THÈSES D'ORSAY

ELISBETH GASSIAT GRENIER

# Déconvolution aveugle

*Thèses d'Orsay*, 1988

<http://www.numdam.org/item?id=BJHTUP11_1988__0232__A1_0>

L'accès aux archives de la série « Thèses d'Orsay » implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

UNIVERSITÉ PARIS SUD

NUMDAM

*Thèse numérisée par la bibliothèque mathématique Jacques Hadamard - 2016*
*et diffusée dans le cadre du programme*
*Numérisation de documents anciens mathématiques*
http://www.numdam.org/

63668

ORSAY

n° d'ordre :

# UNIVERSITE DE PARIS SUD

## Centre d'Orsay

# T H E S E

présentée

Pour obtenir

## Le TITRE de DOCTEUR EN SCIENCES

par

*Elisabeth GASSIAT*

*Sujet :* *DECONVOLUTION AVEUGLE.*

*Soutenue le 7 Janvier 1988 devant le Jury composé de :*

| | | | |
|---|---|---|---|
| *MM.* | *DACUNHA-CASTELLE* | *Didier* | *Président* |
| | *AZENCOTT* | *Robert* | |
| | *BRETAGNOLLE* | *Jean* | |
| | *GUYON* | *Xavier* | |
| | *LEON* | *José* | |

**Abstract:**

Considering a signal X which is a process of random variables identically independently distributed, and the signal Y obtained by filtering X through a linear system s, we study the estimation of s from the observation of Y in the following semi-parametric situation : the law of X is unknown and non Gaussian, and s has an inverse of convolution with finite length. We need no assumption on the phase of the system, i.e. on the causality or non causality of s. We propose an estimation by maximum objective. The estimates are consistent and asymptotically Gaussian ; this result is still available whatever the dimension of the index space of the series is.
We study the asymptotic efficiency of the estimate and, in the causal case, we compare it to the usual minimum square estimates. The output Y being an autoregressive field, we propose a consistent method of identification of the order of the model. We study different types of robustness : robustness to underparametrization, robustness to additive noise on the observations. We also investigate the case where the law of X has infinite moments, and we show that, for "*standardized cumulants*" as objectives, and under assumptions which are in particular verified for laws in the attraction domains of stable laws, the obtained estimates are still consistent, and the speed of convergence is, in the causal case, better than for laws with finite variance.

**Key-words**

LINEAR FIELDS
SEMI PARAMETRIC ESTIMATION
DECONVOLUTION
NON-CAUSAL SYSTEM
IDENTIFICATION
ROBUSTNESS
INFINITE MOMENTS
MINIMUM CONTRAST

**"à Pierre,**

**Paul,**

**Rémy, et . . . "**

C'est au sein de l'équipe d'Orsay que cette thèse a été réalisée, et je souhaite remercier chacun de ses membres - même s'ils ne peuvent être tous nommés ici - pour l'ambiance de travail et les relations amicales qui ont pu s'y créer.

Je tiens à nommer en premier lieu Didier DACUNHA-CASTELLE qui a dirigé mon travail, efficacement et chaleureusement, et a su me donner confiance. C'est pour moi un plaisir de lui exprimer remerciements et amitié.

David DONOHO est à l'origine de plusieurs idées développées dans cette thèse. Je le remercie pour son soutien. Travailler avec lui est un plaisir.

José LEON - qui me fait l'honneur de participer au jury - et Paul DOUKHAN m'ont donné des clés de mélanges ; Abdelkader MOKKADEM m'a plusieurs fois indiqué des pistes précises : merci.

Je remercie Xavier GUYON - qui m'a fait plusieurs fois profiter de son expérience des champs -, Jean BRETAGNOLLE et Robert AZENCOTT d'avoir accepté de participer au jury de cette thèse.

Au cours de mes passages sur Olivetti, Sirius, ou Vax, j'ai souvent sollicité la compétence des initiés, que ce soit Patrice ASSOUAD, Yves MISITI, Jean COURSOL ou Patrick JAKUBOWICZ : je tiens à saluer leur continuelle disponibilité.

Je ne peux oublier Jean-Claude FORT, Georges OPPENHEIM, et beaucoup d'autres, qui m'ont fait bénéficier de fructueuses discussions à des distances plus ou moins grandes (en quel sens ?...) de la déconvolution aveugle.

C'est enfin grâce aux soins de Mme PARVAN et Mme BAILLET que ce document a pu prendre sa forme définitive, je les en remercie.

# INTRODUCTION.

Cette thèse est consacrée au problème de la déconvolution aveugle, qui intervient fréquemment en théorie du signal, notamment dans les domaines suivants : en exploration sismique des sols, pour la restitution des coefficients de réflexion des couches géophysiques, en télécommunication, pour l'identification d'un canal téléphonique sans émission a priori de séquence connue. Les domaines évoqués ci-dessus font intervenir des séries temporelles; nous nous intéresserons tout aussi bien à la déconvolution aveugle de champs, c'est à dire de processus à indexation multidimensionnelle. Le problème concerne alors tous les domaines impliqués par la restoration d'image.

Les modèles de déconvolution sont de la forme suivante :

(1) $Y_t = \sum_{k \in \mathbb{Z}^d} s_k X_{t-k} + \varepsilon_t \ , \quad t \in \mathbb{Z}^d .$

Y est le signal observé, X est le signal à reconstituer, s est le filtre linéaire caractérisant le système, et $\varepsilon$ est un bruit d'observation.

La reconstitution de X se fait alors en deux étapes : l'estimation du filtre s, qui est la partie aveugle de la déconvolution, et la reconstruction de X sur la base des observations Y, s étant connu

réellement ou par son estimation, ce qui constitue la partie non aveugle de la déconvolution.

## 0.1. Partie non aveugle de la déconvolution.

### *0.1.1. Modélisation déterministe.*

Lorsque X est un signal déterministe, la transformation de Fourier est un outil fréquemment utilisé pour la reconstruction de X. Dans le domaine fréquentiel, l'équation de convolution s'écrit : $\hat{Y}(\omega) = \hat{s}(\omega).\hat{X}(\omega) + \hat{\varepsilon}(\omega)$, où $\hat{Y}$, $\hat{s}$, $\hat{X}$, $\hat{\varepsilon}$ sont les transformées de Fourier des séries Y, s, X et $\varepsilon$. L'estimation brutale de $\hat{X}$ par $\hat{Y}/\hat{s}$ est non seulement peu robuste au niveau de bruit, mais surtout particulièrement sensible aux points du spectre où $\hat{s}$ est pratiquement nul. Les méthodes utilisées habituellement pour résoudre ces problèmes sont celles utilisant le filtre de Wiener, que l'on peut retouver dans [13].

Cette méthode produisant des résultats peu satisfaisants dans le cas où X possède certaines propriétés de concentration, plusieurs auteurs ont considéré le problème comme un problème d'extension de spectre ($\hat{X}$ n'étant pas estimé aux fréquences où $\hat{s}$ est trop faible). Les problèmes d'extension de spectre ont été largement étudiés . Une solution particulièrement intéressante été proposée récemment ([16], [27] ) qui utilise comme critère la minimisation de la norme $L_1$ du signal X, du bruit $\varepsilon$, ou d'une combinaison pondérée des deux. Cette dernière méthode a le mérite de conduire à une programmation linéaire, dont plusieurs algorithmes ont été proposés pour réduire temps de calcul et espace de stockage des données .

### *0.1.2. Modélisation stochastique.*

Lorsque X est un signal stochastique, plusieurs techniques de vraisemblance, en particulier gaussienne, peuvent être utilisées. Elles ont donné, par modifications successives, plusieurs techniques de reconstruction. Citons la méthode par filtrage de Kalman ([13], [7], [19]). Goussart et al. ont développé une technique de vraisemblance lorsque le signal X est bernoulli-gaussien ([11]).

## 0.2. Partie aveugle de la déconvolution.

L'estimation du filtre s a été essentiellement étudiée dans le cadre non bruité ($\varepsilon=0$) et en modélisant X par un signal stochastique, la robustesse à l'addition de bruit étant ensuite évaluée.

### *0.2.1. Modélisation non paramétrique.*

Sans aucune restriction sur le filtre s et sur la loi des $X_t$, si ce n'est leur non gaussianité, les premiers résultats précis sont dus, d'une part à Ruget et al. ([2]) qui proposent un algorithme stochastique permettant l'identification du système, d'autre part à Rosenblatt et al. ([17]) qui proposent une méthode spectrale d'estimation de la fonction de transfert du système.

### *0.2.2. Modélisation semi-paramétrique.*

Une modélisation pratique et couramment utilisée d'un système linéaire est la modélisation ARMA du système.

#### *0.2.2.1. Cas causal.*

L'estimation des paramètres d'un système ARMA fait l'objet d'une abondante littérature. ([1], [5]). Classiquement, les estimateurs utilisés sont des variantes d'estimateurs des moindres carrés ou du maximum de vraisemblance sous hypothèse gaussienne .

