qquad (2.27)\\
&= \left(\sum_{i,k=1}^{\theta} c_i c_k \, \mathrm{Cov}(x_{t_i}, x_{t_k}) \right) \Lambda \\
&= ({}^t c \, C \, c) \, \Lambda \\
&= ({}^t 1_\theta \, C^{-1} \, 1_\theta) \, \Lambda \\
&= \nu^2 \, \Lambda \quad .
\end{aligned}
$$

Donc

$$
\text{sous } \Pi_j : \qquad g_s \sim \mathcal{N}_r(\nu^2 \mu_j, \, \nu^2 \Lambda). \qquad (2.28)
$$

Le fait que le vecteur discriminant g_s intervenant dans la classification contextuelle est de la même dimension r que celle de x_s est important dans les applications: l'introduction du contexte dans la classification augmente peu la complexité algorithmique dans cette situation de factorisation.

Le facteur ν^2 s'interprète comme un *facteur de gain* de la règle contextuelle CCS par rapport à la règle aveugle CA. On le voit mieux s'il n'y a que deux textures à discriminer: $K = 2$. Les deux règles CA, CCS deviennent:

$$
\begin{aligned}
\text{Règle CA : } \quad & \hat{\lambda}_s = 2 \quad \text{si} \quad {}^t(\mu_2 - \mu_1) \, \Lambda^{-1} \, x_s \;\geq\; \frac{1}{2} \, {}^t(\mu_2 - \mu_1) \Lambda (\mu_2 + \mu_1) \, , \\
& \hat{\lambda}_s = 1 \quad \text{sinon} \qquad (2.29)
\end{aligned}
$$

et

$$
\begin{aligned}
\text{Règle CCS : } \quad & \hat{\lambda}_s = 2 \quad \text{si} \quad {}^t(\mu_2 - \mu_1) \, \Lambda^{-1} \, g_s \;\geq\; \frac{1}{2} \, \nu^2 \, {}^t(\mu_2 - \mu_1) \Lambda (\mu_2 + \mu_1) \, , \\
& \hat{\lambda}_s = 1 \quad \text{sinon} \; . \qquad (2.30)
\end{aligned}
$$

Les erreurs de classification théoriques (cf. l'équation(2.8)) $e = e_{12} = e_{21}$ et $\tilde{e} = \tilde{e}_{12} = \tilde{e}_{21}$ valent respectivement:

$$
\text{Règle CA : } \quad e = \; \phi(-\frac{D}{2}) \; , \qquad (2.31)
$$

$$
\text{Règle CCS : } \quad \tilde{e} = \; \phi(-\frac{\nu \, D}{2}) \; , \qquad (2.32)
$$

avec ϕ la fonction de répartition de la gaussienne centrée réduite $\mathcal{N}(0,1)$ et D la distance de Mahalanobis entre les deux gaussiennes $\mathcal{N}(\mu_j, \Lambda)$, $j = 1, 2$:

$$
D \stackrel{\text{déf}}{=} \left[\, {}^t(\mu_2 - \mu_1) \, \Lambda^{-1} \, (\mu_2 - \mu_1) \, \right]^{1/2} \quad . \qquad (2.33)
$$

Ainsi dans le cas de deux textures, si la classification aveugle conduit à une erreur de $e = \phi(-D/2) = 20\%$ (resp. 10%) , une classification contextuelle simple

comprenant les quatre plus proches voisins avec une indépendance spatiale , i.e. $\rho(0) = 1$, $\rho(s) = 0$ si $s \neq 0$ (ce qui implique que $\nu = \sqrt{\theta} = \sqrt{5}$), réduit l'erreur à $\tilde{e} = \phi(-\nu\ D/2) = 3\%$ (resp. 0.2%). Nous retrouverons cette décroissance (exponentielle) de l'erreur en fonction de la taille de la fenêtre utilisée dans d'autres situations (paragraphes 2.3 et 2.4).

Terminons ce paragraphe sur un examen plus détaillé du facteur ν^2. D'abord, comme on l'espérait, celui-ci est toujours plus grand que 1. En effet,

$$\nu^2 = {}^t 1_\theta\ C^{-1}\ 1_\theta \geq \theta/\alpha_{\max}\ ,$$

où $\alpha_{\max}$ est la plus grande valeur propre de C. Or

$$\alpha_{\max} \leq \mathrm{tr} C = \theta\ .$$

Donc $\nu \geq 1$.

La dépendance de la valeur de ν en fonction des corrélations spatiales $\{\rho_s\}$ est la suivante: ν varie de 1 à $\sqrt{\theta}$ et de $\sqrt{\theta}$ à l'infini selon que la corrélation spatiale est attractive, indépendante ou répulsive sur la fenêtre V_s. L'exemple suivant illustre ce phénomène.

Exemple. Supposons que la corrélation spatiale ρ prend la forme, avec $s = (a, b)$:

$$\rho(s) = \tau^{|a|+|b|}\ ,\ |\tau| < 1\ .$$

Prenons encore la fenêtre du premier ordre constituée de s et de ses quatre plus proches voisins ($\theta = 5$). On obtient

$$\nu^2 = \frac{5 - 3\tau}{\tau + 1} = \begin{cases} 1 & \text{si } \tau \longrightarrow 1_-\ , \\ 5 & \text{si } \tau = 0\ , \\ +\infty & \text{si } \tau \longrightarrow (-1)_+\ . \end{cases}$$

Ainsi, ν est proche de 1 si la dépendance spatiale locale est importante, de $\sqrt{5}$ si les textures gaussiennes sont voisines d'un bruit blanc, et tend vers $+\infty$ si "ces textures deviennent spatialement répulsives".

Le vecteur discriminant g_s s'écrit:

$$g_s = \frac{1 - 3\tau}{1 + \tau}\ x_s + \frac{4}{1 + \tau}\ \overline{x}_s\ ,$$

où $\overline{x}_s$ est la moyenne sur les quatre voisins de s:

$$\overline{x}_s = \frac{1}{4} \sum_{t:\text{voisin de } s} x_t\ .$$

♠

2.3 Classification suivant les covariances

Supposons que les K textures gaussiennes $\{\Pi_j\}$ possèdent les mêmes niveaux moyens $\mu_j \equiv \mu$, pour tout $j \in \Omega$. Nous voulons toujours comparer la classification contextuelle simple CCS à la classification aveugle CA en évaluant des erreurs de classification.

Lorsque la discrimination porte sur le second ordre, même en situation de covariance factorisante, le calcul explicite des erreurs est difficile. Pour cette raison, nous nous restreignons désormais au cas de deux textures, i.e. $\Omega = \{1,2\}$.

Sans perte de généralité , le niveau moyen commun μ est fixé à 0. Les règles de classification CCS et CA s'écrivent ici (cf. (2.6)-(2.7)) :

$$\text{Règle CA : } \begin{array}{l} \hat{\lambda}_s = 1 \quad \text{si} \quad {}^t x_s \ (\Lambda_2^{-1} - \Lambda_1^{-1}) \ x_s \ \geq \ -\log\det(\Lambda_2\Lambda_1^{-1}) \ , \\ \hat{\lambda}_s = 2 \quad \text{sinon} \ . \end{array} \tag{2.34}$$

et

$$\text{Règle CCS : } \begin{array}{l} \hat{\lambda}_s = 2 \quad \text{si} \quad {}^t y_s \ (\Sigma_2^{-1} - \Sigma_1^{-1}) \ y_s \ \geq \ -\log\det(\Sigma_2\Sigma_1^{-1}) \ , \\ \hat{\lambda}_s = 2 \quad \text{sinon} \ . \end{array} \tag{2.35}$$

L'étude des erreurs de classification se ramène à celle des formes quadratiques

$$\Delta \stackrel{\text{déf}}{=} {}^t x_s \ (\Lambda_2^{-1} - \Lambda_1^{-1}) \ x_s \ , \tag{2.36}$$

et

$$\tilde{\Delta} \stackrel{\text{déf}}{=} {}^t y_s \ (\Sigma_2^{-1} - \Sigma_1^{-1}) \ y_s \ . \tag{2.37}$$

On supposera la structure des covariances factorisante (cf. (2.16)):

$$\Sigma_j = C_j \otimes \Lambda_j \quad j = 1,2 \ .$$

Ainsi, la différence des deux textures peut se manifester "spectralement", i.e. $\Lambda_1 \neq \Lambda_2$ ou spatialement i.e. $C_1 \neq C_2$. Nous allons examiner séparément ces deux situations.

Avant de calculer les erreurs de classification, nous rappelons d'abord un résultat de KHATRI[I–26] qui évalue la distribution d'une forme quadratique des variables gaussiennes. Notons par $\varsigma(A)$ le rang d'une matrice A quelconque. Soit une forme quadratique

$$h = {}^t z \ A \ z + 2 \ {}^t l z + c \ . \tag{2.38}$$

où

$$\begin{array}{rcl} z & \sim & \mathcal{N}_p(\mu, V) \ , \\ \varsigma(V) & = & q \ , \quad q \leq p, \\ l \in \mathbb{R}^p & \text{et} & c \in \mathbb{R} \ . \end{array}$$

Il existe alors un q-vecteur gaussien w tel que $z = Bw + \mu$ et $w \sim \mathcal{N}_q(0, I_q)$ avec une factorisation de V: $V = B\,{}^tB$.

Définissons la décomposition spectrale de la matrice ${}^tB\,A\,B$:

$$ {}^tB\,A\,B = \sum_{j=1}^{m} \lambda_j\, E_j \ , \tag{2.39} $$

où $\lambda_1 > \lambda_2 > \cdots > \lambda_m$, $\lambda_j \neq 0$ sont des valeurs propres non nulles et E_j, $j = 1, \ldots, m$ les projecteurs sur les sous-espaces propres associés. La multiplicité de λ_j est comptée par f_j et égale à $\varsigma(E_j)$. Soit aussi

$$ E_0 = I_q - \sum_{j=1}^{m} E_j \ , \tag{2.40} $$

le projecteur sur le sous-espace propre associé à la valeur propre nulle. Introduisons d'autre part les quantités suivantes:

$$ \begin{aligned} \nu_j &= {}^t(l + A\mu)\,(BE_j{}^tB)\,(l + A\mu) \quad j = 0, 1, \ldots, m \ , \\ c_{(1)} &= {}^t\mu\,A\,\mu + 2{}^tl\mu + c \ , \\ c_{(2)} &= c_{(1)} - \sum_{j=1}^{m} (\nu_j/\lambda_j) \ . \end{aligned} \tag{2.41} $$

Lemme 2 (Khatri[I–26]) *Soit $z \sim \mathcal{N}_p(\mu, V)$. Alors la forme quadratique*

$$ h = {}^tz\,A\,z + 2{}^tlz + c \ \sim \ \sum_{j=1}^{m} \lambda_j \chi'^2(f_j, \nu_j/\lambda_j^2) \ + \ U \ , \tag{2.42} $$

où les $\{\chi'^2\}$ et U sont indépendamment distribués, les χ'^2 sont des chi-deux décentrés de degré de liberté f_j et $U \sim \mathcal{N}(c_{(2)}, 4\nu_0)$. Les λ_j, f_j et $c_{(2)}$ sont définis dans (2.39) et (2.41). ♣

On vérifiera que les $\{\lambda_j\}$ sont aussi les valeurs propres , distinctes et non nulles, de la matrice AV (ou VA) avec les mêmes multiplicités $\{f_j\}$.

Lorsque la variable normale z est centrée (i.e. $\mu = 0$) et si h est une quadratique pure (i.e. $l = c = 0$), h suit simplement la loi d'un mélange des χ^2 indépendants pondérés par des coefficients positifs:

Corollaire 1 *Soit $z \sim \mathcal{N}_p(0, V)$. Alors la forme quadratique*

$$ h = {}^tz\,A\,z \ \sim \ \sum_{j=1}^{m} \lambda_j \chi^2(f_j) \ , \tag{2.43} $$

où les $\{\chi^2\}$, de degré de liberté respectif f_j, sont indépendants. Les λ_j, f_j sont définis dans (2.39) et (2.41). ♣

2.3.1 Différence spectrale et facteur spatial constant

Cette situation est celle où les deux textures gaussiennent possèdent des covariances multispectrales différentes $\Lambda_1 \neq \Lambda_2$, tout en ayant une corrélation spatiale identique, $\rho_1 = \rho_2$. Comme conséquence, $C_1 = C_2 = C$. Rappelons que les niveaux moyens sont fixés à 0 et avec les notations du début du chapitre, nous avons

$$\text{Sous} \quad \Pi_j \ : x_s \ \sim \ \mathcal{N}_r(0, \Lambda_j) \quad \text{et} \quad y_s \ \sim \ \mathcal{N}_{r\theta}(0, \Sigma_j) \ , \tag{2.44}$$

avec $\Sigma_j = C \otimes \Lambda_j$, $j = 1, 2$ et $\theta = |V_s|$.

Nous allons expliciter les lois des formes quadratiques Δ (2.36) et $\tilde{\Delta}$ (2.37). Notons les deux seuils qui apparaissent dans les discriminations (2.34)-(2.35) par:

$$u = -\log\det(\Lambda_2 \Lambda_1^{-1}) \ , \tag{2.45}$$

et

$$\tilde{u} = -\log\det(\Sigma_2 \Sigma_1^{-1}) \ . \tag{2.46}$$

Comme $\Sigma_j = C \otimes \Lambda_j$, $j = 1, 2$, nous avons

$$\tilde{u} = \theta u \ . \tag{2.47}$$

D'autre part, notons $\lambda_j, j = 1, \ldots, m$ les m valeurs propres distinctes et non nulles de la matrice $(I_r - \Lambda_2 \Lambda_1^{-1})$, avec $f_j, j = 1, \ldots, m$ leur multiplicité respective. Utilisant le Corollaire (1) du Lemme 2, on obtient pour les lois des formes quadratiques discriminantes Δ, $\tilde{\Delta}$:

$$\text{Sous} \quad \Pi_2 \ : \ \Delta \ \sim \ \sum_{j=1}^{m} \lambda_j \ \chi^2(f_j) \ , \tag{2.48}$$

où les $\chi^2(f_j)$ sont indépendantes et:

$$\begin{aligned} \text{Sous} \quad \Pi_2 \ : \ \tilde{\Delta} \ &\sim \ \Delta_1 + \Delta_2 + \cdots + \Delta_\theta \\ &\sim \ \sum_{j=1}^{m} \lambda_j \ \chi^2(\theta f_j) \ , \end{aligned} \tag{2.49}$$

où les $\Delta_i, 1 \leq i \leq \theta$ forment un θ-échantillon de Δ.

Ainsi, pour les règles de classification CCS (2.35) et CA (2.34), les erreurs respectives $e_{21}, \tilde{e}_{21}$, i.e. les probabilités d'un classement erroné du pixel s avec le label 1 tandis qu'il appartient au label 2, sont donnés par:

$$e_{21} = \mathbf{P}(\Delta > u) \ , \tag{2.50}$$

et

$$\tilde{e}_{21} = \mathbf{P}(\tilde{\Delta} > \tilde{u}) \ . \tag{2.51}$$

Pour des modèles spécifiés, un calcul précis de ces erreurs peut être effectué (simulation, ou utilisation de tables spécifiques). Sans spécifier le modèle, tentons une comparaison de ces deux erreurs: utilisant une inégalité de grande déviation (cf. [II–5], Tome 1, p.102), on obtient une majoration de $\tilde{e}_{21}$:

$$\tilde{e}_{21} = \mathbf{P}(\tilde{\Delta} > \theta u) \leq e^{-\theta \psi_\Delta(u)} \quad , \tag{2.52}$$

à condition que $u > \mathbf{E}\Delta$. La fonction ψ_Δ est la transformée de Young de la loi de Δ:

$$\psi_\Delta(z) = \sup_{t \in G_\Delta} (zt - \log \varphi_\Delta(t)) \quad , \tag{2.53}$$

qui est positif pour $z > \mathbf{E}\Delta$ et où φ_Δ est quant à lui la transformée de Laplace de la même loi:

$$\varphi_\Delta(t) = \prod_{j=1}^{m} (1 - 2\lambda_j t)^{-f_j/2} \quad , \tag{2.54}$$

définie pour les valeurs de t dans l'ensemble

$$G_\Delta = \{t \in \mathbb{R} \ : \ \lambda_j t < \frac{1}{2}, \ 1 \leq j \leq m\} \ . \tag{2.55}$$

Examinons la condition "$u > \mathbf{E}\Delta$" nécessaire à la positivité de $\psi_\Delta(u)$:

$$\begin{aligned} u - \mathbf{E}\Delta &= -\log \det(\Lambda_2 \Lambda_1^{-1}) - \sum_{j=1}^{m} \lambda_j f_j \\ &= \log \det(\Lambda_1 \Lambda_2^{-1}) + \mathrm{tr}(\Lambda_1 \Lambda_2^{-1} - I_r) \\ &= 2K(\Lambda_2, \Lambda_1) \ , \end{aligned} \tag{2.56}$$

où $K(\Lambda_2, \Lambda_1)$ désigne l'information de Kullback de la loi $\mathcal{N}_r(0, \Lambda_2)$ sur $\mathcal{N}_r(0, \Lambda_1)$. Comme $\Lambda_2 \neq \Lambda_1$, cette information est toujours positive de même que $\mu - \mathbf{E}\Delta$.

Il est intéressant de noter que l'inégalité (2.52) donne lorsque $\theta = 1$ une majoration de l'erreur de la classification aveugle CA:

$$e_{21} = \mathbf{P}(\Delta > u) \leq e^{-\psi_\Delta(u)} \ . \tag{2.57}$$

On peut conclure de cette comparaison que l'erreur de classification décroît exponentiellement en fonction de θ , la taille de la fenêtre V_s utilisée dans la classification contextuelle. La décroissance est d'autant plus rapide que l'information $K(\Lambda_2, \Lambda_1)$ est grande. La dissemblance $K(\Lambda_2, \Lambda_1)$ entre les deux textures gaussiennes joue ici un rôle similaire à celui de la distance de Mahalanobis apparue dans le paragraphe 2.2 lorsque la discrimination portait sur les niveaux moyens.

Notons d'autre part le fait curieux suivant dans ce cadre (2.44): ni la forme explicite de la fenêtre V_s ni la structure exacte de C, la corrélation spatiale sur cette fenêtre, n'influent sur l'erreur contextuelle $\tilde{e}_{21}$. Seule θ, la taille de V_s y intervient.

Nous allons terminer ce paragraphe en considérant un cas particulier où les erreurs de classification peuvent être évaluées d'une manière exacte.

Exemple. Supposons que $\Lambda_1 = \alpha I_r$ et $\Lambda_2 = \beta I_r$, avec $\alpha > \beta > 0$. Ici les variables Δ et $\tilde{\Delta}$ sont proportionnelles respectivement à un $\chi^2(r)$ et à un $\chi^2(\theta r)$. Les erreurs $e_{21}, \tilde{e}_{21}$ valent:

$$e_{21} = \mathbf{P}\left(\chi^2(r) > \frac{r\tau \log \tau}{\tau - 1}\right) \quad , \tag{2.58}$$

et

$$\tilde{e}_{21} = \mathbf{P}\left(\chi^2(\theta r) > \theta \frac{r\tau \log \tau}{\tau - 1}\right) \quad , \tag{2.59}$$

où

$$\tau = \alpha/\beta > 1 \ .$$

Des calculs et comparaisons ont été effectués pour $r = 4$ et $\tau = 4, 2$, ainsi que $1,2$ (proximité spectrale des 2 textures). La transformée de Young, donnée par (avec $u = 4\log(\alpha/\beta) = 4\log\tau$):

$$\psi_\Delta(u) = 2\left(\frac{\tau \log \tau}{\tau - 1} - 1 - \log(\frac{\tau \log \tau}{\tau - 1})\right) \quad ,$$

prend respectivement les valeurs 0.468, 0.120 et 0.008 pour les 3 valeurs de τ. Les trois Tables 2.1- 2.2- 2.3 (qui se trouvent à la fin du chapitre) donnent, pour chaque choix de $\tau = \alpha/\beta$ et les 3 fenêtres considérées, les valeurs calculées de e_{21} ($\theta = |V_s| = 1$, la règle aveugle CS) et de $\tilde{e}_{21}$ ($\theta = |V_s| = 5,9$, la règle contextuelle CCS).

Mis à part le cas où les deux populations restent difficiles à discriminer ($\alpha = 1.2\beta$), on peut conclure que les erreurs diminuent considérablement lorsque la taille de la fenêtre utilisée passe de 1 à 5 ou 9 pixels. ♠

2.3.2 Différence spatiale et facteur spectral constant

Le modèle de corrélation adopté étant toujours du type factorisant spatial⊗multispectral, nous examinons ici la situation "symétrique" de celle du paragraphe précédent: les deux textures possèdent une covariance spectrale identique, $\Lambda_1 = \Lambda_2 = \Lambda$, et une corrélation spatiale différente, $\rho_1 \neq \rho_2$. Alors C_1, C_2, les matrices de corrélations spatiales sur la fenêtre V_s sont distinctes. Dans ce contexte, les deux textures examinées, en un seul pixel, sont indiscernables, puisque $\mu_1 = \mu_2 = 0$ et $\Lambda_1 = \Lambda_2$, de sorte que les erreurs de la classification aveugle valent: $e_{12} = e_{21} = 50\%$. Seule l'introduction du contexte associé à la fenêtre V_s dans la classification contextuelle permet une discrimination des deux textures avec leur différente interaction spatiale C_1 et C_2.

Malgré cette différence, l'évaluation des erreurs de la classification contextuelle est analytiquement similaire à celle du paragraphe précédent: le facteur spectral Λ est remplacé par le facteur spatial C. Rappelons que

$$y_s \ \sim \ \mathcal{N}_{r\theta}(0, \Sigma_j) \quad , \tag{2.60}$$

avec $\theta = |V_s|$ et $\Sigma_j = C_j \otimes \Lambda$, $j = 1, 2$.

Nous voulons expliciter la loi de la forme quadratique $\tilde{\Delta}$ qui apparaît dans la règle contextuelle (2.35). Soient

$$u = -\log\det C_2 C^{-1} \quad , \tag{2.61}$$

et

$$\tilde{u} = -\log\det(\Sigma_2 \Sigma_1^{-1}) = r\, u \quad . \tag{2.62}$$

Soient $\lambda_j, j = 1, \ldots, m$ les $m$valeurs propres distinctes et non nulles de $(I_\theta - C_2 C_1^{-1})$ avec $f_j, j = 1, \ldots, m$ leur multiplicité respective. Notons par Z la somme pondérée des $\chi^2(f_j)$ indépendants:

$$Z = \sum_{j=1}^{m} \lambda_j \ \chi^2(f_j) \quad . \tag{2.63}$$

Utilisant encore le Corrolaire 1 du Lemme 2, on obtient pour la loi de $\tilde{\Delta}$:

$$\begin{aligned} \text{Sous} \quad \Pi_2 \ : \ \tilde{\Delta} &\sim Z_1 + Z_2 + \cdots + Z_r \\ &\sim \sum_{j=1}^{m} \lambda_j \ \chi^2(r f_j) \quad , \end{aligned} \tag{2.64}$$

où les $Z_i, 1 \leq i \leq r$ forment un r-échantillon de Z.

Ainsi, l'erreur $\tilde{e}_{21}$ de la règle CCS (2.35) est donnée par:

$$\tilde{e}_{21} = \mathbf{P}(\tilde{\Delta} > \tilde{u}) = \mathbf{P}(Z_1 + \cdots + Z_r > ru) \quad . \tag{2.65}$$

Soit ψ_z la transformée de Young de la loi de Z . D'une manière similaire à (2.52), nous avons:

$$\tilde{e}_{21} = \mathbf{P}(\tilde{\Delta} > \theta u) \leq e^{-r\,\psi_Z(u)} \quad , \tag{2.66}$$

à condition que $u > \mathbf{E}Z$. Mais

$$\begin{aligned} u - \mathbf{E}Z &= -\log\det(C_2 C_1^{-1}) - \sum_{j=1}^{m} \lambda_j f_j \\ &= 2\, K(C_2, C_1) \quad . \end{aligned} \tag{2.67}$$

où $K(C_2, C_1)$ représente l'information de Kullback de la loi $\mathcal{N}_\theta(0, C_2)$ sur $\mathcal{N}_\theta(0, C_1)$. Comme $C_2 \neq C_1$, la positivité de cette information garantit celle de $\mu - \mathbf{E}Z$.

2.4 Classification mixte

Nous examinons ici la situation mixte: les deux textures se distinguent à la fois par leurs niveaux moyens: $\mu_1 \neq \mu_2$ et leur corrélation spectrale: $\Lambda_1 \neq \Lambda_2$. Nous

supposons cependant qu'une caractérisque commune existe: la corrélation spatiale est identique sur les deux textures: $\rho_1 = \rho_2$ et $C_1 = C_2 = C$.

Les classification CA et CCS s'écrivent alors: (cf. (2.6)-(2.7)) :

$$\text{Règle CA : } \begin{array}{l} \hat{\lambda}_s = 1 \quad \text{si} \quad {}^t x_s \ (\Lambda_2^{-1} - \Lambda_1^{-1}) \ x_s + 2\, {}^t l x + c \ \geq \ u \ , \\ \hat{\lambda}_s = 2 \quad \text{sinon} \ . \end{array} \tag{2.68}$$

et

$$\text{Règle CCS : } \begin{array}{l} \hat{\lambda}_s = 2 \quad \text{si} \quad {}^t y_s \ (\Sigma_2^{-1} - \Sigma_1^{-1}) \ y_s + 2\, {}^t \tilde{l} x + \tilde{c} \ \geq \ \tilde{u} \ , \\ \hat{\lambda}_s = 2 \quad \text{sinon} \ . \end{array} \tag{2.69}$$

avec les notations suivantes:

$$\left\{ \begin{array}{rcl} c & = & {}^t\mu_2 \, \Lambda_2^{-1} \, \mu_2 - {}^t\mu_1 \, \Lambda_1^{-1} \, \mu_1 \ , \\ \tilde{c} & = & {}^t\tilde{\mu}_2 \, \Sigma_2^{-1} \, \tilde{\mu}_2 - {}^t\tilde{\mu}_1 \, \Sigma_1^{-1} \, \tilde{\mu}_1 \ , \\ \Sigma_j & = & C \otimes \Lambda_j, \quad j = 1, 2 \\ l & = & \Lambda_1^{-1}\mu_1 - \Lambda_2^{-1}\mu_2 \\ \tilde{l} & = & \Sigma_1\tilde{\mu}_1 - \Sigma_2\tilde{\mu}_2 \ , \\ u & = & -\log\det(\Lambda_2\Lambda_1^{-1}) \ , \\ \tilde{u} & = & -\log\det(\Sigma_2\Sigma_1^{-1}) \ . \end{array} \right. \tag{2.70}$$

L'étude des erreurs de classification se ramène ainsi à celle des formes quadratiques

$$\Delta = {}^t x_s \ (\Lambda_2^{-1} - \Lambda_1^{-1}) \ x_s + 2\, {}^t l x + c \ , \tag{2.71}$$

et

$$\tilde{\Delta} = {}^t y_s \ (\Sigma_2^{-1} - \Sigma_1^{-1}) \ y_s + 2\, {}^t \tilde{l} x + \tilde{c} \ . \tag{2.72}$$

Les lois respectives de Δ et $\tilde{\Delta}$ sont obtenues par une nouvelle application du Lemme 2. Comme dans le paragraphe 2.3.1, notons $\lambda_j, j = 1, \ldots, m$, les m valeurs propres distinctes et non nulles de $(I_r - \Lambda_2\Lambda_1^{-1})$ avec $f_j, j = 1, \ldots, m$ leur multiplicité respective. Alors:

1.

$$\text{Sous} \quad \Pi_2 \ : \quad \Delta \ \sim \ \sum_{j=1}^{m} \lambda_j \chi'^2(f_j, \nu_j/\lambda_j^2) \ + \ U \ , \tag{2.73}$$

avec

- les χ'^2 et U sont indépendants;
- $U \sim \mathcal{N}(c_{(2)}, 4\nu_0)$,
- les $\nu_j, j = 0, 1, \ldots, m$ et $c_{(2)}$ sont définis dans (2.39) et (2.41) avec $V = \Lambda_2, A = \Lambda_2^{-1} - \Lambda_1^{-1}$ et $\mu = \mu_2$.

2.

$$\text{Sous} \quad \Pi_2 \; : \quad \tilde{\Delta} \quad \sim \quad \sum_{j=1}^{m} \lambda_j \chi'^2(\theta f_j, \nu^2 \, \nu_j/\lambda_j^2) \; + \; \tilde{U} \; , \qquad (2.74)$$

avec

- les $\nu_j, j = 0, 1, \ldots, m$ et $c_{(2)}$ sont comme ci-dessus;
- les χ'^2 et $\tilde{U}$ sont indépendants;
- $U \sim \mathcal{N}(\tilde{c}_{(2)}, 4\tilde{\nu}_0)$, avec $\tilde{\nu}_0 = \nu^2 \, \nu_0$, $\tilde{c}_{(2)} = \nu^2 c_{(2)}$ et $\nu^2 = {}^t 1_\theta \, C \, 1_\theta$.

Remarquons que si $\sum_{j=1}^{m} f_j = 0$, i.e. $\Lambda_1 = \Lambda_2$, les discriminations quadratiques basées sur Δ, $\tilde{\Delta}$ dégénèrent en discriminations linéaires: on retrouve alors le problème de classification suivant les niveaux moyens (cf. paragraphe 2.2); tandis que $\sum_{j=1}^{m} f_j = r$ entrainera la disparition des composantes "linéaires normales" U et $\tilde{U}$ et avec $\mu_1 = \mu_2$ le problème de discrimination suivant les covariances (paragraphe 2.3) réapparaît.

L'étude explicite des erreurs de classifications dans ce contexte général est difficile à cause du mélange de χ'^2 et des variables gaussiennes $U, \tilde{U}$ qui apparaissent dans la lois des quadratiques discriminantes. Néanmoins, on retient le rôle du facteur ν^2 comme un facteur de gain de la classification contextuelle simple CCS par rapport à la classification aveugle CA.

On remarquera aussi que le cas particulier d'indépendance spatiale entre les pixels d'une fenêtre V_s, i.e. $C = I_\theta, \nu = \sqrt{\theta}$ peut être traité d'une manière similaire à celle employée au paragraphe 2.3.1 conduisant encore à une majoration des erreurs de la règle contextuelle à décroissance exponentielle en θ, la taille de la fenêtre V_s utilisée.

2.5 Récapitulatif

Dans ce chapitre, nous avons entrepris une étude quantitative de comparaison des deux procédures de classification, la classification aveugle CA et la classification contextuelle simple CCS, en évaluant leur probabilité d'erreur. Ces calculs sont rarement faisables dans le cadre général; ils sont cependant possibles, comme nous l'avons montré, lorsque les textures multispectrales à discriminer peuvent être considérées comme réalisations des processus gaussiens stationnaires et factorisants.

Lorsque la différence des textures ne se manifeste que par leur niveau de gris moyen, nous avons rappelé les premiers résultats obtenus par MARDIA[I–31]. Nous avons effectué une analyse du même esprit mais pour une situation plus complexe: la structure du second ordre n'est pas uniforme parmi les textures. Cette différence peut apparaître soit uniquement dans leur corrélation spatiale ou dans leur structure de covariance spectrale, soit d'une manière mixte, i.e. simultanément avec une différence du niveau de gris moyen parmi les textures. En exploitant des techniques différentes, nous avons mis (ou remis) en évidence le fait que l'erreur de classification, pour une procédure contextuelle simple, décroît exponentiellement en fonction de la taille de la fenêtre utilisée.

Bien que la probabilité d'erreur ne soit pas toujours le meilleur critère possible dans l'évaluation qualitative d'une procédure de classification des pixels, les calculs que nous avons menés ont permis, dans un modèle spécifique, de quantifier l'amélioration apportée par la prise en compte du contexte spatial de l'image.

$\theta = \lvert V_s \rvert$	1	5	9
$\tilde{e}_{21}$ (e_{21}) calculée =	12 %	1 %	0,2 %

Table 2.1: Comparaison des erreurs: $\alpha = 4\beta$

$\theta = \lvert V_s \rvert$	1	5	9
$\tilde{e}_{21}$ (e_{21}) calculée =	24 %	12 %	5 %

Table 2.2: Comparaison des erreurs: $\alpha = 2\beta$

$\theta = \lvert V_s \rvert$	1	5	9
$\tilde{e}_{21}$ (e_{21}) calculée =	37 %	35 %	34 %

Table 2.3: Comparaison des erreurs: $\alpha = 1.2\beta$

Légende V_s: Fenêtre utilisée
$\theta = 1$: Classification aveugle CA
$\theta = 5,9$: Classification contextuelle CCS

Chapitre 3

Etude expérimentale comparative

Ce chapitre reprend les résultats de [I–43].

3.1 L'objectif et la mise en œuvre de l'étude

L'étude expérimentale que nous proposons ici a pour objectif d'étudier les comportements des différentes méthodes de segmentation précédemmment examinées, soient:

1. la classification aveugle CA (cf. paragraphe 1.1.1, (1.1));
2. la classification contextuelle simple CCS(cf. paragraphe 1.1.1, définition 3);
3. la classification contextuelle géometrique CCG (cf. paragraphe 1.1.1, définition 2);
4. la segmentation par le maximum a posteriori MAP (cf. paragraphe 1.2.1);
5. la segmentation par les maximums des marginales a posteriori MPM (cf. paragraphe 1.2.2);
6. la segmentation par des modes conditionnels itérés ICM (cf. paragraphe 1.2.3).

La classification aveugle CA par maximum de vraisemblance en chaque pixel est le point de départ de l'algorithlme ICM.

Nous traiterons une image test binaire recouverte, sur ses zones de label constant, par différentes textures bruitées par simulation: ce cadre "d'école" est restrictif, mais fournit un environnement bien défini dans lequel la comparaison escomptée peut être convenablement réalisée. Plus précisément, sur l'ensemble des pixels $S = [1, 32] \times [1, 32]$, les labels $\lambda_s, s \in S$ sont binaires (image noir-blanc):

$\Lambda_s \in \Omega = \{0,1\}$. Les textures gaussiennes $\{x_s, s \in S\}$ vérifient l'indépendance conditionnelle suivante:

$$\begin{cases} \mathbf{P}(x_s, s \in S | \lambda_s, s \in S) &= \Pi_{s \in S} \mathbf{P}(x_s | \lambda_s) \ , \\ \forall\, s \in S, \quad \mathbf{P}(x_s | \lambda_s = j) &= \mathcal{N}(0, \sigma_j^2) \ , j \in \{0,1\}. \end{cases} \tag{3.1}$$

Les paramètres sont donc $\{\sigma_0^2, \sigma_1^2\}$.

La Figure 3.1-(a) montre la carte image binaire λ_0: la lettre β est dessinée sur un lattice 32×32 où "M" représente le label 0 et "." le label 1. Il s'agit donc de retrouver cette carte à partir des textures observées x qui la recouvrent et qui sont obtenues par simulation.

Nous avons utilisé deux types de modélisation a priori pour λ:

- pour la classification contextuelle géométrique CCG, il s'agit du modèle *p-q-r* (cf. paragraphe 1.1.1) combiné avec une loi a priori $\{\pi_0, \pi_1\}$ sur $\Omega = \{0,1\}$.
- pour les segmentations globales MAP, ICM et MPM, il s'agit du modèle d'Ising:

$$\mathbf{P}(\lambda) = Z^{-1} \exp\left(\beta \sum_{\langle s,t \rangle} \delta(\lambda_s, \lambda_t)\right) \quad , \beta > 0 \ , \tag{3.2}$$

  où la relation de voisinage $\langle , \rangle$ retenue est celle des quatre plus proches voisins.