Ces méthodes nécessitent intrinsèquement une restriction importante sur le système : la causalité du filtre s. Cette restriction n'est pas acceptable physiquement dans les domaines d'application précités (géophysique, télécommunications, images).

*0.2.2.2. Cas non causal.*

Lorsque la condition de causalité n'est pas imposée sur le système une distinction doit être faite entre signal stationnaire et signal non stationnaire. En effet, en situation causale, signaux stationnaires et non stationnaires coincident asymptotiquement. Ceci n'est plus vrai en situation non causale.

Pour un signal non stationnaire et en modélisation auto-régressive, un travail de Lai et Wei ( [15] ) montre que l'estimation par moindres carrés fournit un estimateur consistant.

Lorsque X et Y sont stationnaires, on ne peut plus utiliser les méthodes gaussiennes. Par ailleurs, le problème est insoluble lorsque le signal X est gaussien. Sous la seule hypothèse de non gaussianité et en modélisation auto-régressive, Donoho a proposé une estimation des paramètres par méthode de contraste ( [9] ).
C'est cette méthode que nous développons et étudions en détails dans cette thèse.

## 0.3. Le modèle.

Notations : Pour un élément t de $\mathbb{Z}^d$, nous utiliserons alternativement la notation t ou $(t_1,\ldots,t_d)$. L'élément nul de $\mathbb{Z}^d$ sera noté 0. Une série indexée par $\mathbb{Z}^d$ sera notée en majuscule pour les processus, en minuscule pour les filtres :

ainsi Z désigne le processus $(Z_t)$, $t\in\mathbb{Z}^d$, et a désigne le filtre $(a_t)$, $t\in\mathbb{Z}^d$.

De façon générale, on notera $Z = a*U$ le processus défini par :

$$Z_t = \sum_{k\in\mathbb{Z}^d} a_k U_{t-k} .$$

Le produit de monomes $z_1^{k_1}\dots z_d^{k_d}$ dans $\mathbb{C}^d$ sera noté $z^k$, si bien qu'un polynome $\sum_k b_{k_1,\dots,k_d}.z_1^{k_1}\dots z_d^{k_d}$ sera noté $\sum_k b_k.z^k$ .

On appellera bruit blanc une suite de variables aléatoires indépendantes équidistribuées.

Pour un filtre b ayant un nombre fini de coefficients non nuls, on notera $P_b$ le polynome à d variables complexes défini par :

$$P_b(z) = \sum_k b_k z^k, \quad z\in\mathbb{C}^d .$$

Le modèle est le suivant :

$$(2) \quad Y_t = \sum_{k\in\mathbb{Z}^d} s_{t-k}X_k, \quad t\in\mathbb{Z}^d .$$

On suppose que les $X_t$ sont des variables aléatoires réelles indépendamment identiquement distribuées de loi commune F non nécessairement centrée.

On observe $Y_t$, $t\in T_n$, où $T_n = \{ t\in\mathbb{Z}^d / \forall i\leqslant d,\ 1\leqslant t_i\leqslant n \}$.

On fait sur F et s des hypothèses suffisantes pour garantir par (2) la définition de Y, qui sont par exemple l'existence d'un moment d'ordre 1 pour F et la sommabilité de la série s. On suppose aussi que s possède un inverse de convolution b, permettant d'inverser la relation (2) en :

$$(3) \quad \sum_{k\in\mathbb{Z}^d} b_k Y_{t-k} = X_t .$$

Dans le début de notre travail étudiant la consistance de l'estimateur proposé nous ne supposerons pas nécessairement que b possède un nombre fini de coefficients non nuls. Cette hypothèse sera par contre toujours supposée vérifiée par la suite.

Par simple translation de l'indexation, on peut alors supposer que les coefficients non nuls de b sont les $b_k$ où $0 \leq k_i \leq p_i$, $i = 1 \ldots d$, avec $b_0 \neq 0$.

Y apparait alors comme un champ stationnaire auto-régressif d'ordre $p = (p_1, \ldots, p_d)$.

Les hypothèses sur le modèle seront toujours les suivantes :

(M0) La loi commune F des $X_t$ n'est pas gaussienne.

(M1) Le polynome $P_b$ n'a pas de racines sur le tore $|z_1| = \ldots = |z_d| = 1$.

. Il est connu que, sans aucune connaissance sur s, le problème d'identification de s est insoluble lorsque la loi F est gaussienne. En effet, si $Y = s*X$ et $Z = u*X$ pour deux filtres distincts s et u, les processus Y et Z ne diffèrent que par leur fonction d'autocovariance, ce qui ne suffit pas, nous le verrons plus loin, à déterminer s.

. (M1) est la condition sous laquelle la série stationnaire Y solution de (3) est bien définie et non explosive.

Rappelons quelques résultats sur la modélisation auto-régressive. Les résultats suivants sont classiques :

(a). La relation (3) a une unique solution stationnaire et régulière Y si et seulement si $P_b$ n'a pas de racines sur le tore $|z_i| = 1$ $i = 1 \ldots d$.

(b). Lorsque la condition (M1) est vérifiée, la relation (3) s'inverse en la relation (2), où $\sum_k s_k z^k$ est le développement en série de Laurent de $1/P_b$.

(c). La densité spectrale (série de Fourier de la fonction d'autocovariance) du processus Y est :

$$f(\lambda) = \frac{1}{|P_b(e^{-i\lambda})|^2} \quad , \lambda \in \mathbb{R}^d \text{ et } e^{-i\lambda} \text{ désignant } (e^{-i\lambda_j})_{j=1\ldots d}.$$

(d). Quel que soit le polynome Q, $Q(z) = \sum_{k=0}^{q} c_k z^k$ , vérifiant $f(\lambda) = \dfrac{1}{|Q(e^{-i\lambda})|^2}$ , il existe un bruit blanc $\varepsilon$ tel que $c*Y = \varepsilon$.

($\varepsilon \neq X$ dès que $b \neq c$).

En particulier, dans le cas d=1, la relation $|z-u|=|u|.|z-1/\bar{u}|$ pour $|z|=1$ permet de constuire toute une famille de polynomes de même module aux points z de module 1, dont les racines sont en partie les mêmes et en partie inverses l'une de l'autre.

(e). Dans le cas où d=1, la relation (2) est causale, c'est à dire $s_k=0$ pour $k<0$, si et seulement si $P_b$ a toutes ses racines à l'extérieur du disque unité de $\mathbb{C}$. Le processus X est alors l'innovation du processus Y.

Dans le cas des champs (d>1), on a des résultats du même type; on pourra par exemple pour d=2 consulter le cours de X.Guyon ([12]). L'estimation des paramètres d'un processus ARMA a fait l'objet de nombreux travaux : on pourra consulter [1] et [5] pour une bibliographie plus complète et une vue d'ensemble sur le problème. Classiquement, on a étudié des estimateurs des moindres carrés,

qui ne font intervenir le processus que par sa densité spectrale, et donc, au vu de (d), ne permettent pas de distinguer les filtres b dont les polynomes associés ont le même module sur le tore. On est alors conduit, pour choisir parmi ces filtres, à faire l'hypothèse indispensable de causalité, ce qui n'est pas acceptable dans les domaines d'applications cités plus haut.
Les méthodes d'estimation gaussiennes pour un champ ARMA (d>1) sont étudiées dans [28] .

### 0.4. Plan détaillé de la thèse.

Le plan que nous avons adopté est le suivant :

## I.DECONVOLUTION AVEUGLE OU ESTIMATION DES PARAMETRES D'UN PROCESSUS AUTO-REGRESSIF DE LOI INCONNUE NON GAUSSIENNE.

Dans cette partie, nous étudions la méthode d'estimation de b, inverse de s, proposée par D.Donoho dans [9].

I.1. Description de la méthode, consistance de l'estimateur.
La méthode développée est en fait une méthode d'estimation par minimum de contraste, que nous rebaptisons maximum d'objectif en remplaçant le contraste par son opposé que l'on appelle alors objectif.
La méthode en elle-même est classique, la partie intéressante étant de caractériser les fonctions J pouvant servir d'objectif. De par la structure de l'équation de convolution, on est conduit à choisir J parmi les fonctions invariantes par changement d'échelle. On donne dans ce paragraphe, après les rappels essentiels sur les contrastes, une caractérisation des objectifs par rapport au préordre de la divisibilité des lois. Ce préordre apparait naturellement en considérant les processus c*Y et en remarquant que la solution est

en c=b. On obtient alors un estimateur consistant.

I.2 Exemples.

Nous donnons alors quelques exemples d'objectifs, le plus remarquable par la simplicité des calculs auxquels il conduit étant le cumulant standardisé. Pour les autres, information de Fisher, néguentropie, métrique idéale, nous indiquons des conditions sur la loi des $X_t$ permettant éventuellement de les utiliser comme objectifs en situation adaptative.

I.3 Comportement asymptotique.

Une première partie étudie de façon générale la normalité asymptotique de l'estimateur de maximum d'objectif. La méthode employée est celle, classique, de démonstration en utilisant le développement de Taylor à l'ordre 1 avec reste intégral du gradient de l'objectif autour de la vraie valeur du paramètre. Nous obtenons ainsi la variance limite de l'estimateur.

Une deuxième partie étudie plus en détails le cas où J est une fonction de fonctionnelles linéaires de la loi F des $X_t$ , i.e. de la forme : $J(F)=L(E(\varphi(X_0)))$, où L est une fonction réelle et $\varphi$ une fonction réelle q-dimensionnelle. Ce paragraphe s'applique en particulier immédiatement au cas des cumulants standardisés.

Pour obtenir la normalité asymptotique, nous sommes amenés, du fait de la structure dépendante des observations, à utiliser des techniques de mélange; dans le cas où d=1, nous montrons que nos processus sont des mixingales au sens de Mac-Leish ([18]), sans hypothèse supplémentaire sur le modèle autre que l'existence de certains moments; dans le cas où d>1, nous utilisons un théorème de Rosen-

blatt ([25]) et sommes amenés à supposer que le processus est un champ mélangeant.

I.4 Comparaison avec les méthodes usuelles.

Lorsque l' on suppose que le filtre est causal, on dispose des méthodes usuelles d'estimation par maximum de vraisemblance gaussienne approché. Nous comparons notre méthode avec celles-ci, en définissant un domaine de lois pour lesquelles la méthode par objectif conduit à une meilleure variance limite. Cette étude est illustrée par des simulations.

I.5 Problèmes d'efficacité.

Le problème d'efficacité asymptotique est envisagé ici. Nous calculons tout d'abord la borne de Cramer-Rao (information de Fisher asymptotique) appliquée à notre modèle. Nous envisageons ensuite la possibilité de construire adaptativement un estimateur efficace. Nous proposons, par analogie avec les méthodes développées par Bickel ([3]), une condition d'existence de tels estimateurs : la causalité du modèle.

I.6 Identification : estimation de l'ordre p.

Si l'on plonge le modèle dans un modèle d'ordre plus grand, on obtient une valeur plus grande pour le maximum d'objectif. On ne peut donc utiliser l'objectif non modifié pour obtenir, par maximisation sur l'ordre et sur le paramètre b, un estimateur de l'ordre p. On définit alors, par analogie avec la vraisemblance compensée, un objectif compensé qui permet d'obtenir un estimateur consistant de p.

### I.7 Robustesse

Dans cette partie, nous étudions deux types de robustesse de la méthode : lorsque l'on a sous estimé l'ordre du modèle, et lorsque les observations sont bruitées par un bruit gaussien. Ce deuxième cas est étudié pour les objectifs "cumulants standardisés".
On démontre dans les deux cas un résultat de type "continuité" de la méthode garantissant une certaine robustesse.

### I.8 Moments infinis et objectifs cumulants.

L'étude asymptotique nous a conduit a faire sur le processus des hypothèses de moments. Si l'on ne suppose plus ces moments finis, l'objectif est infini asymptotiquement; on est alors conduit à penser que l'optimum n'en est que mieux défini asymptotiquement. Pour les cumulants et pour des lois de $X_0$ vérifiant certaines propriétés, en particulier pour des lois dans le domaine d'attraction de lois stables, nous étudions consistance et vitesse de convergence de l'estimateur du maximum d'objectif; cette vitesse est meilleure que lorsque la loi est de variance finie.

## II. RECHERCHE DE CONTRASTES ET ALGORITHME STOCHASTIQUE POUR L'IDENTIFICATION DE S.

Cette partie est une analyse de l'article de Benveniste, Ruget et Goursat ([2] ) portant sur le même problème. Leur méthode repose aussi sur un contraste V.

### II.1 Etude de la fonction V

Pour construire l'algorithme, il faut vérifier certaines conditions sur le contraste que nous présentons ici.

II.2 Algorithme avec contrainte.

Les auteurs développent un algorithme stochastique, et l'améliorent par une contrainte que nous examinons.

II.3 Recherche des contrastes.

C'est cette étude qui a retenu toute notre attention. Les auteurs proposent une caractérisation des contrastes présentant la forme particulière $J(c)=E\psi(c*s*X)$ et pour une classe de lois de $X_0$ : les lois sous-gaussiennes et les lois super-gaussiennes. Nous proposons une interprétation de cette caractérisation en lien avec l'étude menée en I.

## III. DECONVOLUTION AVEUGLE PAR ESTIMATION SPECTRALE.

Rosenblatt et Lii ont proposé une méthode d'estimation de la fonction spectrale complète du système. Nous la présentons ici et comparons les vitesses de convergence obtenues avec celle obtenue en I.

# I. DECONVOLUTION AVEUGLE , OU ESTIMATION DES PARAMETRES D'UN PROCESSUS AUTO-REGRESSIF DE LOI INCONNUE NON GAUSSIENNE.

## I.1. Description de la méthode, consistance de l'estimateur.

Dans ce paragraphe, nous nous placerons dans le cadre plus général où l'inverse b de s n'a pas nécessairement un nombre fini de coefficients non nuls.

Pour résoudre le problème, nous considérons les processus construits sur les observations $Y_t$ : $X(c)=(X(c)_t)$ avec $X(c)_t=\Sigma_k c_k Y_{t-k}$, où c varie dans l'espace des paramètres, et l'on cherche à restituer $X(b)=X$.

La méthode proposée ici est une méthode d'estimation de b par minimum de contraste, méthode dépendant d'une fonctionnelle $J(c)$ qui atteint son minimum au point $c=b$.

Rappelons tout d'abord quelques résultats relatifs aux estimateurs de minimum de contraste ([8]), et comment ils s'appliquent à notre problème.

Considérons un modèle statistique $(\Omega,\mathcal{A},(P_\theta)_{\theta\in\Theta})$, $\Theta\subset\mathbb{R}^d$. Une fonction de contraste de ce modèle relativement à $\theta$ est une fonction réelle de la variable $\alpha$ $K(\theta,\alpha)$ qui a un minimum strict pour $\alpha=\theta$. Lorsque les observations sont $Y_t, t\in T_n$, et si $\mathbb{F}_t$ est la filtration définie

par : $\mathbb{F}_t = \sigma(\{Y_u / \forall i \leqslant d,\ u_i \leqslant t_i\})$, un contraste relativement à $\theta$ et à K est une fonction réelle U : $(\alpha,t,\omega) \longrightarrow U(\alpha,t,\omega)$ qui vérifie les propriétés :

a). $\forall\ (\alpha,t) \in \times\ \mathbb{Z}^d$, la variable aléatoire $U_t(\alpha)$ : $\omega \longrightarrow U_t(\alpha,\omega)$ est $\mathbb{F}_t$-mesurable.

b). $U_t(\alpha) \longrightarrow K(\theta,\alpha)$ en probabilité sous $P_\theta$ pour $|t| = \min_i t_i \longrightarrow \infty$ .

Un estimateur $\hat{\theta}_t$ du minimum de contraste est donné par :

$U_t(\hat{\theta}_t) = \inf\ \{\ U_t(\alpha), \alpha \in \Theta\}$ .

On a alors le résultat :

**THEOREME** 1 : $\hat{\theta}_t$ est un estimateur consistant de $\theta$ (i.e. $\hat{\theta}_t \longrightarrow \theta$ en probabilité sous $P_\theta$) si les conditions suivantes sont vérifiées

(C1) : $\alpha \longrightarrow K(\theta,\alpha)$ et $\alpha \longrightarrow U_t(\alpha,\omega)$ sont continues .

(C2) : La convergence de $U_t(\alpha)$ vers $K(\theta,\alpha)$ est uniforme sur $\Theta$.

Démonstration : Soit $E_t(\varepsilon) = \{\ \sup_{\alpha\in\Theta}\ |U_t(\alpha)-K(\theta,\alpha)| < \varepsilon\ \}$ .

$P_\theta(E_t(\varepsilon)) \longrightarrow 1$ pour $|t| \longrightarrow \infty$ par (C2). $K(\theta,\alpha)$ est localement continument inversible autour de $\alpha=\theta$ à cause de l'extrémalité stricte de $K(\theta,\theta)$.

Donc : $\forall\varepsilon>0,\ \exists\eta>0$ tel que $|K(\theta,\alpha)-K(\theta,\beta)| \leqslant \eta \Rightarrow \|\alpha-\beta\| \leqslant \varepsilon$.

Pour $\varepsilon>0$ et sur $E_t(\eta/2)$ on a :

$K(\hat{\theta}_t,\theta) - \eta/2 \leqslant U_t(\hat{\theta}_t) \leqslant U_t(\theta) \leqslant K(\theta,\theta) + \eta/2$

$|K(\hat{\theta}_t,\theta) - K(\theta,\theta)| \leqslant \eta$, donc $\|\theta-\hat{\theta}_t\| \leqslant \varepsilon$, et $P(\|\theta-\hat{\theta}_t\| \leqslant \varepsilon) \geqslant P(E_t(\eta/2))$

Remarque : Lorsque $\Theta$ est compact, (C2) est équivalente, par le théorème de Prokhorov (voir [4]) à une condition sur les oscillations de $U_t(\alpha)$ :

$\forall\varepsilon>0$, $\forall\eta>0$, $\exists\delta>0$ tel que $P(\sup_{\|\alpha-\beta\|<\delta} |U_t(\alpha)-U_t(\beta)|>\eta) \leqslant \varepsilon$.