Les modélisations (3.1)–(3.2) donnent la loi a posteriori:

$$\mathbf{P}(\lambda | x) = \tag{3.3}$$
$$c(x) \ \exp\left[\beta \sum_{<s,t>} \delta(\lambda_s, \lambda_t) \ + \ \frac{1}{2} \sum_{s \in S} \lambda_s \left(\log \frac{\sigma_0^2}{\sigma_1^2} + \frac{(x_s - \mu_0)^2}{\sigma_0^2} - \frac{(x_s - \mu_1)^2}{\sigma_1^2}\right)\right].$$

L'estimation de paramètres de ces modèles à priori est directement calculée sur la carte test λ_0. Soit

- $\{(\hat{p}, \hat{q}, \hat{r}), (\hat{\pi}_0, \hat{\pi}_1)\}$ pour les classifications contextuelles ([I–25]);
- $\hat{\beta}$ pour les méthodes globales (estimation par pseudo-vraisemblance [I–2]-[II–12]) et méthode de Possolo ([I–37])

On trouve en particulier, $\hat{\pi}_0 = 0.39$ et $\hat{\pi}_1 = 0.61$.

D'autre part, nous nous sommes intéressé à une situation où les deux textures gaussiennes possèdent un nême niveau moyen, $\mu_0 = \mu_1 = 0$, la discrimination ne portant que sur le second ordre. Le degré de différenciation entre les textures est repéré par $u = \sigma_1^2/\sigma_0^2 > 1$. Pour chacune des 5 valeurs de $u = 2, 3, 4, 6, 8$ et $\sigma_0^2 = 1$, on a réalisé 10 simulations, ce qui nous permettra d'évaluer empiriquement, pour

chacune des méthodes, *le pourcentage d'erreur de classification PEC* qui sera ici notre critère de qualité.

Il faut noter que pour ces différentes valeurs de u (cf. exemples de textures simulées, Figure 3.1 (b)-(c)), même pour $u = 8$ correspondant à la plus grande différenciation des textures, l'observation visuelle de x rend très difficile la segmentation des λ.

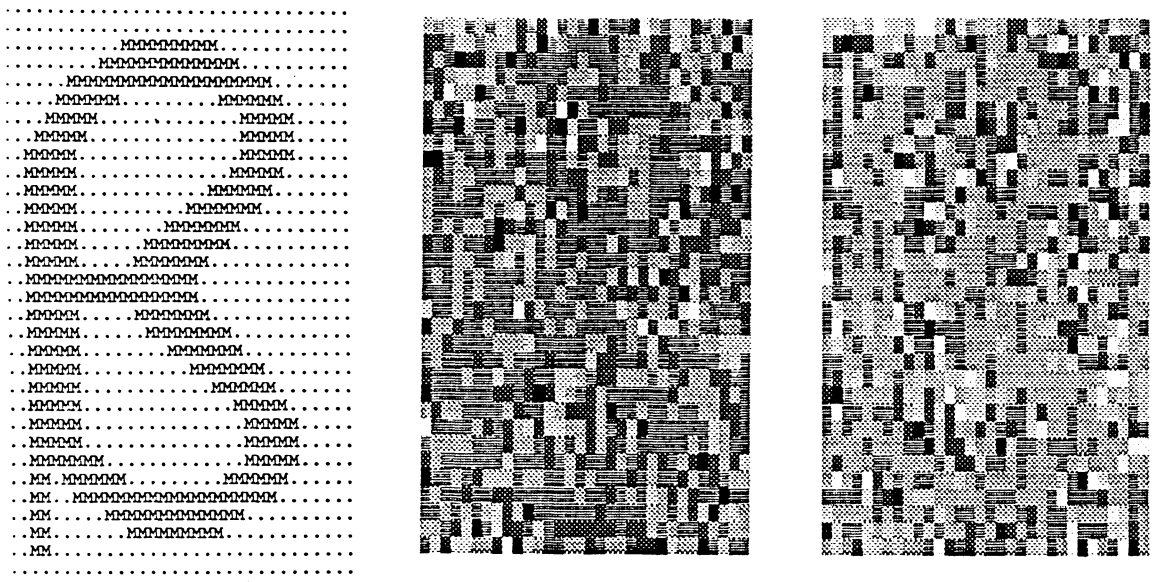


Figure 3.1: (a) = Image originale λ_0 (32×32).
(b) = Textures simulées: $\sigma_0^2 = 1$ et $\sigma_1^2 = 4$.
(c) = Textures simulées: $\sigma_0^2 = 1$ et $\sigma_1^2 = 8$.

Pour ce modèle, la classification aveugle CA conduit à une erreur théorique:

$$e = 2\pi_0\phi\left(-\sqrt{\frac{u\log u}{u-1}}\right) + \pi_1\left[1 - 2\phi\left(-\sqrt{\frac{\log u}{u-1}}\right)\right] \tag{3.4}$$

ϕ étant la fonction de répartition de $\mathcal{N}(0,1)$ et $\{\pi_0, \pi_1\}$ une loi a priori.

La Table 3.1 donne ces erreurs théoriques, associés à λ_0 avec les valeurs estimées $\hat{\pi}_0, \hat{\pi}_1$ indiquées ci-dessus, pour les 5 valeurs de u considérées . Le bruitage, très important pour $u = 2$ décroît lorsque u augmente tout en restant significatif en $u = 8$.

u	2	3	4	6	8
e (%)	46	41	38	33	30

Table 3.1: Erreur théorique de classification CA par maximum de vraisemblance. $u = \sigma_1^2/\sigma_0^2$ est le degré de différentiation des 2 textures.

3.2 Résultats et interprétation

3.2.1 Sur les classifications contextuelles

La Table 3.2 rassemble l'ensemble des *PEC* qui correspondent aux classifications aveugles CA et contextuelles CCS, CCG dont les tracés sont illustrés par la Figure 3.2 (ces courbes ainsi que les autres illustrations qui vont suivre se trouvent à la fin du chapitre). Nous rappelons que pour les classifications contextuelles (comme pour les segmentations globales) le contexte considéré est celui du premier ordre:

$$V_s = \{s\} + \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix} \right\} .$$

La classification CCG utilise les estimations suivantes des paramètres:

$$\begin{pmatrix} \hat{p} \\ \hat{q} \\ \hat{r} \end{pmatrix} = \begin{pmatrix} 0.17 \\ 0.56 \\ 0.27 \end{pmatrix} \quad \text{et} \quad \begin{pmatrix} \hat{\pi}_0 \\ \hat{\pi}_1 \end{pmatrix} = \begin{pmatrix} 0.39 \\ 0.61 \end{pmatrix} \quad . \tag{3.5}$$

Les dessins(c)-(d)-(e) de la Figure 3.10 montrent une expérimentation des classifications CA, CCS et CCG, où l'image x bruitée est un échantillon simulé avec $u = \sigma_1^2/\sigma_0^2 = 6$ et $\sigma_0^2 = 1$ (les classifications contextuelles CCS, CCG ont été suivies d'un post-traitement rapide de lissage, le "vote par majorité"). L'amélioration par la prise en compte du contexte est significative.

Bien que leurs pourcentages d'erreur de classification soient proches, la classification CCG semble, visuellement, de meilleure qualité que la classification CCS. Le problème des bords (pixels "frontière") en est la principale cause. L'hypothèse d'une complète "continuité spatiale" employée dans la classification CCS implique une moins bonne classification de ces pixels.

D'autre part, nous avons étudié la robustesse de la classification CCG par rapport au choix des paramètres (p, q, r). Une expérimentation exhaustive a été

u Méthode	2	3	4	6	8
CA	45.4	40.4	36.9	33.6	29.9
CCS	33.8	24.4	19.5	14.5	12.4
CCG	31.1	23.4	18.8	14.5	11.3

Table 3.2: Les *PEC*(%) des méthodes locales évalués par simulation.

u = Niveau de différenciation des 2 textures
CA = La classification aveugle
CCS = La classification contextuelle simple
CCG = La classification contextuelle géométrique

menée avec les 21 choix différents suivants de (p, q, r):

$$\begin{pmatrix} p \\ q \\ r \end{pmatrix} \in \left\{ \frac{1}{5} \begin{pmatrix} i \\ j \\ k \end{pmatrix} : i, j, k \in \mathbb{N} \text{ et } i + j + k = 5 \right\} . \tag{3.6}$$

La Figure 3.3 regroupe les 22 *PEC* correspondant à ces 21 triplets (p, q, r) plus la valeur estimée $(\hat{p}, \hat{q}, \hat{r})$ en fonction du paramètre du bruitage u. Elles forment un faisceau compact dont la variation maximale du diamètre vaut environs 1.8%. Le triplet $(\hat{p}, \hat{q}, \hat{r})$ est à peu près l'enveloppe inférieure du faisceau (les valeurs optimales de *PEC* resultant de ces 21 choix valent successivement 31%, 23%, 19%, 14% et 11% pour les 5 valeurs de u et peuvent être comparées avec celles obtenues en $(\hat{p}, \hat{q}, \hat{r})$). Cette étude montre donc que, par rapport aux paramètres (p, q, r) modélisant localement l'image binaire λ_0, la classification contextuelle géométrique CCG est robuste. Par contre, par rapport aux paramètres modélisant la texture x (ici u), il sera nécessaire de disposer d'une bonne estimation résultant directement de l'observation x.

3.2.2 Sur les segmentations globales

L'objet λ_0 étant binaire, un algorithme exact du MAP déduit d'une adaptation de l'algorithme de Ford-Fulkerson, est proposé dans PORTEOUS & al.[I–36]. C'est cet algorithme que nous avons utilisé. Pour le MPM et l'ICM, respectivement 500 balayages (1 balayage=1 parcours complet) et 10 balayages ont été éffectués. Pour toutes ces méthodes, une question fondamentale se pose: quel choix faire pour le paramètre de régularisation β dans (3.2)?

Les estimations de β par pseudo-vraisemblance ([I–2], [II–12]) et par la méthode

de Possolo[I–37] basées sur λ_0 donnent respectivement 2.5 et 1.5 (voir également le Chapitre 4 concernant les estimations du type EM [I–5] et par maximum de vraisemblance [I–45] (aussi sur la base de λ_0). Toutes les méthodes globales appliquées en $\beta = 1.5$ fournissent de mauvais résultats.

Afin d'étudier l'influence de ce paramètre de régularisation, nous avons expérimenté les 3 segmentations globales pour β variant de 0.1 à 1.5 avec un pas constant de 0.1. Les Figures 3.4, 3.5 et 3.6 donnent respectivement pour le MPM, le MAP et l'ICM et pour les cinq niveaux de bruitage, cinq courbes d'évolution de *PEC* en fonction de β. En ce qui concerne le MAP, à partir de $\beta = 1.5$ les cinq courbes admettent une même asymptote $PEC = 0.39$ associée au dessin restauré uniforme ($\hat{\lambda}_s = 0, \forall s$).

Le phénomène commun est qu'il existe bien un choix optimal de β, ici subjectif puisque λ_0 est connu, et ce choix dépend à la fois de la méthode de segmentation adoptée M et du niveau de bruitage u. Ces valeurs optimales observées $\beta(M, u)$ sont résumées par la Table 3.3. $\beta(M, u)$ est croissant en u pour une méthode M donnée. Un avantage des méthodes contextuelles locales par rapport aux méthodes globales, outre leur moindre coût en temps de calculs, est leur relative robustesse par rapport aux choix des paramètres des modèles (locaux) utilisés.

Les *PEC* des trois méthodes appliquées avec ces valeurs optimales $\beta(M, u)$ sont données dans la Table 3.4 (comparer avec la Table 3.2). dont les tracés sont illustrés par la Figure 3.7 Pour un fort bruitage, le MPM est préférable au MAP (voir également l'étude de MARROQUIN & al.[I–32]) et lorsque le bruitage devient faible, le MAP semble légèrement meilleur que le MPM. Signalons cependant que le critère retenu, le *PEC* , ne prend pas en compte la bonne régularité géométrique d'une reconstruction et donc délaisse en partie la qualité (mais aussi le défaut) régularisante du MAP si le bruitage est faible (fort).

u / Méthode	2	3	4	6	8
MPM	0.6	0.7	0.8	0.9	0.9
MAP	0.3	0.4	0.5	0.7	0.8
ICM	0.2	0.3	0.4	0.6	0.8

Table 3.3: Méthodes globales: valeurs optimales $\beta(M, u)$.
M = Méthode
u = Niveau de différenciation des 2 textures

La Figure 3.8 compare les différents comportements des trois méthodes globales par rapport au β et pour le niveau $u = 4$ (lorsque β décroît vers 0, les *PEC* tendent

u / Méthode	2	3	4	6	8
MPM	33.5	25.0	19.3	14.1	11.2
MAP	36.3	26.3	19.3	13.9	10.8
ICM	43.5	35.1	28.7	23.2	17.1

Table 3.4: Les *PEC*(%) des méthodes globales évalués par simulation.

u = Niveau de différenciation des 2 textures
MPM = Maximums des marginales a posteriori
MAP = Maximum a posteriori
ICM = Modes conditionnels itérés

tous vers celui de la classification aveugle CA). On remarque que l'ICM reste le plus robuste vis à vis de β mais aussi le moins performant (selon le *PEC*) et que le MAP se montre plus sensible que le MPM au choix de β. On notera que, quelque soit la méthode, le β optimal observé est très inférieur au β estimé "modélisant" λ_0. Ces résultats complètent et confirment ceux de [I–36]. Ils illustrent aussi la remarque de RIPLEY B.D.[I–38] concernant le choix de β pour l'ICM.

Les Figures 3.10 (f)-(g)-(h) sont les restaurations par le MPM, le MAP et l'ICM du même échantillon 3.10 (b).

La dernière observation concerne la Table 3.5 qui rassemble les temps CPU nécessaires à chacune des méthodes (sur un VAX 750). L'ICM reste le plus rapide parmi les trois méthodes globales. Le fait qu'ici le MAP consomme moins de temps que le MPM s'explique par deux raisons: d'abord l'algorithme utilisé pour le MAP est un algorithme exact et non itératif; ensuite, le MPM a été exécuté avec 500 balayages. En général, le MAP via le recuit simulé sera plus long que le MPM.

3.2.3 Remarques finales

L'étude expérimentale que nous venons de réaliser a permis d'apprécier différentes caractéristiques des méthodes de classification-segmentation considérées. L'examen des *PEC*, i.e. pourcentages d'erreur de classification rassemblés dans la Figure 3.9 ainsi que celui des dessins restaurés. (Figure 3.10) prouvent que, par rapport aux méthodes globales, les deux classifications contextuelles CCS et CCG fournissent des résultats satisfaisants. De plus ces deux méthodes, travaillant pixel par pixel sur des vecteurs de faible dimension et ne demandant pas d'algorithme itératif, sont économiques (cf. Table 3.5). D'autre part le fait, observé par MARROQUIN & al.[I–32], que la performance du MPM est meilleur que celle du MAP pour des

niveaux de bruit importants est confirmé.

Méthode	CA	CCS	CCG	ICM	MAP	MPM
temps CPU (s)	0.2	0.4	2	0.5	94	167

Table 3.5: Temps CPU consommés. VAX 750 et image 32×32

Cette étude, pour être opérationnelle sur des données réelles, doit évidemment être complétée par une réponse à la question essentielle suivante: au vu de l'observation x, comment choisir les paramètres (ici β ou (p, q, r), (π_0, π_1) relatifs à l'image à restaurer, u relatif au modèle de bruitage). En effet, pour les méthodes globales, nous avons constaté la forte sensibilité des restaurations au choix du paramètre β pour l'image à restaurer (par ordre croissant de sensibilité: l'ICM, le MPM et le MAP). Les valeurs optimales retenues (au sens du *PEC*) se trouvent nettement plus petites que les valeurs des estimations statistiques, ces dernières fournissant de mauvaises restaurations. Le Chapitre 4 qui suit poursuit l'examen de cette question.

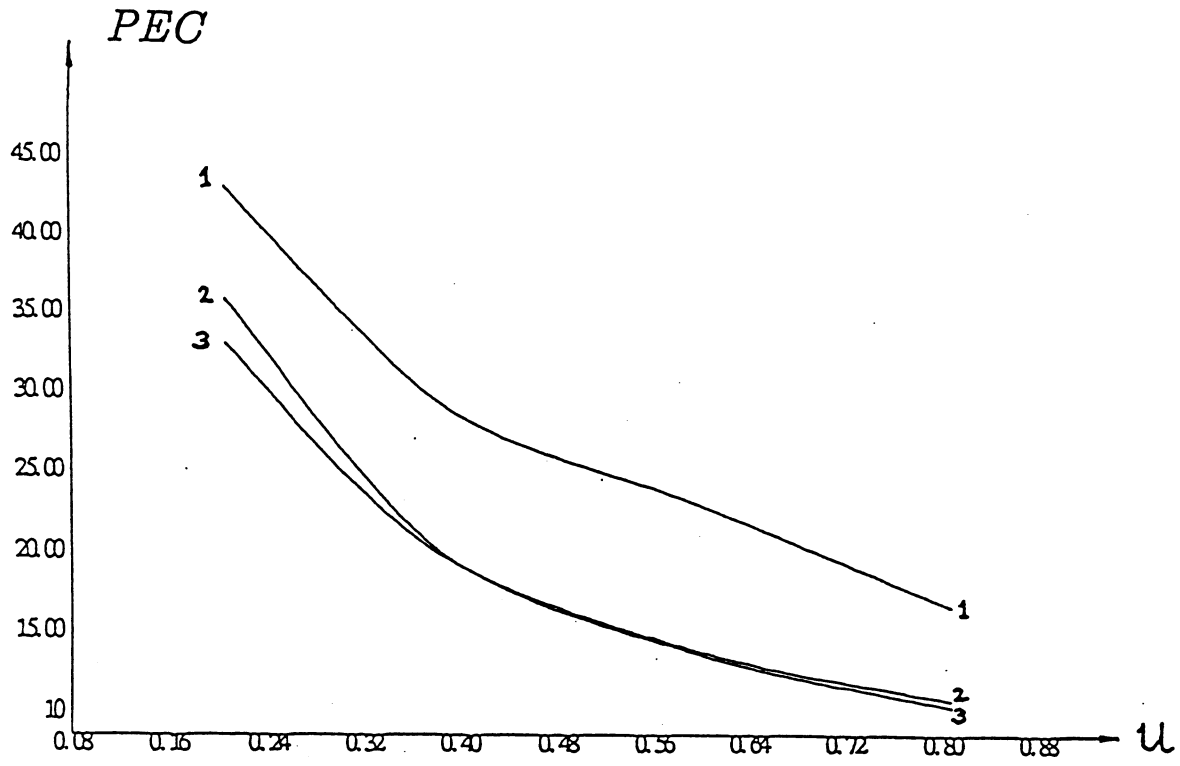


Figure 3.2: *PEC* des méthodes locales en fonction du niveau u de différenciation des 2 textures.
1 = CA, 2 = CCS, 3 = CCG.

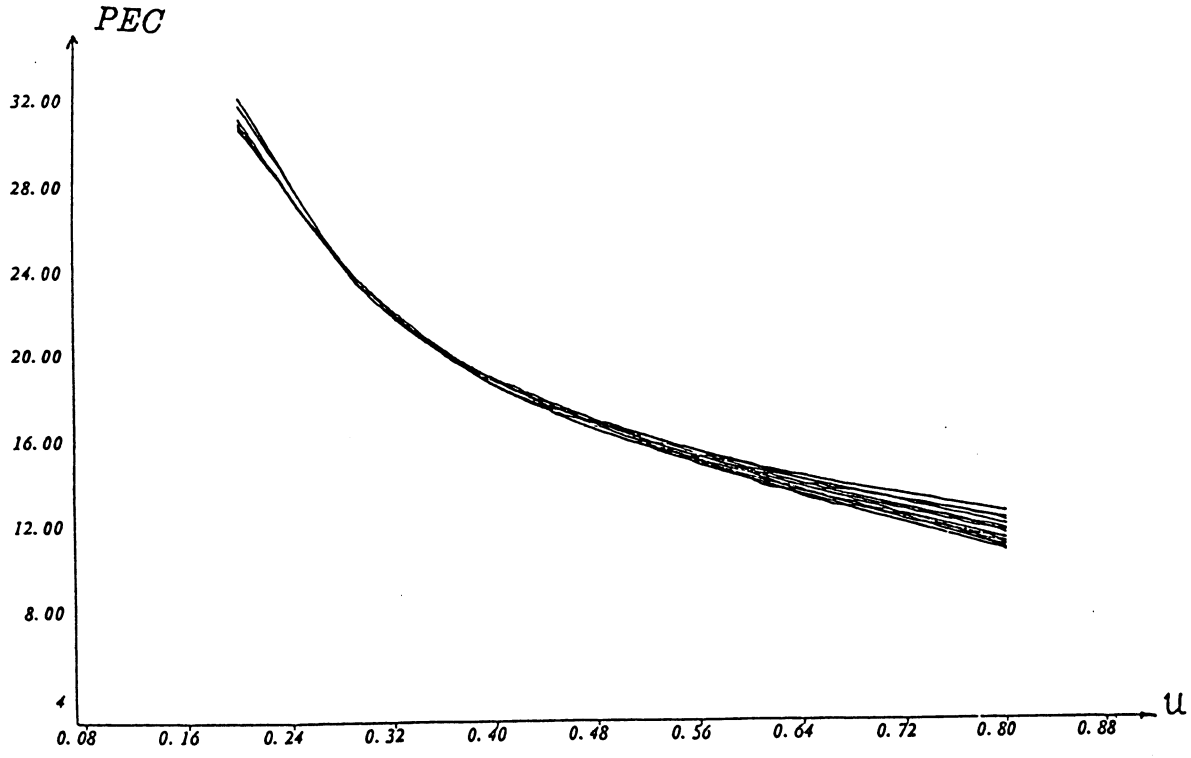


Figure 3.3: Robustesse en (p, q, r) de la classification CCG (le faisceau correspond au balayage des 21 choix possibles de (p, q, r) donnés par (3.6)).

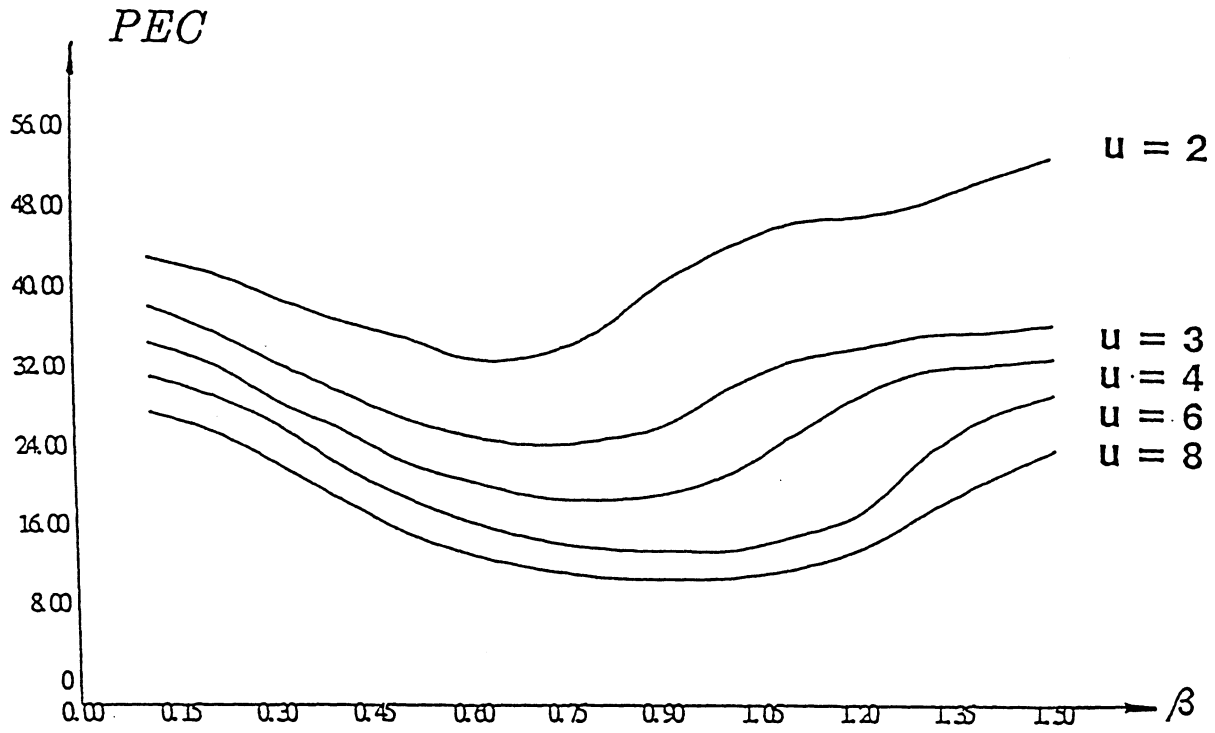


Figure 3.4: MPM: variation du *PEC* suivant le paramètre β. Sensibilité du MPM en β pour différents niveaux de différenciation des 2 textures u.

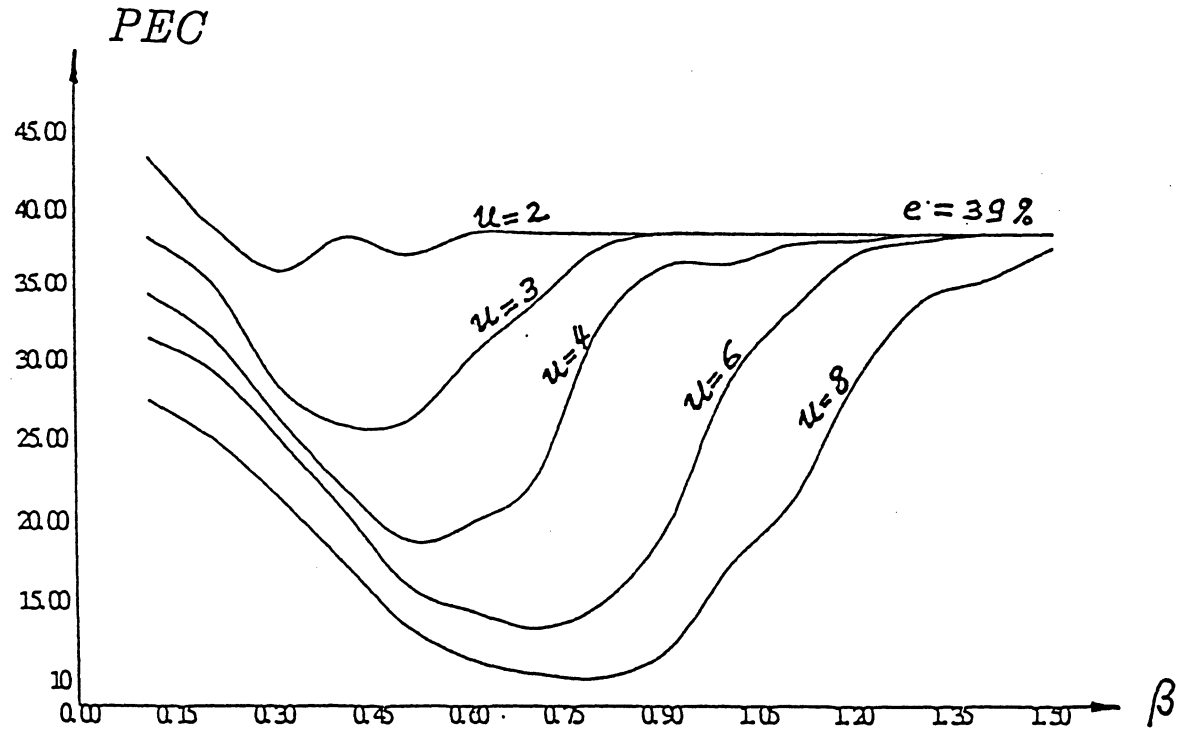


Figure 3.5 MAP: variation du *PEC* suivant le paramètre β. Mauvaise robustesse en β du MAP lorsque la différenciation u est grande (bruitage faible).

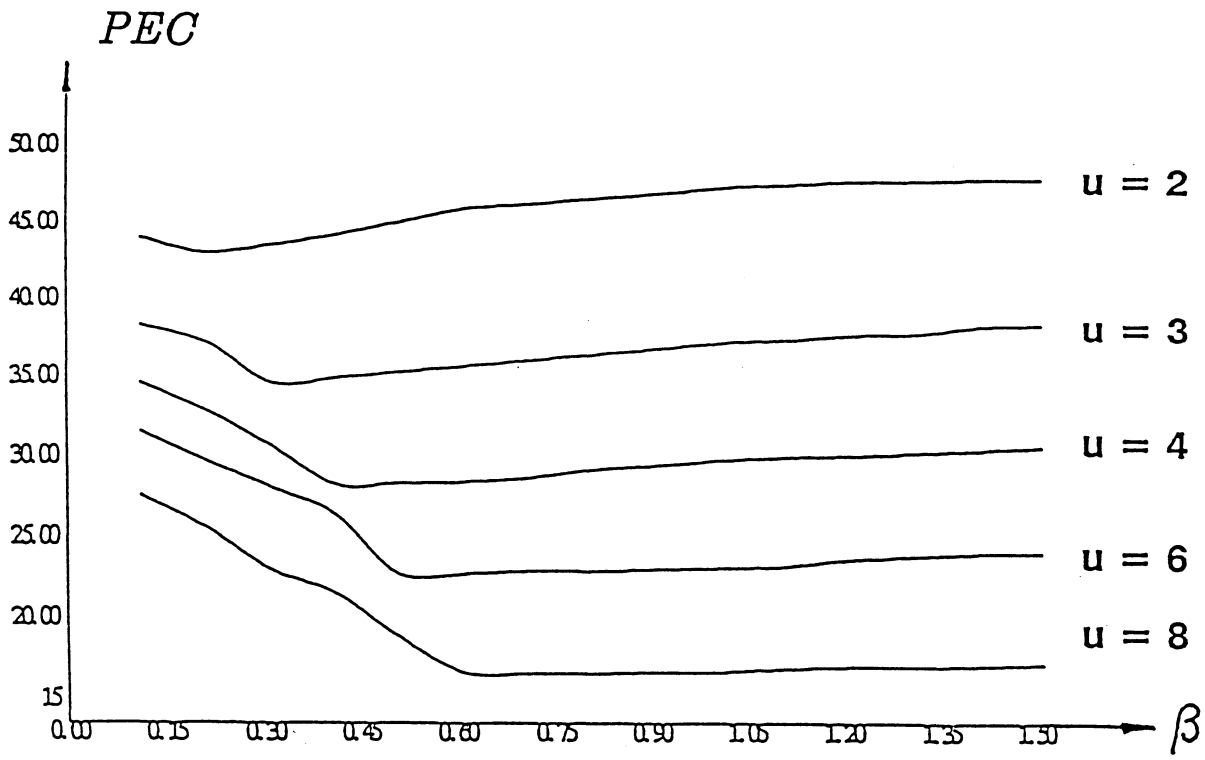


Figure 3.6: ICM: variation du *PEC* suivant le paramètre β.
Phénomène de plateau lorsque β augmente.

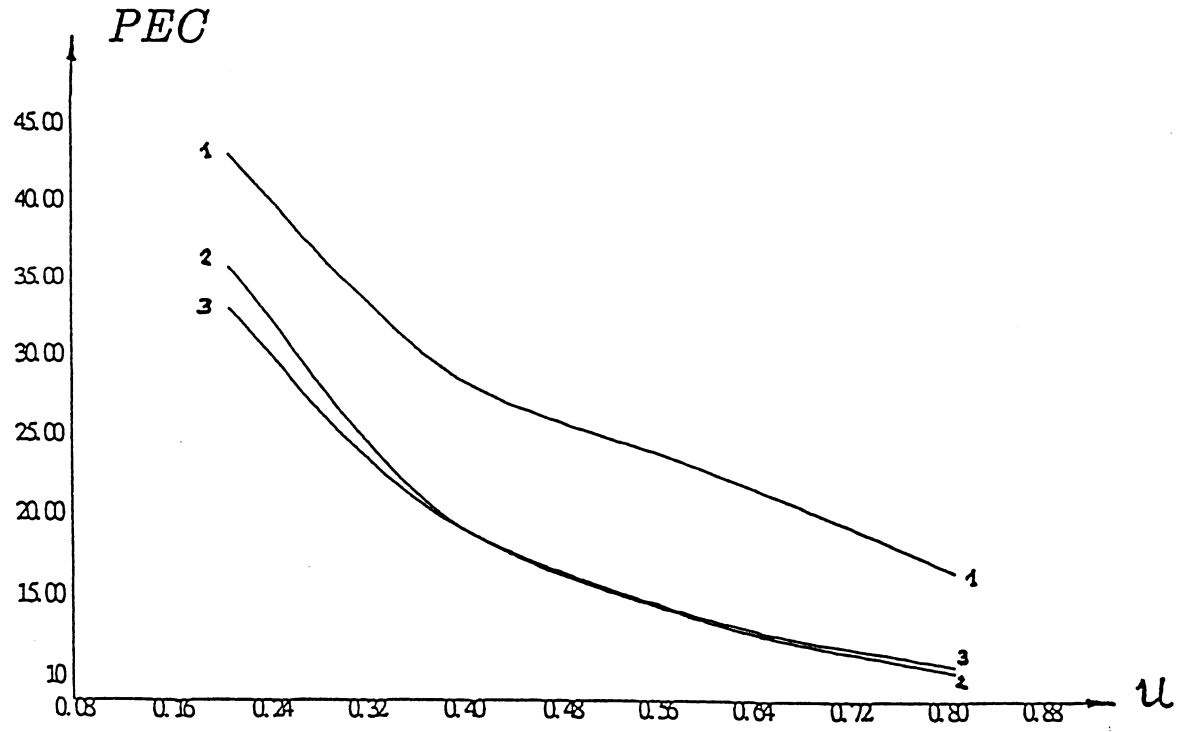


Figure 3.7: *PEC* des méthodes globales en fonction du niveau u
(de différenciation des 2 textures) au $\beta(M,u)$ optimal.
1 = ICM, 2 = MAP, 3 = MPM.

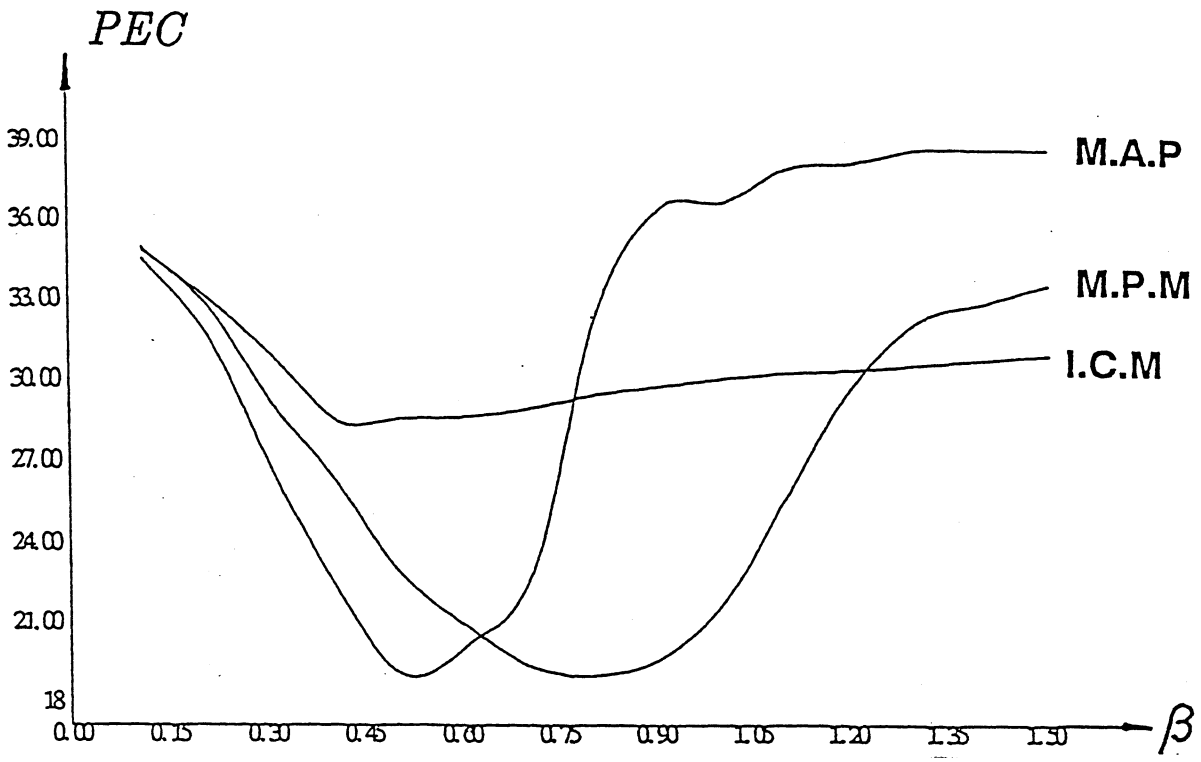


Figure 3.8: Comparaison des robustesses en β des méthodes globales.