Dans notre problème, $\theta=b$ ; nous appellerons fonctions d'objectif $J$ et objectifs $J_n$ les opposés de fonctions de contraste $K$ et les opposés de contrastes $U_{(n,\ldots n)}$, et nos estimateurs seront donc estimateurs de maximum d'objectif.

Nous étudierons ici une classe très générale d'objectifs, ceux qui vérifient :

(O1) : $J(c)$ (resp. $J_n(c)$) dépend seulement de la loi commune $F(c)$ des variables $X(c)_t$ (resp. la loi empirique $F(c)_n$ des $X(c)_t$ construite sur les observations).

(O2) : $J$ et $J_n$ sont invariants par changement d'échelle, c'est à dire $J(\lambda c)=J(c)$ et $J_n(\lambda c)=J_n(c)$ pour tout réel non nul $\lambda$.

L'invariance par changement d'échelle est une condition naturelle dans la mesure où l'échelle du paramètre ne peut pas être identifiée, puisque si l'on multiplie $s$ par $\lambda$ et si l'on divise les $X_t$ par $\lambda$, le processus observé $Y$ ne change pas.

En vertu de (O1), on peut définir la fonctionnelle sur des lois de variables aléatoires; si $Z$ est une variable aléatoire de loi $F_Z$, on notera $J(F_Z)$ la valeur de $J$ ainsi définie.

Si $V$ est un ensemble de lois non gaussiennes stable par convolution et changement d'échelle sur lequel $J$ est défini, on a le résultat suivant :

**THEOREME** 2 : J est une fonction de contraste pour le modèle défini par (1) pour F, loi commune des $X_t$, dans V et pour tout filtre sommable si et seulement si :

Quelle que soit la variable aléatoire X de loi dans V, quel quel que soit la série sommable $(a_i)$, $i\in\mathbb{Z}$, si les $(X_i)$, $i\in\mathbb{Z}$ sont des variables aléatoires indépendantes de même loi que X, alors :

(O3) $J(F_{\Sigma_i a_i X_i}) \leqslant J(F_X)$ , avec égalité si et seulement si tous les $a_i$ sont nuls sauf un.

Ce théorème est immédiat dès que l'on voit que $F(c) = F_{\Sigma_k (c*s)_k X_{-k}}$.

Son intérêt est de donner une caractérisation simple des fonctions d'objectif dans le problème qui nous intéresse, caractérisation soulignant l'extrémalité de la loi gaussienne où le problème est insoluble. On peut en effet définir une relation d'ordre partiel sur V $\cup$ { gaussiennes }, ordre pour lequel la gaussienne est le plus petit élément, et (O3) signifie alors que J respecte l'ordre ainsi défini. Pour des détails et commentaires sur cette caractérisation, voir [9].

Dans tous les paragraphes qui suivent, on supposera que le paramètre est fini-dimensionnel, i.e. $b_k=0$ dès que l'un des $k_i$ vérifie $k_i<0$ ou $k_i>p_i$.

A cause de la condition d'invariance (O2), on choisit pour le paramètre des conditions de normalisation, et l'on définit deux espaces de paramètre :

$\mathbb{T} = \{B \in \mathbb{R}^p / \ \Sigma_k b_k^2 = 1\}$, $p = (p_1+1)\ldots(p_d+1)$

et $\mathbb{U} = \{B \in \mathbb{R}^p \ / \ b_{(0,\ ,0)} = 1\}$.

$b \in \mathbb{U}$ est la normalisation usuelle pour les modèles ARMA, $b \in \mathbb{T}$ est une normalisation par appartenance à un compact.

On a alors le résultat suivant :

**THEOREME** 2bis :

Considérant $J$ et $J_n$ qui satisfont les conditions (O1), (O2), (O3), et (C1), (C2), les estimateurs $\hat{b}_n$ et $\hat{\hat{b}}_n$ maximisant $J_n(c)$ respectivement sur $\mathbb{T}$ et $\mathbb{U}$ sont consistants.

## I.2. Exemples.

On peut montrer que les objectifs suivants vérifient (O1), (O2) et (O3) :

(E1) : Cumulants standardisés.

$$J(F_X) = \left| \frac{C_m(X)}{(C_2(X))^{m/2}} \right|$$ , $m>2$, que l'on définit sur l'espace des variables ayant des moments jusqu'à l'ordre m,

où $C_m(X) = (-i \frac{d}{dt})^m \operatorname{Log} E(\exp itX)|_{t=0}$ (cumulant d'ordre m de X).

On démontre aisément que $U(\Sigma a_i X_i) = \left| \frac{\Sigma\ a_i^m}{(\Sigma\ a_i^2)^{m/2}} \right| . U(X)$

et donc que U vérifie (O3) dès que le cumulant standardisé de X est non nul.

(E2) : Néguentropie.

$J(F_X) = K(X,\varphi)$ où K désigne l'information de Kullback et $\varphi$

est la gaussienne centrée de même variance que X, que l'on définit sur l'espace des variables ayant un moment d'ordre 2 et dont la loi possède une densité d'entropie finie.

Notons S(X) l'entropie de la variable aléatoire X, soit $S(X)=-\int p(x) \text{ Log } p(x)\, dx$ où p désigne la densité de X alors $J(F_X)= -S(X) + S(\varphi)$.

On sait (voir [24]) que $\forall X$ et $\forall Y$ pour lesquels S est définie : $\exp(2S(X+Y)) \geqslant \exp(2S(X))+\exp(2S(Y))$, l'inégalité étant stricte dès que X et Y n'ont pas même loi.

On a donc, pour tout n et pour toute suite $(a_i)$ de longueur n : $\exp(2S(\Sigma_{i=1}^n a_i X_i)) \geqslant \Sigma_{i=1}^n \exp(2S(a_i X_i))$, les $X_i$ étant i.i.d. de même loi que X donnée.

$S(a_i X_i)=S(X)+\text{Log } a_i$, on obtient donc $S(\Sigma a_i X_i) \geqslant S(X)+1/2.\text{Log}(\Sigma a_i)^2$, l'inégalité étant stricte dès que au moins deux des $a_i$ sont non nuls et que X n'est pas gaussienne.

D'autre part, si $\varphi$ est gaussienne centrée de même variance que X et $\gamma$ gaussienne centrée de même variance que $\Sigma a_i X_i$,

$J(F_{\Sigma a_i X_i})= -S(\Sigma a_i X_i)+S(\gamma)$ et $J(F_X) = -S(X)+S(\varphi)$ ;

or $S(\varphi)=S(\gamma)+1/2.\text{Log}(\Sigma a_i)^2$ ; en combinant ceci avec l'inégalité précédente on obtient :

$J(F_{\Sigma a_i X_i}) < J(F_X)$ pour toute série de longueur finie ayant au moins deux termes non nuls.

Pour étendre ce résultat à une série infinie de carré sommable, il suffit de se placer sous des conditions où S(Z) est continue au sens $L_2$ par rapport à sa variable Z sur l'espace de Hilbert engendré par les combinaisons linéaires finies de répétitions indépendantes de X.

On a alors, en notant $Z_n=\Sigma_1^n a_i X_i$ : $J(F_{Z_n})<J(F_{Z_{n_0}})<J(X)$, où $n_0$ est le premier indice pour lequel au moins deux des $a_i$, $i\leqslant n_0$, sont non nuls.

Une condition suffisante permettant d'écrire $S(Z)=\lim_n S(Z_n)$ est la suivante :

(NEG) : $\exists p>0$ tel que $|\varphi|\in L_p$, $\varphi$ étant la fonction caractéristique de la variable X.

En effet, si l'on note $p_n$ la densité de la loi de $Z_n$ (qui existe puisque celle de X existe et que la loi de $Z_n$ est une convolution finie de lois ayant une densité), si l'on note $\varphi_n$ la fonction caractéristique de $Z_n$, on a : $|\varphi_n|\in L_{p/n}$.

D'autre part, $\exists N$ tel que $p/N\leqslant 1$, et on a alors : $\forall n\geqslant N$, $|\varphi_n|\in L_1$. Ceci entraine que $p_N$ est borné, puis, par convolution, que :

$\forall n\geqslant N$, $\forall x$, $p_n(x)\leqslant \|p_N\|_\infty$

$\forall t$, $|\varphi_n(t)|\leqslant|\varphi_N(t)|$ .

Ceci permet de déduire, par le théorème de convergence dominée de Lebesgue, que $p_n(x)\longrightarrow q(x)$, où q est la densité de la loi de Z.

Les $Z_n$ étant uniformément intégrables et les $p_n$ uniformément bornées, on peut appliquer le résultat de Mokkadem ([21] ) et $\lim_n S(Z_n)=S(Z)$.

(E3) : <u>Information de Fisher</u>.

$J(F_X)=I(X)=\int(p')^2/p\ dx$ où p désigne la densité de la loi X .

C'est un résultat classique que l'information de Fisher ainsi définie décroit par convolution. Rappelons en rapidement une démonstration :

Si p et q sont deux densités ayant une information de Fisher, si $h=p*q$ :

on a $h(x)=\int p(x-y)q(y)dy = \int q(y)dy\int^x p'(t-y)dt$

$\int q(y)\int^x |p'(t-y)|\ dy \leqslant \int |p'(x)|dx$, donc $h'(t)=\int p'(t-y)q(y)dy$

Par Schwarz, $(\int (p'(t-y)/\sqrt{p(t-y)}).\sqrt{p(t-y)}.q(y)dy)^2$

$$\leqslant h(t).\int (p'/p)^2(t-y)p(t-y)q(y)dy$$

Donc $I(h)\leqslant I(p)$.

On a donc immédiatement que, dès que X a une densité p ayant une information de Fisher, $\Sigma_1^n a_i X_i$ en a une vérifiant :

$I(\Sigma a_i X_i) < I(X)$ (On n'a l'égalité que si un seul des $a_i$ est non nul, ceci par l'utilisation de l'inégalité de Schwarz dans la démonstration précédente).

D'autre part, pour une suite infinie $(a_n)$, on a :

$\lim_n I(\Sigma_1^n a_i X_i) \leqslant I(\Sigma_1^{n_0} a_i X_i) < I(X)$ où deux au moins des $a_i$ sont non nuls pour $i\leqslant n_0$.

Par le lemme de Fatou, avec les mêmes notations que dans (E2) :

$$\int \underline{\lim}\ \frac{p'^2_n}{p_n}\,dx \leqslant \underline{\lim} \int \frac{p'^2_n}{p_n}\,dx \quad (< I(X) \text{ par ce qui précède}).$$

Il suffit alors de se placer sous des hypothèses assurant la convergence de $p'^2_n/p_n$ vers $q'^2/q$. L'hypothèse suivante est suffisante :

(FISH) : $\exists\varepsilon>0$ et $\exists p>0$ tels que $t^\varepsilon\varphi$ soit élément de $L_p$.

De la même manière que pour la négentropie, il existe N tel que si n est supérieur à N, $t\varphi_n$ est élément de $L_1$ et $|t\varphi_n|\leqslant|t\varphi_N|$.

On a alors $p_n \longrightarrow q$ et $p'_n \longrightarrow q'$.

(E4) : <u>Métrique idéale</u>.

Si $\mu_s$ est une métrique idéale d'ordre s, on peut définir :

$J(F_X) = \mu_s(X,G)$, où G est une variable aléatoire gaussienne et où $s>2$.

Rappelons la définition d'une métrique idéale : (Pour plus de précisions, voir [29])

Soit $\xi$ un ensemble de variables aléatoires défini sur un espace probabilisé et prenant leur valeur dans un Banach séparable. On désigne par $(\xi)_k$ l'espace des distributions jointes de tous les k-uples $(X_1;\ldots;X_k)$ de variables de $\xi$.

$\mu$ est <u>une métrique simple</u> si $\mu$ est une fonction de $(\xi)_1\times(\xi)_1$ dans $[0;+\infty]$ qui vérifie : symétrie, inégalité triangulaire, et identification, c'est à dire $P_X=P_Y \Rightarrow \mu(X,Y)=0$, où $P_X$ désigne la loi de X. Une métrique simple $\mu$ est une <u>métrique idéale</u> si elle vérifie:

1. $\forall\ X,Y,Z \in \xi$ , $\mu(P_X*P_Z,P_Y*P_Z) \leqslant \mu(P_X,P_Y)$

2. $\forall\ X,Y \in \xi$ et $\forall\ c\in\mathbb{R}^*$, $\mu(cX,cY)=|c|^s\mu(X,Y)$. s est l'ordre de $\mu$

Un exemple simple de métrique idéale est le suivant :

$\mu_s(X,Y) = \sup_{f\in D(s,K)} |E(f(X)-f(Y))|$, où $s=m+\alpha$, $m\in\mathbb{N}$ et $\alpha\in]0,1]$, et où D(s,K) est l'ensemble des fonctions réelles f continues m fois dérivables dont la dérivée d'ordre m vérifie : $|f^{(m)}(x)-f^{(m)}(y)|\leqslant K|x-y|^{\alpha}$.

Il s'agit ensuite de définir une suite de fonctions aléatoires $J_n$ vérifiant (C1) et (C2). On peut choisir $J_n(c) = J(F_n(c))$. Dans le cas des cumulants, on fait ce choix. Il est ensuite facile de montrer que si X possède des moments finis jusqu'à l'ordre m, alors $J_n$ converge p.s. uniformément vers J.

Pour les 3 derniers exemples, un tel choix n'est pas possible. En effet, J fait intervenir la loi inconnue de X. On notera $J(c)=J(p,c)$, p désignant la densité de la loi de X.

On peut alors envisager une procédure de type "adaptative" :

1). Faire une estimation $b_n$ préliminaire de $b$

2). Faire une estimation régulière $p_n$ de la densité de X à l'aide de la loi empirique de $b_n*Y$; on peut faire par exemple une estimation par noyau, soit $p_n(x)=1/n^d.\ \Sigma_k K_n(x-(b_n*Y)_k)$ où $K_n$ est un noyau régularisant, par exemple $K_n(x)=\psi(x/h_n)$, $h_n \longrightarrow 0$ et $\psi$ est la densité gaussienne centrée réduite.

3). Poser $J_n(c) = J(p_n, F_n(c))$.

## I.3. Comportement asymptotique.

Dans ce paragraphe, on suppose que J et $J_n$ vérifient les conditions suffisantes de consistance pour $\hat{\hat{b}}_n$ et $\hat{b}_n$. Les résultats qui suivent sont valides quel que soit d>0.

Nous utiliserons les hypothèses suivantes :

(A0) : $J(c)$ et $J_n(c)$ sont deux fois continument différentiables.

(A1) : $J_n(c) \xrightarrow{p.s.} J(c)$

$\text{Grad } J_n(c) \xrightarrow{p.s.} \text{Grad } J(c)$

$D^2 J_n(c) \xrightarrow{p.s.} D^2 J(c) = A(c),\ A(b)=A.$

(A2) : $n^{d/2} \text{Grad } J_n(b) \xrightarrow{\lambda} N(0,\Lambda)$ , où $\xrightarrow{\lambda}$ désigne la convergence en loi et $N(0,\Lambda)$ la gaussienne centrée de covariance $\Lambda$.

(A3) : $\exists r>0$ tq $(D^2J_n(b+\alpha)-D^2J_n(b)) \xrightarrow{P} 0$ uniformément pour $|\alpha|\leq r$

<u>Remarque</u> : Ces hypothèses n'imposent rien sur la régularité de F, ni même sur l'existence d'une densité. Par exemple, lorsque J est fonction de fonctionnelles linéaires de F(en particulier lorsque J est un cumulant standardisé), F peut être quelconque sous réserve que les fonctionnelles linéaires intervenant soient finies.

**LEMME**1 : Si (A0) et (A1), alors Ker A = $\{\lambda B/\lambda\in\mathbb{R}\}$.

<u>Démonstration</u> : Soit $\mathcal{D}=\{\lambda b/\lambda\in\mathbb{R}\}$. Pour $u\in\mathbb{R}$ et $c\notin\mathcal{D}$, soit h(c)(u) la fonction donnée par $h(c)(u)=J(b+uc)$. $h(c)'(u)=\langle \mathrm{Grad}\ J(b+uc),c\rangle$, $h(c)''(u)={}^tc.A(b+uc).c$. h(c) a un maximum strict en u=0, donc : $h(c)'(0)=0$ et $h(c)''(0)\neq 0$, soit ${}^tc.A.c\neq 0$ pour $c\notin\mathcal{D}$.
Prenant maintenant c dans $\mathcal{D}$ : $h(\lambda b)$ est constante grace à (O2), et $h(c)''$ est identiquement nulle, donc ${}^tc.A.c=0$.

Pour les théorèmes qui vont suivre, nous aurons besoin de notations particulières :
A est diagonalisable dans une base orthonormée. Prenons $b/\|b\|$ comme premier vecteur de cette base. Dans cette base, soit D (resp.$\Delta$) la transformée de A (resp.$\Lambda$), soit $\hat{c}_n$ le transformé de $\hat{b}_n$.
Si M est une matrice de $\mathbb{R}^{p^2}$, on notera (M) la matrice de $\mathbb{R}^{(p-1)^2}$ obtenue en retirant à M sa première ligne et sa première colonne.
Pour un vecteur m, on notera (m) le vecteur obtenu en retirant à m sa première coordonnée.
On a alors les résultats suivants :

**THEOREME** 3 : Si (A0)...(A3) :

1. $n^{d/2}((\hat{c}_n)) \xrightarrow{\lambda} N(0,\Sigma)$, avec $\Sigma=(D)^{-1}(\Delta)(D)$.