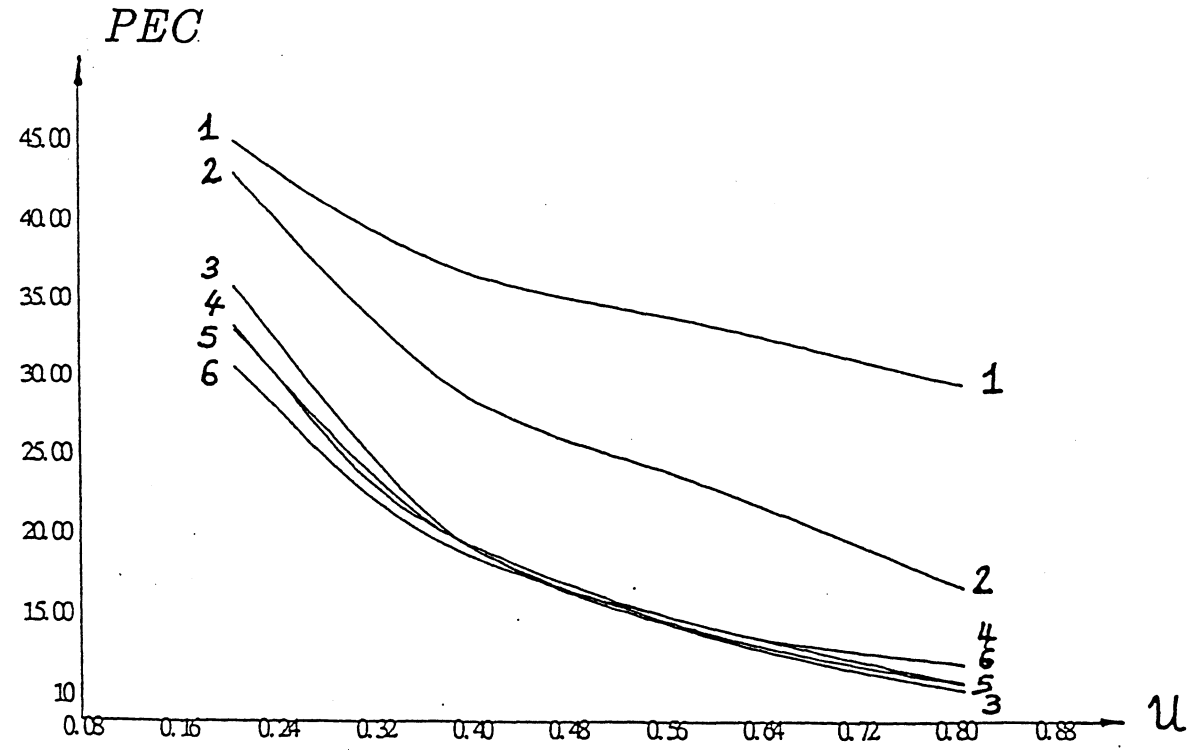
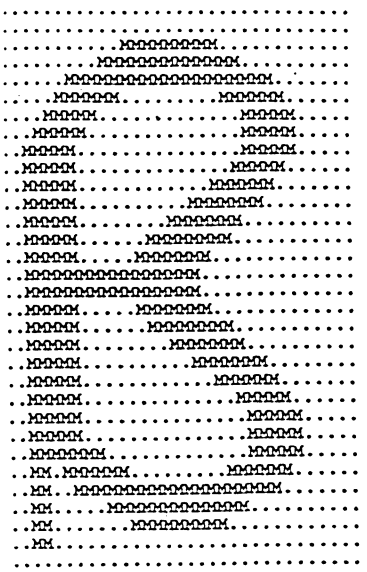
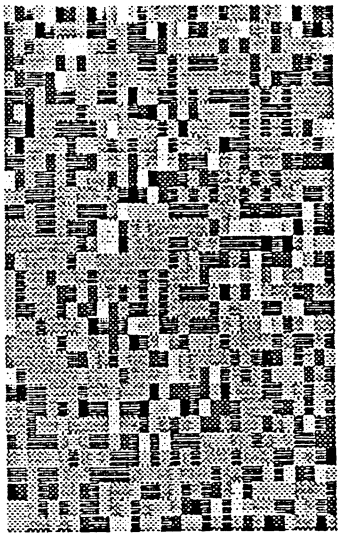


Figure 3.9: Ensemble des *PEC* des méthodes locales et globales (au $\beta(M,u)$ optimal).
1 = CA, 2 = ICM, 3 = MAP,
4 = CCS, 5 = MPM, 6=CCG.



a



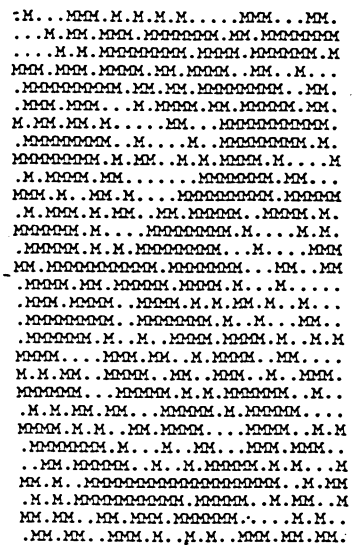
b

Figure 3.10:
(a). Image originale (32 × 32).

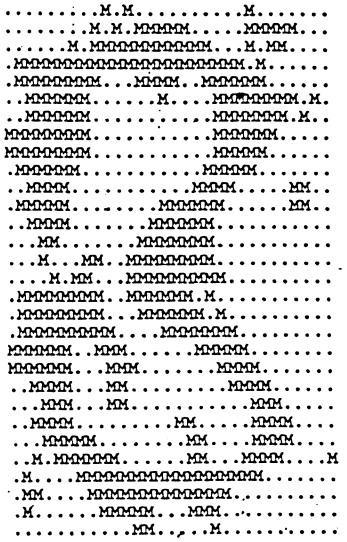
(b). Textures simulées à partir de (a): $\sigma^2_{M^*} = 1, \sigma^2_{\cdot^*} = 6$.

Segmentations de (b):

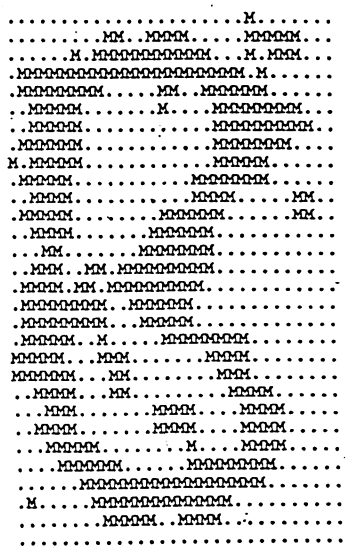
(c) = CA, (d) = CCS,
(e) = CCG, (f) = MPM,
(g) = MAP, (h) = ICM.



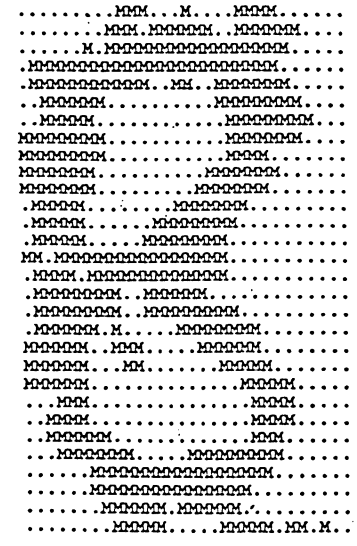
c



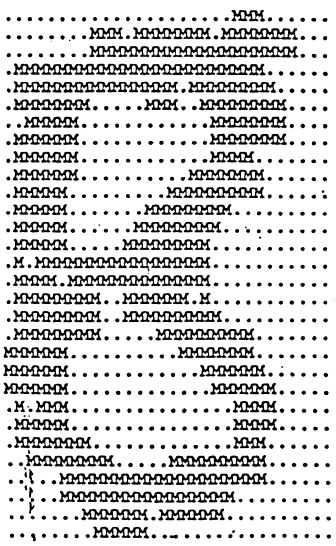
d



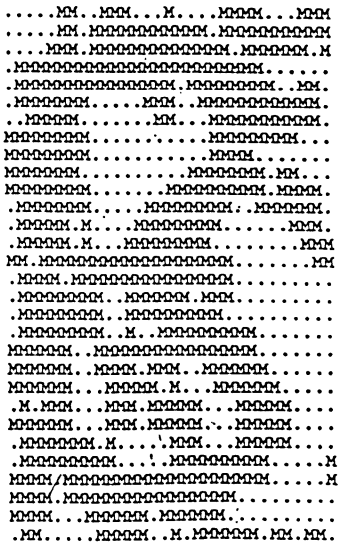
e



f



g



h

Chapitre 4

Sur le choix du paramètre de régularisation

Ce chapitre reprend [I–15], une collaboration de DINTEN J.M. & GUYON X. avec l'auteur.

Dans ce papier, nous apportons une étude plus approfondie sur le problème du choix du paramètre de régularisation déjà relevé dans le Chapitre 3.

On the choice of the regularization parameter: the case of binary images in the Bayesian restoration framework

J.M. Dinten* X. Guyon† J.F. Yao

October 1988

*C.N.R.S Statistique Appliquée UA 743, Université Paris-Sud, Bât. 425, 91405 ORSAY Cedex, France *and* C.E.A Centre d'Etude de Vaujours, BP N^o 7, 77181 COURTRY, France

†C.N.R.S Statistique Appliquée UA 743, Université Paris-Sud,Bât. 425, 91405 ORSAY Cedex, France

Abstract

We study the problem of the influence and of the choice of the regularization parameter β in the Bayesian image restoration framework. Binary and geometrically regular images are examined. The noise degradation process which leads to the observed record can be either additive gaussian or a binary symmetric channel noise of transmission. MAP is not robust with respect to β, MPM and ICM are more robust. For the three methods, a good choice of β depends strongly on the noise level. On the basis of the observed record, two possible choices of β are examined: if the statistical one seems reasonable at a low noise level, it isn't the case for higher noise for which the cross-validation criterion still gives good results.

Keys words: Bayesian restoration of images,
Regularization methods,
Choice of regularization parameter.

A.M.S subject clssification :
Primary : $60G35$, $62M20$
Secondary : $62M30$, $62M99$

1 Introduction: the context of our study

Let $x = \{x_s, s \in S\}, S = \{1, 2 \ldots, n\}^2$ an image to be reconstructed from an observation $y = \{y_s, s \in S\} = \Phi(x, \eta), \Phi$, the noise mechanism, know as well as the noise η. Suppose that $E(x)$ is an appropriate energy on the configuration x which summarizes the prior information about it and $D(y, x)$ is some fidelity distance between y and x. Then classical regularization methods estimate x by $\hat{x}(\beta) = \arg Min_x H(x|y;\beta)$, where the posterior energy H depends on a so-called regularization parameter β as following

$$H(x|y;\beta) = D(y, x) + \beta E(x)$$

Standard M.A.P restoration ([6]) are exactly of this kind with $(-H)$ the logarithm of the posterior probability $\Pr(x|y, \beta)$, β being the parameter of a prior law on x. M.P.M method ([11]) is also defined from such a scheme by choosing, in each pixel s : $\hat{x}_s = \arg\ Max_{x_s}\ \Pr(x_s|y;\beta)$, where this marginal probability in x_s is deduced from H. The prior E can be chosen as a Markovian energy, an entropy or a regularization function as for example, flexion in the spline smoothing context. The distance D is always directly derived from both the degradation process Φ and the noise η.

A crucial question is: *how to choose the regularization parameter β on the basis of the record y ?* Our paper will give some hints for such a choice based on experimental results in a specific and well defined context. We first examine the dependency of the reconstruction $\hat{x}(\beta)$ in β (Section 2) as well as some choice for β like statistical estimation , cross-validation choice and joint M.A.P estimate on (x, θ)— θ is the parameter of the y–model (Section 3). This will be studied for two kinds of binary images: the first one being the realization of some Markov random field (M.R.F); the other one a hand-drawn picture. The noise degradation can be a binary symmetric channel (B.S.C) noise of transmission, an additive gaussian one or a textural variance noise.

By *well defined context* we mean that we are able to give some answers to the following questions:

1. What particular family is being studied in the large botanic of images?

2. For a known degradation model (Φ,η), what is the level of the noise: low, medium or high?

3. What is the quality criterion for restoration?

4. What is the regularization method used?

Without reasonable answers to such questions, there can be no reasonable results, experimental or theoretical, to the problem of choice of the regularization parameter.

Test images are chosen to be binary and regular in a natural geometric sense. In particular we will discuss the influence of the second and the fourth questions above on the problem. The noise level remain high giving from 20% to 40% error rate in the pixel by pixel maximum likelihood restoration. The mean percentage of pixel calassification error rate is our restoration criterion (we note it by PEC through all the paper) and M.A.P, M.P.M and I.C.M methods will be experimented.

In the literature, theoretical results are obtained in the following two situations:

- regularization by spline functions ([4],[14],[18]) or approximate solutions for integral equations of the first kind ([9],[16],[17])

- smoothing techniques in image restoration in a $\mathcal{L}^2$ framework ([8]).

In each of these situations, the mathematical context is well defined and theoretical answers can be derived to help us in the choice of the regularization parameter: cross-validation choice (or a more easily computable variant) in the first context; for the second one where it is assumed that the variance σ^2 of the noise is low, one must choose the regularization parameter to be proportional to σ^2 whereas all classical choices suggest to take it proportional to σ which is too regularizing.

Let us describe now more precisely the object x and its noisy observation y. Two *true* images x are to be recontructed:

I1 The first one is the realization of a binary isotropic M.R.F (with $H(x) = \beta E(x)$) :

$$H(x) = -\beta \sum_{<s,t>} \delta(x_s, x_t) \tag{1}$$

Here the neighbourhood system $\{< s,t >\}$ is the four nearest one. Figure (1-a) gives such a realization on a 64×64 lattice, using free boundray conditions with $\beta = \beta_0 = 1.149$ and the Gibbs sampler ([6]). This one is run during 80 raster scans starting at a 0–1 uniform i.i.d configuration. Figure (3-a) gives another 128×128 realization for $\beta_0 = 1.40$.

I2 The second one is a hand-drawn (looking like a β) 32×32 image given in figure (6-a)[1]

The noise degradation is assumed pixel-independent:

$$\Pr(y|x) = \prod_{s \in S} \Pr(y_s|x_s)$$

We will examine the following three kinds of noise:

N1 Binary symetrical channel noise (B.S.C.,[11]):

$$\forall s \in S, \ \Pr(y_s|x_s) = \begin{cases} \epsilon & \text{if } y_s \neq x_s \\ 1-\epsilon & \text{if not} \end{cases} \tag{2}$$

Experiments are realized with $\epsilon = 0.2$, 0.3 and 0.4 .

N2 Additive gaussian noise:

$$y = x + \eta \tag{3}$$

Here η is a gaussian white noise with mean zero and variance σ^2. σ is chose as $0.594, 0.953$ and 1.974 which corresponds to error rate $20\%, 30\%$ and 40% in maximum likelihood classification .

N3 Texture of variance noise :

$$\forall s \in S, \ \Pr(y_s|x_s) \sim \mathcal{N}(0, \sigma^2_{x_s}) \tag{4}$$

Let us define $u = \sigma_1^2/\sigma_0^2$ and suppose that it is greater than 1. For a fixed σ_0^2, the smaller u, the greater the noise. With $\sigma_0^2 = 1.0$ and $u = 2, 4, 8$, the M.L classification error rate are respectively $46\%, 38\%$ and 30%.

[1]For reconstruction, the same energy (1)) is used.

2 Influence of the regularization parameter β on the restored image

Using the prior (1) for x and the noise degradation (2)–(3)–(4) we obtain the three posterior energies $H_i(x|y)$, $i = 1, 3$:

$$\begin{cases} H_i(x|y) = & -\alpha_i \sum_s (2y_s - 1)x_s - \beta \sum_{<s,t>} \delta(x_s, x_t), \quad i = 1, 2 \\ & \alpha_1 = \ln \frac{(1-\epsilon)}{\epsilon}, \quad \alpha_2 = \frac{1}{2\sigma^2} \\ H_3(x|y) = & -\frac{1}{2} \sum_s x_s \left[\ln \frac{\sigma_0^2}{\sigma_1^2} + y_s^2 \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right) \right] - \beta \sum_{<s,t>} \delta(x_s, x_t) \end{cases} \tag{5}$$

On the basis of such posterior probabilities, MAP, MPM and ICM are used to reconstruct the image x. For the MAP and to avoid the dependency on a particular cooling schedule, we use the exact maximization algorithm proposed in [12] for binary images. Referring to this work, we can see that a pratical simulated annealing can differ strongly from the exact MAP and that it is also strongly dependent on β (see the test image **A** of [12]). The MPM solution is obtained via the Gibbs sampler after fifty raster scans (with the first ten sweeps run to reach stationarity). The ICM restoration is obtained after 8 sweeps.

2.1 Markov ramdom field image

2.1.1 M.R.F image 64×64 for $\beta_0 = 1.149$ (see figure (1-a))

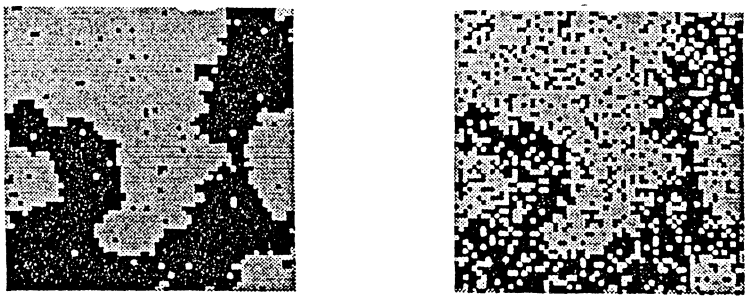
Experiments carried out in ([11]) prove that for low noise level, MAP and MPM for $\beta = \beta_0$ give good restorations[2]. For higher noise level, it seems that such property is not preserved and we have performed experimentations to examine the subjective criterion $PEC(\beta)$ in β.

[2] In [11] MAP restorations are obtained via simulated annealing with a logarithmic schedule.

First note that, based on the true original image (1-a), pseudo-maximum likelihood estimation ([1],[7]) for β gives $\hat{\beta}_{PML} = 1.01$ and the stochastic algorithm for maximum likelihood estimation proposed by ([21],[22]) gives $\hat{\beta}_{ML} = 0.92$. From a B.S.C noisy record with the error rate $\epsilon = 30\%$, the Gibbsian E.M. algorithm developped by B. Chalmond([3]) gives $\hat{\beta}_{EM} = 1.00$ (and $\hat{\epsilon} = 0.28$). Such results shows that x is in accordance with a markovian realization of model (1) for $\beta_0 = 1.149$.

Figure 1.b-c-d are noisy records y for B.S.C noise levels $20\%, 30\%$ and 40%.

The curves shown in Figure 2.a-b-c illustrates the variations of the PEC in β for the M.A.P, the M.P.M and the I.C.M respectively. Each point of the curve was a mean point based on five independent realizations of the noise.



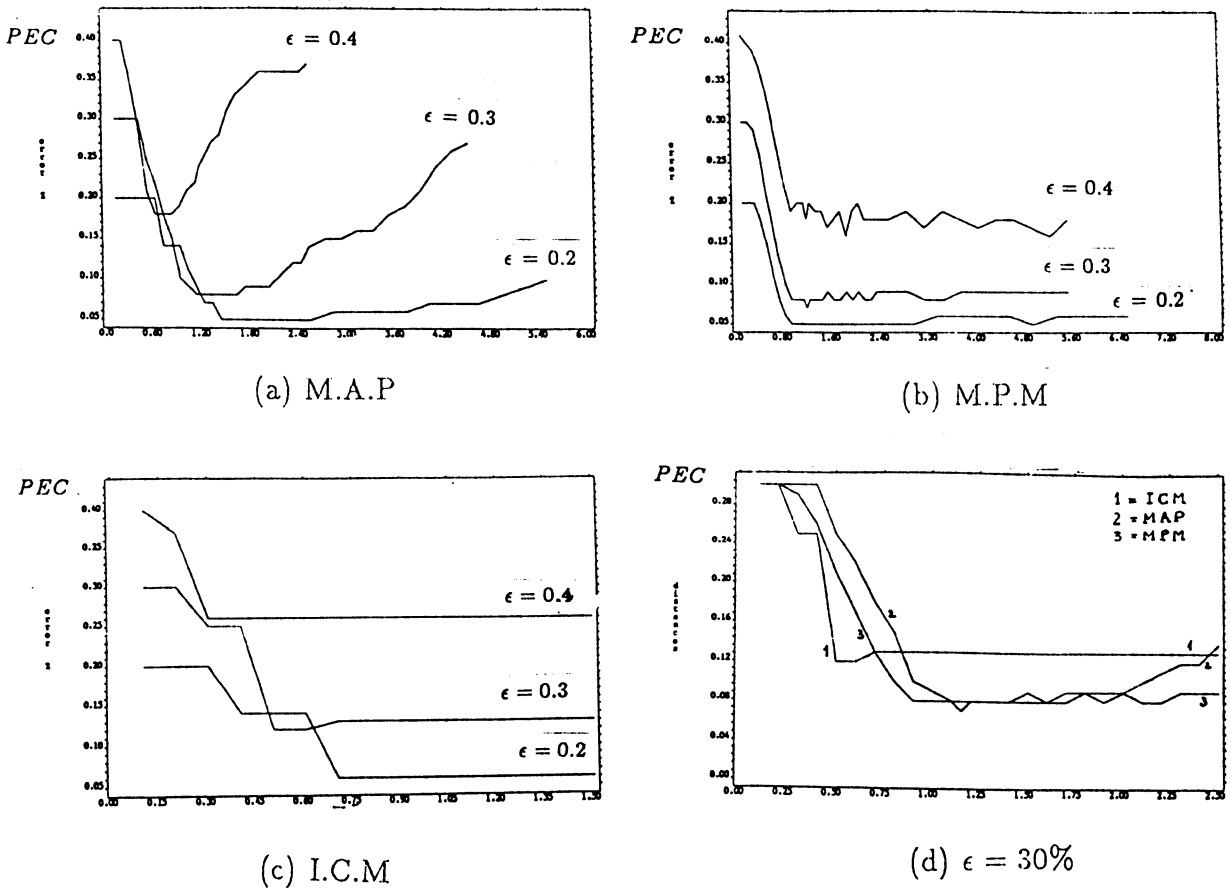
(a) original (b) $\epsilon = 20\%$

(c) $\epsilon = 30\%$ (d) $\epsilon = 40\%$

Figure 1: Original picture 64×64 realization of a M.R.F with $\beta_0 = 1.149$ and B.S.C noisy images

For the Figure 2-b (M.P.M method) a stochastic behavior of $PEC(\beta)$ appears: this is a consequence of the Monte-Carlo algorithm used to compute the marginal mode and based on a relatively little number of iterations (50 in fact). Such behavior fluctuation doesn't exist in Figure 2-a or 2-c, because for the M.A.P and the I.C.M the performed algorithms are deterministic. Figure 2-d gives, for $\epsilon = 30\%$, PEC for the three methods.

For the I.C.M reconstruction, a plateau phenomenon appears (See Appendix A): if β exceeds some threshold $\beta(\epsilon)$ depending on the noise level ϵ, the I.C.M and the Iterated Modal Filtering are equivalent. Such a behavior can also be observed for the M.P.M.



(a) M.A.P

(b) M.P.M

(c) I.C.M

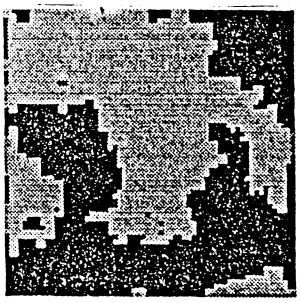
(d) $\epsilon = 30\%$

Figure 2: The curves of $PEC(\beta)$ at various noise for MAP, MPM and ICM

Figures 3.a-b-c give the M.A.P, M.P.M and I.C.M reconctructions of the noisy image 1-d with the (subjective) optimal choice of β whereas Figures 3.d-e-f give such reconstructions for $\beta = \beta_0$. Table 1 gives the (subjective) optimal choices for each method and each noise level.

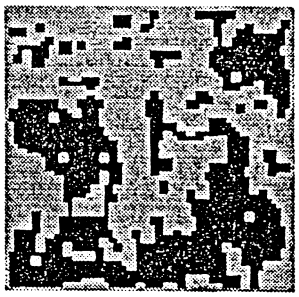
Method	$\epsilon = 0.2$	$\epsilon = 0.3$	$\epsilon = 0.4$
ICM	$[0.7, \infty)$	$[0.5, \infty)$	$[0.3, \infty)$
MAP	$[1.4, 2.5]$	$[1.1, 1.6]$	$[0.6, 0.8]$
MPM	$[1.1, 1.7]$	$[1.1, 1.7]$	$[1.5, 1.6]$

Table 1. Optimal values for β



(a) $\beta = 0.7(18\%)$

(b) $\beta = 1.8(16\%)$

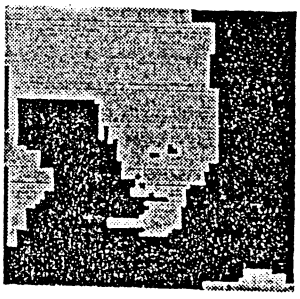


(c) $\beta = 0.3(26\%)$

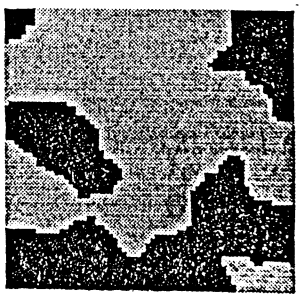
M.A.P

M.P.M

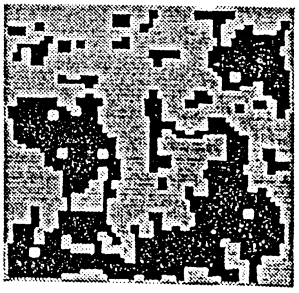
I.C.M



(d) $\beta = \beta_0(24\%)$



(e) $\beta = \beta_0(18\%)$

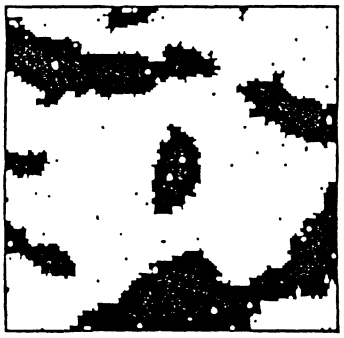


(f) $\beta = \beta_0(26\%)$

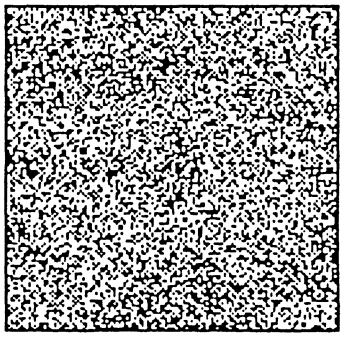
Figure 3: Restorations of image 1-d ($\epsilon = 40\%$). (a)-(b)-(c) with $\beta_{optimal}$; (d)-(e)-(f) with β_0. For each restoration, the PEC is given in brackets.

2.1.2 M.R.F. Image 128×128 for $\beta_0 = 1.4$ (See Figure 4-a)

A similar study was done for a 128×128 realization of the M.R.F with energy function (1) and $\beta_0 = 1.4$. Figure 4.b is the noisy image y with B.S.C noise level $\epsilon = 40\%$. Figures 4.c-d give the optimal M.A.P and M.P.M reconstruction whereas Figures 4.e-f are the same restorations for $\beta_0 = 1.4$. Estimations on the basis of x give $\hat{\beta}_{PMV} = 1.42$, $\hat{\beta}_{ML} = 1.19$.



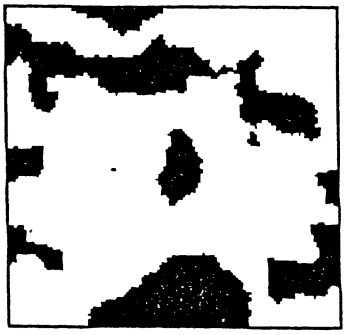
(a) original



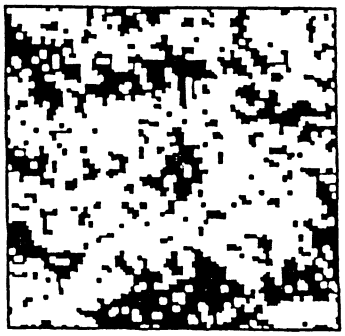
(b) $\epsilon = 40\%$

Figure 4: (a). Original image: 128×128 M.R.F with $\beta_0 = 1.4$ (b) B.S.C record with $\epsilon = 40\%$; (c)-(d)-(e) with optimal choice of β; (f)-(g)-(h) with β_0.

(c) M.A.P $\beta = 0.8(11\%)$

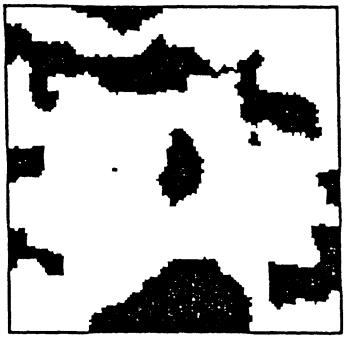


(d) M.P.M $\beta = 1.4(12\%)$

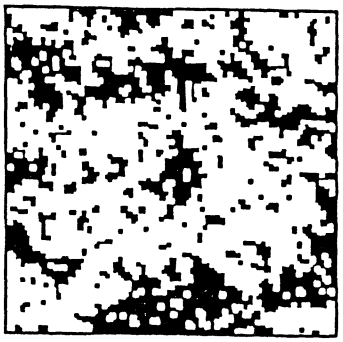


(e) I.C.M $\beta = 0.3(22\%)$

(f) M.A.P $\beta = \beta_0(26\%)$



(g) M.P.M $\beta = \beta_0(12\%)$



(h) I.C.M $\beta = \beta_0(22\%)$

Experiments on y based on other independent realizations of the noise lead to the same results. The PEC curves in β are given in Figure 5. On this graphic, we have also show the variation of another (more contextual) criterion, the percentage of misclassified windows:

$$PEC_2(\beta) \stackrel{\text{def}}{=} \frac{1}{c} \sum_{s \in S} \delta(x_{W_s}, \hat{x}_{W_s})$$

where $\{W_s\}$ is the nearest neighbour system and c a normalizing constant.

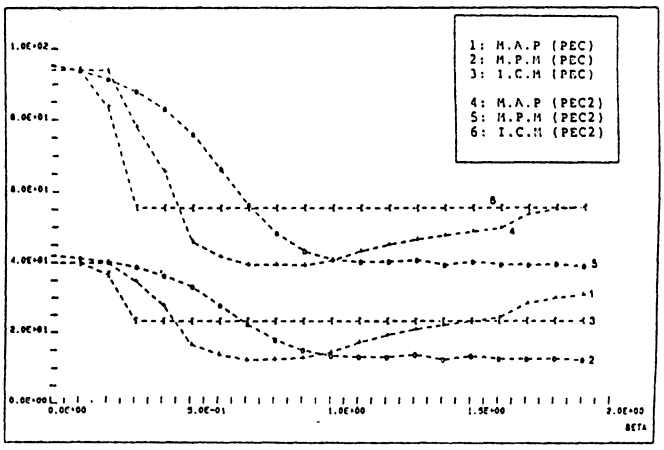


Figure 5: Restoration of the 128×128 record 4-b. The curves of $PEC(\beta)$ and $PEC_2(\beta) + 0.25$.

As can be seen here, the two optimal choices relative to both criterions are very close.

From these observed results,one can note that

- The M.A.P restoration is strongly sensible to β, all the more if the noise is high. The (subjective) optimal value of β (see Table 1) is decreasing when the noise increases. At the "error rate" $\epsilon = 40\%$, this value becomes significantly less than to β_0.
- For the MPM, optimal value β_{MPM} of β are nearer to β_0 and for

$\beta \geq \beta_{MPM}$, there is a plateau phenomenon giving good robustness in β in the following sense: too large choices for β are not very dangerous.

- An accentuation of this phenomenon appears for the ICM, the plateau for the PEC function begin at a threshold β_{ICM} which is independent of β_0 for B.S.C noise ($\beta_{ICM} = \alpha/2$, see Appendix A).
- For the three methods(M), the optimal value $\beta_M(\epsilon)$ is decreasing in ϵ. We have also noted (see fig.5), using two distinct criterions, that the dependency on these criterions is not strong.

2.2 Hand-drawn image 32×32 (see Fig.6-a)

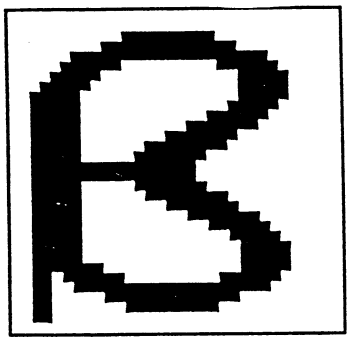
We have performed various estimations for the parameter β of the a priori energy(1) for this image: from x itself we obtained:

$$\hat{\beta}_{PML} = 2.74 \quad \hat{\beta}_{ML} = 0.86 \quad \hat{\beta}_{POS} = 1.58$$

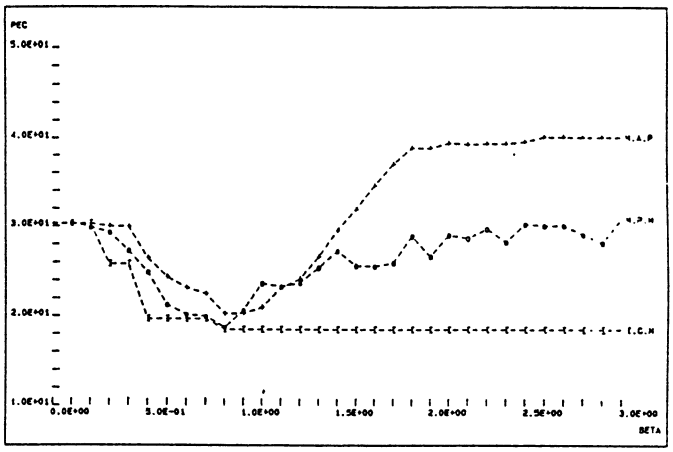
where $\hat{\beta}_{POS}$ is the logistic estimation proposed by A.Possolo for binary images([13]). From the record y and the Gibbsian algorithm([3]) we obtain:

- For a B.S.C at $\epsilon = 20\%$: $\hat{\beta}_{EM} = 1.15$, $(\hat{\epsilon}_{EM} = 0.20)$.
- For an additive Gaussian noise,$\sigma = 0.594$: $\hat{\beta}_{EM} = 0.99$, $(\hat{\sigma}_{EM} = 0.565)$.