2. $\hat{c}^2_{n(0,\ ,0)} = 1-\Sigma_j\hat{c}^2_{nj}$ , d'où la loi asymptotique de $\hat{c}_n$.

**THEOREME** 4 : Si (A0)...(A3) :

$$n^{d/2}((\hat{b}_n) - (b)) \xrightarrow{\lambda} N(0,\Sigma^*), \text{ avec } \Sigma^* = (A)^{-1}(\Lambda)(A).$$

Pour démontrer ces deux théorèmes, on écrit le développement de Taylor à reste intégral à l'ordre 2 pour le gradient de $J_n$:

$$\text{Grad } J_n(c\ ) = \text{Grad } J_n(b) + {}^t(c\ -b)\int_0^1 D^2 J_n(b+u(c\ -b))\ du \qquad (T)$$

$$= 0 \text{ pour } c=\hat{b}_n \text{ et pour } c=\hat{\hat{b}}_n\ ,$$

car grâce à (O2), un extremum de $J_n$ sur $\mathbb{T}$ ou sur $\mathbb{U}$ est un extremum sur tout l'espace. Le théorème 4 est une conséquence immédiate de (T), (A2) et du lemme1.

<u>Démonstration du théorème 3</u> :

Réécrivons (T) dans la base choisie diagonalisant A :

(1) $n^{d/2}\ {}^t(\hat{c}_n - c).D_n = E_n$, où $c = (1,0,...0)$, soit b écrit dans la nouvelle base.

(2)

$$D_n \xrightarrow{P} \begin{bmatrix} 0 & & & \\ & \lambda_1 & & \\ & & \ddots & \\ & & & \lambda_{p-1} \end{bmatrix} \qquad \lambda_j \neq 0$$

(3) $E_n \xrightarrow{\lambda} N(0,\Delta)$

(4) $\Sigma_j\ \hat{c}_{nj}^2 = 1$

(2) : par (A1) et (A3) on obtient facilement que

$$\int_0^1 D^2 J_n(b+u(\hat{b}_n - b))\ du \xrightarrow{P} A$$

Par changement de base on obtient le résultat.

(3) : (A2) écrit dans la nouvelle base.

(4) : car le changement de base est orthonormé.

On remarque que (4) peut se réécrire en : $(\hat{c}_0^n - 1)(\hat{c}_0^n + 1) + \sum_{j\neq 0} \hat{c}_j^{n\,2} = 0$

Notons $v_n = n^{d/2}(\hat{c}^n - c)$

$A(n,M)=\{\ \omega\ /\ \|v_n\|>M\ \}$

$B(n,M)=\{\ \|J_n\|>M\}$

$C(n,\varepsilon)=\{\ \sup_{i,j}|(D_n)_{i,j}|>\varepsilon\ \}$

$D(n,\beta)=\{\ |\hat{c}_0^n-1|>\beta\ \}$

On sait que $\forall\varepsilon,\beta\ \lim_{M\longrightarrow\infty}\overline{\lim_n}\ P(B(n,M))=0,\ \lim_n P(C(n,\varepsilon))=0$

$\lim_n P(D(n,\beta))=0.$

On en déduit facilement que $\forall\alpha>0,\forall\varepsilon>0$ et $\forall\beta>0$,

$\lim_M \overline{\lim}\ P(\mathcal{Q}(n,M,\alpha,\varepsilon,\beta)) = \lim_M \overline{\lim}\ P(A(n,M))$

où $\mathcal{Q}(n,M,\alpha,\varepsilon,\beta) = A(n,M)\cap B^c(n,\alpha M)\cap C^c(n,\varepsilon)\cap D^c(n,\beta)$

On fixe $\beta_0< 2/p+1$, $\varepsilon_0= 1/2(p+1).\mathrm{Sup}_j|\lambda_j|$, $\alpha_0=1/4(p+1).\inf_j|\lambda_j|$.

Soit $\omega\in\mathcal{Q}(n,M,\alpha_0,\varepsilon_0,\beta_0)$, $j$ la coordonnée telle que $|v_{nj}|=\sup_i|v_{ni}|$

On a $|v_{nj}|\geqslant M/p+1$.

.Si $j\neq 0$ on écrit : $\sum_{i\neq j}v_{ni}D_{nij}=-v_{nj}D_{njj}+U_{nj}$.

Dans $B^c(n,\alpha M)$, $\|U_n\|<\alpha M$ donc $|U_{nj}|<\alpha M$

$|\sum_{i\neq j}v_{ni}D_{nij}|\geqslant |v_{nj}D_{njj}|-\alpha M \geqslant |v_{nj}|(|\lambda_j|-\varepsilon)-\alpha M$ car on est dans $C^c(n,\varepsilon_0)$

D'autre part : $|\sum_{i\neq j}v_{ni}D_{nij}|\leqslant p.|v_{nj}|\varepsilon$.

On a donc : $|v_{nj}|.|\lambda_j|/2\leqslant|v_{nj}|(|\lambda_j|-(p+1)\varepsilon)\leqslant\alpha M\leqslant|\lambda_j|.M/4(p+1)$

D'où $|v_{nj}|\leqslant M/2(p+1)$, ce qui est impossible.

.Si $j=0$ on écrit $\sum_{j\neq 0} u_{ni}^2= - u_{n0}.n^{d/2}(\hat{c}_0^n+1)$ .

Donc $p|u_{0n}|^2\geqslant|u_{0n}|.n^{d/2}(2-\beta_0)$, ce qui donne $\beta_0\geqslant 2/p+1$, ce qui est contradictoire.

Donc $P(\mathcal{Q}(n,M,\alpha_0,\varepsilon_0,\beta_0))=0$, donc $\lim_M\overline{\lim_n}\ P(A(n,M))=0$,

et $n^{d/2}(\hat{c}^n-c)$ est tendu. Le théorème s'en déduit immédiatement.

On donne maintenant un lemme très simple, qui sera utile plus tard pour simplifier quelques expressions de matrices.

**LEMME** 2 : Si (A0)...(A3), alors $b \in \mathrm{Ker}\Lambda$.

Démonstration : Faire le produit scalaire de (T) avec b, multiplier par $n^{d/2}$, appliquer le lemme 1 et le théorème 1.

**Cas particulier important : cas où J est fonction de fonctionnelles linéaires de F.**

Dans cette partie, on suppose que $J$ et $J_n$ s'écrivent :

$J(c) = L(E\varphi_1(X(c)_0);\ldots;E\varphi_q(X(c)_0))$

$J_n(c) = L(1/n^d \sum_{t=(1,\ldots,1)}^{(n,\ldots,n)} \varphi_1(X(c)_t);\ldots;1/n^d \sum \varphi_q(X(c)_t))$

$L$ et $\varphi = (\varphi_1,\ldots,\varphi_q)$ étant des fonctions réelles.

Remarque : Dans le cas des cumulants, $J=|K|$ et $J_n=|K_n|$ où $K$ et $K_n$ sont de la forme précédente, mais il est très facile de voir que tous les résultats qui suivent sont encore valables.

Considérons les hypothèses suivantes :

(H1) $\varphi$ et $L$ sont deux fois continument dérivables.

(H2) Les espérances des fonctions suivantes sont finies :

$\varphi(X(c)_0)$, $\mathrm{Var}\,\varphi(X_0)$, $\varphi''_k(X(c)_0^\alpha\cdot(X(c)_0)$, $X_0^\alpha\psi^2(X_0)$ $(\alpha = 0,1,2)$,

$\varphi'_k(X(c)_0)\cdot(X(c)_0)^\beta$, $\varphi_k(X_0)\psi(X_0)X_0^\beta$ $(\beta = 0,1)$,

et, si $d>1$ tout produit de 4 termes formés à partir des termes suivants :

$X_0^\beta\psi(X_0)$ et $\varphi_k(X_0)$, $\beta=0,1$, $k=0,\ldots,q$.

(H3) $E\psi(X_0) = 0$ et $EX_0\psi(X_0) = 0$

(H4) $d>1$ et $(Y_t), t \in \mathbb{Z}^d$ est un champ fortement mélangeant.

Où $\psi$ est donnée par :

$\psi(u) = \sum_{k=1}^q \varphi'_k(u)\ D_kL(E\varphi(X(c)_0)$.

Remarque : Pour les cumulants, (H2) est une hypothèse sur les moments de $X_0$.

Sous (H1), on a :

$$\text{Grad } J_n(c)_i = 1/n^d \sum \psi_n(X(c)_t)Y_{t-i}$$

$$D^2J_n(c)_{i,j} = 1/n^d \sum \psi_n'(X(c)_t)Y_{t-i}Y_{t-j} + 1/n^{2d} \sum \gamma_n(X(c)_t, X(c)_u)Y_{t-i}Y_{u-j}$$

avec $\psi_n(u) = \Sigma_1^q \varphi_k'(u)\ D^kL(E_n\ \varphi(X(C))$

$$\gamma_n(u,v) = \Sigma_1^q \varphi_k'(u)\ \varphi_l'(v)\ D^kD^lL(E_n\varphi(X(C)).$$

$\psi_n$ et $\gamma_n$ ne sont pas indicées par c pour plus de lisibilité lorsque cela ne porte pas à confusion. $E_n$ désigne l'espérance sous la loi empirique .