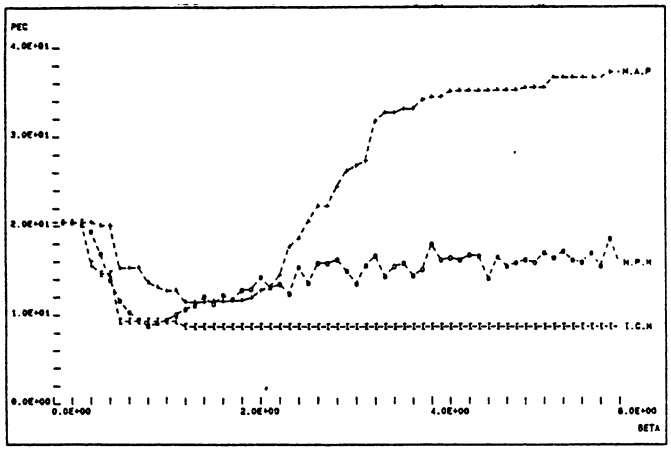
As it was observed for M.R.F Images, the structure of the noise is very well recovered by the E.M estimation, but here, the four estimations proposed for β are widely scatered. Two reasons can explain this: the small size of the image, but also some inadequacy of this image to be the realization of a M.R.F driven by (1).



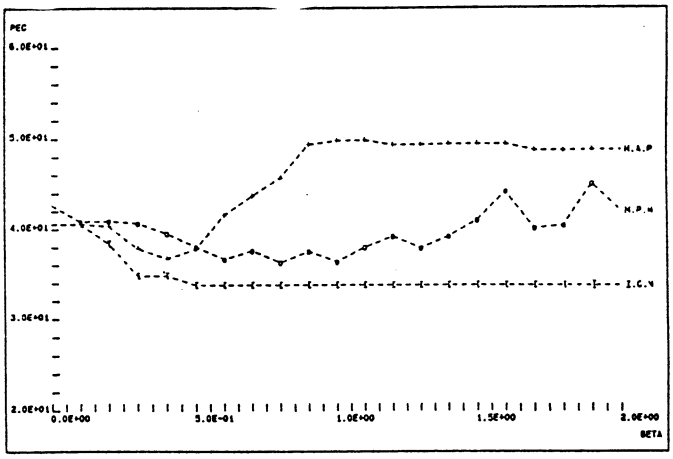
(a) original



(c) $\epsilon = 30\%$



(b) $\epsilon = 20\%$



(d) $\epsilon = 40\%$

Figure 6: The original handrawn image and the curves of $PEC(\beta)$ (B.S.C noise)

2.2.1 B.S.C Noise degradation

Figures 6.b-c-d give $PEC(\beta)$ at the three levels $20\%, 30\%, 40\%$. For $\beta \geq \beta_{ICM}(\epsilon)$, ICM is equivalent to Iterative Modal Filtering and superior to any MAP or MPM. This phenomenon is specific to this kind of noise as we shall see later for additive Gaussian or variance texture noise. As we have explained before, the fluctuation of the MPM-curves results from the small number of iterations in the Monte Carlo algorithm.

$\beta_{MPM}(\epsilon)$, the optimal choice of β , has little sensitivity to ϵ whereas $\beta_{MAP}(\epsilon)$ has a great sensibility in ϵ (Table 2)

Method	*B.S.C Noise*			*Gaussian noise*		
	$\epsilon = 0.2$	$\epsilon = 0.3$	$\epsilon = 0.4$	$\sigma = 0.594$	$\sigma = 0.953$	$\sigma = 1.974$
MAP	1.7	0.9	0.4	1.6	1.3	0.5
MPM	1.0	0.9	0.8	1.1	0.9	0.7

Table 2. Optimal β

2.2.2 Additive Gaussian noise:

For additive Gaussian noise at level $20\%, 30\%, 40\%$ ($\sigma = 0.594, 0.953, 1.974$ respectively) figures 7.a-b-c give the PEC in β. Here, MPM at β_{MPM} is optimal, there is still a plateau phenomenon for ICM and MPM; but the optimal choice β_{ICM} gives a better result than Iterative Modal Filtering.

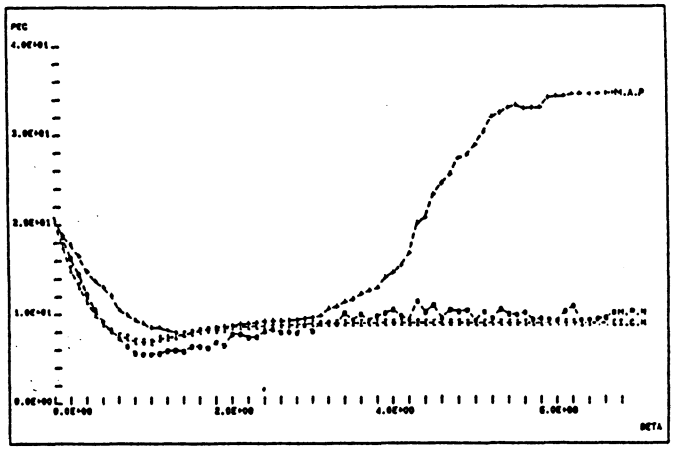
Table 2 shows the relative stability of β_{MPM} in ϵ whereas β_{MAP} is quite variable.

2.2.3 Variance texture noise:

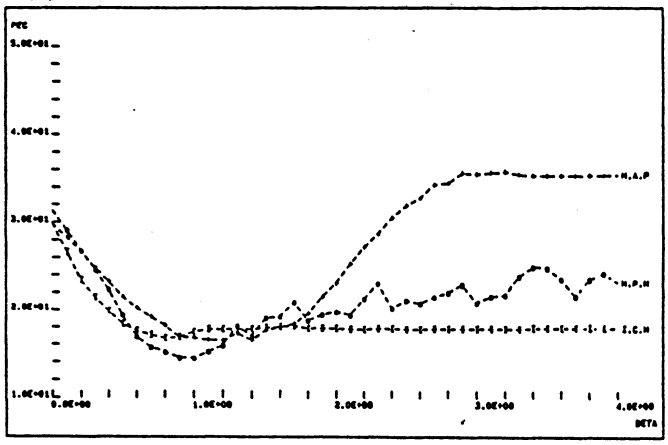
Figure 8-a gives such a degradation at level $u = \sigma_1^2/\sigma_0^2 = 4$ ($\sigma_0^2 = 1$) and curves 8.b-c give respectively PEC in β for the MAP at various levels and PEC at level $u = 4$ (error rate 35%) for MAP, MPM and ICM.

If $\beta > 1.50$ all MAP restorations are uniformly white: this is a negative effect of the strong spatial regularization property of the MAP restoration if β is too large. For MPM, robustness in β is better than for MAP but it is not very good indeed; ICM is still robust for $\beta > \beta_{ICM}$ but its quality is

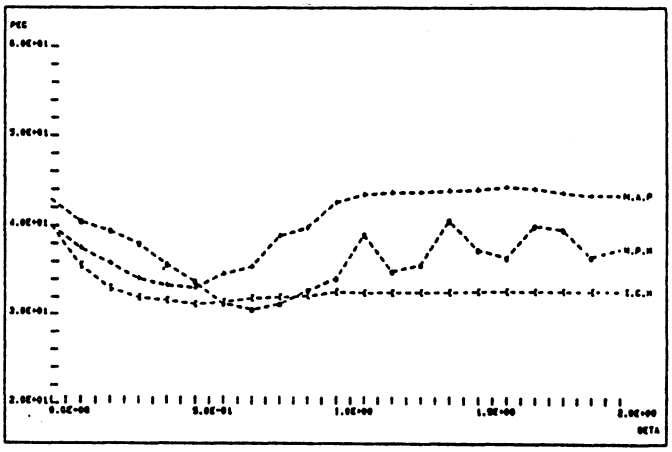
relatively poor. (See also [20] where comparison between global and local bayesian restoration is studied).



(a) $\sigma = 0.594$

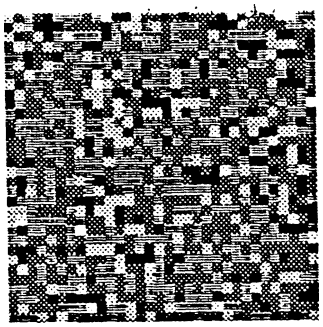


(b) $\sigma = 0.953$

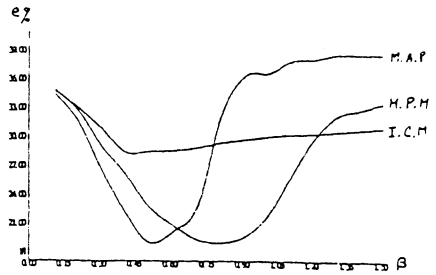


(c) $\sigma = 1.974$

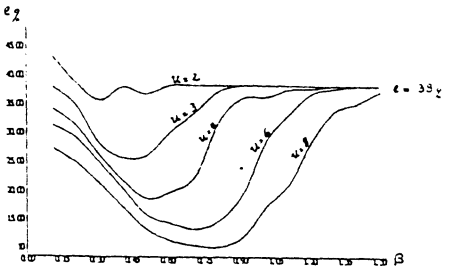
Figure 7: The curves of $PEC(\beta)$ (Gaussian noise on the original image of Fig. 6-a).



(a) noisy image ($u = 4$)



(b)$PEC(\beta)$ curves of M.A.P



(c) $PEC(\beta)$ curves of M.A.P, M.P.M, I.C.M ($u = 4$)

Figure 8: Variance texture record (a) of the image of Fig. 6-a. and PEC curves for various reconstructions and various noise level u.

3 On the choice of the regularization parameter β

We shall examine here three possible choices for β on the basis of the record y: 1. the statistical choice; 2. the cross-validation choice; 3. the jointly MAP choice on (θ, x) for both θ (the model parameters : $\theta = (\epsilon, \beta)$ or (σ^2, β)) and x .

3.1 Statistical choice:

We have first used the Gibbsian E.M algorithm([3]): in this algorithm, the pseudo-likelihhod([1],[7]) of the Gibbsian field is substituted to the exact likelihood and in the E-step, expectations are calculated by the Monte-Carlo method with simulations given through the Gibbs sampler. These estimation results are the following:

- M.R.F 64×64 image (see Fig. 1-a, $\beta_0 = 1.149$) and for B.S.C noise:

$$\begin{array}{lll} \text{for} & \epsilon = 0.02 & (\hat{\beta}_{EM}, \hat{\epsilon}) = (1.28, 0.03) \\ \text{for} & \epsilon = 0.20 & (\hat{\beta}_{EM}, \hat{\epsilon}) = (1.15, 0.20) \\ \text{for} & \epsilon = 0.30 & (\hat{\beta}_{EM}, \hat{\epsilon}) = (1.00, 0.28) \end{array}$$

- Handdrawn Image 32×32 : (see Fig. 6-a)

  B.S.C noise

$$\begin{array}{lll} \text{for} & \epsilon = 0.02 & (\hat{\beta}_{EM}, \hat{\epsilon}) = (1.43, 0.02) \\ \text{for} & \epsilon = 0.20 & (\hat{\beta}_{EM}, \hat{\epsilon}) = (1.15, 0.20) \end{array}$$

  Additive gaussian noise

$$\begin{array}{lll} \text{for} & \sigma = 0.594 & (\hat{\beta}_{EM}, \hat{\sigma}) = (0.99, 0.565) \\ \text{for} & \sigma = 0.953 & (\hat{\beta}_{EM}, \hat{\sigma}) = (0.56, 0.797) \end{array}$$

For higher level of the noise ($\epsilon = 0.4$ or $\sigma = 1.974$), the Gibbsian E.M doesn't give satisfactory results.

Secondly, moment estimations proposed by A. Frigessi & M. Piccioni([5]) in the B.S.C noise case are used. We have approximated the involved elliptic integrals by polynomals of degree 5 (error uniformly bounded by 10^{-8})[3]. The estimation results are as following:

- M.R.F 64×64 image (see Fig. 1-a, $\beta_0 = 1.149$) and for B.S.C noise:

$$\begin{array}{lll} \text{for} & \epsilon = 0 & (\hat{\beta}_{MT}, \hat{\epsilon}) = (0.98, 0.00) \\ \text{for} & \epsilon = 0.20 & (\hat{\beta}_{MT}, \hat{\epsilon}) = (0.88, 0.16) \\ \text{for} & \epsilon = 0.30 & (\hat{\beta}_{MT}, \hat{\epsilon}) = (1.02, 0.31) \\ \text{for} & \epsilon = 0.40 & (\hat{\beta}_{MT}, \hat{\epsilon}) = (0.94, 0.38) \end{array}$$

- Handdrawn Image 32×32 : (see Fig. 6-a)

$$\begin{array}{lll} \text{for} & \epsilon = 0 & (\hat{\beta}_{MT}, \hat{\epsilon}) = (0.84, -0.04) \\ \text{for} & \epsilon = 0.20 & (\hat{\beta}_{MT}, \hat{\epsilon}) = (0.77, 0.11) \\ \text{for} & \epsilon = 0.30 & (\hat{\beta}_{MT}, \hat{\epsilon}) = (0.99, 0.32) \\ \text{for} & \epsilon = 0.40 & (\hat{\beta}_{MT}, \hat{\epsilon}) = (0.38, 0.25) \end{array}$$

Despite the fact that energy(1) is a good measure of geometric regularity, the remoteness of the object x from a realization of a M.R.F can lead , in real images (here the hand-drawn one) to important fluctuations in the various estimations of β. Further, such a statistical choice of β is independent of the restoration method: this is not satisfactory when we take into account the strong dependency of the subjective optimal choice β_M on the method M. Note that other methods of estimation using incomplete data are available (see L. Younes, [21],[22]).

Restoration experiments with these estimated values $\hat{\beta}_M$ and $\hat{\beta}_{MT}$ are given in the Figure 10 and 11.

3.2 Cross-validation choice:

In some sense, this non parametric choice takes into account all the factors defining the context of the problem: the method of reconstruction, the quality criterion, and the process of degradation.

[3]See Abramowitz & I. Stegun: Handbook of Mathematical Functions.

Given one pixel $s \in S$ let $\hat{x}^{(s)}$ be the restoration obtained when all the record but the value y_s is observed. The Cross-Validation (C.V) distance is defined by:

$$d_{CV}(\beta) \stackrel{def}{=} \frac{1}{a}\sum_s (y_s - \hat{x}_s^{(s)})^2$$

where a is a convenient normalizing factor depending only on S (see [15], [4], [9], [14], [16], [17]). The C.V choice for β is:

$$\beta_{CV} \stackrel{def}{=} \arg\ Min_\beta\ d_{CV}(\beta)$$

Theoretical results are obtained in some linear and hilbertian context for the object x (see for example [4]): if we note

$$d(\beta) \stackrel{def}{=} \frac{1}{a}\sum_s (x_s - \hat{x}_s)^2$$

the distance to the true object x, then under good conditions and when the density of observations increases to infinity, then $d_{CV}(\beta)$ and $d(\beta)$ reach their minimum at "the same value" $\beta = \beta_{CV}$. Note that for a binary image x, $d(\beta)$ and $PEC(\beta)$ are proportional.

When the context for x doesn't stay linear (convex or not convex constraints on x), no theoretical results are available, but for some problems, the C.V choice still seems reasonable (see [17]). The following experiments confirm this opinion: values of $PEC(\beta)$ and $d_{CV}(\beta)$ have been computed from $\beta = 0.2$ to $\beta = 2$ with an increment of 0.1 for MAP, MPM, ICM methods, for BSC or additive Gaussian noise at levels 20% and 30%. Table 3 gives values of β minimising d_{CV} and PEC, showing the proximity of the two determinations.

Method	*B.S.C Noise*		*Gaussian noise*	
ICM	0.70—**0.70**	0.50—**0.50**	1.13—**1.11**	0.76—**0.75**
MPM	1.21—**0.89**	0.92—**0.69**	1.27—**1.04**	0.95—**0.61**
MAP	1.66—**1.74**	1.00—**1.10**	1.39—**1.28**	0.82—**0.95**

Table 3. PEC and **C.V** choices for β

These minima values of β are deduced from the smoothing curves $PEC(\beta)$ and $d_{CV}(\beta)$, obtained themselves by polynomial regressions of degree 3.

Figure 9 gives a example of smoothed PEC and d_{CV} curves after a convenient change of scale for d_{CV}.

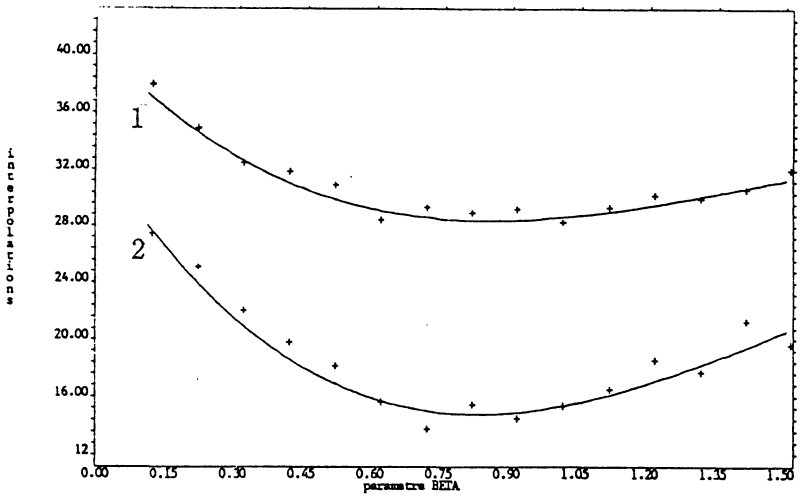
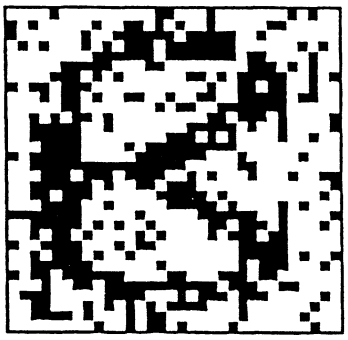


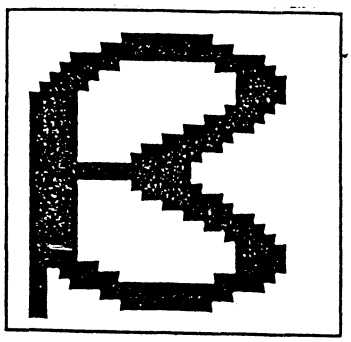
Figure 9: Exact and smoothed d_{CV} (curve 1) and PEC (curve 2).
Additive gaussian noise record ($\sigma = 0.953$) and M.P.M restoration.

The set of figure 10 and 11 allows us to say that overall there is a good quality of reconstruction with β_{CV}. Statistic selection lead to important fluctuations in this quality.

Nevertheless, at our stage, C.V. criterion cannot be used due to over lengthy computing time unless the restoration method is very rapid (for example ICM): here, one exact MAP restoration takes 30 seconds (CPU time) on a VAX 750 (less for small values of β), MPM with 50 loops in the Monte-Carlo estimation needs 7 seconds (the same for Simulated Annealing with the same characteristics), ICM needs only 0.2 second.

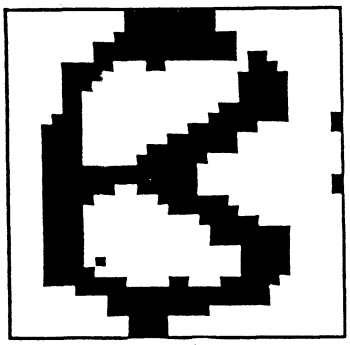


noisy image

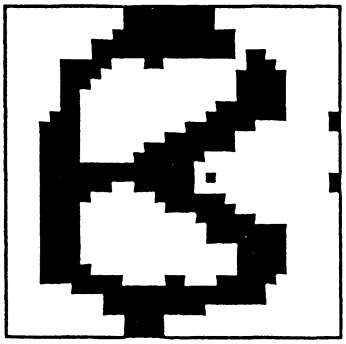


original

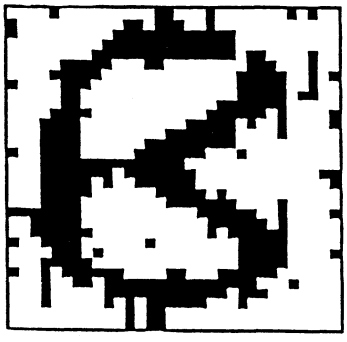
M.A.P



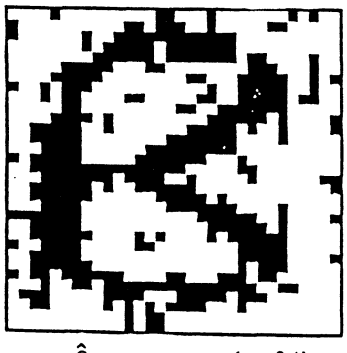
$\beta_{PEC} = 1.66(10\%)$



$\beta_{CV} = 1.74(10\%)$

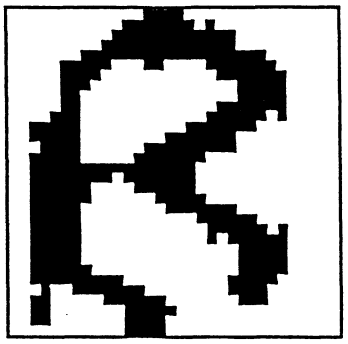


$\hat{\beta}_{EM} = 1.15(12\%)$

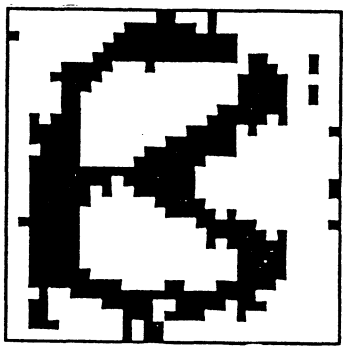


$\hat{\beta}_{MT} = 0.77(14\%)$

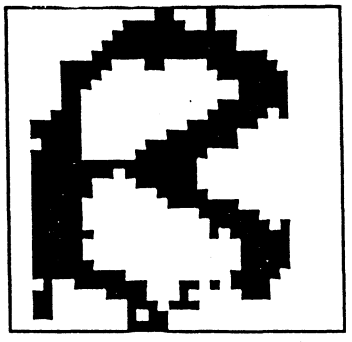
M.P.M



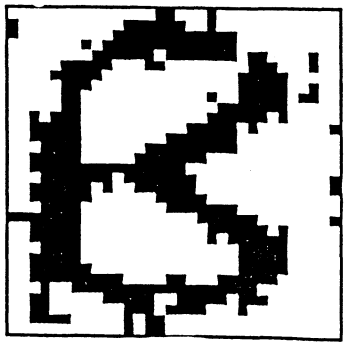
$\beta_{PEC} = 1.21(8\%)$



$\beta_{CV} = 0.89(8\%)$

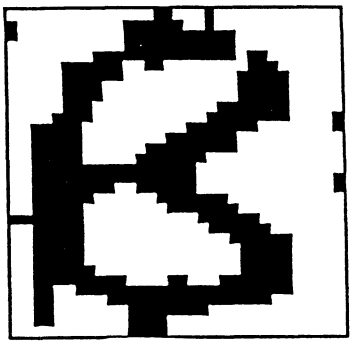


$\hat{\beta}_{EM} = 1.15(8\%)$



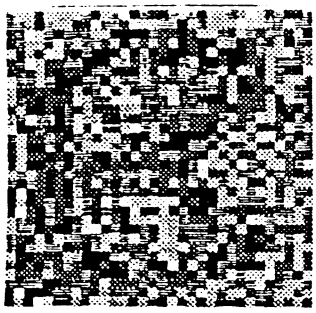
$\hat{\beta}_{MT} = 0.77(9\%)$

I.C.M

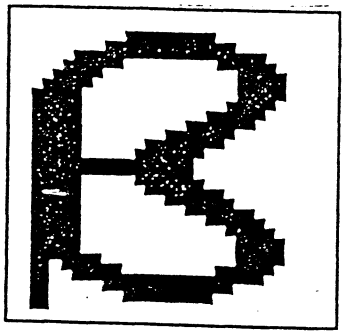


$\beta_{PEC} = 0.70(8\%)$
$\hat{\beta}_{CV} = 0.70(8\%)$
$\hat{\beta}_{EM} = 1.15(8\%)$
$\hat{\beta}_{MT} = 0.77(8\%)$

Figure 10: Various restorations of the B.S.C. record with $\epsilon = 20\%$. for various choices of β: subjective optimal choice, C.V. choice, E.M. choice and reconstructions.

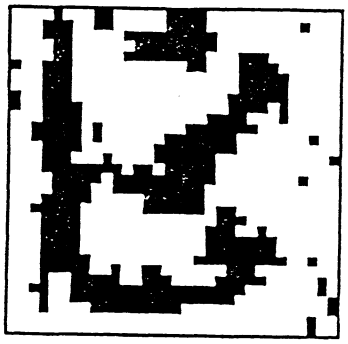


noisy image

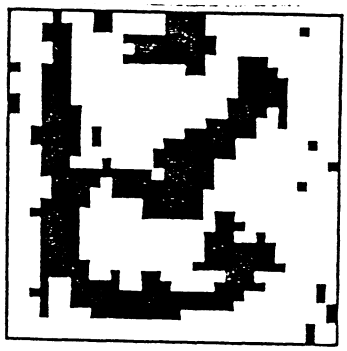


original

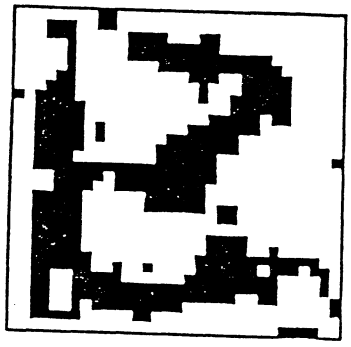
M.A.P



$\beta_{PEC} = 0.82(16\%)$

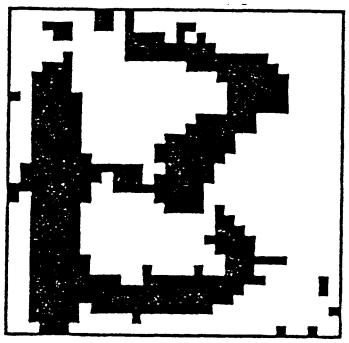


$\beta_{CV} = 0.85(16\%)$

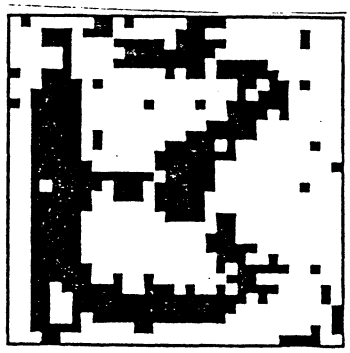


$\hat{\beta}_{EM} = 0.56(18\%)$

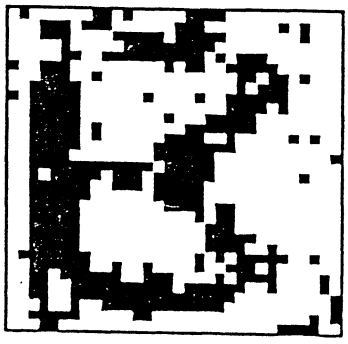
M.P.M



$\beta_{PEC} = 0.95(14\%)$

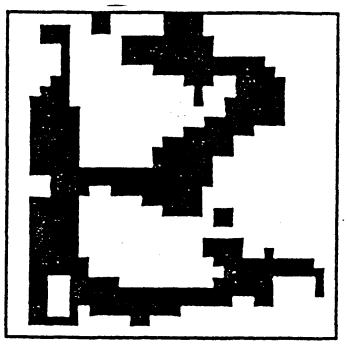


$\beta_{CV} = 0.61(14\%)$

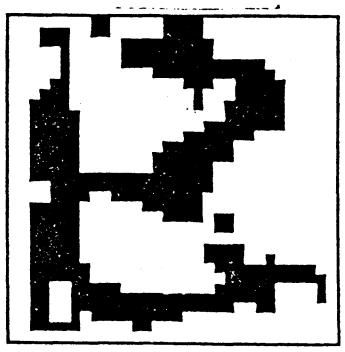


$\hat{\beta}_{EM} = 0.56(15\%)$

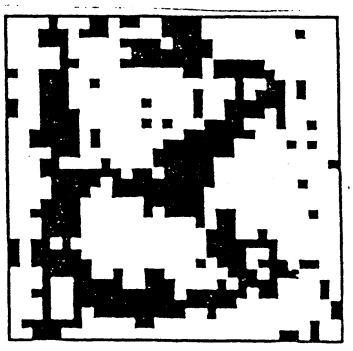
I.C.M



$\beta_{PEC} = 0.76(13\%)$



$\beta_{CV} = 0.75(13\%)$



$\hat{\beta}_{EM} = 0.56(13\%)$

Figure 11: Similar experiments as Fig. 10 for additive gaussian noise with $\sigma = 0.953$.

3.3 Joint MAP on both parameter (to be estimated) and object (to be restored):

It is known that, frequently, maximizing the conditional likelihood $\Pr(x|y,\beta)$ in both parameters β and object x leads to some degenerated solutions, see for example ([10],[19]). We shall briefly illustrate this behaviour on three examples of filtering (see appendix B for the sketch of the proofs).

Example 1: Let $x = \{x_1, x_2, \ldots, x_n\}$ a binary signal taking values in $\{-1, 1\}$ and $y = \{y_1, y_2, \ldots, y_n\}$ the record: $y_i = x_i + \epsilon_i$ where $\{\epsilon_i\}$ are i.i.d $\mathcal{N}(0, \sigma^2)$ variables. Suppose x has the (one dimensional) prior given by (1). Then the likelihood of the complete data is:

$$\Pr(x, y; \sigma^2) = K(\sigma^2, \beta, y) \exp\left[\beta n(x) + \frac{1}{\sigma^2} < y, x >\right]$$

where $n(x) = \sum_{<i,j>} \delta(x_i, x_j)$ and $< x, y >$ is the scalar product in $\mathcal{R}^n$.

Let $T(y), T(|y|)$ be the sums of the components of y, $|y| = \{|y_1|, |y_2|, \ldots, |y_n|\}$ respectively, $\eta = \inf_i |y_i|$,and denote $s(z)$ the sign of a real z. Let P^y be the conditional likelihood. Then we have the following result:

- If $\sigma^2 \to 0$, $\beta \to \infty$ and $\beta\sigma^2 \geq C > 2T(|y|)$, then for $\hat{x} = \{s, s, \ldots, s\}$ with $s = s(T(y))$, $P^y \to 1$.
- If $\sigma^2 \to 0$ and $\beta \to \infty$ with $\beta\sigma^2 \leq C < 2\eta/n$, then with $\hat{x} = \{s(y_i), i = 1, n\}$, $P^y \to 1$.

In both case, a degeneration appears: the estimated parameter value goes to the boundary of the domain of definition and restoration will be driven uniquely by the regularization component ($\beta\sigma^2$ is large) or by the fidelity to the data y ($\beta\sigma^2$ is small).

The same behaviour appears exactly if x is recorded via a B.S.C with parameter α_1 (see (5)): complete regularization if $\beta\alpha_1^{-1} \geq C > n/2$, complete fidelity to the record if $\beta\alpha_1^{-1} \leq C < 2/n$ (α_1^{-1} is increasing with level of noise).

Example2 Examine now the following filtering of an $AR(1)$ process.

The model is:

$$\begin{cases} y_i = x_i + \epsilon_i & \\ & i = 1, 2, \ldots, n \\ x_i = \rho x_{i-1} + \eta_i, & \rho \geq 0 \end{cases}$$

with $\{\epsilon_i\}$, $\{\eta_i\}$ i.i.d variables, and $\epsilon_i \sim \mathcal{N}(0, \tau^2)$, $\eta_i \sim \mathcal{N}(0, \sigma^2)$. The parameter is $\theta = (\tau^2, \sigma^2, \rho)$. We have:

$$P^y = \mathcal{L}(x|y, \theta) = \mathcal{N}_n\left((I + \tau^2\Sigma_x^{-1})^{-1}y, \Sigma_x(I - (I + \tau^2\Sigma_x^{-1})^{-1})\right)$$

from which it can be deduced:

- (ρ, σ^2) beeing fixed, $\tau^2 \to 0_+$ and $\hat{x} = y$ give the maximum of P^y.
- τ^2 and $\sigma_x^2 \stackrel{\text{def}}{=} \frac{\sigma^2}{1-\rho^2}$ beeing fixed, $\sigma \to 0_+$ and $\rho \to 1_-$, $\hat{x} = (\overline{y}, \ldots, \overline{y})$ gives the maximum of P^y.

Example 3: This third example is taken from an exercise in the book by Ch. Gourieroux.[4] Let $x = (x_1, x_2, \ldots, x_n)$, $x_i \in \overline{\mathcal{R}}$ with no prior knowledge and independent observations $y_i = \begin{pmatrix} y_{i_1} \\ y_{i_2} \end{pmatrix}$, $i = 1, n$, where $\{y_{i_1}\}$ and $\{y_{i_2}\}$ are independent Bernoulli's variables:

$$\begin{cases} \Pr(y_{i_1} = 1) = F(x_i) & \\ \Pr(y_{i_2} = 1) = F(x_i + \beta) & i = 1, n \\ \text{with } F(x) = (1 + e^{-x})^{-1} & \text{(Logistic distribution)} \end{cases}$$

Then the M.L (estimation $\times$ restoration) for β and x leads to: if $y_{i_1} + y_{i_2} = 0, \hat{x}_i = -\infty$ and if $y_{i_1} + y_{i_2} = 2, \hat{x}_i = +\infty$.

Let $\overline{n}$ be the cardinal of the set $\{i/1 \leq i \leq n, y_{i_1} + y_{i_2} = 1\}$ and $\overline{\overline{y_2}}$ the arithmetic mean of the $\{y_{i_2}\}$ for the preceeding index set, then:

- If $y_{i_1} + y_{i_2} = 1, \hat{x}_i = -\hat{\beta}/2$ where $\hat{\beta} = 2\ln\left(\frac{1-\overline{\overline{y_2}}}{\overline{\overline{y_2}}}\right)$

  is the joint M.L estimator of β . It is easy to check that $\hat{\beta} \xrightarrow{\text{a.s.}} 2\beta$. Then, if we may consider as satisfactory the restoration of x , we see that it is associated to a biased estimation of the unknown parameter.

[4] Econometrie des variables qualitatives (Ed. Economica, Paris 1984), Ex. 19, p. 109.

4 Discussion and further work

We have seen that it is difficult to give a general answer to the crucial question: *How to choose the parameter (say β in a regularization reconstruction method?* First, the general context of the problem need to be conveniently defined: in which family is the object to be reconstructed?, what is the level of the noise degradation?, but also the choice of regularization parameter is dependent on the quality criterion selected and on the reconstruction method. We have focussed this experimental study on geometrically regular binary image x and record y resulting from a noise degradation process of relatively high level. The robustness in β is decreasing from I.C.M to M.A.P via M.P.M, and it is observed that for high level of the noise, optimal values of β can differ strongly on true (M.R.F images) or estimated (modelisation of a non M.R.F image by a M.R.F) β, particulary for M.A.P.

If statistical criterion (for ex. E.M estimation) of choice of β seems reasonable at low noise level, this is not the case in medium or high noise level. In this case, cross-validation criterion gives satisfactory results, but, at our stage of study, its numerical implementation is too expensive to be used in practice. Next we note that joint M.L on both parameter and object leads to degenerated solutions.

Future works arise naturally: how such results are relevant (or not) to other families of objects?, Are they some theoretical results that validate the C.V choice, in the context of constrained regularisation? What are numerically computable variants of C.V criterion?

Aknowledgement: The authors are grateful to B. Chalmond and L. Younes for helpful discussions in the use of the Gibbsian E.M. and stochastic M.L. algorithms, and to Sylvia and Marie Yvonne for the revison of the english version of the paper.

References

[1] BESAG J., 1974, *Spatial interaction and the statistical analysis of lattice systems*, J.R.S.S.B, 36, 192-236.

[2] BESAG J., 1986, *On the statistical analysis of dirty pictures,* J.R.S.S.B, 48, 259-302.

[3] CHALMOND B., 1988, *An iterative Gibbsian technique for reconstruction of M-ary images*, to appear in Pattern Recognition.

[4] CRAVEN P. and WAHABA G., 1979, *Smoothing noisy data by spline functions*, Num. Math, 31, 377-403.

[5] FRIGESSI A. and PICCIONI M., 1988, *Parameter estimation for two-dimensional Ising fields corrupted by a noise*, Quaderno n^0 10, Inst. Appl. del Calculo, Roma.

[6] GEMAN S. et GEMAN D., 1984, *Stochastic relaxation, Gibbs distributions and the bayesian restoration of Images*, IEEE-PAMI, 6, 721-741.

[7] GUYON X., 1987, *Estimation d'un champ par pseudo-vraisemblance conditionnelle: étude asymptotique et application au cas markovien*, in "Spatial processes and spatial time series analysis", Proc. 6 th Franco-Belgian Meeting of Statistics, Droesbecke F. editor, Bruxelles.

[8] HALL P. and TITTERINGTON D.M, 1986, *On some smoothing techniques in image restoration*, J.R.S.S.B, 48, 330-343.

[9] HILGERS J.W, 1976, *On the equivalence of regularization and certain reproducing kernel Hilbert space approximation for solving first kind problems*, S.I.A.M J. Numer. Anal., 13, 172-174.

[10] LITTLE R.J.A. and RUBIN D.B., 1983, *On jointly estimating parameters and missing values by maximising the complete data likelihood*, Amer. Statist., 37, 218-220.

[11] MARROQUIN J., MITTER S. and POGGIO T., 1987, *Probabillistic solution of ill-posed problems in computational vision*, JASA 79, n^0 387, 584-589.

[12] PORTEOUS B.T., GREIG D.M. and SEHEULT A.H, 1987, *Exact M.A.P estimation for binary images*, Preprint, University of Durham.

[13] POSSOLO A., 1986, *Estimation of binary Markov Random Fields*, Tech. report n^o 77, Dept. of Stat., University of Washington.

[14] SILVERMAN B.W., 1984, *A fast efficient cross-validation method for smoothing parameter choice in spline regression*, JASA 79, n^o 387, 584-589.

[15] STONE M., 1974, *Cross-validatory choice and assessment of statistical prediction*, J.R.S.S.B, 36, 111-147.

[16] WAHBA G., 1977, *Practical approximate solutions to linear operators equations when the data are noisy*, S.I.A.M, J.Numer.Anal., 14, 651-667.

[17] WAHBA G., 1982, *Constrained regularization for ill-posed linear operator equations with application in meteorology and medecine*, in "Statistical Design Theory and Related Topics: III, vol 2", Eds. S. GUPTA and J.D BERGER, N.Y., Academic Press.

[18] WENDELBERGER J., 1981, *The computation of Laplacian smoothing splines with examples*, Techn. report 648, Stat. Dept, Univ. of Wisconsin.

[19] WOODWARD W.A, PARR W.C, SCHUCANY W.R and LINDSEY H., 1984, *A comparison of minimum distance and maximum likelihood estimation of a mixture proportion*, J. Amer. Stat. Assoc., 79, 590-598.

[20] YAO J.F., 1988, *Methodes bayesiennes en segmentation d'images (comparaison des methodes globales et contextuelles)*, in "Bayesian Statistics", Proc. 8th. Franco-Belgian Meeting of Statistics, Univ. of Louvain, Belgique.

[21] YOUNES L., 1988, *Estimation and Annealing for Gibbsian Fields*, Ann. Inst. Henri Poincare, Vol. 2, 269-294.

[22] YOUNES L., 1988, *Problèmes d'estimation paramétriques pour des champs de Gibbs Markoviens. Applications en traitement d'images*, These de Docteur en Sciences, Universite Paris XI, Sept. 1988.

APPENDIX

A Equivalence for large β between I.C.M and the iterative modal filtering

Up to a constant, we have (see (5), $i = 1, 2$):

$$H_i(x|y) = -\alpha_i \sum_s (2y_s - 1)x_s - \beta \sum_{<s,t>} \delta(x_s, x_t)$$

which leads to the local conditional probalities:

$$\ln P_s(x_s|x_t, t \neq s, y) = a_s + \alpha x_s(2y_s - 1) + \beta \sum_{t:<s,t>} \delta(x_s, x_t)$$

In the I.C.M context, let $x^{(k)}$ be the configuration after the k^{th} iteration and $s = s(k+1)$ be the pixel visited at time $k+1$, then:

$$x_s^{(k+1)} = \arg\ Max_{x_s}\ P_s(x_s|x_t^{(k)}, t \neq s, y)$$

So, if $v = v_s^{(k)} = \sum_{t:<s,t>} x_t^{(k)}$,

$$x_s^{(k+1)} = \begin{cases} 1 & \text{if } v > 2 - \alpha(2y_s - 1)/2\beta \\ 0 & \text{if not} \end{cases}$$

First case: B.S.C noise. Here the above updating rule becomes

$$\text{if } y_s = 1 \quad x_s^{(k+1)} = \begin{cases} 1 & \text{if } v > 2 - \alpha/2\beta \\ 0 & \text{if not} \end{cases}$$
$$\text{if } y_s = 0 \quad x_s^{(k+1)} = \begin{cases} 0 & \text{if } v < 2 + \alpha/2\beta \\ 1 & \text{if not} \end{cases}$$

Then suppose that $\alpha/2\beta < 1$,

$$\text{if } y_s = 1 \quad x_s^{(k+1)} = \begin{cases} 1 & \text{if } v \geq 2 \\ 0 & \text{if not} \end{cases}$$
$$\text{if } y_s = 0 \quad x_s^{(k+1)} = \begin{cases} 0 & \text{if } v \leq 2 \\ 1 & \text{if not} \end{cases}$$

$x_s^{(k+1)}$ is exactly obtained by modal choice. For ϵ =0.2, 0.3 and 0.4 the threshold β_{ICM} are respectively 0.693, 0.424 and 0.202.

Second case: Additive gaussian noise. Now the record values $\{y_s\}$ don't belong to a finite set as above. However, there is the same plateau phenomenon when $\alpha/2\beta$ is small.

For M.P.M restorations, similar behavior occurs: the reason is that the Gibbs sampler uses the same local conditional laws $\{ P_s(x_s|x_t, t \neq s, y) \}$.

Then if β is large, the Gibbs sampler leads to the modal choice with strong probility p. In the B.S.C noise case, simple calculations show that for $a \in (0,1)$,

$$\text{if } b \stackrel{\text{def}}{=} max(1/2 \ln \frac{a(1-\epsilon)}{\epsilon(1-a)}, 1/2 \ln \frac{a\epsilon}{(1-\epsilon)(1-a)}), \text{ then } \beta \geq b \Longrightarrow p \geq a$$

As example for $\epsilon = 0.3$, $\beta \geq 1.90$ ensures that $p \geq 0.95$.

B Degeneracy of joint M.L estimate on θ the parameter and x the object

B.1 Example 1

Choose β,σ^2 such that $\beta\sigma^2 > 2T(|y|)$: if we suppose that $T(y) > 0$, then $\hat{x} = (1, 1 \ldots, 1) = \hat{x}(\beta, \sigma^2)$ gives the maximum for $\Pr(x|y; \beta, \sigma^2)$. Looking at function ψ of β, σ^2: $(\beta, \sigma^2) \mapsto \Pr(\hat{x}|y; \beta, \sigma^2)$, we have the estimation

$$\psi \geq [1 + \exp(-2T(y)/\sigma^2) + 2^n \exp(-\beta + 2T(|y|)/\sigma^2)]^{-1}$$

Then if $\beta\sigma^2 = C > 2T(|y|)$ and $\sigma^2 \to 0_+$, this probability goes to 1.

Suppose now that $\beta\sigma^2 = C < 2\eta/n$. Then it's easy to see that $\hat{x} = \{s(y_i), i = 1, n\}$ is the (β, σ^2)-MAP estimate. Then for the function ψ

$$\psi \geq [1 + 2^n \exp(n\beta + 2\eta/\sigma^2)]^{-1}$$

So if $\beta\sigma^2 = c < 2\eta/n, \psi \to 1$ as $\sigma^2 \to 0_+$.

When we are in the presence of a B.S.C noise, we have

$$\Pr(x, y; \alpha, \beta) = K(\alpha, \beta) \exp\{\beta n(x) + \alpha n(x, y)\}$$

with

$$n(x) \stackrel{\text{def}}{=} \sum_{<i,j>} \delta(x_i, x_j) \text{ and } n(x,y) \stackrel{\text{def}}{=} \sum_i \delta(x_i, y_i)$$

If $2\beta\alpha^{-1} > n, \hat{x} = \{a, a \ldots, a\}$ where $a = s(T(y))$ (the mode of the sequence $\{y_i, i = 1, n\}$) is the MAP in x and the estimation

$$\Pr(\hat{x}|y;\beta,\alpha) \geq [1 + \exp(\alpha[n(-\hat{x},y) - n(\hat{x},y)] + 2^n \exp(-\beta + n\delta/2)]^{-1}$$

shows that this probability goes to 1 if $\beta\alpha^{-1} = C > n/2$ and $\alpha \to \infty$.
If $n\beta\alpha < 2, \hat{x} = y$ is the MAP and

$$\Pr(\hat{x}|y;\beta,\alpha) \geq [1 + 2^n \exp(n\beta/2 - \delta)]^{-1}$$

So $\beta\alpha^{-1} = c < 2/n$ and $\alpha \to 0$ gives limit value of 1 to this probability.

B.2 Example 2

Straightforward calculation from the gaussian form of the conditional law $\Pr(x|y)$.

B.3 Example 3

$$\ln \ell(y_{i_1}, y_{i_2}; x_i, \beta) = x_i(y_{i_1} + y_{i_2} - 2) - \beta(1 - y_{i_2}) - \ln[(1 + e^{-x_i})(1 + e^{-(x_i+\beta)})]$$

and so we obtain the annonced result if $y_{i_1} + y_{i_2} = 0$ or 2. If $y_{i_1} + y_{i_2} = 1$, maximization in x_i gives: $\exp(-\hat{x}_i) = \exp(\beta/2)$. Then, the informative part for β is:

$$\begin{aligned} \ell(\beta) &= \Pi_{i:y_{i_1}+y_{i_2}=1} \ell(y_{i_1}, y_{i_2}; \hat{x}_i, \beta) \\ &= \exp[\beta\overline{n}(\overline{\overline{y_2}} - 1/2)][2 + \exp(\beta/2) + \exp(-\beta/2)]^{-\overline{n}} \end{aligned}$$

and maximization in β gives: $\hat{\beta} = 2\ln(\frac{\overline{\overline{y_2}}}{1-\overline{\overline{y_2}}})$
Examine now the conditional law $\Pr(y_{i_2}|y_{i_1} + y_{i_2} = 1)$

$$\Pr(y_{i_2}|y_{i_1} + y_{i_2} = 1) = \frac{1}{1 + e^{-\beta}}$$

Then, by the law of large number, $\overline{\overline{y_2}} \stackrel{a.s}{\to} \frac{1}{1+e^{-\beta}}$ and so : $\hat{\beta} \stackrel{a.s}{\to} 2\beta$.

Partie II

Sur l'estimation par rabotage des modèles spatiaux

Introduction

La deuxième partie de ce travail traite le problème d'estimation spectrale par rabotage des processus stationnaires spatiaux, i.e. indexés par $\mathbb{Z}^d$.

Considérons un processus stationnaire réel et r-vectoriel sur $\mathbb{Z}^d$, $X(t) = [X_a(t)]$, $t \in \mathbb{Z}^d, a = 1, \ldots, r$, observé sur un rectangle de $\mathbb{Z}^d$, $P_n = [1, n_1] \times [1, n_2] \ldots \times [1, n_d]$ de taille $|n| := |P_n| = n_1 \times n_2 \ldots \times n_d$ ($n = (n_1, \ldots, n_d)$ est un multi-indice). Un des problèmes des statistiques spatiales est celui des effets de bords. Si par exemple nous utilisons pour estimateurs empiriques des covariances $R(u) := \mathbf{E}X(0)^t X(u)$ (le processus sera supposé centré pour simplification),

$$\gamma_n(u) = \frac{1}{|n|} \sum_{s, s+u \in P_n} X(s)^t X(s+u),$$

lorsque $P_n \uparrow \mathbb{Z}^d$ en gardant ses côtés proportionnels, le biais est de l'ordre de $O(|n|^{-1/d})$ qui, pour $d \geq 2$, n'est pas plus petit que la déviation standard en $O(|n|^{-1/2})$ (Guyon [II–11]). Des phénomènes similaires apparaissent pour des estimations paramétriques, en particulier pour celles associées au contraste de Whittle qui fait intervenir le périodogramme. Guyon [II–11] propose une solution alternative à ce problème en introduisant les estimateurs *modifiés* des covariances

$$\gamma_n^*(u) = \prod_{k=1}^{d} (n_k - |u_k|)^{-1} \sum_{s, s+u \in P_n} X(s)^t X(s+u),$$

qui sont toujours sans biais.

Comme l'ont remarqué Dahlhaus-Künsch [II–8], ces estimateurs *modifiés* $\{\gamma_n^*(u)\}$ possèdent certains inconvénients. Notamment, $\{\gamma_n^*(u)\}$ ainsi que les estimations de la densité spectrale qui en découlent ne sont plus nécessairement du type positif. Ceci entraîne entre autres que les estimateurs de Whittle n'existent pas toujours.

Afin de contourner à la fois le problème des effets de bords et la non positivité des estimateurs *modifiés*, Dahlhaus-Künsch [II–8] proposent d'utiliser des *rabots* ("tapers" en Anglais) pour l'estimation spectrale ou paramétrique. Ils obtiennent pour ceux-ci de bonnes propriétés asymptotiques dans le cas des processus scalaires.

L'estimation *par rabotage* possède une autre propriété attrayante. En analyse des séries chronologiques, lorsque le spectre comporte de grands pics, des données *non rabotées* conduisent parfois à de sévères sur-estimations du spectre à d'autres fréquences que ces grands pics. Par exemple, des pics moins importants peuvent être ainsi dissimulés. Ce phénomène est d'autant plus sensible que la taille d'échantillon est petite. Dans ce genre de situation, des estimations par rabotage atténuent cet effet de sur-estimation (voir par exemple Dahlhaus [II–6]).

Ce travail donne suite à celui de Dahlaus-Künsch [II–8] qui traite des processus univariés. Plus précisément:

- Nous établissons d'abord, dans le Chapitre 5, des résultats généralisant ceux de [II–8] pour des processus multivariés en nous inspirant des travaux de Dahlhaus [II–6][II–7]. Nous montrons que le rabotage conduit à un biais de l'ordre de $O(|n|^{-2/d})$ (Théorème 1) qui nous permet, dans le cas des lattices de dimension d inférieure ou égale à 3 et avec d'autres hypothèses plus classiques, d'avoir un théorème de limite centrale (Théorème 3) pour des estimateurs spectrographiques rabotés.

- Dans le Chapitre 6, d'abord nous établissons à l'aide des résultats du Chapitre 5, la consistance et la normalité asymptotique de l'estimateur de Whittle par rabotage. Partant de cet estimateur, nous développons un test d'hypothèse emboîtée basé sur la différence du contraste de Whittle raboté. Le test obtenu est applicable à des processus stationnaires généraux et constitut ainsi une généralisation du résultat classique qui concerne, soit des processus gaussiens, soit des processus linéaires. Le chapitre termine avec une application aux champs de Markov (MRF) gaussiens.

- Dans le Chapitre 7, nous donnons une étude de simulation expérimentant les estimateurs de Whittle rabotés dans le contexte d'un MRF gaussien.

Chapitre 5

Etude des estimateurs spectrographiques rabotés

Soit $\{X(t), t \in P_n\}$ une observation du processus X sur un rectangle P_n de $\mathbb{Z}^d$. Les *données rabotées* sont des quantités

$$X^R(t) = [X_a^R(t)] := [\underline{h}_a(t;n)X_a(t)] \ , 1 \leq a \leq r, \ \ t \in P_n, \tag{5.1}$$

où les $\{\underline{h}_a(t;n)\}$ sont des *poids de rabotage.* Les poids sont "spatialement" identiques, i.e.

$$\underline{h}_a(t;n) = h_a(t_1;n_1)\dots h_a(t_d;n_d), \tag{5.2}$$

avec $\{h_a(.;.)\}$ les poids unidimentionnels pour la direction a. Il est par contre souvent intéressant de garder "spectralement" ces poids distincts, i.e

$$h_a \neq h_b, \quad \text{si } a \neq b.$$

Soit $d^n(\alpha)$ la transformée de Fourier discrète des données rabotées, i.e.

$$d^n(\alpha) := \sum_{t \in P_n} X^R(t)e^{-i<\alpha,t>} \ , \ \alpha \in \Pi^d \tag{5.3}$$

avec $\Pi := [-\pi, \pi)$. Le *périodogramme* associé ,

$$I^n(\alpha) := [I_{ab}^n(\alpha)], \ \ a,b = 1,\dots,r,$$

est défini par

$$I_{ab}^n(\alpha) := \frac{d_a^n(\alpha)d_b^n(-\alpha)}{(2\pi)^d \underline{H}_{ab}^n(0)}, \tag{5.4}$$

avec

$$\underline{H}_{ab}^n(\alpha) := \sum_{t \in P_n} \underline{h}_a(t;n)\underline{h}_b(t;n)e^{-i<\alpha,t>}. \tag{5.5}$$

Les estimateurs $C^n(u) := [C^n_{ab}(u)], u \in \mathbb{Z}^d$, de covariances sont identifiés par l'égalité

$$I^n(\alpha) = \frac{1}{(2\pi)^d} \sum_{u \in \mathbb{Z}^d} C^n(u) e^{i<\alpha,u>},$$

c-à-d

$$C^n_{ab}(u) = \frac{\sum_{t,t+u \in P_n} X^R_a(t) X^R_b(t+u)}{\underline{H}^n_{ab}(0)}. \tag{5.6}$$

Ce sont des estimateurs asymptotiquement sans biais.

Posons

$$\underline{K}^n_{ab}(\alpha) := \frac{\underline{H}^n_a(\alpha)\underline{H}^n_b(-\alpha)}{(2\pi)^d \underline{H}^n_{ab}(0)}, \tag{5.7}$$

avec

$$\underline{H}^n_a(\alpha) := \sum_{t \in P_n} \underline{h}_a(t;n) e^{i<\alpha,t>} \quad , \quad 1 \leq a \leq r. \tag{5.8}$$

Ce sont les *fenêtres spectrales* ou *fréquentielles* relatives au rabotage (cf. Lemme 4 ci-dessous).

Il est utile d'introduire l'équivalent de ces fonctions dans le cas $d = 1$ (série chronologique):

$$T \in \mathbb{N}, \quad \sigma \in \Pi, \quad H^T_{a,b}(\sigma) := \sum_{t=1}^{T} h_a(t;T) h_b(t;T) e^{-i\sigma t}, \tag{5.9}$$

$$H^T_a(\sigma) := \sum_{t=1}^{T} h_a(t;T) e^{-i\sigma t}, \tag{5.10}$$

$$K^T_{ab}(\sigma) := \frac{H^T_a(\sigma) H^T_b(-\sigma)}{2\pi H^T_{ab}(0)}. \tag{5.11}$$

On vérifie facilement les identités multiplicatives suivantes :

$$\alpha \in \Pi^d, \quad \underline{H}^n_{a,b}(\alpha) = \prod_{k=1}^{d} H^{n_k}_{a,b}(\alpha_k), \tag{5.12}$$

$$\underline{H}^n_a(\alpha) = \prod_{k=1}^{d} H^{n_k}_a(\alpha_k), \tag{5.13}$$

$$\underline{K}^n_{ab}(\alpha) = \prod_{k=1}^{d} K^{n_k}_{ab}(\alpha_k). \tag{5.14}$$

Donnons d'abord un lemme facile qui sera utile ultérieurement. Dorénavant, $f := [f_{ab}]_{1\le a,b\le r}$ notera *la densité spectrale (matricielle)* du processus:

$$f(\alpha) := \sum_{u\in\mathbb{Z}^d} R(u)e^{i<\alpha,u>}, \quad \alpha\in\Pi^d. \tag{5.15}$$

Lemme 3 *Pour* $1\le a,b\le r$ *et* $\alpha\in\Pi^d$

$$\mathbf{E}I^n_{ab}(\alpha) = (\underline{K}^n_{ab} * f_{ab})(\alpha) = \int_{\Pi^d} \underline{K}^n_{ab}(\beta) f_{ab}(\alpha-\beta)d\beta \tag{5.16}$$

Preuve. Par vérification directe:

$$\begin{aligned}
&\mathbf{E}I^n_{ab}(\alpha)\\
&= \{(2\pi)^d \underline{H}^n_{ab}(0)\}^{-1}\\
&\quad \mathbf{E}\left[\sum_{s\in P_n} \underline{h}_a(s;n)X_a(s)e^{-i<s,\alpha>}\right]\left[\sum_{l\in P_n} \underline{h}_b(l;n)X_b(l)e^{i<l,\alpha>}\right]\\
&= \{(2\pi)^d \underline{H}^n_{ab}(0)\}^{-1} \sum_u \left[\sum_{s,s+u\in P_n} \underline{h}_a(s;n)\underline{h}_b(s+u;n)\right] R_{ab}(u)e^{i<\alpha,u>}\\
&= \sum_u R_{ab}(u)\widehat{\underline{K}^n_{ab}}(u)e^{i<\alpha,u>}.
\end{aligned}$$

♠

Nous allons fixer une classe de rabotages de la manière suivante. Pour tout $1\le a\le r$, soit $w_a : [0,1]\to\mathbb{R}$, non décroissant et vérifiant $w_a(0)=0$ et $w_a(1)=1$. Nous définissons d'abord les *fonctions de rabotage*, ou simplement les *rabots* $\{h_a\}_{1\le a\le r}$ comme suit (cf. Dahlhaus-Künsch [II–8]):

Définition 6 *Pour tout* $1\le a\le r$,

$$h_a(u) := \begin{cases} w_a(\frac{2u}{\rho_a}) & 0\le u<\frac{\rho_a}{2}\\ 1 & \frac{\rho_a}{2}\le u\le\frac{1}{2}\\ h_a(1-u) & \frac{1}{2}<u\le 1\end{cases}. \tag{5.17}$$

Les poids de rabotage sont définis quant à eux comme valeurs discrétisées des rabots:

$$h_a(t;T) \stackrel{\text{def}}{=} h_a(\frac{t-\frac{1}{2}}{T}), \quad 1\le a\le r. \tag{5.18}$$

Par convention, on posera $h_a(u) \equiv 0$ si $u \notin [0,1]$. ρ_a est le *degré* de rabotage a-directionnel vérifiant $0 < \rho_a \leq 1$. En effet, $100(1-\rho_a)\%$ des données $\{X_a(t),\ t \in P_n\}$ de la direction spectrale a restent inchangées après le rabotage.

Il importe de signaler que la classe de rabots que nous venons d'introduire contient la majorité des robots usuels (cf. Brillinger [II–4], Table 3.3.1).

Remarquons que l'on a pour $u \notin [\frac{\rho_a}{2}, 1 - \frac{\rho_a}{2}]$,

$$h_a'(u) = \frac{2}{\rho_a} w_a'(\frac{2u}{\rho_a}), \tag{5.19}$$

et

$$\mathcal{L}(h_a') \leq (\frac{2}{\rho_a})^2 \mathcal{L}(w_a'), \tag{5.20}$$

où (et dans toute la suite) $\mathcal{L}(\psi)$ désigne le coefficient de Lipschitz d'une fonction ψ lorsque celui-ci existe.

Les *spectrographes* sont des quantités

$$J_{ab}^n(\phi) := \int_{\Pi^d} \phi(\alpha) f_{ab}(\alpha) \mathrm{d}\alpha, \tag{5.21}$$

où ϕ: $\Pi^d \to \mathbf{C}$ est une fonction appropriée. Naturellement, on l'estime par

$$J_{ab}^n(\phi) := \int_{\Pi^d} \phi(\alpha) I_{ab}^n(\alpha) \mathrm{d}\alpha = \sum_{u \in \mathbb{Z}^d} \hat{\phi}(u) C_{ab}^n(u). \tag{5.22}$$

Intuitivement, la perte d'information ainsi que la variance de ces estimateurs spectrographiques seront d'autant plus petites que les degrés de rabotage $\{\rho_a\}$ sont proches de 0. L'efficacité asymptotique sera en effet obtenue si $\rho_a = \rho_a(n)$ et $\rho_a \to 0$ à une bonne vitesse. Nous renvoyons aux [II–7], [II–8] pour la discussion concernant l'efficacité asymptotique. D'autre part, on ne peut faire décroître trop vite les $\{\rho_a\}$ sous peine de retrouver les inconvénients de non-rabotage (i.e. $\rho_a = 0$ pour tout $1 \leq a \leq r$). Une décroissance appropriée sera indiquée dans le Théorème 1 (paragraphe 5.1) qui déterminera le biais des estimateurs spectrographiques. Les paragraphes suivants (5.2 et 5.3) établiront respectivement les variances-covariances asymptotiques et le théorème de limite centrale pour ces estimateurs.

5.1 Biais aymptotique

Le résultat central de ce paragraphe est le Théorème 1 qui donne l'expression asymptotique du biais

$$B_{ab}(\phi) = \mathbf{E} J_{ab}^n(\phi) - J_{ab}(\phi). \tag{5.23}$$

Dans toute la suite, nous noterons C_j des constantes qui interviendront dans diverse formules indépendamment de la taille d'observation n. Notons aussi une identité algébrique utile:

$$\sum_{k=p}^{q} v_k \theta^k = \frac{\theta}{1-\theta} \sum_{k=p-1}^{q} (v_{k+1} - v_k)\theta^k, \quad \theta \neq 1, \tag{5.24}$$

où on a posé $v_{p-1} = v_{q+1} = 0$.

Le lemme et les deux propositions qui suivent donnent les propriétés des noyaux $\{\underline{K}^n_{ab}\}_{1\leq a,b\leq r}$ qui sont les clés du Théorème 1. La symétrie des rabots $\{h_a\}_{1\leq a\leq r}$ par rapport à $\frac{1}{2}$ est essentielle pour le lemme.

Lemme 4 *On a pour* $1 \leq a, b \leq r$,

1.) $K^T_{ab}(\sigma)$ *est une fonction à valeurs réelles , paire et de masse totale 1*

2.) *Il existe une constante* C_1 *ne dépendant que de* h_a, h_b, *telle que*

$$\forall \sigma \in \Pi, \quad |K^T_{ab}(\sigma)|^2 \leq C_1 K^T_{aa}(\sigma) K^T_{bb}(\sigma). \tag{5.25}$$

Les propriétés sont similaires pour $\underline{K}^n_{ab}(\alpha)$.

Preuve.

1.) L'identité de Parseval donne

$$\begin{aligned}
&\int_\Pi H^T_a(\sigma) H^T_b(-\sigma) d\sigma \\
&= \int_\Pi (\sum_{s=1}^{T} h_a(\frac{s-\frac{1}{2}}{T}) e^{-is\sigma})(\sum_{l=1}^{T} h_b(\frac{l-\frac{1}{2}}{T}) e^{il\sigma}) d\sigma \\
&= 2\pi \sum_{s=1}^{T} h_a(\frac{s-\frac{1}{2}}{T}) h_b(\frac{s-\frac{1}{2}}{T}) \\
&= 2\pi H^T_{ab}(0).
\end{aligned}$$

D'autre part, par définition,

$$K^T_{ab}(-\sigma) = \overline{K^T_{ab}(\sigma)}.$$

Il suffit donc de prouver que $K^T_{ab}(\sigma)$ est réelle.

$$H^T_a(\sigma) H^T_b(-\sigma) = \sum_{s,l=1}^{T} h_a(\frac{s-\frac{1}{2}}{T}) h_b(\frac{l-\frac{1}{2}}{T}) e^{-i(s-l)\sigma}.$$

En inversant l'ordre de sommation par changement d'indices $\mu = T - s + 1$, $\nu = T - l + 1$, on trouve

$$H_a^T(\sigma)H_b^T(-\sigma) = \sum_{\mu,\nu=1}^{T} h_a(\frac{T-\mu-\frac{1}{2}}{T})h_b(\frac{T-\nu-\frac{1}{2}}{T})e^{i(\mu-\nu)\sigma}.$$

Or h_a, h_b sont symétriques par rapport à $\frac{1}{2}$, donc

$$\begin{aligned} H_a^T(\sigma)H_b^T(-\sigma) &= \sum_{\mu,\nu=1}^{T} h_a(\frac{\mu-\frac{1}{2}}{T})h_b(\frac{\nu-\frac{1}{2}}{T})e^{i(\mu-\nu)\sigma} \\ &= \overline{H_a^T(\sigma)H_b^T(-\sigma)}. \end{aligned}$$

2.) Par définition,

$$|K_{ab}^T(\sigma)|^2 = K_{aa}^T(\sigma)K_{bb}^T(\sigma)\frac{H_{aa}^T(0)H_{bb}^T(0)}{(H_{ab}^T(0))^2},$$

et le dernier facteur tend vers

$$\frac{(\int_0^1 h_a^2(x)dx)(\int_0^1 h_b^2(x)dx)}{(\int_0^1 h_a(x)h_b(x)dx)^2}$$

quand $T \to \infty$.

♠

Proposition 2 *Pour* $1 \leq a, b \leq r$

$$\int_\Pi K_{ab}^T(\sigma)\sin^2\frac{\sigma}{2}\mathrm{d}\sigma = \frac{1}{4}T^{-2}\left(\frac{\int_0^1 h_a'(x)h_b'(x)\mathrm{d}x}{\int_0^1 h_a(x)h_b(x)\mathrm{d}x}\right)(1+o(1)). \tag{5.26}$$

ρ_a, ρ_b pourraient dépendre de T de sorte que $(\rho_a \wedge \rho_b)^{-1} = o(T^{1/3})$. Le terme $o(1)$ est uniforme en ρ_a et ρ_b.

Preuve. Notons par ν l'intégrale à évaluer. Par définition de H_a^T et en utilisant la transformation d'Abel (équation 5.24) et le fait que $h_a(u) \equiv 0$ si $u \notin [0,1]$, on obtient pour $\sigma \neq 0$,

$$H_a^T(\sigma) = (e^{i\sigma}-1)^{-1}\sum_{s=0}^{T}\left[h_a(\frac{s+\frac{1}{2}}{T}) - h_a(\frac{s-\frac{1}{2}}{T})\right]e^{-is\sigma}. \tag{5.27}$$

Notons

$$\mu_a^T(s) := h_a(\frac{s+\frac{1}{2}}{T}) - h_a(\frac{s-\frac{1}{2}}{T}), \quad s = 0, \ldots, T,$$

on trouve

$$\sin(\frac{\sigma}{2})H_a^T(\sigma) = \frac{e^{-i\frac{\sigma}{2}}}{2i}\sum_{s=0}^{T}\mu_a^T(s)e^{-is\sigma}.$$

Cette dernière relation est aussi valable pour $\sigma = 0$ du fait que $\sum_{s=0}^{T}\mu_a^T(s) = 0$. Les expressions sont similaires pour $H_b^T(\sigma)$:

$$\sin(\frac{\sigma}{2})H_b^T(\sigma) = \frac{e^{-i\frac{\sigma}{2}}}{2i}\sum_{s=0}^{T}\mu_b^T(s)e^{-is\sigma},$$

avec

$$\mu_b^T(s) := h_b(\frac{s+\frac{1}{2}}{T}) - h_b(\frac{s-\frac{1}{2}}{T}), \quad s = 0,\ldots,T.$$

Or,

$$\begin{aligned}\nu &= (2\pi H_{ab}^T(0))^{-1}\int_\Pi \sin^2(\frac{\sigma}{2})H_a^T(\sigma)H_b^T(-\sigma)\mathrm{d}\sigma \\ &= (2\pi H_{ab}^T(0))^{-1}\int_\Pi (2i\sin(\frac{\sigma}{2})e^{i\frac{\sigma}{2}}H_a^T(\sigma))\overline{(2i\sin(\frac{\sigma}{2})e^{i\frac{\sigma}{2}}H_b^T(\sigma))}\mathrm{d}\sigma.\end{aligned}$$

D'où d'après l'identité de Parseval,

$$\nu = (4H_{ab}^T(0))^{-1}\sum_{s=0}^{T}\mu_a^T(s)\mu_b^T(s). \tag{5.28}$$

Soient les intervalles $A_s = [\frac{s-\frac{1}{2}}{T}, \frac{s+\frac{1}{2}}{T}), s = 0,\ldots,T$ et $\{s_j\}_{j=1,2,3,4}$ les 4 valeurs de s telles que les A_{s_j}, $j = 1,2,3,4$ contiennent respectivement les points éventuels de discontinuité de h_a', h_b'. En d'autre termes, soit

$$\frac{\rho_a}{2} \in A_{s_1}, \ 1-\frac{\rho_a}{2} \in A_{s_2}, \ \frac{\rho_b}{2} \in A_{s_3}, \ 1-\frac{\rho_b}{2} \in A_{s_4}.$$

Alors h_a (resp. h_b) est continûment différentiable sur A_s pour $s \neq 0, T, s_1, s_2$ (resp. $s \neq 0, T, s_3, s_4$). Soit $D = \{0, T, s_1, s_2, s_3, s_4\}$. Pour $s \notin D$, par le théorème de la valeur moyenne, il existe $\bar{y}_s, \bar{z}_s \in A_s$, tels que

$$\mu_a^T(s) = T^{-1}h_a'(\bar{y}_s), \quad \mu_b^T(s) = T^{-1}h_b'(\bar{z}_s).$$

Ecrivons

$$\begin{aligned}\mu_a^T(s)\mu_b^T(s) &= T^{-2}h_a'(\bar{y}_s)h_b'(\bar{z}_s) \\ &= T^{-1}\int_{A_s} h_a'(x)h_b'(x)\mathrm{d}x \\ &\quad + T^{-1}\int_{A_s}\left[h_a'(\bar{y}_s)h_b'(\bar{z}_s) - h_a'(x)h_b'(x)\right]\mathrm{d}x.\end{aligned}$$

Posons $\gamma(a,b) = \|h_a'\|_\infty \mathcal{L}(h_b') + \|h_b'\|_\infty \mathcal{L}(h_a')$. Alors

$$\left|h_a'(\bar{y}_s)h_b'(\bar{z}_s) - h_a'(x)h_b'(x)\right| \leq \gamma(a,b)T^{-1}, \quad \forall x \in A_s,$$

et

$$\left|\sum_{s\notin D} T^{-1}\int_{A_s}\left[h_a'(\bar{y}_s)h_b'(\bar{z}_s) - h_a'(x)h_b'(x)\right]\mathrm{d}x\right| \leq \gamma(a,b)T^{-2}.$$

Ce terme est de l'ordre de $o(T^{-1})$ soit si $\gamma(a,b)$ est indépendant de T, soit dans le cas contraire si $\gamma(a,b) = o(T)$. Ce qui est vrai dans le deuxième cas, puisque

$$\gamma(a,b) \leq C_2\rho_a^{-2}\rho_b^{-1} + C_3\rho_b^{-2}\rho_a^{-1} \leq C_4(\rho_a \wedge \rho_b)^{-3},$$

et que $(\rho_a \wedge \rho_b)^{-1} = o(T^{1/3})$ par hypothèse. Finalement,

$$\sum_{s\notin D}\mu_a^T(s)\mu_b^T(s) = \left[\int_{\bigcup_{s\notin D}A_s} h_a'(x)h_b'(x)\mathrm{d}x\right]T^{-1}(1+o(1)).$$

Compte tenu du fait que

$$T^{-1}H_{ab}^T(0) \longrightarrow \int_0^1 h_a(x)h_b(x)\mathrm{d}x, \quad T\to\infty,$$

l'expression 5.26 serait prouvée si

$$-T^{-1}(\sum_{s\in D}\int_{A_s} h_a'(x)h_b'(x)\mathrm{d}x) + \sum_{s\in D}\mu_a^T(s)\mu_b^T(s) = o(T^{-1}).$$

Ce qui est immédiat puisque D est fini et que quelque soit s ,

$$\begin{aligned}\left|T^{-1}\int_{A_s} h_a'(x)h_b'(x)\mathrm{d}x\right| &\leq T^{-2}\|h_a'\|_\infty\|h_b'\|_\infty \\ &\leq C_5T^{-2}(\rho_a\rho_b)^{-1},\end{aligned}$$

et

$$\begin{aligned}|\mu_a^T(s)\mu_b^T(s)| &\leq T^{-2}\|h_a'\|_\infty\|h_b'\|_\infty \\ &\leq C_5T^{-2}(\rho_a\rho_b)^{-1},\end{aligned}$$

sachant que $(\rho_a\rho_b)^{-1} \leq (\rho_a\wedge\rho_b)^{-2} = o(T^{2/3})$.