Par ergodicité, on obtient alors immédiatement (A1), (A0) est donnée par (H1),et pour (A3), la démarche est analogue à la consistance. En particulier,(A3) est vraie pour les cumulants, un calcul des oscillations le montrant facilement.

Le calcul donne alors :

$$A_{i,j} = \Sigma_1^q E(\varphi_k''(X_0)Y_{-i}Y_{-j})\ D^kL(E\varphi(X_0))$$
$$+ \Sigma_1^q E(\varphi_k'(X_0)Y_{-i})\ E(\varphi_l'(X_0)Y_{-j})\ D^kD^lL(E\varphi(X_0))$$

On a le théorème :

**THEOREME** 3 : Sous ((H1)...(H3) si d=1 ou (H1)...(H4) si d>1) on a :

$$n^{d/2}\text{Grad } J_n(b) \xrightarrow{\lambda} N(0,\Lambda).$$

$$\Lambda_{i,j} = E(\psi\ (X_0)Y_{-i}Y_{-j}) + \Sigma_1^q a_{k,i}E(\varphi_k(X_0)\psi(X_0)Y_{-j})$$
$$+ \Sigma_1^q a_{k,j}E(\varphi_k(X_0)\psi(X_0)Y_{-i}) + \Sigma_1^q a_{k,i}a_{l,j}\text{Cov}(\varphi_k(X_0),\varphi_l(X_0))$$

où $a_{k,i} = \Sigma_{l=1}^q D^kD^lL(E\varphi(X_0)).E(\varphi_l'(X_0)Y_{-i})$ .

De ce théorème on déduit la loi asymptotique pour $\hat{b}_n$ et $\hat{\hat{b}}_n$.

Remarques sur les hypothèses :

1). Lorsque d=1, on n'a fait aucune hypothèse de régularité sur la loi de X.

2). La condition de champ mélangeant n'a pas été étudiée pour des processus linéaires. Des conditions de mélange fort ont été données pour des processus linéaires indexés dans $\mathbb{Z}$ par Gorodetski ([10] ) et Mokkadem ([22]).

Démonstration du théorème :

$$n^{d/2}\,\mathrm{Grad}\, J_n(b) = 1/n^{d/2}\Sigma\psi(X_t)Y_{t-i} + \Sigma_{k,l}a^n_{k,l,i}\; n^{d/2}(E_n\,\varphi_l(X)-E\varphi_l(X)).$$
$$\left\{ D^kD^lL(E\varphi(X)) + \int_0^1 [D^kD^lL(E\varphi+t(E_n\,\varphi-E\varphi))-D^kD^lL(E\varphi)]\,dt \right\} \qquad (\alpha)$$

où . $a^n_{kli} = 1/n^d\ \Sigma\ \varphi_k'(X_t)Y_{t-i} \xrightarrow{P} E\varphi_k'(X_0)Y_{-i}$

. Le reste intégral tend en probabilité vers 0 .

Il reste à montrer :

$$\begin{cases} 1/n^{d/2}\cdot\ \Sigma\ \psi(X_t)Y_t \\ \vdots \\ i/n^{d/2}\cdot\ \Sigma\ \psi(X_t)Y_{t-p} \\ n^{d/2}(E_n\,\varphi_1(X)-E\varphi_q(X)) \\ \vdots \\ n^{d/2}(E_n\,\varphi_q(X)-E\varphi_q(X)) \end{cases} \xrightarrow{\lambda} N(0,\zeta)$$

On prend une combinaison linéaire des coordonnées du vecteur étudié multiplié par $n^{d/2}$.

$$Z_u = \Sigma_0^p\alpha_i(\Sigma_{t=(1,\ ,1)}^{(u1,\ ,ud)}\psi(X_t)Y_{t-i}) + \Sigma_1^q\beta_i\Sigma_t(\varphi_i(X_t)-E\varphi_i(X))$$

$Z_u$ est un processus stationnaire centré (par (H3)) sur $\mathbb{Z}^d$.

$$EZ_u^2 = \Sigma\ \alpha_i\alpha_j\Sigma_{t1,t2}E[\psi(X_{t1})\psi(X_{t2})Y_{t1-i}Y_{t2-j}]$$
$$+ \Sigma\ \beta_i\beta_j\Sigma_{t1,t2}E[(\varphi_i(X_{t1})-E\varphi_i(X_{t1}))\cdot(\varphi_j(X_{t2})-E\varphi_j(X_{t2}))]$$
$$+ \Sigma\ \alpha_i\beta_j\Sigma_{t1,t2}E[\psi(X_{t1})Y_{t1-i}\cdot(\varphi_j(X_{t2})-E\varphi_j(X_{t2}))]$$

Lorsque $u = (u_1;\ldots;u_d)$ varie dans $\mathbb{Z}^d$ on obtient :

$EZ_u^2 = \prod_{i=1}^d u_i \cdot [\sum \beta_i^2 Var\varphi_i(X) + \sum \alpha_i^2 E(\psi^2(X) Y_{-i}^2) + \sum \alpha_i \beta_j E[\psi(X) Y_{-i}(\varphi_j(X) - E\varphi_j(X))]$

$= \prod_{i=1}^d u_i \cdot (\alpha,\beta)\xi^t(\alpha,\beta)$ $\xi$ définie de manière évidente ci-dessus, et qui est donc la variance limite du vecteur étudié lorsque n tend vers l'infini ($u_i = n^d$).

1). <u>Lorsque</u> $d=1$ :

On utilise le principe d'invariance de Mac-Leish ([18]). Les hypothèses à vérifier sont les suivantes :

Si $\mathbb{F}_n$ est une séquence non-décroissante de sigma-algèbres, $V_n$ un processus :

.Il existe une suite $g_k$ positive tendant vers 0 telle que pour $i>0$, $k \geqslant 0$, $\|E_{i-k}V_i\|_2 \leqslant g_k$

$\|V_i - E_{i+k}V_i\|_2 \leqslant g_{k+1}$

où $E_i$ désigne l'espérance conditionnelle à la tribu $\mathbb{F}_i$.

.Il existe une suite positive non décroissante $L_k$ telle que :

$\sum 1/nL_n < \infty$, $L_n - L_{n-1} = O(L_n/n)$, $g_n = o(1/\sqrt{n}.L_n)$

.Les $V_n$ sont uniformément intégrables

.$E_{k-m}[(S_{n+k}-S_k)^2/n] \longrightarrow \sigma^2$ quand $m,k,n \longrightarrow \infty$

Alors $S_n/\sqrt{n}$ tend en loi vers $N(0,\sigma)$, avec $S_n = \sum_{i=1}^n V_i$.

On prend $V_n = \sum_0^p \alpha_i \psi(X_n) Y_{n-i} + \sum_1^q \beta_i(\varphi_i(X_n) - E\varphi_i(X))$

$\mathbb{F}_n$ : sigma algèbre engendrée par les $X_p$, $p \leqslant n$.

On a $S_n = Z_n$

.$E_{n-k}(V_n) = 0$

.$\|E_{n+k}(V_n) - V_n\|^2 = \sum \alpha_i \alpha_j \sum_{l_1 < -i-k, l_2 < -j-k} s_{l_1} . s_{l_2} . E(\psi(X_n)^2 X_{n-i-l_1} X_{n-j-l_2})$

on prend $g_{k+1} = (EX^2 . E\psi^2)^{1/2} . (\sum |\alpha_i|) . \sum_{l \geqslant 0} |s_{-k-l}|$ qui décroit exponentiellement vers 0. En prenant $L_n = Cte.n$, les 2 premières hypothèses sont vérifiées.

.Le processus $V_n$ étant stationnaire, les $V_n^2$ sont uniformément intégrables.

.Le processus étant stationnaire et les $s_k$ décroissant exponentiellement vers 0, il est immédiat que la dernière hypothèse est vérifiée avec $\sigma^2 = (\alpha,\beta)\xi^t(\alpha,\beta)$.

On a donc la conclusion.

2). <u>Lorsque</u> $d>1$ :

On utilise un théorème de Rosenblatt ([25] p.77) :

Soit $(Z_n)_{n\in\mathbb{Z}^d}$ un champ centré strictement stationnaire et fortement mélangeant tel que :

$$E|\sum_{n_i=c_i}^{d_i} Z_{n_1,\dots,n_d}|^2 = h(d_1-c_1;\dots;d_d-c_d) \quad \text{où } h\longrightarrow\infty \text{ quand } d_i-c_i\longrightarrow\infty,$$

où $h(\alpha_1;\dots;\alpha_d)=o(h(\beta_1;\dots;\beta_d))$ dès que $\alpha_i\longrightarrow\infty$ $\forall i$ avec $\alpha_i=O(\beta_i)$ mais $\alpha_j=o(\beta_j)$ pour un certain $j$.