♠

Proposition 3 *Soit* $0 < \eta \leq \pi$. *Pour* $1 \leq a \leq r$,

$$\int_{|\sigma|>\eta} K_{aa}^T(\sigma)\mathrm{d}\sigma = O(T^{-3})\varsigma(\eta)\gamma(\rho_a), \tag{5.29}$$

avec

$$\varsigma(\eta) = \int_{|\sigma|>\eta} |e^{i\sigma} - 1|^{-4} \mathrm{d}\sigma = \frac{1}{4}\left[\cot(\frac{\eta}{2}) + \frac{1}{3}\cot^3(\frac{\eta}{2})\right],$$

$$\gamma(\rho_a) = \sup(\mathcal{L}(w_a'), \|w_a'\|_\infty)\left[\rho_a^4 \int_0^1 h_a^2(x)\mathrm{d}x\right]^{-1}.$$

Le terme $O(T^{-3})$ est uniforme en ρ_a.

Preuve. De l'expression (5.27) pour $\sigma \neq 0$,

$$H_a^T(\sigma) = (e^{i\sigma} - 1)^{-1} \sum_{s=0}^{T} \left[h_a(\frac{s+\frac{1}{2}}{T}) - h_a(\frac{s+\frac{1}{2}}{T})\right] e^{-is\sigma},$$

et

$$h_a(\frac{s+\frac{1}{2}}{T}) - h_a(\frac{s-\frac{1}{2}}{T}) = \int_{(s-\frac{1}{2})/T}^{(s+\frac{1}{2})/T} h_a'(x)\mathrm{d}x = T^{-1}\int_0^1 h_a'(\frac{s-\frac{1}{2}+x}{T})\mathrm{d}x.$$

D'où

$$H_a^T(\sigma) = [T(e^{i\sigma} - 1)]^{-1} \int_0^1 \left[\sum_{s=0}^{T} h_a'(\frac{s-\frac{1}{2}+x}{T}) e^{-is\sigma}\right] \mathrm{d}x.$$

Une nouvelle application de la transformation d'Abel (équation 5.24), compte tenu de

$$h_a'(\frac{-1-\frac{1}{2}+x}{T}) = h_a'(\frac{T+1-\frac{1}{2}+x}{T}) = 0 \quad \forall x \in [0,1],$$

donne pour $\sigma \neq 0$,

$$\sum_{s=0}^{T} h_a'(\frac{s-\frac{1}{2}+x}{T}) e^{-is\sigma} = (e^{i\sigma} - 1)^{-1} \sum_{s=-1}^{T} \left[h_a'(\frac{s+\frac{1}{2}+x}{T}) - h_a'(\frac{s-\frac{1}{2}+x}{T})\right] e^{-is\sigma},$$

et

$$H_a^T(\sigma) = T^{-1}(e^{i\sigma} - 1)^{-2} \int_0^1 \left(\sum_{s=-1}^{T} \left[h_a'(\frac{s+\frac{1}{2}+x}{T}) - h_a'(\frac{s-\frac{1}{2}+x}{T})\right] e^{-is\sigma}\right) \mathrm{d}x.$$

Etant donné $x \in [0,1]$, il existe 4 valeurs de s telles que h_a' possède un point éventuel de discontinuité dans l'intervalle $[(s-\frac{1}{2}+x)/T, (s+\frac{1}{2}+x)/T)$. Pour ces valeurs de s,

$$|h_a'(\frac{s+\frac{1}{2}+x}{T}) - h_a'(\frac{s-\frac{1}{2}+x}{T})| \leq 2\|h'\|_\infty.$$

Et pour s différent de ces 4 valeurs,

$$|h_a'(\frac{s+\frac{1}{2}+x}{T}) - h_a'(\frac{s-\frac{1}{2}+x}{T})| \leq T^{-1}\mathcal{L}(h')$$

Donc

$$|\sum_{s=-1}^{T}\left[h_a'(\frac{s+\frac{1}{2}+x}{T}) - h_a'(\frac{s-\frac{1}{2}+x}{T})\right]e^{-is\sigma}| \leq \frac{T+2-4}{T}\mathcal{L}(h') + 4*2\|h'\|_\infty.$$

Comme $\frac{2}{\rho_a} \leq (\frac{2}{\rho_a})^2$ on obtient

$$|H_a^T(\sigma)| \leq T^{-1}|e^{i\sigma}-1|^{-2}C_6(\frac{2}{\rho_a})^2 \sup(\mathcal{L}(w_a'), \|w_a'\|_\infty).$$

D'autre part, quand $T \to \infty$,

$$\frac{1}{T}H_{aa}^T(0) = \frac{1}{T}\sum_{s=1}^{T}\left(h_a(\frac{s-\frac{1}{2}}{T})\right)^2 \longrightarrow \int_0^1 h_a^2(x)\mathrm{d}x.$$

Donc

$$K_{aa}^T(\sigma) = O(T^{-3})\sup(\mathcal{L}(w_a'), \|w_a'\|_\infty)\left[\rho_a^4|e^{i\sigma}-1|^4\int_0^1 h_a^2(x)\mathrm{d}x\right]^{-1}.$$

Le terme $O(T^{-3})$ est indépendant de ρ_a et de σ. En intégrant,

$$\int_{|\sigma|>\eta} K_{aa}^T(\sigma)\mathrm{d}\sigma = O(T^{-3})\sup(\mathcal{L}(w_a'), \|w_a'\|_\infty)\left[\rho_a^4\int_0^1 h_a^2(x)\mathrm{d}x\right]^{-1}\varsigma(\eta),$$

avec

$$\varsigma(\eta) = \int_{|\sigma|>\eta}|e^{i\sigma}-1|^{-4}\mathrm{d}\sigma = \frac{1}{4}\left[\cot(\frac{\eta}{2}) + \frac{1}{3}\cot^3(\frac{\eta}{2})\right].$$

♠

Il résulte de la Proposition 3 et du Lemme 4 que $\{\underline{K}_{ab}^n\}$ est une appoximation de l'unité, pour tout $1 \leq a, b \leq r$ (cf. Edwards [II–10], §3.2.2). Comme par les Lemmes 3 et 4 le biais $B_{ab}(\phi)$ s 'écrit encore

$$B_{ab}(\phi) = \int_{\Pi^d}\phi(\alpha)\Delta_{ab}(\alpha)\mathrm{d}\alpha, \tag{5.30}$$

avec

$$\Delta_{ab}(\alpha) := \mathbf{E}I_{ab}^n(\alpha) - f_{ab}(\alpha) = \int_{\Pi^d}\underline{K}_{ab}^n(\beta)\left(f_{ab}(\alpha-\beta) - f_{ab}(\alpha)\right)\mathrm{d}\beta.$$

D'où, si $f_{ab} \in \mathcal{L}_p(\Pi^d)$ et $\phi \in \mathcal{L}_{p'}(\Pi^d)$, p et p' étant des exposants conjugués, $\Delta_{ab} \longrightarrow 0$ dans $\mathcal{L}_p$ et par l'inégalité de Hölder, le biais $B_{ab} \longrightarrow 0$. Le théorème suivant précise ce biais asymptotique avec des conditions supplémentaires. Celui-ci a été démontré par Dahlhaus-Künsch [II–8] dans le cas des processus univariés, i.e. $r = 1$ (leur démonstration publiée reste toujours difficile à lire).

Théorème 1 *Soit $\{X(t)\}, t \in \mathbb{Z}^d$ un processus centré r-vectoriel et stationnaire. Supposons que*

1. *l'élément f_{ab} de la densité spectrale f admette des dérivées secondes continues;*
2. *les fonctions de rabotage w_a, w_b possèdent une dérivée Lipschitzienne.*

Alors , si $n_k \to \infty$ et $n_k/n_1 \to \lambda_k \in (0,\infty)$, $k = 1,\ldots,d$, on a[1]

$$B_{ab}(\phi) = \left\{ \frac{1}{2}|n|^{-2/d}(\lambda_1 \cdots \lambda_d)^{2/d} \frac{\int_0^1 h_a'(x)h_b'(x)\mathrm{d}x}{\int_0^1 h_a(x)h_b(x)\mathrm{d}x} \left(\sum_{k=1}^{d} \lambda_k^{-2} \int_{\Pi^d} \phi(\alpha)(\partial_{kk}^2 f_{ab})(\alpha)\mathrm{d}\alpha \right) \right\} (1 + o(1)).$$

ρ_a et ρ_b pourraient dépendre de $|n|$, de sorte que $(\rho_a \wedge \rho_b)^{-1} = o(|n|^{1/(4d)})$. Le terme $o(1)$ est uniforme en ρ_a et ρ_b.

Preuve. Rappelons que

$$B_{ab}(\phi) = \int_{\Pi^d} \phi(\alpha)\Delta_{ab}(\alpha)\mathrm{d}\alpha,$$

avec

$$\Delta_{ab}(\alpha) = \int_{\Pi^d} \underline{K}_{ab}^n(\beta)\,(f_{ab}(\alpha - \beta) - f_{ab}(\alpha))\,\mathrm{d}\beta.$$

Comme f_{ab} est deux fois continûement dérivable,

$$f_{ab}(\alpha - \beta) - f_{ab}(\alpha) = - \sum_{k=1}^{d} \beta_k (\partial_k f_{ab})(\alpha) + 2 \sum_{k,l=1}^{d} \sin(\frac{\beta_k}{2}) \sin(\frac{\beta_l}{2})(\partial_{kl}^2 f_{ab})(\alpha) + R(\alpha,\beta), \quad (5.31)$$

où

$$|R(\alpha,\beta)| = o\left\{ \sum_{k=1}^{d} \sin^2(\frac{\beta_k}{2}) \right\} \quad \text{(uniforme en } \alpha),$$

[1] $\partial^m_{k_1,k_2,\ldots,k_m} = \frac{\partial^m}{\partial\alpha_{k_1}\partial\alpha_{k_2}\cdots\partial\alpha_{k_m}}$

lorsque $\beta \to 0$. Du fait que $\underline{K}_{ab}^n$ est paire et que $\int_\Pi K_{ab}^{n_k} \mathrm{d}\alpha_k = 1$, $\forall k = 1, \ldots, d$ (Lemme 4), on a

$$\begin{aligned}\Delta_{ab}(\alpha) &= 2\int_{\Pi^d} \left(\sum_{k=1}^{d} (\partial_{kk}^2 f_{ab})(\alpha) \sin^2(\frac{\beta_k}{2}) \right) \underline{K}_{ab}^n(\beta)\mathrm{d}\beta + \int_{\Pi^d} R(\alpha,\beta)\underline{K}_{ab}^n(\beta)\mathrm{d}\beta \\ &= 2\sum_{k=1}^{d} (\partial_{kk}^2 f_{ab})(\alpha) \left(\int_\Pi K_{ab}^{n_k} \sin^2(\frac{\beta_k}{2})\mathrm{d}\beta_k \right) + \int_{\Pi^d} R(\alpha,\beta)\underline{K}_{ab}^n(\beta)\mathrm{d}\beta \quad (5.32)\end{aligned}$$

Grâce à la Proposition 2 , le terme principal du biais $B_{ab}(\phi)$ est donné par celui du premier terme de $\Delta_{ab}(\alpha)$, compte tenu du mode de croissance des rectangles P_n,

$$\frac{n_k}{n_1} \longrightarrow \lambda_k \in (0, \infty) \quad k = 1, \ldots, d.$$

La même Proposition indique aussi que ce terme est de l'ordre de $O(n_1^{-2} + \cdots + n_d^{-2})$. Pour achever la preuve, nous allons montrer que

$$\int_{\Pi^d} R(\alpha,\beta)\underline{K}_{ab}^n(\beta)\mathrm{d}\beta = o(n_1^{-2} + \cdots + n_d^{-2}). \quad (5.33)$$

Soit ε positif. Alors

$$\exists \eta > 0, \ |\beta|_\infty := \sup_k |\beta_k| \leq \eta \Longrightarrow |R(\alpha,\beta)| \leq \varepsilon \left\{ \sum_{k=1}^{d} \sin^2 \frac{\beta_k}{2} \right\},$$

uniformément en α. On en déduit,

$$\left| \int_{\Pi^d} R(\alpha,\beta)\underline{K}_{ab}^n(\beta)\mathrm{d}\beta \right| \leq \varepsilon \int_{\Pi^d} |\underline{K}_{ab}^n(\beta)| \left\{ \sum_{k=1}^{d} \sin^2(\frac{\beta_k}{2}) \right\} \mathrm{d}\beta + C_7 \int_{|\beta|_\infty > \eta} |\underline{K}_{ab}^n(\beta)|\mathrm{d}\beta. \quad (5.34)$$

Du Lemme 4, il existe C_1,

$$|\underline{K}_{ab}^n(\beta)| \leq C_1 \underline{K}_{aa}^n(\beta)\underline{K}_{bb}^n(\beta)$$

Par l'inégalité de Cauchy-Schwarz et la Proposition 2,

$$\int_{\Pi^d} |\underline{K}_{ab}^n(\beta)| \left\{ \sum_{k=1}^{d} \sin^2(\frac{\beta_k}{2}) \right\} \mathrm{d}\beta = O(n_1^{-2} + \cdots + n_d^{-2}). \quad (5.35)$$

D'autre part,

$$|\beta|_\infty > \eta \Longrightarrow \exists l, \ 1 \leq l \leq d, \ |\beta_l| > \eta,$$

et

$$\begin{aligned}\int_{|\beta|_\infty>\eta} |\underline{K}^n_{ab}(\beta)|\mathrm{d}\beta &\leq \int_{|\beta_l|>\eta} |\underline{K}^n_{ab}(\beta)|\mathrm{d}\beta \\ &\leq \sqrt{C_1}\left(\int_{|\beta_l|>\eta} |\underline{K}^n_{aa}(\beta)|\mathrm{d}\beta\right)^{\frac{1}{2}}\left(\int_{|\beta_l|>\eta} |\underline{K}^n_{bb}(\beta)|\mathrm{d}\beta\right)^{\frac{1}{2}} \\ &= \sqrt{C_1}\left(\int_{|\beta_l|>\eta} |K^{n_l}_{aa}(\beta_l)|\mathrm{d}\beta_l\right)^{\frac{1}{2}}\left(\int_{|\beta_l|>\eta} |K^{n_l}_{bb}(\beta_l)|\mathrm{d}\beta_l\right)^{\frac{1}{2}}.\end{aligned}$$

D'où par la Proposition 3,

$$\begin{aligned}\int_{|\beta|_\infty>\eta} |\underline{K}^n_{ab}(\beta)|\mathrm{d}\beta &= O(n_l^{-3})\varsigma(\eta)\sqrt{\gamma(\rho_a)\gamma(\rho_b)} \\ &= O(n_l^{-3})\varsigma(\eta)(\rho_a\rho_b)^{-2}.\end{aligned}$$

Maintenant faisons dépendre (ε,η) de n, et $\varepsilon_n,\eta_n \to 0$. Lorsque $\eta \to 0$, $\varsigma(\eta) = O(\eta^{-3})$. Choisissons $\varepsilon_n \to 0$ suffisamment lentement tel que

$$\eta_n^{-3}(\rho_a\rho_b)^{-2} = o(n_l^{-1}),$$

(remarquons qu'ici la condition $(\rho_a \wedge \rho_b)^{-1} = o(|n|^{\frac{1}{4d}})$ est essentielle). Ainsi,

$$\int_{|\beta|_\infty>\eta_n} |\underline{K}^n_{ab}(\beta)|\mathrm{d}\beta = o(n_l^{-2}) = o(n_1^{-2}+\cdots+n_d^{-2}),$$

et il en est de même pour $|\int_{\Pi^d} R(\alpha,\beta)\underline{K}^n_{ab}(\beta)\mathrm{d}\beta|$ puisque $\varepsilon_n \to 0$ (équations 5.34 et 5.35).

♠

Remarques.

- Le biais $|n|^{-\frac{2}{d}}$ est plus petit que $|n|^{-\frac{1}{2}}$ pour $d = 1,2,3$. D'autre part, des calculs directs montrent que ,

$$\frac{\int_0^1 h'_a(x)h'_b(x)\mathrm{d}x}{\int_0^1 h_a(x)h_b(x)\mathrm{d}x} = O((\rho_a \wedge \rho_b)^{-1})$$

  Ainsi, $B_{ab}(\phi)$ est d'ordre plus petit que $|n|^{-\frac{1}{2}}$ si $(\rho_a \wedge \rho_b)^{-1} = o(|n|^{2/d-1/2})$, a fortiori si $(\rho_a \wedge \rho_b)^{-1} = o(|n|^{1/(4d)})$ pour $d = 1,2,3$.

- Dorénavant, on fixera le mode de croissance des rectangles $P_n \uparrow \mathbb{Z}^d$ par celui évoqué ici: $n_k/n_1 \longrightarrow \lambda_k \in (0,\infty), k = 1,\ldots,d$.

5.2 Asymptotique probabiliste

Soient ϕ_1 et ϕ_2 deux fonctions de $\Pi^d \to \mathrm{C}$, $1 \leq a_1, b_1, a_2, b_2 \leq r$, et $J_{a_1 b_1}(\phi_1), J_{a_2 b_2}(\phi_2)$ (resp. $J^n_{a_1 b_1}(\phi_1), J^n_{a_2 b_2}(\phi_2)$) les spectrographes (resp. leurs estimations) associés. Nous voulons ici donner une estimation asymptotique de la covariance

$$\mathbf{Cov}\,\{J_{a_1 b_1}(\phi_1), J_{a_2 b_2}(\phi_2)\}$$

Dahlhaus [II–6] a complètement résolu ce problème pour les séries chronologiques ($d = 1$) scalaires ou vectoriels. Le Théorème 2 ci-dessous généralise ce résultat (Dahlhaus[II–6], Lemma 6) au cas spatial ($d \geq 1$). Sa preuve sera omise car la généralisation est directe.

Rappelons quelques définitions sur les cumulants et leurs fonctions spectrales. Le cumulant d'ordre m, d'indices $a_1, \ldots, a_m$, lorsqu'il existe, est défini par

$$C_{a_1,\ldots,a_m}(u_1, \ldots, u_{m-1}) := \mathrm{cum}\,(X_{a_1}(u_1), \ldots, X_{a_{m-1}}(u_{m-1}), X_{a_m}(0))$$

et la fonction spectrale correspondante par

$$f_{a_1,\ldots,a_m}(\alpha_1, \ldots, \alpha_{m-1}) := \sum_{u_1,\ldots,u_{m-1} \in \mathbb{Z}^d} C_{a_1,\ldots,a_m}(u_1, \ldots, u_{m-1})\ e^{-i \sum_{j=1}^{k-1} u_j \alpha_j}$$

avec $\alpha_1, \ldots, \alpha_{m-1} \in \Pi^d$. Lorsque $m = 2$, il s'agit de la densité spectrale. Nous supposerons que la fonction spectrale $f_{a_1 b_1 a_2 b_2}$ d'ordre 4 d'indices $a_1 b_1 a_2 b_2$ vérifie la condition suivante:

Condition 1

$$(a). \exists \delta > 0,\ \int_{\Pi^{3d}} |f_{a_1 b_1 a_2 b_2}(\alpha, \beta, \gamma)|^{1+\delta}\, \mathrm{d}\alpha \mathrm{d}\beta \mathrm{d}\gamma < \infty$$

$$(b). \lim_{\beta \to 0} \int_{\Pi^{2d}} |f_{a_1 b_1 a_2 b_2}(\alpha, -\alpha + \beta, \gamma) - f_{a_1 b_1 a_2 b_2}(\alpha, -\alpha, \gamma)|\, \mathrm{d}\alpha \mathrm{d}\gamma = 0$$

Nous aurons aussi besoin des constantes suivantes relatives au rabotage:

$$H_{a_1,\ldots,a_m} := \int_0^1 \left(\prod_{j=1}^{m} h_{a_j}(x) \right) \mathrm{d}x \tag{5.36}$$

Théorème 2 *Soit $\{X(t)\}, t \in \mathbb{Z}^d$ un processus centré r-vectoriel et stationnaire jusqu'à l'ordre 4. Supposons que*

1. *les éléments* $f_{a_1a_2}, f_{b_1b_2}, f_{a_1b_2}, f_{b_1a_2}$ *de la densité spectrale soient de carrés - intégrables;*

2. $f_{a_1b_1a_2b_2}$ *remplisse la Condition 1;*

3. ϕ_1, ϕ_2 *soient à variation bornée.*

Alors , lorsque $P_n \uparrow \mathbb{Z}^d$,

$$\mathrm{Cov}\left\{J^n_{a_1b_1}(\phi_1), J^n_{a_2b_2}(\phi_2)\right\} = \tag{5.37}$$

$$\left(2\pi \frac{H_{a_1b_1a_2b_2}}{H_{a_1b_1}H_{a_2b_2}}\right)^d |n|^{-1} \left\{ \begin{array}{l} \int_{\Pi^d} \phi_1(\alpha)\overline{\phi_2(\alpha)} f_{a_1a_2}(\alpha) f_{b_1b_2}(-\alpha)\mathrm{d}\alpha \\ + \int_{\Pi^d} \phi_1(\alpha)\overline{\phi_2(-\alpha)} f_{a_1b_2}(\alpha) f_{b_1a_2}(-\alpha)\mathrm{d}\alpha \\ + \int_{\Pi^d}\int_{\Pi^d} \phi_1(\alpha)\overline{\phi_2(-\beta)} f_{a_1b_1a_2b_2}(\alpha,-\alpha,\beta)\mathrm{d}\alpha\mathrm{d}\beta \end{array} \right\} (1+o(1)).$$

Observons que le facteur $\frac{H_{a_1b_1a_2b_2}}{H_{a_1b_1}H_{a_2b_2}}$ est toujours plus grand que 1. En effet, la classe $\mathcal{H}$ des rabots que nous avons fixés possède 2 propriétés suivantes: si $g_1, g_2 \in \mathcal{H}$, alors $g_1g_2 \in \mathcal{H}$ et [2]

$$\int_0^1 g_1(x)g_2(x)\mathrm{d}x \geq \left(\int_0^1 g_1(x)\mathrm{d}x\right)\left(\int_0^1 g_2(x)\mathrm{d}x\right).$$

On en déduit en particulier que le rabotage augmente la variance des estimateurs spectrographiques par rapport au cas du non-rabotage ($h_{a_j} \equiv 1$ et $\frac{H_{a_1b_1a_2b_2}}{H_{a_1b_1}H_{a_2b_2}} = 1$), ce qui explique la perte d'information dûe au rabotage.

Signalons que dans le cas de non-rabotage et non-spatial ($d = 1$), cette structure de covariance se trouve dans Brillinger [II–4] (équation (7.6.7)).

5.3 Le théorème de limite centrale

Nous nous intéressons désormais aux processus strictement stationnaires. Les résultats de limite centrale nécessitent des hypothèses du type de " mélange spatial–faible dépendance" sur le processus étudié. Nous utiliserons les coefficients de mélange spatial $\{\alpha_{k,l}(m)\}$ introduits dans Bolthausen [II–3]: pour $\Lambda \in \mathbb{Z}^d$, soit $\sigma(\Lambda)$ la σ-algèbre engendrée par $X_a(t), t \in \Lambda, 1 \leq a \leq r$. Si $\Lambda_1, \Lambda_2 \subset \mathbb{Z}^d$, soit $d(\Lambda_1, \Lambda_2) = \inf\{d(t_1,t_2) : t_1 \in \Lambda_1, t_2 \in \Lambda_2\}$. Alors

[2] Par une inégalité de Tchebyshev. Voir par exemple *"Inequalities"* de Hardy G.H. & Littlewood J.E. & Pólya G., 1964. Cambridge: London

Définition 7 *si* $m \in \mathbb{N}, k, l \in \mathbb{N} \cup \{\infty\}$,

$$\alpha_{k,l}(m) = \sup\{|\mathbf{P}(A_1 \cap A_2) - \mathbf{P}(A_1)\mathbf{P}(A_2)| : A_i \in \sigma(\Lambda_i), |\Lambda_1| \leq k, |\Lambda_2| \leq l, d(\Lambda_1, \Lambda_2) \geq m\}. \tag{5.38}$$

L'hypothèse de mélange utilisée est la suivante:

Hypothèse 1 *Il existe* $\delta > 0$, *tel que*

1. $\mathbf{E}|X_a(s)|^{4+2\delta} < \infty$, $s \in \mathbb{Z}^d$, $1 \leq a \leq r$,
2. $\sum_{m=1}^{\infty} m^{d-1} \alpha_{2,\infty}^{\delta/(2+\delta)} < \infty$.

Des processus gaussiens[3], des processus linéaires[4] ou des champs de Gibbs stationnaires (voir Guyon[II–12]) sont des exemples de processus qui vérifient cette hypothèse de mélange.

Le résultat sur le biais asymptotique (Théorème 1) et sur la structure asymptotique des covariances (Théorème 2) nous amène au

Théorème 3 *Soit* $\{X(t)\}, t \in \mathbb{Z}^d$ *un processus centré r-vectoriel et strictement stationnaire vérifiant l'hypothèse 1 de mélange ci-dessus. Supposons que*

1. *la densité spectrale admette des dérivées secondes continues; quels que soient* $1 \leq a, b, c, d \leq r$, *la fonction spectrale des cumulants d'ordre 4 d'indice abcd,* f_{abcd} *remplisse la Condition* 1;
2. *les fonctions de rabotage* $\{w_a\}$, $1 \leq a \leq r$, *admettent une dérivée première Lipschitzienne;*
3. *les fonctions* $\phi_1, \ldots, \phi_p$ *soient continues et à variation bornée.*

Alors pour $d = 1, 2, 3$, *lorsque* $P_n \uparrow \mathbb{Z}^d$, *la suite des vecteurs*

$$\sqrt{|n|}\left[J^n_{a_j b_j}(\phi_j) - J_{a_j b_j}(\phi_j)\right], \quad j = 1, \ldots, p$$

converge en distribution vers un vecteur gaussien centré $[\eta(\phi_j)]$, $j = 1, \ldots, p$, *de*

[3]Voir par exemple "*Processus Aléatoires Gaussiens*" d'Ibrahimov & Rozanov, 1974. Editions Mir: Moscou

[4]Voir par exemple "*The Statistical Analysis of Times Series*" d'Anderson, 1972. John Wiley & Sons: New York

covariance

$$\mathbf{Cov}\{\eta(\phi_j),\eta(\phi_k)\} =$$

$$\left(2\pi\frac{H_{a_jb_ja_kb_k}}{H_{a_jb_j}H_{a_kb_k}}\right)^d \left\{ \begin{array}{l} \int_{\Pi^d}\phi_j(\alpha)\overline{\phi_k(\alpha)}f_{a_ja_k}(\alpha)f_{b_jb_k}(-\alpha)\mathrm{d}\alpha \\ +\int_{\Pi^d}\phi_j(\alpha)\overline{\phi_k(-\alpha)}f_{a_jb_k}(\alpha)f_{b_ja_k}(-\alpha)\mathrm{d}\alpha \\ +\int_{\Pi^d}\int_{\Pi^d}\phi_j(\alpha)\overline{\phi_k(-\beta)}f_{a_jb_ja_kb_k}(\alpha,-\alpha,\beta)\mathrm{d}\alpha\mathrm{d}\beta \end{array} \right\} (5.39)$$

Les $(\rho_a,\ 1 \leq a \leq r)$ *pourraient dépendre de* n *de sorte que* $(\min_a \rho_a)^{-1} = o(|n|^{\frac{1}{4d}})$ *et dans ce cas, les rapports* $H_{a_jb_ja_kb_k}/(H_{a_jb_j}H_{a_kb_k})$ *sont remplacés par leur valeurs limites.*

La preuve complète de ce théorème ne sera pas donnée ici. Il s'agit d'une généralisation d'un procédé classique connu en analyse de séries chronologiques (voir par exemple Dahlhaus [II–7], Lemma 4.1). Elle consiste, dans un premier temps, à approximer uniformément les $\{\phi_j\}$ par leurs sommes de Cesàro. Ensuite, on détermine l'asymptotique de ces somme partielles grâce à une application du Théorème 2 aux fonctions de test $\{e^{i<\alpha,u>}, u \in \mathbb{Z}^d\}$ qui, avec l'hypothèse de mélange, implique directement un résultat de limite centrale pour les estimateurs de covariances $\{C_{ab}^n(u)\}$. Le résultat s'obtient enfin en faisant tendre ces sommes partielles vers leur limite compte tenu du fait que les biais d'estimation sont d'ordre plus petits que $|n|^{-\frac{1}{2}}$.

Chapitre 6

Estimation paramétrique de Whittle par rabotage

6.1 Introduction

On considère $\{\mathbf{P}_\theta\}_{\theta\in\Theta}$, une famille de probabilités globales des processus stationnaires réels et r-vectoriels sur $\mathbb{Z}^d$. L'ensemble des paramètres Θ est un compact de l'espace $\mathbb{R}^p$ (p entier ≥ 1). Sachant a priori que le processus examiné X appartient à cette famille et disposant d'une observation de celui-ci sur un rectangle P_n de $\mathbb{Z}^d$, il s'agit d'estimer la valeur du paramètre correspondant, notons-le θ_0. Un procédé classique consiste à exploiter la famille des fonctionnelles de Whittle. Celles-ci sont des log-vraisemblances approchées lorsque les processus sont gaussiens; sinon, elles permettent encore de bonnes estimations (voir par exemple: Whittle[II–17], Guyon[II–11], Azencott & Dacunha-Castelle[II–2]). Dahlhaus[II–6] et Dahlhaus & Künsch[II–8] ont examiné cette estimation dans le cas des observations rabotées, respectivement pour des séries chronologiques ($d = 1$) et des processus spatiaux ($d > 1$) scalaires. Nous proposons une approche étendue aux processus spatiaux vectoriels.

Les foctionnelles de Whittle sont particulièrement intéressantes en statistique spatiale. En effet, l'absence d'un ordre naturel du lattice $\mathbb{Z}^d$ implique que des procédés connus sur $\mathbb{Z}$, tels que l'algorithme d'aller-retour de Box-Jenkins ne fonctionne plus.

La croissance des rectangles $P_n \uparrow \mathbb{Z}^d$, les rabots $\{h_a\}$ ainsi que les notations du chapitre précédent sont conservés. L'aléatoire et la structure probabiliste sous-jacente seront manifestés par l'introduction de ω.

Les fonctionnelles de Whittle $\{L_n(\theta,\omega)\} : \Theta \to \mathbb{R}$ sont définies comme suit (lorsque celles-ci existent et à une constante additive près):

$$L_n(\theta,\omega) := -\frac{|n|}{2}(2\pi)^{-d}\int_{\Pi^d}\left(\log\det[f_\theta(\alpha)] + \mathrm{tr}\left[f_\theta^{-1}(\alpha)I^n(\alpha,\omega)\right]\right)\mathrm{d}\alpha, \qquad (6.1)$$

où $I^n(\alpha,\omega)$ est le périodogramme associé aux observations rabotées de X sur P_n. Définissons $V_n(\theta,\omega) := -|n|^{-1}(L_n(\theta,\omega) - L_n(\theta_0,\omega))$, c-à-d,

$$V_n(\theta,\omega) := \frac{1}{2}(2\pi)^{-d}\int_{\Pi^d}\left(\mathrm{tr}\left[(f_\theta^{-1}(\alpha) - f_{\theta_0}^{-1}(\alpha))I^n(\alpha,\omega)\right] - \log\det[f_\theta^{-1}(\alpha)f_{\theta_0}(\alpha)]\right)\mathrm{d}\alpha. \tag{6.2}$$

Supposons que nous puissions appliquer le Théorème 2 pour le processus X. Alors, pour tout $\theta \in \Theta$, $V_n(\theta,\omega)$ converge en $\mathbf{P}_{\theta_0}$-probabilité vers

$$K(\theta,\theta_0) = \frac{1}{2}(2\pi)^{-d}\int_{\Pi^d}\left(\mathrm{tr}\left[f_\theta^{-1}(\alpha)f_{\theta_0}(\alpha) - Id\right] - \log\det[f_\theta^{-1}(\alpha)f_{\theta_0}(\alpha)]\right)\mathrm{d}\alpha, \tag{6.3}$$

où Id est la matrice identité. Si A est une matrice hermitienne définie positive, toutes ses valeurs propres $\{\mu_j\}$ sont des réels strictement positifs. Et

$$\mathrm{tr}(A - Id) - \log\det A = \sum_j(\mu_j - 1 - \log\mu_j) \geq 0.$$

L'inégalité est stricte sauf si $\forall j,\ \mu_j = 1$, i.e., $A = Id$. Ainsi, $K(\theta,\theta_0) \geq 0$, $\forall\theta \in \Theta$, et $K(\theta,\theta_0) = 0$ si et seulement si f_θ, f_{θ_0} ne se différencient que sur un ensemble négligeable. Nous supposerons vérifiée dans toute la suite, la condition suivante d'identifiabilité du modèle:

Condition 2 *Le modèle est identifiable, c'est à dire:*

$$\text{si}\quad \theta \neq \theta' \text{alors}\quad \{\alpha \in \Pi^d / f_\theta(\alpha) \neq f_{\theta'}(\alpha)\}\ \text{est non négligeable.}$$

Ainsi,

$$K(\theta,\theta_0) = 0 \Longleftrightarrow \theta = \theta_0. \tag{6.4}$$

K sera appelé la fonction de contraste (ou le contraste) relatif à θ_0 (si les processus sont gaussiens, K est à un facteur près, l'information asymptotique de Kullback de $\mathbf{P}_\theta$ sur $\mathbf{P}_{\theta_0}$). L'optique développée ici est standard et bien connue (cf. [II–5]): $\{V_n(\theta,\omega)\}$ est un processus de contraste associé à K et relatif à θ_0 ([II–5], définition 3.2.7). Soit $\sigma(P_n)$ la tribu engendrée par les $\{X_a(s), s \in P_n, 1 \leq a \leq r\}$. L'estimateur de Whittle, défini comme une solution $\sigma(P_n)$-mesurable de

$$\hat{\theta}_n(\omega) = \mathrm{Arg}\max_{\theta\in\Theta} L_n(\theta,\omega) = \mathrm{Arg}\min_{\theta\in\Theta} V_n(\theta,\omega), \tag{6.5}$$

est un estimateur du minimum de contraste.

La suite de ce chapitre est divisée comme suit: nous établissons d'abord la consistance et la normalité aymptotique de ces estimateurs $\{\hat{\theta}_n\}$ sous des conditions appropriées (3.2); ensuite, nous construisons un test asymptotique de surparamétrisation (3.3); enfin, nous développons ces estimateurs pour des champs de Markov gaussiens.

6.2 La consistance et la normalité aymptotique de l'estimateur $\{\hat{\theta}_n\}$

Nous commençons par la consistance.

Théorème 4 *On suppose que*

1. *la famille des fonctions* $\{\theta \mapsto f_\theta(\alpha)\}_{\alpha \in \Pi^d}$ *est équicontinue sur* $\overset{\circ}{\Theta}$;
2. $\forall(\theta, \alpha) \in \Theta \times \Pi^d$, $f_\theta(\alpha)$ *est régulière;*
3. $\forall \theta \in \Theta$, $\{\alpha \mapsto f_\theta(\alpha)\}$ *est de variation bornée;*
4. *le processus* $\{X(t)\}$ *est stationnaire jusqu'à l'ordre* 4 *et les fonctions spectrales des cumulants d'ordre* 4, $\{f_{\theta_0,abcd}\}$ *remplissent la* Condition 1 *pout tout* $1 \leq a, b, c, d \leq r$.

Alors, si $\theta_0 \in \overset{\circ}{\Theta}$, $\hat{\theta}_n$ *est un estimateur consistant.*

Preuve. L'hypothèse 2 est fondamentale pour la définition du processus de contraste $\{V_n(\theta, \omega)\}$. L'hypothèse 3, jointe au fait qu'à θ fixé , det f_θ est minoré et f_θ bornée sur Π^d, implique que $\{\alpha \mapsto f_\theta^{-1}(\alpha)\}$ sont aussi des fonctions de variation bornée.

D'autrepart, comme ces fonctions sont aussi bornées, nous avons (voir commentaire avant Théorème 1):

$$\mathbf{E}_{\theta_0} V_n(\theta, \omega) \longrightarrow K(\theta, \theta_0).$$

Aussi, l'hypothèse 4 permet l'application du théorème 2 qui donne

$$\mathbf{Var}_{\theta_0} V_n(\theta, \omega) = O(|n|^{-1}).$$

D'où par l'inégalité de Tchebyshev,

$$V_n(\theta, \omega) \longrightarrow K(\theta, \theta_0) \quad \text{en } \mathbf{P}_{\theta_0} \text{ probabilité .} \tag{6.6}$$

Enfin, l'équicontinuité de la famille $\{\theta \mapsto f_\theta(\alpha)\}_{\alpha \in \Pi^d}$ implique que $K(\theta, \theta_0)$ et à ω fixé $V_n(\theta, \omega)$ sont des fonctions continues en θ.

Notons pour $\eta > 0$, les v.a.

$$w_n(\eta,\omega) := \sup_{\|\theta-\theta'\|_\infty \le \eta} \left|V_n(\theta,\omega) - V_n(\theta',\omega)\right|, \quad \theta,\theta' \in \Theta.$$

Suivant le théorème 3.2.8 de [II–5], la consistance sera établie si l'on prouve l'existence d'une suite (ε_k) qui décroît vers 0, telle que pour tout k,

$$\lim_{P_n \uparrow \mathbb{Z}^d} \mathbf{P}_{\theta_0}\left[w_n(\frac{1}{k},\omega) \ge \varepsilon_k\right] = 0.$$

Nous allons construire une suite (ε_k) de ce type. L'équicontinuité de la famille $\{\theta \mapsto f_\theta(\alpha)\}_{\alpha\in\Pi^d}$ implique celle des deux autres familles $\{\theta \mapsto \log\det f_\theta(\alpha)\}_{\alpha\in\Pi^d}$, et $\{\theta \mapsto f_\theta^{-1}(\alpha)\}_{\alpha\in\Pi^d}$ puisque les $\{f_\theta(\alpha)\}$ sont régulières (on munira les vecteurs et matrices de la norme uniforme $\|.\|_\infty$). Notons φ,ψ les modules de continuité respectifs de ces deux familles. Observons d'abord que , si A est une matrice $(r\times r)$ quelconque et B hermitienne $(r\times r)$ " multiplicative" (il existe un vecteur complexe v tel que $B = v^t\bar{v}$), alors

$$|\mathrm{tr}(AB)| \le r\|A\|_\infty \mathrm{tr}(B).$$

L'application à $A = (f_\theta^{-1} - f_{\theta'}^{-1})(\alpha)$ et $B = I^n(\alpha,\omega)$ donne

$$\begin{aligned}
&2(2\pi)^d \left|V_n(\theta,\omega) - V_n(\theta',\omega)\right| \\
&\le \int_{\Pi^d} \left(\left|\log\det f_\theta(\alpha) - \log\det f_{\theta'}(\alpha)\right| + \left|\mathrm{tr}\left[(f_\theta^{-1}(\alpha) - f_{\theta'}^{-1}(\alpha))I^n(\alpha,\omega)\right]\right|\right) \mathrm{d}\alpha \\
&\le \int_{\Pi^d} \left(\varphi(||\theta-\theta'||) + r\psi(||\theta-\theta'||)\mathrm{tr}I^n(\alpha,\omega)\right) \mathrm{d}\alpha \\
&= (2\pi)^d\varphi(||\theta-\theta'||) + r\psi(||\theta-\theta'||)\int_{\Pi^d} \mathrm{tr}I^n(\alpha,\omega)\mathrm{d}\alpha.
\end{aligned}$$

Posons $\xi(u) := \frac{1}{2}\max\{\varphi(u), r\psi(u)\}$, $u \in \mathbb{R}^+$, on obtient

$$\left|V_n(\theta,\omega) - V_n(\theta',\omega)\right| \le \xi(||\theta-\theta'||)\left(1 + (2\pi)^{-d}\int_{\Pi^d} \mathrm{tr}I^n(\alpha,\omega)\mathrm{d}\alpha\right).$$

D'où, par définition de $w_n(\frac{1}{k},\omega)$,

$$w_n(\frac{1}{k},\omega) \le \xi(\frac{1}{k})\left(1 + (2\pi)^{-d}\int_{\Pi^d} \mathrm{tr}I^n(\alpha,\omega)\mathrm{d}\alpha\right). \tag{6.7}$$

Notons $a(\theta_0) = (2\pi)^{-d} \int_{\Pi^d} \mathrm{tr} f_{\theta_0}(\alpha) \mathrm{d}\alpha$. Selon des arguments similaires à ceux conduisant à l'équation 4.6, nous avons

$$\lim_{P_n \uparrow \mathbb{Z}^d} (2\pi)^{-d} \int_{\Pi^d} \mathrm{tr} I^n(\alpha, \omega) \mathrm{d}\alpha = a(\theta_0) \quad \text{en } \mathbf{P}_{\theta_0} \text{ probabilité.}$$

Donc,

$$\mathbf{P}_{\theta_0} \left((2\pi)^{-d} \int_{\Pi^d} \mathrm{tr} I^n(\alpha, \omega) \mathrm{d}\alpha \geq a(\theta_0) + 1 \right) \longrightarrow 0. \tag{6.8}$$

Prenant $\varepsilon_k = \xi(\frac{1}{k})(2 + a(\theta_0))$, d'une part , il est clair que ε_k est décroissant et tend vers 0 quand $k \to \infty$; d'autre part, par le choix de $\{\varepsilon_k\}$ et l'inégalité (4.7), on a pour tout k fixé ,

$$\mathbf{P}_{\theta_0} \left(w_n(\frac{1}{k}) \geq \varepsilon_k \right) \leq \mathbf{P}_{\theta_0} \left(\int_{\Pi^d} \mathrm{tr} I^n(\alpha, \omega) \mathrm{d}\alpha \geq a(\theta_0) + 1 \right)$$

qui tend vers 0.

♠

Remarque. L'équicontinuité de la famille $\{\theta \mapsto f_\theta(\alpha)\}_{\alpha \in \Pi^d}$ est évidemment vérifiée si $\{(\theta, \alpha) \mapsto f_\theta(\alpha)\}$ est continue sur le compact $\Theta \times \Pi^d$.

Le résultat essentiel est le théorème suivant qui établit la normalité asymptotique des estimateurs de Whittle par rabotage. Les conditions requises sont essentiellement celles du Théorème 3 qui garantissent la normalité asymptotique des estimateurs spectrographiques.

Théorème 5 *Supposons que*

1. *le processus* $\{X(t)\}, t \in \mathbb{Z}^d$ *,ses fonctions spectrales* $\{f_{\theta_0}, f_{\theta_0, abcd}\}$ *ainsi que les rabots* $\{h_a\}$ *vérifient les hypothèses du Théorème 3;*
2. $\forall (\theta, \alpha) \in \Theta \times \Pi^d, f_\theta(\alpha)$ *est régulière;* $\{\theta \mapsto f_\theta(\alpha)\}_{\alpha \in \Pi^d}$ *est équicontinue sur* $\overset{\circ}{\Theta}$ *;* $\forall \theta, \{\alpha \mapsto f_\theta(\alpha)\}$ *est à variation bornée;*
3. $\forall \omega$ *et* $\forall \alpha \in \Pi^d$, $\{\theta \mapsto f_\theta(\alpha)\}$ *admet des dérivées secondes continues dans un voisinage de* θ_0.

Si $\theta_0 \in \overset{\circ}{\Theta}$, *alors pour* $d = 1, 2, 3$, *lorsque* $P_n \uparrow \mathbb{Z}^d$,

$$\left[\sqrt{|n|} (\hat{\theta}_n - \theta_0) \right]$$

converge en distribution vers un vecteur gaussien centré $[\eta_i]_{i=1,\dots,p}$ *de covariance* $\Gamma(\theta_0)^{-1}C(\theta_0)\Gamma(\theta_0)^{-1}$, *avec*[1] *(pour* $1 \le i,j \le p$*),*

$$\Gamma_{ij}(\theta_0) = \frac{1}{2}(2\pi)^{-d}\int_{\Pi^d} \mathrm{tr}\left[f_{\theta_0}^{-1}(\alpha)\mathrm{D}_j f_{\theta_0}(\alpha) f_{\theta_0}^{-1}(\alpha)\mathrm{D}_i f_{\theta_0}(\alpha)\right]\mathrm{d}\alpha, \tag{6.9}$$

et

$$C_{ij}(\theta_0) =$$

$$\frac{1}{4}(2\pi)^{-d}\sum_{a,b,c,d=1}^{r}\left(\frac{H_{abcd}}{H_{ab}H_{cd}}\right)^d \left\{\begin{array}{l} \int_{\Pi^d} \mathrm{D}_i f_{\theta_0}^{-1}(\alpha)_{ba}\overline{\mathrm{D}_j f_{\theta_0}^{-1}(\alpha)_{dc}} f_{\theta_0}(\alpha)_{ac} f_{\theta_0}(-\alpha)_{bd}\mathrm{d}\alpha \\ + \int_{\Pi^d} \mathrm{D}_i f_{\theta_0}^{-1}(\alpha)_{ba}\overline{\mathrm{D}_j f_{\theta_0}^{-1}(-\alpha)_{dc}} f_{\theta_0}(\alpha)_{ad} f_{\theta_0}(-\alpha)_{bc}\mathrm{d}\alpha \\ + \int_{\Pi^d}\int_{\Pi^d} \mathrm{D}_i f_{\theta_0}^{-1}(\alpha)_{ba}\overline{\mathrm{D}_j f_{\theta_0}^{-1}(-\beta)_{dc}} f_{\theta_0,abcd}(\alpha,-\alpha,\beta)\mathrm{d}\alpha\mathrm{d}\beta. \end{array}\right. \tag{6.10}$$

Les $(\rho_a,\ 1 \le a \le r)$ *pourraient dépendre de* n *de sorte que* $(\min_a \rho_a)^{-1} = o(|n|^{\frac{1}{4d}})$ *et dans ce cas, les rapports* $H_{abcd}/(H_{ab}H_{cd})$ *sont remplacés par leur valeur limite.*

Preuve. Observons d'abord que les hypothèses 1, 2 impliquent entre autre la consistance des estimateurs. Ensuite pour chaque n,

$$\mathrm{grad}V_n(\hat{\theta}_n,\omega) = 0.$$

Le développement de Taylor d'ordre 1 (avec reste intégral) donne pour tout $1 \le j \le p$:

$$0 = \mathrm{D}_j V_n(\hat{\theta}_n,\omega) = \mathrm{D}_j V_n(\theta_0,\omega) + \sum_{i=1}^{p}\left(\hat{\theta}_n^i - \theta_0^i\right)\mathrm{D}_{ij}^2 V_n(\theta_0,\omega) + \Delta_j,$$

avec

$$\Delta_j = \sum_{i=1}^{p}\left(\hat{\theta}_n^i - \theta_0^i\right)\int_0^1 \left[\mathrm{D}_{ij}^2 V_n(\theta_0 + t(\hat{\theta}_n - \theta_0),\omega) - \mathrm{D}_{ij}^2 V_n(\theta_0,\omega)\right]\mathrm{d}t.$$

L'hypothèse 3 et la consistance de $\hat{\theta}_n$ impliquent que Δ_j tend vers 0 en $\mathbf{P}_{\theta_0}$-probabilité. Le résultat annoncé sera prouvé si on établit les 2 propriétés suivantes:

1. $\sqrt{|n|}\,[\mathrm{grad}V_n(\theta_0,\omega)] \xrightarrow{\mathcal{L}(\mathbf{P}_{\theta_0})} \mathcal{N}_p(0, C(\theta_0))$;

2. $\left[\mathrm{D}_{ij}^2 V_n(\theta_0,\omega)\right] \xrightarrow{\mathbf{P}_{\theta_0}} [\Gamma_{ij}(\theta_0)]$.

[1] $\mathrm{D}^m_{k_1,k_2,\dots,k_m} = \frac{\partial^m}{\partial\theta_{k_1}\partial\theta_{k_2}\cdots\partial\theta_{k_m}}$.

Avant de procéder à la démonstration de ces deux propriétés, rappelons 3 identités. Si $A(\theta)$ est une fonction matricielle et hermitienne positive, alors on a [2]:

1. $\mathrm{D}_j \log\det A = \mathrm{tr}(A^{-1}\mathrm{D}_j A)$;
2. $(\mathrm{D}_j A)A^{-1} + A(\mathrm{D}_j A^{-1}) = 0$;
3. $(\mathrm{D}^2_{ij} A)A^{-1} + (\mathrm{D}_j A)(\mathrm{D}_i A^{-1}) + (\mathrm{D}_i A)(\mathrm{D}_j A^{-1}) + A(\mathrm{D}^2_{ij} A^{-1}) = 0$.

Maintenant, pour tout $\theta \in \overset{\circ}{\Theta}$,

$$\mathrm{D}_j V_n(\theta,\omega) = \frac{1}{2}(2\pi)^d \int_{\Pi^d} \mathrm{tr}\left[(\mathrm{D}_j f_\theta^{-1}(\alpha)) I^n(\alpha,\omega) + f_\theta^{-1}(\alpha)\mathrm{D}_j f_\theta(\alpha)\right] \mathrm{d}\alpha. \qquad (6.11)$$

L'application des identités 1-2 ci-dessus donne pour θ_0,

$$\mathrm{D}_j V_n(\theta_0,\omega) = \frac{1}{2}(2\pi)^d \int_{\Pi^d} \mathrm{tr}\left[(\mathrm{D}_j f_{\theta_0}^{-1}(\alpha))(I^n(\alpha,\omega) - f_{\theta_0}(\alpha))\right] \mathrm{d}\alpha.$$

La propriété 1 en découle directement par application du théorème 3, grâce à la continuité des dérivées premières $\{\mathrm{D}_j f_{\theta_0}^{-1}(\alpha)\}$ en α.

Une seconde dérivation de l'équation (4.11) conduit à

$$\mathrm{D}^2_{ij} V_n(\theta,\omega) = \frac{1}{2}(2\pi)^d \int_{\Pi^d} \mathrm{tr}\left[(\mathrm{D}^2_{ij} f_\theta^{-1}(\alpha)) I^n(\alpha,\omega) + \mathrm{D}_i f_\theta^{-1}(\alpha)\mathrm{D}_j f_\theta(\alpha) + f_\theta^{-1}(\alpha)\mathrm{D}^2_{ij} f_\theta(\alpha)\right] \mathrm{d}\alpha, \qquad (6.12)$$

qui grâce au théorème 2 tend vers, en $\mathbf{P}_{\theta_0}$-probabilité et pour $\theta = \theta_0$:

$$\Lambda_{ij} = \frac{1}{2}(2\pi)^d \int_{\Pi^d} \mathrm{tr}\left[(\mathrm{D}^2_{ij} f_{\theta_0}^{-1}(\alpha)) f_{\theta_0}(\alpha) + \mathrm{D}_i f_{\theta_0}^{-1}(\alpha)\mathrm{D}_j f_{\theta_0}(\alpha) + f_{\theta_0}^{-1}(\alpha)\mathrm{D}^2_{ij} f_{\theta_0}(\alpha)\right] \mathrm{d}\alpha. \qquad (6.13)$$

Utilisant les identités 2-3 ci-dessus, ce terme est égal à:

$$\Gamma_{ij}(\theta_0) = \frac{1}{2}(2\pi)^d \int_{\Pi^d} \mathrm{tr}\left[-\mathrm{D}_j f_{\theta_0}^{-1}(\alpha)\mathrm{D}_i f_{\theta_0}(\alpha)\right] \mathrm{d}\alpha.$$

C'est la propriété 2.

♠

Remarque. La matrice de covariance $\Gamma(\theta_0)^{-1} C(\theta_0) \Gamma(\theta_0)^{-1}$ prend une forme plus familière lorsque les valeurs (resp. valeurs limites) des $\{H_{abcd}, H_{ab}\}$ sont indépendantes des directions spectrales a, b, c, d. C'est le cas par exemple si les

[2] La première se démontre par vérification directe et les deux dernières par dérivation de la relation $AA^{-1} = Id$.

$\{w_a\}, 1 \leq a \leq r$ sont identiques et si les $\{\rho_a\}$ sont les mêmes (resp. tendent vers une même valeur limite). Alors, soit $\lambda = (H_{abcd}/(H_{ab}H_{cd}))^d$ (resp. sa valeur limite) pour tout $1 \leq a, b, c, d \leq r$, un calcul simple montre que

$$C(\theta_0) = \lambda \left(\Gamma(\theta_0) + B(\theta_0)\right), \tag{6.14}$$

avec

$$B_{ij}(\theta_0) = \frac{1}{4}(2\pi)^{-d} \sum_{a,b,c,d=1}^{r} \int_{\Pi^d}\int_{\Pi^d} \mathrm{D}_i f_{\theta_0}^{-1}(\alpha)_{ba} \overline{\mathrm{D}_j f_{\theta_0}^{-1}(-\beta)_{dc}} f_{\theta_0,abcd}(\alpha, -\alpha, \beta) \mathrm{d}\alpha \mathrm{d}\beta. \tag{6.15}$$

6.3 Sur un test d'hypothèse emboîtée

Jusqu'à présent, l'ensemble des paramètres $\theta = (\theta^1, \ldots, \theta^p)$ est un certain espace Θ de $\mathbb{R}^p$. Souvent on se demande si le paramètre θ_0 se trouve en fait dans un espace S plus petit. Nous allons construire un test asymptotique de l'hypothèse emboîtée "$\theta_0 \in S$" sur la base du contraste de Whittle raboté. Nous supposerons que S admet un paramétrage régulier suivant: $S = \psi(\Omega)$ avec $\Omega \subset \mathbb{R}^q$, $q < p$, et ψ une application dont le différentiel ($\alpha \in \Omega$)

$$R(\alpha) = \begin{bmatrix} \frac{\partial \psi_1}{\partial \alpha_1} & \cdots & \frac{\partial \psi_1}{\partial \alpha_q} \\ \vdots & \ddots & \vdots \\ \frac{\partial \psi_p}{\partial \alpha_1} & \cdots & \frac{\partial \psi_p}{\partial \alpha_q} \end{bmatrix} \tag{6.16}$$

existe sur Ω et est de rang plein égal à q. Avant de construire le test, considérons un exemple de S.

Exemple de S. Souvent l'hypothèse emboîtée se présente sous la forme de "$\varphi(\theta) = 0$" où φ est une application différentiable de Θ dans $\mathbb{R}^\nu$, $\nu < p$. C'est-à-dire,

$$S = \{\theta \in \Theta : \varphi(\theta) = 0\}.$$

Supposons qu'il existe des indices $1 \leq m_1 < m_2 < \cdots < m_\nu \leq p$ de sorte que le Jacobien partiel

$$\det\left[\mathrm{D}_{m_i}\varphi_j(\theta) \ , \ i, j = 1, \ldots, \nu\right]$$

soit non nul pour tout $\theta \in S$. Alors par le théorème des fonctions implicites, S admet un paramétrage régulier du type ci-dessus avec $q = p - \nu$. Le cas typique de cette situation est fourni par l'hypothèse (origine de l'adjectif "emboîtée")

$$\theta^1 = \theta^2 = \cdots = \theta^\nu = 0,$$

dont la fonction du paramétrage ψ s'identifie au plongement naturel de $\mathbb{R}^q$ dans

$\mathbb{R}^p$:

$$\alpha = \begin{pmatrix} \theta^{\nu+1} \\ \theta^{\nu+2} \\ \vdots \\ \theta^p \end{pmatrix} \mapsto \theta = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \theta^{\nu+1} \\ \vdots \\ \theta^p \end{pmatrix}.$$

Revenons sur le test lui-même. Il sera construit sur la base de la différence du contraste de Whittle

$$T_n = 2|n| \left(V_n(\bar{\theta}_n) - V_n(\hat{\theta}_n) \right), \tag{6.17}$$

où $\hat{\theta}_n$ est l'estimateur du minimum du contraste précédemment étudié et $\bar{\theta}_n$ celui évalué sous la contrainte "$\theta \in S$":

$$\bar{\theta}_n(\omega) = \operatorname{Arg} \min_{\theta \in S} V_n(\theta, \omega). \tag{6.18}$$

Evidemment si

$$\bar{\alpha}_n(\omega) = \operatorname{Arg} \min_{\alpha \in \Omega} V_n(\psi(\theta), \omega), \tag{6.19}$$

on a $\bar{\theta}_n = \psi(\bar{\alpha}_n)$.

En suivant une démarche classique (voir par exemple Amemiya[II–1]), nous allons étudier la loi asymptotique de $\{T_n\}$ en nous plaçant désormais dans le cadre du théorème 5. Ecrivons

$$T_n = 2|n| \left(V_n(\theta_0) - V_n(\hat{\theta}_n) \right) - 2|n| \left(V_n(\theta_0) - V_n(\bar{\theta}_n) \right).$$

Le développement de Taylor donne pour le premier terme,

$$2|n| \left(V_n(\theta_0) - V_n(\hat{\theta}_n) \right) = |n| \, {}^t(\theta_0 - \hat{\theta}_n) \left[D^2_{ij} V_n(\theta_0) \right] (\theta_0 - \hat{\theta}_n) + \delta_1 \ ,$$

avec

$$\delta_1 = |n| \, {}^t(\theta_0 - \hat{\theta}_n) \left[D^2_{ij} V_n(\theta_n^*) - D^2_{ij} V_n(\theta_0) \right] (\theta_0 - \hat{\theta}_n)$$

où θ_n^* appartient au segment joignant θ_0 et $\hat{\theta}_n$. Le fait que la matrice $\left[D^2_{ij} V_n(\theta_n^*) - D^2_{ij} V_n(\theta_0) \right]$ tende vers 0 en $\mathbf{P}_{\theta_0}$-probabilité (cf. la preuve du théorème 5) et que $\sqrt{|n|}(\theta_0 - \hat{\theta}_n)$ converge en loi implique que $\delta_1 \xrightarrow{\mathbf{P}_{\theta_0}} 0$. D'autre part (cf. encore la preuve du théorème 5),

$$\left[D^2_{ij} V_n(\theta_0) \right] \xrightarrow{\mathbf{P}_{\theta_0}} \Gamma(\theta_0)$$

et

$$\sqrt{|n|} \, (\theta_0 - \hat{\theta}_n) = -\sqrt{|n|} \, \Gamma(\theta_0)^{-1} \operatorname{grad} V_n(\theta_0) + \delta_2$$

avec $\delta_2 \xrightarrow{\mathbf{P}_{\theta_0}} 0$. Ainsi ,

$$2|n|\left(V_n(\theta_0) - V_n(\hat{\theta}_n)\right) = |n|\,{}^t\mathrm{grad}V_n(\theta_0)\Gamma(\theta_0)^{-1}\mathrm{grad}V_n(\theta_0) + \delta_3 \ , \tag{6.20}$$

avec $\delta_3 \xrightarrow{\mathbf{P}_{\theta_0}} 0$.

Maintenant, mettons-nous sous l'hypothèse "$\theta_0 \in S$". Asymptotiquement, on peut faire une étude similaire par rapport au paramètre α en considérant le contraste $\tilde{V}_n(\alpha) = V_n(\psi(\alpha))$. Par ailleurs, soit α_0 tel que $\theta_0 = \psi(\alpha_0)$ et $R_0 = R(\alpha_0)$. On obtient alors un résultat analogue à l'équation (4.20):

$$\begin{aligned} 2|n|\left(V_n(\theta_0) - V_n(\bar{\theta}_n)\right) &= 2|n|\left(\tilde{V}_n(\alpha_0) - \tilde{V}_n(\bar{\alpha}_n)\right) \\ &= |n|\,{}^t\mathrm{grad}\tilde{V}_n(\alpha_0)\Gamma(\alpha_0)^{-1}\mathrm{grad}\tilde{V}_n(\alpha_0) + \delta_4 \ , \end{aligned} \tag{6.21}$$

avec $\delta_4 \xrightarrow{\mathbf{P}_{\theta_0}} 0$ ($\mathbf{P}_{\alpha_0} = \mathbf{P}_{\theta_0}$). Or par définition,

$$\mathrm{grad}\tilde{V}_n(\alpha_0) = {}^tR_0\,\mathrm{grad}V_n(\theta_0), \tag{6.22}$$

et

$$\Gamma(\alpha_0) = {}^tR_0\,\Gamma(\theta_0)R_0 \ . \tag{6.23}$$

Avec les équations (4.20)-(4.21)-(4.22), on obtient

$$\begin{aligned} T_n &= |n|\,{}^t\mathrm{grad}V_n(\theta_0)\left[\Gamma(\theta_0)^{-1} - {}^tR_0\,\Gamma(\alpha_0)^{-1}\,R_0\right]\mathrm{grad}V_n(\theta_0) + \delta_5 \\ &= {}^t\xi_n P(\theta_0)\xi_n + \delta_5 \ , \end{aligned}$$

où $\delta_5 \xrightarrow{\mathbf{P}_{\theta_0}} 0$ avec

$$\xi_n = \sqrt{|n|}\,\Gamma(\theta_0)^{-\frac{1}{2}}\mathrm{grad}V_n(\theta_0) \ ,$$

et

$$P(\theta_0) = I_p - \Gamma(\theta_0)^{\frac{1}{2}}R_0\,\Gamma(\alpha_0)^{-1}\,{}^tR_0\,\Gamma(\theta_0)^{\frac{1}{2}}. \tag{6.24}$$

Remarquons que la relation (4.23) montre que la matrice $P(\theta_0)$ est idempotente de rang $(p-q)$.

D'autre part (cf. toujours la preuve du théorème 5), nous avons

$$\xi_n \xrightarrow{\mathcal{L}(\mathbf{P}_{\theta_0})} \mathcal{N}_p(0, Q(\theta_0)) \ ,$$

avec

$$Q(\theta_0) = \Gamma(\theta_0)^{-\frac{1}{2}}C(\theta_0)\Gamma(\theta_0)^{-\frac{1}{2}}. \tag{6.25}$$

Il en résulte que la loi asymptotique de la statistique T_n est celle d'une forme quadratique positive des variables normales:

$$T_n \xrightarrow{\mathcal{L}(\mathbf{P}_{\theta_0})} {}^t z\, P(\theta_0)\, z \ , \tag{6.26}$$

où z est un vecteur normal de loi $\mathcal{N}(0, Q(\theta_0))$. Finalement, suivant le Corrolaire 1 du Lemme 2 du Chapitre 2, nous venons de prouver la

Proposition 4 *On se place dans le cadre du théorème 5. Alors la statistique de la différence du contraste raboté*

$$T_n = 2|n| \left(V_n(\bar{\theta}_n) - V_n(\hat{\theta}_n) \right)$$

converge en loi, sous l'hypothèse emboîtée "$\theta_0 \in S$", vers celle d'un mélange de $\chi^2(1)$ pondérés

$$Z = \sum_{j=1}^{p-q} a_j \chi^2(1) \ , \tag{6.27}$$

où les $\chi^2(1)$ sont indépendants et les $\{a_j\}$ sont les $(p-q)$ valeurs propres non nulles (non nécessairement distinctes) de la matrice $P(\theta_0)Q(\theta_0)$. $Q(\theta_0)$ (équation 4.25) est définie positive; $P(\theta_0)$ (équation 4.24) est une matrice idempotente de rang $(p-q)$.

Ce résultat permet de construire un test asymptotique effectif d'une hypothèse emboîtée. En effet, il existe des algorithmes numériques efficaces permettant d'établir, pour un mélange de $\chi^2(1)$ pondérés par des coefficients positifs, la fonction de répartition et à fortiori les divers quantiles (cf. Shah [II–14]). De plus, ces quantités sont tabulées lorsque le nombre des coefficients $\{a_j\}$ (ici $p-q$) ne dépasse pas 10. Evidemment, ce test nécessite l'étude des sous-espaces propres associés aux valeurs propres $\{a_j\}$. En pratique, cette étude se scindera en deux étapes: l'estimateur $\bar{\theta}_n$ est calculé sous la contrainte "$\theta_0 \in S$"; ensuite on en déduit une estimation des valeurs propres $\{a_j = a_j(\bar{\theta}_n)\}$.

Whittle[II–17] a introduit initialement ce test pour des processus spatiaux gaussiens et établi la convergence asymptotique du test vers un $\chi^2(p-q)$. Il s'est aussi demandé si cette limite aymptotique de $\chi^2(p-q)$ subsiste lorsque le processus n'est plus gaussien. Notre résultat fournit une réponse complète à cette question: généralement, la loi limite de la différence du contraste n'est pas un $\chi^2(p-q)$ mais un mélange de $\chi^2(1)$ pondérés par des coefficients positifs. Guyon[II–11] a étendu ce test aux processus spatiaux $(d \geq 2)$ gaussiens ou linéaires et abouti dans ce cas aussi à un test asymptotique de $\chi^2(p-q)$.

Notre résultat est plus général. Celui-ci opère une généralisation en deux directions:

1. il s'applique à tous les processus spatiaux (gaussiens ou non, linéaires ou non) pourvu qu'ils remplissent les conditions du théorème 5;

2. notre test utilise des estimations par rabotage.

D'autre part, l'analyse développée restant à la fois classique et générale, notre résultat peut être encore généralisé à un contraste quelconque si l'estimateur du minimum de contraste correspondant possède un théorème de limite centrale semblable à celui du théorème 5.

Comme illustration de la proposition ci-dessus, nous allons considérer 2 situations où le test est (sans surprise parfois) un χ^2 asymptotique.

Première situation (Whittle[II–17]): le processus est gaussien et asymptotiquement, les rabots $\{h_a\}$ sont spectralement identiques (cf. la remarque juste après le théorème 5 et les équations (4.14)-(4.15)). Comme les cumulants d'ordre 4 sont nuls, il en est de même pour leurs fonctions spectrales génératrices $\{f_{\theta_0,abcd}, 1 \leq a,b,c,d \leq r\}$. Donc $B(\theta_0) = 0$ et $C(\theta_0) = \lambda\Gamma(\theta_0)$ avec $\lambda = (H_{abcd}/(H_{ab}H_{cd}))^{\frac{1}{d}}$. Ainsi $Q(\theta_0) = \lambda I_p$ et $P(\theta_0)Q(\theta_0) = \lambda P(\theta_0)$. D'où $a_j \equiv \lambda$, $j = 1,\ldots,p-q$ et par conséquent $\lambda^{-1}T_n$ converge en loi vers un $\chi^2(p-q)$.

Deuxième situation (Voir aussi Guyon[II–11]): le processus est linéaire (nous décrivons le cas scalaire i.e. $r = 1$, par simplicité) :

$$X(t) = \sum_{s\in\mathbb{Z}^d} b_s\eta(t-s), \tag{6.28}$$

où $\{\eta(t)\}$ est un bruit blanc de variance 1. Les densités spectrales $\{f_\theta\}$ sont paramétrées de la façon suivante:

$$\theta = (\phi, \sigma^2), \tag{6.29}$$

et

$$f_\theta(\beta) = \sigma^2 g_\phi(\beta), \tag{6.30}$$

où $\theta^p = \sigma^2$ est *le paramètre d'échelle* (qui est aussi la variance d'une innovation associée par rapport à un "demi-plan") défini par la normalisation

$$\log\sigma^2 = \frac{1}{(2\pi)^d}\int_{\Pi^d} \log f_\theta(\beta)\mathrm{d}\beta. \tag{6.31}$$

$\phi = (\theta^1,\ldots,\theta^{p-1})$ s'appelle parfois *le paramètre de structure*, car un changement d'échelle le laisse invariant. Nous allons montrer que si le test d'adéquation du modèle "$\theta_0 \in S$" n'affecte pas σ^2, c-à-d porte uniquement sur ϕ, il est alors un χ^2 asymptotique. Une conséquence de la normalisation précédente est

$$\int_{\Pi^d} \log g(\beta)\mathrm{d}\beta = 0 \ .$$

Les matrices $B(\theta_0)$ (cf. l'équation (4.15)) et $\Gamma(\theta_0)$ sont, avec $\theta_0 = (\phi_0, \sigma_0^2)$:

$$B(\theta_0) = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{4}\kappa_4(\theta_0)\sigma_0^{-4} \end{bmatrix} , \quad \Gamma(\theta_0) = \begin{bmatrix} \Lambda & 0 \\ 0 & \frac{1}{2}\sigma_0^{-4} \end{bmatrix} ,$$

où $\kappa_4(\theta_0)$ est le cumulant d'ordre 4 de l'innovation et Λ une matrice $(p-1)\times(p-1)$ avec

$$\Lambda_{ij} = \frac{1}{2}(2\pi)^{-d}\int_{\Pi^d} \mathrm{D}_i \log g(\beta)\mathrm{D}_j \log g(\beta)\mathrm{d}\beta , \quad 1 \le i,j \le p-1.$$

Alors (cf. équation (4.25))

$$\begin{aligned} Q(\theta_0) &= \Gamma(\theta_0)^{-\frac{1}{2}} [\lambda(\Gamma(\theta_0 + B(\theta_0))] \Gamma(\theta_0)^{-\frac{1}{2}} \\ &= \lambda \left[I_p + \Gamma(\theta_0)^{-\frac{1}{2}} B(\theta_0) \Gamma(\theta_0)^{-\frac{1}{2}} \right] \\ &= \lambda \begin{bmatrix} I_{p-1} & 0 \\ 0 & 1 + \frac{1}{2}\kappa_4(\theta_0) \end{bmatrix} , \end{aligned}$$

avec $\lambda = \lim \left[\int_0^1 h^4(x)\mathrm{d}x / (\int_0^1 h^2(x)\mathrm{d}x)^2 \right]^d$.

La dernière composante α_q de α étant égale à σ^2, le Jacobien R_0 prend une forme diagonale par bloc suivante:

$$R_0 = \begin{bmatrix} W & 0 \\ 0 & 1 \end{bmatrix},$$

avec une matrice W appropriée. Ainsi, on trouve

$$P(\theta_0) = \begin{bmatrix} I_{p-1} - \Lambda^{\frac{1}{2}} W ({}^t W \Lambda W)^{-1} {}^t W \Lambda^{\frac{1}{2}} & 0 \\ 0 & 0 \end{bmatrix}.$$

Ici encore, on a $P(\theta_0)Q(\theta_0) = \lambda P(\theta_0)$. En conséquence, $\lambda^{-1}T_n$ converge en loi vers un $\chi^2(p-q)$.

Examinons maintenant le cas où le processus est r-vectoriel:

$$X(t) = \sum_{s \in \mathbb{Z}^d} B_s \eta(t-s), \tag{6.32}$$

où $\{\eta(t)\}$ est un bruit blanc r-vectoriel de matrice de covariance I_r. Si G est la matrice de covariance d'une innovation associée, nous avons

$$\log\det G = \frac{1}{(2\pi)^d}\int_{\Pi^d} \log\det f_\theta(\beta)\mathrm{d}\beta. \tag{6.33}$$

Alors, un raisonnement similaire au cas scalaire conduit encore à un test de $\chi^2(p-q)$ asymptotique. Nous résumons les résultats de cette deuxième situation par le

Corollaire 2 *On se place encore dans le cadre du théorème 5. Supposons en plus que le processus $\{X(t)\}$ est linéaire avec G la matrice de covariance d'une innovation associée. Si l'hypothèse emboîtée "$\theta_0 \in S''$ n'affecte pas G et si les rabots $\{h_a\}$ sont asymptotiquement identiques, alors la statistique de la différence du contraste raboté*

$$\lambda^{-1}T_n = 2\lambda^{-1}|n|\left(V_n(\bar{\theta}_n) - V_n(\hat{\theta}_n)\right)$$

converge en loi, sous l'hypothèse "$\theta_0 \in S$", vers celle d'un $\chi^2(p-q)$ avec $\lambda = (H_{abcd}/(H_{ab}H_{cd}))^d$.

6.4 Application aux champs de Markov gaussiens sur Z^d

Un champ markovien (MRF) gaussien stationnaire est défini par

$$\begin{aligned} \mathbf{E}\left(X(t)|X(s), s \neq t\right) &= \sum_{s\in V} \beta_s X(t+s), \qquad (6.34) \\ \mathbf{Var}\left(X(t) - \sum_{s\in V} \beta_s X(t+s)\right) &= \Lambda, \end{aligned}$$

pour tout $t \in \mathbb{Z}^d$ avec Λ une matrice de covariance de rang plein r. La moyenne du champ est prise égale à 0. Les $\{\beta_s\}$ sont des matrices réelles et V un ensemble des voisins de l'origine possédant la symétrie suivante:

$$t \in V \Longleftrightarrow -t \in V. \qquad (6.35)$$

Si un tel processus existe, la densité spectrale est de la forme:

$$f(\alpha) = \left(I_r - \sum_{s\in V} \beta_s e^{i<\alpha,s>}\right)^{-1} \Lambda. \qquad (6.36)$$

Le théorème ci-dessous qui assure l'existence des MRF gaussiens est dû à Mardia [II–13]. Fixons d'abord une norme matricielle. Si A est une matrice réelle quelconque,

$$\|A\| = \left\{ \text{ la plus grande valeur propre de } \ {}^tAA \right\}^{1/2}.$$

Théorème 6 (Mardia [II–13])
Supposons que les $\{\beta_s, s \in V\}$ et Λ vérifient

1. $\beta_s \Lambda = \Lambda {}^t\beta_{-s}$,
2. $\sum_{s\in V} ||\Lambda^{-1/2}\beta_s\Lambda^{1/2}|| < 1$,

où $\Lambda^{1/2}$ est la racine carrée de Λ. Alors, il existe un MRF gaussien stationnaire défini par (4.34) avec la densité spectrale (4.36).

La première condition exprime en fait une "symétrie" des coéfficients $\{\beta_s\}$ qui est nécessaire pour que la densité spectrale soit hermitienne.

Ces deux conditions se réduisent à une forme plus simple lorsque le processus possède une structure de covariance factorisante (cf. chapitre 2), i.e.

$$R(u) = \xi(u)R(0), \quad u \in \mathbb{Z}^d, \tag{6.37}$$

où ξ est une fonction réelle de corrélation, parfois appelée *spatiale*, vérifiant $\xi(0) = 1$. Il est utile de noter que dans le cas d'un MRF gaussien, la factorisation ci-dessus est caractérisée par la proportionnalité suivante des $\{\beta_s\}$:

$$\beta_s = b_s I_r, \quad s \in V,$$

où les $\{b_s\}$ sont des scalaires réels. Alors, les conditions dans le théorème 6 ci-dessus deviennent:

1. $b_s = b_{-s}, \quad s \in V,$
2. $\sum_{s \in V} |b_s| < 1.$

En particulier, on y retrouve une condition suffisante d'existence pour des processus scalaires ($r = 1$).

Nous allons développer les estimateurs de Whittle pour les MRF gaussiens remplissant les deux conditions du théorème 6.

Les coefficients de Fourier $\{A_s\}$ de l'inverse de la densité spectrale, $f^{-1}(\alpha)$ sont (cf. équation (4.36)):

$$A_s = \begin{cases} \Lambda^{-1}, & \text{si } s = 0 \\ -\Lambda^{-1}\beta_s, & \text{si } s \in V \\ 0, & \text{sinon.} \end{cases} \tag{6.38}$$

La "symétrie" des $\{\beta_s\}$ devient pour les $\{A_s\}$:

$$A_s = {}^t A_{-s}, \quad s \in V. \tag{6.39}$$

Ceci nous conduit à une paramétrisation naturelle du modèle en choisissant une moitié V^+ de V, i.e. $V =$ la réunion disjointe de V^+ et $\{-V^+\}$, et en prenant le paramètre

$$\theta = (A_s, s \in V^+ \cup \{0\}), \tag{6.40}$$

et son espace

$$\Theta = \left\{ \theta : \sum_{s \in V^+} ||A_0^{-1/2} A_s A_0^{-1/2}|| < \frac{1}{2} \quad \text{et } A_0 \text{ définie positive} \right\}. \tag{6.41}$$

Ici la dimension paramétrique vaut $|V^+|r^2+r(r+1)/2$. En fonction du paramètre θ, nous avons

$$f^{-1}(\alpha) = A_0 + \sum_{s \in V^+} \left(A_s e^{i<\alpha,s>} + {}^t A_s e^{-i<\alpha,s>} \right).$$

Rappelons la forme du contraste V_n:

$$V_n(\theta,\omega) = \frac{1}{2}(2\pi)^d \int_{\Pi^d} \left(\log\det\left[f_\theta(\alpha)\right] + \mathrm{tr}\left[f_\theta^{-1}(\alpha) I^n(\alpha,\omega)\right] \right) \mathrm{d}\alpha + \varpi(\theta_0,\omega),$$

où $\varpi(\theta_0,\omega)$ est indépendant de θ. Par l'identité de Parseval, le terme

$$\int_{\Pi^d} \mathrm{tr}\left[f_\theta^{-1}(\alpha) I^n(\alpha,\omega)\right] \mathrm{d}\alpha,$$

est égal à

$$\mathrm{tr}\left[\sum_{s \in V \cup \{0\}} A_s C^n(s) \right] = \mathrm{tr}\left[A_0 C^n(0) + 2 \sum_{s \in V^+} A_s C^n(s) \right],$$

où les $\{C^n(s)\}$ sont des estimateurs par rabotage des covariances $\{R_\theta(s)\}$. Notons, pour une fonction réelle ϕ dépendant d'une variable matricielle $M = [M_{ij}]$, la dérivée partielle

$$\frac{\partial \phi}{\partial M} = \left[\frac{\partial \phi}{\partial M_{ij}} \right].$$

Des calculs simples montrent que

$$2(2\pi)^d \frac{\partial V_n(\theta,\omega)}{\partial A_0} = 2C^n(0) - \mathrm{diag}C^n(0) - (2R_\theta(0) - \mathrm{diag}R_\theta(0)), \tag{6.42}$$

$$2(2\pi)^d \frac{\partial V_n(\theta,\omega)}{\partial A_s} = 2C^n(s) - 2R_\theta(s), \quad s \in V^+, \tag{6.43}$$

où $\mathrm{diag}B$ est la matrice des élements diagonaux d'une matrice B. En annulant ces dérivées, nous arrivons à la

Proposition 5 *Pour des MRF gaussiens, l'estimateur par rabotage de Whittle*

$$\hat{\theta}_n = (\hat{A}_s, \quad s \in V^+ \cup \{0\}) \tag{6.44}$$

est obtenu en résolvant le système d'équations suivant

$$R_\theta(s) = C^n(s), \quad s \in V^+ \cup \{0\}. \tag{6.45}$$

Remarque. Le cas vectoriel confirme le cas scalaire: les covariances estimées doivent être égales à des covariances théoriques correspondant au paramètre estimé.

Il importe de noter que, la forme des équations ci-dessus change lorsque la paramétrisation du modèle est modifiée. Comme exemple, nous allons examiner le cas intéressant des processus factorisants. Ici, $\beta_s = b_s I_r, s \in V$ et $b_s = b_{-s}$. On peut paramétrer le modèle comme suit:

$$\theta = (b, \Lambda) \ , \ \text{avec} \ b = (b_s, s \in V^+), \tag{6.46}$$

et l'espace des paramètres

$$\Theta = \left\{ b, \Lambda : \Lambda \text{ définie positive et } \sum_{s \in V^+} |b_s| < \frac{1}{2} \right\}. \tag{6.47}$$

La dimension paramétrique est réduite, par rapport au modèle général (4.34), à $|V^+| + r(r+1)/2$.

Si le processus est factorisant, la densité spectrale aussi se factorise:

$$f_\theta(\alpha) = g_b(\alpha)\Lambda, \tag{6.48}$$

avec

$$g_b(\alpha) = \left(1 - 2 \sum_{s \in V^+} b_s \cos < \alpha, s > \right)^{-1}. \tag{6.49}$$

Alors les covariances $\{R_\theta(s)\}$ se factorisent en

$$R_\theta(s) = \gamma_b(s)\Lambda, \quad s \in \mathbb{Z}^d, \tag{6.50}$$

avec

$$\gamma_b(s) = \int_{\Pi^d} g_b(\alpha) e^{-i<\alpha,s>} \mathrm{d}\alpha. \tag{6.51}$$

On déduit directement des équations (4.42)-(4.43), le

Corollaire 3 *Pour des MRF gaussiens factorisant, l'estimateur par rabotage de Whittle*

$$\hat{\theta}_n = (\hat{b}, \ \hat{\Lambda}) \tag{6.52}$$

est obtenu en résolvant le système d'équations suivant

$$\begin{cases} \gamma_b(s)/\gamma_b(0) &= r^{-1} \operatorname{tr} [C^n(s)/C^n(0)], \quad s \in V^+, \\ \Lambda &= C^n(0)/\gamma_b(0). \end{cases} \tag{6.53}$$

A ce stade, on peut construire pour des MRF gaussiens un test de factorisation du modèle. Si $\hat{\theta}_n = (\hat{A}_s,\ s \in V^+ \cup \{0\})$ est l'estimateur solution du système (4.45) et $\hat{\alpha}_n = (\hat{b},\ \hat{\Lambda})$ l'estimateur solution du système (4.53), c-à-d sous l'hypothèse de factorisation, alors la statistique de la différence du contraste

$$\lambda^{-1}|n|\left(V_n(\psi(\hat{\alpha}_n)) - V_n(\hat{\theta}_n)\right),$$

avec

$$\psi(\hat{\alpha}_n) = (-\hat{b}_s\hat{\Lambda}^{-1}, s \in V^+ \cup \{0\}, \hat{\Lambda}^{-1}),$$

converge en loi vers un χ^2 avec le degré de liberté égal à $(r^2-1)|V^+|$. La constante λ vaut toujours $(H_{abcd}/(H_{ab}H_{cd}))^d$ en supposant que les rabots sont asymptotiquement identiques.

Chapitre 7

Etude par simulation de l'estimateur de Whittle raboté

7.1 Modèle d'expérimentation

Dans ce chapitre, nous proposons une étude de simulation expérimentant l'estimateur de Whittle par rabotage. Le premier problème auquel on est confronté relève du choix d'un modèle. L'évaluation de l'estimateur de Whittle nécessite en général des algorithmes numériques itératifs du type de Newton-Raphson ou de gradient conjugué. Le choix d'une "bonne" estimation initiale pour ces algorithmes itératifs joue un rôle important (nous renvoyons sur ce sujet à Dzhaparidze &. Yaglom [II–9]). Cette subtilité nous a incité à prendre comme modèle des ARU, i.e. processus autoregressifs unilatéraux (par rapport à un "demi-plan") pour lesquels l'évaluation de l'estimateur de Whittle se réalise par un calcul direct. Notre propos principal étant la comparaison de l'estimation par rabotage à celle de non rabotage, ce choix du modèle de simulation nous permet ainsi d'éviter complètement les erreurs que pourraient causer un algorithme numérique itératif.

Plus précisément, des processus ARU plans ($d = 2$) et scalaires ($r = 1$) ont été considérés:

$$\sum_{s \in L} a_s X(t-s) = \varepsilon(t), \quad t \in \mathbb{Z}^2, \tag{7.1}$$

où

$$L = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} = \{s_0, s_1, s_2, s_3\}, \tag{7.2}$$

avec $\{\varepsilon(t)\}$ un bruit blanc de variance σ^2 et $a_0 \equiv 1$ (dans tout ce chapitre, $a_{s_j}, s_j \in L, j = 0,1,2,3$ sont parfois notés simplement par $a_j, j = 0,1,2,3$). Les

paramètres du modèle sont au nombre de 4, $\theta = (a_1, a_2, a_3, \sigma^2)$. La condition

$$|a_1| + |a_2| + |a_3| < 1, \tag{7.3}$$

étant suffisante pour garantir l'existence et la régularité de ces processus, l'espace des paramètres Θ est fixé à

$$\Theta = \left\{\theta : |a_1| + |a_2| + |a_3| < 1 \text{ et } \sigma^2 > 0\right\}. \tag{7.4}$$

Les simulations sont menées sur des carrés $P_n = [1, m] \times [1, m]$. La proposition suivante donne l'évaluation directe de l'estimateur de Whittle. Quelques nouvelles notations y sont introduites (où les $\{C^n(s)\}$ sont les estimateurs par rabotage des covariances $\{R_\theta(s)\}$) :

$$C_j = C^n(s_j), \quad (s_j \in L \quad \text{pour} \quad j = 0,1,2,3) \quad \text{et} \quad s_4 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \tag{7.5}$$

et

$$a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, \quad v = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix}, \quad W_4 = \begin{bmatrix} C_0 & {}^t v \\ v & W_3 \end{bmatrix}, \quad W_3 = \begin{bmatrix} C_0 & C_4 & C_2 \\ C_4 & C_0 & C_1 \\ C_2 & C_1 & C_0 \end{bmatrix}. \tag{7.6}$$

Proposition 6 *Pour un processus ARU (1.1), l'estimateur de Whittle*

$$\hat{\theta}_n = (\hat{a}, \hat{\sigma}^2),$$

est donné par

$$\begin{cases} \hat{a} &= -W_3^{-1} v \\ \hat{\sigma}^2 &= C_0 + \sum_{j=1}^{3} C_j \hat{a}_j \end{cases}. \tag{7.7}$$

Preuve. La densité spectrale d'un ARU (1.1) s'écrit:

$$f_\theta(\alpha) = \frac{\sigma^2}{(2\pi)^2 g_a(\alpha)}, \quad \alpha \in \Pi^2,$$

où $g_a(\alpha)$ est le polynôme trigonométrique positif suivant

$$g_a(\alpha) = \left| \sum_{s \in L} a_s e^{i<\alpha, s>} \right|^2.$$

Le contraste de Whittle

$$V_n(\theta, \omega) = (8\pi^2)^{-1} \int_{\Pi^2} [\log f_\theta(\alpha) + I^n(\alpha)/f_\theta(\alpha)] \, d\alpha,$$

vaut

$$V_n(\theta,\omega) = \frac{1}{2}\left[\log\sigma^2 + \frac{1}{\sigma^2}\sum_s C^n(s)\,\hat{g}_a(s)\right], \tag{7.8}$$

où $\{\hat{g}_a(s)\}$ sont les coefficients de Fourier de g_a qui valent

$$\hat{g}_a(s) = \sum_{t,t+s\in L} a_t a_{t+s}.$$

$\hat{g}_a(s) = 0, s = (s_1,s_2)$ si $\max(|s_1|,|s_2|) > 1$ et $\hat{g}_a(s) = \hat{g}_a(-s)$. L'équation (5.8) montre que la minimisation de $V_n(\theta,\omega)$ en $\theta = (a,\sigma^2)$ se réalise séparément en a et σ^2:

$$\hat{a} = \mathrm{Arg}\min_a \sum_s C^n(s)\,\hat{g}_a(s), \tag{7.9}$$

et

$$\hat{\sigma}^2 = \sum_s C^n(s)\,\hat{g}_{\hat{a}}(s). \tag{7.10}$$

Un calcul simple montre que

$$\sum_s C^n(s)\,\hat{g}_a(s) = [1, {}^t a]\, W_4 \begin{bmatrix} 1 \\ a \end{bmatrix},$$

est une forme quadratique positive en a. Le résultat annoncé en découle immédiatement par sa minimisation en a.

7.2 Simulation et résultats

Nous avons choisi un rabot "sinusoïdal":

$$h(u) := \begin{cases} \sin(\frac{\pi u}{\rho}) & 0 \le u < \frac{\rho}{2} \\ 1 & \frac{\rho}{2} \le u \le \frac{1}{2} \\ h(1-u) & \frac{1}{2} < u \le 1 \end{cases}. \tag{7.11}$$

La Figure 5.1 donne l'allure de h pour un degré de rabotage $\rho = 50\%$. A titre de référence, l'un des rabots historiques, le " cosine bell taper" de Tukey[II–15] vaut le carré du nôtre.

D'autre part, 3 jeux de paramètres (vraies valeurs) ont été élaborés:

jeu A: $\theta_0 = (0.4,\ -0.4,\ -0.15,\ 1)$

jeu B: $\theta_0 = (0.7,\ 0.15,\ 0.1,\ 1)$

jeu C: $\theta_0 = (0.6,\ 0.3,\ 0.15,\ 1)$

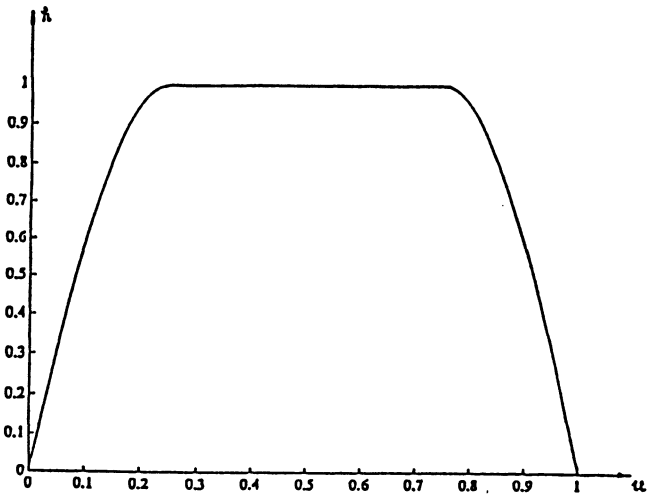


Figure 7.1: Allure du rabot sinusõidal: $\rho = 50$ %.

La variance σ^2 reste toujours égale à 1 afin de fixer l'échelle. Pour chaque jeu de θ_0, 100 simulations indépendantes sont menées sur 2 types de réseaux carrés P_n respectivement de taille 16×16 et 64×64. Le biais b et l'écart-type v de chaque estimateur, avec rabotage ($\rho \neq 0$) ou non ($\rho = 0$), sont évalués à partir de ces 100 simulations. Aussi, nous avons fixé systématiquement le degré de rabotage $\rho = 50\%$.

La Table 5.1 rassemble en 2 blocs verticaux, pour les différentes estimations, le biais b et l'erreur quadratique moyenne $e = \sqrt{b^2 + v^2}$ où v est l'écart-type correspondant. Horizontalement, les trois blocs représentent des résultats respectivent pour les 3 jeux différents des vraies valeurs de θ_0 que nous avons fixées. Enfin, la taille du réseau carré vaut ici 16×16 et tous les chiffres rapportés sont égaux à 1000 fois leur vraie valeur.

De cette expérimentation, on peut constater que

1. le rabotage permet une réduction importante du biais: ici pour le réseau 16×16, le gain varie de 88% ($\hat{\alpha}$, jeu A) à 44% ($\hat{\alpha}$, jeu C). Voici une confirmation du résultat obtenu précédemment (théorème 1) sur la réduction du biais par rabotage.

2. le rabotage augmente par ailleurs sensiblement l'écart-type de l'estimateur par rapport au non rabotage. L'erreur quadratique moyenne des estimateurs rabotés, tout en se trouvant à un niveau comparable à celui des estimateurs non rabotés, montre, la plupart du temps, une légère réduction.

	Biais b					erreur quadratique e				
	paramètre	a_1	a_2	a_3	σ^2	paramètre	a_1	a_2	a_3	σ^2
jeu A	*vraie valeur*	*400*	*-400*	*150*	*1000*	*vraie valeur*	*400*	*-400*	*150*	*1000*
	$\rho = 0$	32	-22	-98	177	$\rho = 0$	62	58	117	221
	$\rho = 50\%$	4	3	-33	55	$\rho = 50\%$	60	59	70	132
jeu B	*vraie valeur*	*700*	*150*	*100*	*1000*	*vraie valeur*	*700*	*150*	*100*	*1000*
	$\rho = 0$	-52	-13	-14	105	$\rho = 0$	73	63	60	146
	$\rho = 50\%$	-27	-3	-5	25	$\rho = 50\%$	58	68	59	116
jeu C	*vraie valeur*	*600*	*300*	*150*	*1000*	*vraie valeur*	*600*	*300*	*150*	*1000*
	$\rho = 0$	-39	-23	-14	82	$\rho = 0$	60	63	57	125
	$\rho = 50\%$	-22	-13	-7	19	$\rho = 50\%$	60	70	70	105

Table 7.1: Biais et Erreur Quadratique Moyenne. Réseau 16 × 16.

La Table 1.2 donne des résultats du même type mais pour un réseau de 64 × 64. L'ampleur de la réduction des biais par rabotage se trouve ici sensiblement amplifiée (la veleur 0 peut apparaître par suite de l'arrondi au millième). D'autre part, une diminution de l'erreur quadratique moyenne des estimateurs rabotés apparaît plus nette que sur le réseau précédent.

Cette étude de simulation prouve que l'emploi du rabotage permet d'avoir des estimateurs moins biaisés. Bien que l'écart-type se trouve en même temps augmenté par rapport au cas non raboté, l'erreur quadratique moyenne qui en résulte est généralement plus faible. Cet avantage s'amplifie lorsque le réseau devient plus grand.

	Biais b					erreur quadratique e				
	paramètre	a_1	a_2	a_3	σ^2	paramètre	a_1	a_2	a_3	σ^2
jeu A	*vraie valeur*	*400*	*-400*	*150*	*1000*	*vraie valeur*	*400*	*-400*	*150*	*1000*
	$\rho = 0$	14	-12	-33	51	$\rho = 0$	21	18	36	56
	$\rho = 50\%$	3	0	-4	4	$\rho = 50\%$	17	16	20	27
jeu B	*vraie valeur*	*700*	*150*	*100*	*1000*	*vraie valeur*	*700*	*150*	*100*	*1000*
	$\rho = 0$	-11	-3	-4	28	$\rho = 0$	16	16	16	35
	$\rho = 50\%$	-2	1	0	2	$\rho = 50\%$	13	20	21	26
jeu C	*vraie valeur*	*600*	*300*	*150*	*1000*	*vraie valeur*	*600*	*300*	*150*	*1000*
	$\rho = 0$	-11	-2	-3	20	$\rho = 0$	17	15	17	30
	$\rho = 50\%$	-2	1	0	2	$\rho = 50\%$	15	18	20	29

Table 7.2: Biais et Erreur Quadratique Moyenne. Réseau 64×64.

Références bibliographiques

I. Segmentation d'image

[I–1] ABEND K., HARLEY T.J.,& KANAL L.N., 1965. Classification of binary random patterns. *IEEE Trans. Inform. Th.* **11**, 538-544

[I–2] BESAG J.E., 1974. Spatial interaction and the statistical analysis of lattice systems. *J. R. Statist. Soc.* **B-36**, 192-236

[I–3] BESAG J.E., 1986. On the statistical analysis of dirty pictures. *J. R. Statist. Soc.* **B-48**, 259-302

[I–4] CHALMOND B., 1988. Image restauration using an estimed Markov model. *Signal Processing* **15**, 115-129.

[I–5] CHALMOND B., 1989. An iterative Gibbsian technique for reconstruction of M-ary images. *Pattern Recognition* **22** (6).

[I–6] CHALMOND B., 1988. *Reconstruction et restauration d'image: utilisation d'outils stochastiques.* Thèse de Docteur en Sciences, Université de Paris – Sud.

[I–7] CHOW C.K., 1962. A recognition method using neighbor dependence. *IRE Trans. Electronic Computers* **11**, 683-690

[I–8] COHEN F.S. & COOPER D.B., 1987. Simple parallel hierarchical and relaxation algorithmes for segmenting noncausal markovian random fields. *IEEE Trans. on Pattern Anal. and Machine Intell.* **9**, 195-219.

[I–9] CROSS & JAIN, 1983. Markov random field texture models. *IEEE Trans. on Pattern Anal. and Machine Intell.* **5** (1), 25-39.

[I–10] DERIN H. & ELLIOT H., 1987. Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Trans. on Pattern Anal. and Machine Intell.* **9**, 39-55.

[I–11] DERIN H., ELLIOT H., CHRISTI R. & GEMAN D., 1984. Bayes smoothing algorithms for segmentation of binary images modelled by Markov random fields. *IEEE Trans. on Pattern Anal. and Machine Intell.* **6**, 707-720.

[I–12] DEVIJVER P.A. & DEKESEL M.M., 1987. Learning the parameters of a hidden Markov random field image model: a simple example. Dans *Pattern Recognition: Theory and Applications* (eds. DEVIJVER P.A. & KITTLER J.). NATO ASI Series **F30**, 141-163. Springer-Verlag.

[I–13] DEVIJVER P.A., 1989. Real–time modeling of image sequences — based on Hidden markov mesh random field models. *Manuscript* **M–307**, Philips Research Laboratory Brussels.

[I–14] DINTEN J.M., 1988. Tomographic reconstruction with a limited number of projections: regularization using a Markov model. Prépublication **88–42**, Université de Paris-Sud et soumis à la publication.

[I–15] DINTEN J. M., GUYON X. & YAO J. F. 1988. On the choice of the regularization parameter: the case of binary images in the Bayesian restoration framework. *Proc. AMS-IMS-SIAM Joint Conference on Spatial Statistics and Imaging*, à paraître dans: Lectures Notes — Monograph Series of Inst. Math. Stat. (éditeur SERFLING R.).

[I–16] GEMAN S. & GEMAN D., 1984. Stochastic relaxation, Gibbs distributions and the bayésian restauration of images. *IEEE Trans. on Pattern Anal. and Machine Intell.* **6**, 721-741.

[I–17] GEMAN S. & GRAFFIGNE C., 1986. Markov random fields image models and their applications to computer vision. *Proc. Int. Congr. of Math. 1986*, AMS: Providence.

[I–18] GEMAN S., GEMAN D. & GRAFFIGNE C., 1986. Locating textures and objets boundries. Dans *Pattern Recognition: Theory and Applications* (eds. DEVIJVER P.A. & KITTLER J.). NATO ASI Series **F30**. Springer-Verlag.

[I–19] GEMAN S. & MCCLURE , 1987. Statistical methods for tomographic image reconstruction. Dans *Proceedings of the 46th Sessions of the International Statistical Institute*, Bulletin of the ISI **52**, 1-17.

[I–20] GRAFFIGNE C., 1987. Experiments in textures analysis and segmentation. *Ph. D. Dissertation*, Brown University.

[I–21] GUYON X. & YAO J.F., 1987. Analyse Discriminante Contextuelle. Dans *Actes des 5-ièmes Journées Internationales en Analyse des Données et Informatiques* (INRIA, France), North Holland

[I–22] HAJEK B., 1985. Cooling Schedules for optimal annealing. *Mathematics of Operation Research* **13** (2), 311-329.

[I–23] HASLETT J., 1985. Maximum likelihood discriminant analysis on the plane using a Markovian model of spatial context. *Pattern Recognition* **18**, 287-296.

[I–24] HASSNER M. & SKLANSKY J., 1980. The use of Markov random fields as models of texture. *Computer Graphics and Image Processing* **12**, 357-370.

[I–25] HJORT N.L., 1985. Neighbourhood based classification of remotely sensed data based on geometric probability models. *Technical Reporty* **10**, Stanford University.

[I–26] KHATRI C.G., 1980. Quadratic forms in normal variables. Dans *Handbook of Statistics* (ed. KRISHNAIAH P.R.), Volume **1** 443-469. North-Holland.

[I–27] KIRPATRICK S., GELATT C.D. & VECCHI M.P., 1983. Optimization by simulated annealing. *Science* **220**, 671-680.

[I–28] KITTLER J. & FÖGLEIN J., 1984. Contextual classification of multispectral pixel data. *Image and Vision Computing Journal* **2**, 13-39.

[I–29] KOTZ S., JOHANSON N.L. & BOYD D.W., 1967. Series representations of quadratic forms in normal variables: *I.* Central case. *Ann. Math. Statist.* **38**, 823-837.

[I–30] KOTZ S., JOHANSON N.L. & BOYD D.W., 1967. Series representations of quadratic forms in normal variables: *II.* Non central case. *Ann. Math. Statist.* **38**, 838-848

[I–31] MARDIA K.V., 1984. Spatial discrimination and classification maps. *Commn. Stat. Theor. Math.* **13** (18), 2184-2197.

[I–32] MARROQUIN J., MITTER S. & POGGIO T., 1987. Probabilistic solution of ill-posed problems in computational vision. *J. Amer. Stat. Ass.* **82** (397), 76-89.

[I–33] OWEN A., 1984. A neighbourhood-based classifier for LANDSAT data. *Can. J. Stat.* **12**, 191-200.

[I–34] PICKARD D.K., 1977. A curious binary lattice process. *J. Appl. Prob.* **14**, 717-731.

[I–35] PICKARD D.K., 1980. Unilateral Markov fields. *Adv. Appl. Prob.* **12**, 655-671.

[I–36] GREIG D.M., PORTEOUS B.T. & SEHEULT A.H, 1989. Exact Maximum A Posteriori estimation for binary images. *Journal of the Royal Statistical Society* **B-51** (2), 271-279.

[I–37] POSSOLO A., 1986. Estimation of binary Markov Random Fields. *Technical Report* **77**, Departement of Statitics, University of Washington.

[I–38] RIPLEY B.D., 1986. Statistics, images and pattern recognition. *Can. J. Stat.* **14**, 83-111.

[I–39] SAEBO H.V., BRATEN K., HJORT N.L., LLEWELLYN B. & MOHN E., 1985. Contextual classification of remotely sensed data: statistical methods and development of a system. *Report* **768**, Norvegian Computing Center, Oslo.

[I–40] SWAIN P.H., VARDEMAN S.B. & TILTON J.C., 1981. Contextual classification of multispectral data. *Pattern Recognition* **13**, 429-441.

[I–41] SWITZER P., 1980. Extension of linear discriminant analysis for statistical classification of remoteley sensed satellite imagery. *Math. Geol.* **12**, 367-376.

[I–42] VAN LAARHOVEN P.J.M. & AARTS E.H.L., 1987. *Simulated annealing: theory and applications.* D. Reidel Pub. Company.

[I–43] YAO J.F., 1989. Segmentation bayésienne d'image: comparaison des méthodes contextuelle et globale. *Cahiers du Centre d'études de Recherche Opérationnelle* **30** (4) (Université Libre de Bruxelles), 269-290.

[I–44] YOUNES L., 1988. Estimation and Annealing for Gibbsian Fields. *Annales de l'Institut Henri Poincaré*, 269-294.

[I–45] YOUNES L., 1988. *Problèmes d'estimation paramétriques pour des champs de Gibbs Markoviens. Applications en traitement d'images.* Thèse de Docteur en Sciences, Université de Paris – Sud.

[I–46] YU T.S & FU K.S., 1983. Recursive contextual classification using a spatial stochastic model. *Pattern Recognition* **16**, 89-108

[I–47] WELCH J.R. & SALTER K.G., 1971. A context algorithme for pattern recognition and image interpretation. *IEEE Trans.* **SMC–1**, 24-30.

II. Estimation spectrale

[II–1] AMEMIYA T., 1985. *Advanced Econometrics.* Basil Blackwell Ltd.: Oxford.

[II–2] AZENCOTT R. & DACUNHA-CASTELLE D., 1984. *Séries d'Observations Irrégulières.* Masson: Paris.

[II–3] BOLTHAUSEN E., 1982. On the central limit theorem for stationary mixing random fields. *The Annals of Probability* **10** (4), 1047-1050.

[II–4] BRILLINGER D.R., 1975. *Times Series: Data Analysis and Theory.* Holt, Rinehart and Winston: New York.

[II–5] DACUNHA-CASTELLE D. & DUFLO M., 1983. *Probabilités et Statistiques*, Volume 1 et 2. Masson: Paris.

[II–6] DAHLHAUS R., 1983. Spectral analysis with tapered data, *Journal of Time Series Analysis* **4** (3), 163-175.

[II–7] DAHLHAUS R., 1984. Parameter estimation of stationary processses with spectra containing strong peaks. Dans *Robust and Nonlinear Time Series Analysis* (eds. FRANKE, HARDLE & MARTIN), L.N.S. **26**, 50-67.

[II–8] DAHLHAUS R. & KÜNSCH H., 1987. Edge effects and efficient parameter estimation for stationary random fields. *Biometrika* **74** (4), 877-882.

[II–9] DZHAPARIDZE K.O. & YAGLOM A.M., 1982. Spectrum parameter estimation in time series analysis. Dans *Developments in Statistics* **4** (ed. KRISHNAIAH P.R.), 1-96. Academic Press: New York.

[II–10] EDWARDS R.E., 1979. *Fourier Series.* Volume 1 (second édition). Springer-Verlag: New York.

[II–11] GUYON X., 1982. Parameter estimation for a stationary process on an d-dimensional lattice, *Biometrika* **69**, 95-105.

[II–12] GUYON X., 1987. Estimation d'un champ par pseudo-vraisemblance conditionnelle: Etude asymptotique et application au cas markovien. Dans *Spatial processes and spatial time series analysis — Proc. 6th. Franco-Belgian Meeting of Statisticians* (ed. DRŒSBEKE F.). Bruxelles

[II–13] MARDIA K.V., 1988. Multi-dimensional multivariate gaussian Markov random fields with application to image processing. *Journal of Multivariate Analysis* **24**, 265-284.

[II–14] SHAH B.K., 1986. The distribution of positive definite quadratic forms. Dans *Selected Tables in Mathematical Statistics* **10** (ed. Inst. Math. Stat.). A.M.S: R.I. Providence.

[II–15] TUKEY J.W., 1967. An introduction to the calculations of numerical spectrum analysis, Dans *Spectral Analysis of Time Series* (ed. HARRIS B.), 25-46. Wiley: New York.

[II–16] WALKER A.M., 1964. Asyptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time serie. *J. Aust. Math. Soc.* **4**, 363-384

[II–17] WHITTLE P., 1954. On stationary processes in the plane. *Biometrika* **41**, 434-449

Table des matières