Si $\exists\ \delta>0$ tel que $E|\sum Z_n|^{2+\delta} = O(h(d_1-c_1;\dots;d_d-c_d)^{1+\delta/2})$

alors $\sum_{n_i=1}^{N} Z_n$ convenablement normalisé converge en loi vers $N(0,Id)$ quand $d_i-c_i\longrightarrow\infty$ $\forall i$.

Ici, $h(\alpha_1;\dots;\alpha_d)=\lambda\prod\alpha_i$, qui vérifie bien les conditions requises.

Nous allons vérifier la dernière condition pour $\delta=2$.

$$\begin{aligned}
EZ_t^4 = {} & \sum\alpha_{i1}\cdot\cdot\alpha_{i4}\sum E(\psi(X_{t1})\dots\psi(X_{t4})Y_{t1-i1}\cdots Y_{t4-i4} \\
& +\sum\beta_{i1}\cdot\cdot\beta_{i4}\sum E(\varphi_{i1}(X_{t1})-E\varphi_{i1}(X))\dots(\varphi_{i4}(X_{t4})-E\varphi_{i4}(X)) \\
& +4\sum\alpha_i\beta_{j1}\cdot\cdot\beta_{j3}\sum E(\psi(X_{t1})Y_{t1-i}\cdot(\varphi_{j1}(X_{t2})-E\varphi_{j1}(X))\cdot\cdot(\varphi_{j3}(X_{t4})-E\varphi_{j3}(X)) \\
& +4\sum\alpha_{i1}\cdot\cdot\alpha_{i3}\beta_j\sum E(\psi(X_{t1})Y_{t1-i1}\cdots\psi(X_{t3})Y_{t3-i3}(\varphi_j(X_{t4})-E\varphi_j(X)) \\
& +6\sum\alpha_{i1}\alpha_{i2}\beta_{j1}\beta_{j2}\sum E(\psi(X_{t2})Y_{t2-i2}\psi(X_{t2})Y_{t2-i2}(\varphi_{j1}(X_{t3})-E\varphi_{j1}(X)). \\
& \qquad (\varphi_{j2}(X_{t4})-E\varphi_{j2}(X))
\end{aligned}$$

Pour chaque terme, on regarde suivant que les $t_i$ sont distincts ou non et en développant les Y sur les X;l'on obtient, sachant (H3) :

.$t_1,\dots,t_4$ tous distincts : 0

.$t_1=\dots=t_4$ : $Cte.\prod(d_i-c_i)$

.deux des $t_i$ distincts : $\leqslant Cte.[\prod(d_i-c_i)]^2$

.3 des $t_i$ distincts : seul le 4ème terme de la somme n'est pas nul, on obtient des termes de la forme :

$$4\sum\alpha_{i1}\dots\alpha_{i3}\beta_j.Cte.E\psi^2(X_0)\ .EX_0^2\psi(X_0).(EX_0\varphi_j(X_0)-EX_0.E\varphi_j(X_0)).$$

$$\sum s_{-i3}s_{t-t3-i2}s_{t-t4-i1}+s_{-i3}s_{t-t3-i1}s_{t-t4-i2}+s_{t-t3-i1}s_{t-t3-i2}s_{t3-t4-i3}$$

qui est borné par :

$$C1.|s_{-i3}|.(\sum|s_k|)^2.\prod(d_i-c_i)\ +\ C2.(\sum|s_ks_{k+i1-i2}|).(\sum|s_k|).\prod(d_i-c_i).$$

On a donc $EZ_t^4 = O(\prod(d_i-c_i)^2)$ . □

Maintenant, en appliquant les lemmes 1 et 2, on peut simplifier les expressions de A et $\Lambda$ pour obtenir :

$$A_{i,j}=\sum_{k\neq 0}s_{k-i}s_{k-j}VarX_0E\psi'(X_0)+\ (\sum_{k\neq 0}s_{k-i}).(\sum_{k\neq 0}s_{k-j}).(EX_0^2).$$

$$\{\ \sum E\varphi_m'(X_0)E\varphi_l(X_0)\ D^mD^lL(E\varphi(X_0))\ +E\psi'(X_0)\ \}$$

$$\Lambda_{i,j}=\sum_{k\neq 0}s_{k-i}s_{k-j}VarX_0E\psi^2(X_0)+\ (\sum_{k\neq 0}s_{k-i}).(\sum_{k\neq 0}s_{k-j}).(EX_0^2).$$

$$\{\ E\psi^2(X_0)+\ \sum D^{k1}D^{l1}L(E\varphi(X_0))D^{k2}D^{l2}L(E\varphi(X_0)).E\varphi_{l1}'(X_0).$$

$$E\varphi_{l2}'(X_0).Cov(\varphi_{k1}(X_0),\varphi_{k2}(X_0))\ +\ 2\sum E\varphi_l'(X_0)E(\varphi_k(X_0)\psi(X_0))$$

$$D^kD^lL(E\varphi)\ \}.$$

<u>Etude du cas où</u> X <u>est centré</u>.

On a alors, en notant $T_{i,j}=\sum_{k\neq 0}s_{k-i}s_{k-j}$ :

$A = varX_0.E\psi'(X_0).T$

$\Lambda = varX_0\ E\psi^2(X_0).T$

<u>Remarque</u>: $\Lambda$ et A diagonalisent dans la même base, donc, en regardant la démonstration du théorème 3, on voit que les $((\hat{c}_j^n))$ sont asymptotiquement indépendantes.

On obtient $\Sigma^* = \dfrac{E\psi^2(X_0)}{\mathrm{Var}X_0(E\varphi'(X_0))^2} \cdot T^{-1} = V(\psi, X_0) \cdot T^{-1}$

On obtient pour $\Sigma^*$ la borne inférieure donnée par le résultat suivant :

**<u>THEOREME</u>** 6 : Sous des conditions suffisantes de régularité de F :

$V(\psi, X_0) \geqslant \dfrac{1}{\mathrm{Var}X_0 I(X_0) - 1}$ où $I(X)$ est l'information de Fisher de la loi de X.

<u>Démonstration</u>.

Soit $A(\psi, X) = \sup_{E\psi=0,\, |E\psi'|<\infty,\, EX\psi=0,\, E\psi^2<\infty} \dfrac{(E\psi'(X)^2)}{E\psi^2(X_0)}$

Posant $\theta(u) = \psi(u) - \dfrac{EX_0\psi(X_0)}{\mathrm{Var}X_0} \cdot (u - EX_0)$, on voit facilement que :

$A(\psi, X) = \sup_{E\theta(X)=0,\, |E\theta'(X)|<\infty,\, E\theta^2(X)<\infty} \dfrac{(E\theta'^2(X))}{E\theta^2(X)}$

$E\theta'(X) = E\psi'(X) - \dfrac{EX\psi(X)}{\mathrm{Var}X} = -\left( E\psi(X)\dfrac{f'(X)}{f(X)} + \dfrac{EX\psi(X)}{\mathrm{Var}X}\right)$ sous des conditions suffisantes de régularité de la densité f de la loi de X.

$E\theta'(X) = -E\left(\psi(X) - \dfrac{X-EX}{\mathrm{Var}X} EX\psi(X)\right) \cdot \left(f'/f(X) + \dfrac{X-EX}{\mathrm{Var}X}\right)$ car $E\psi(X)=0$.

Ecrivons l'inégalité de Schwarz :

$$(E\ \theta'(X))^2 \leqslant (E\theta^2(X)) \cdot E[f'/f(X) + (X-EX)/\mathrm{Var}X]^2$$

soit $A(\psi, X) \leqslant E[f'/f(X) + (X-EX)/\mathrm{Var}X]^2 = I(X) - 1/\mathrm{Var}X$ ,

et cette borne est atteinte uniquement pour :

$\psi(u) - \dfrac{u-EX}{\mathrm{Var}X} \cdot EX\psi(X) = \alpha \cdot \left(\dfrac{f'}{f}(u) + \dfrac{u-EX}{\mathrm{Var}X}\right)$, soit $\psi = \alpha \cdot \dfrac{f'}{f} + \beta \cdot \dfrac{u-EX}{\mathrm{Var}X}$,

la contrainte $EX\psi(X)=0$ entrainant $\alpha=\beta$. □

Si l'on s'intéresse maintenant aux fonctionnelles J permettant d'atteindre la borne du théorème 6, on s'aperçoit, en "intégrant" $\psi$ que c'est lorsque :

$$J(c) = \int f_c \text{ Log } f \, dx + 1/2 \text{ Log Var} X_0(c),$$

$f_c$ désignant la densité de la loi de $X_0(c)$

$$J_n(c) = 1/n^d \sum \text{Log } f(X_t(c)) + 1/2 \text{ Log Var}_n X_0(c).$$

Ce sont bien des fonctionnelles linéaires présentant au point cherché un maximum strict global, mais elles ne sont pas invariantes par changement d'échelle.

De plus, elles font intervenir la loi inconnue de X.
On pourrait alors chercher à construire des estimateurs adaptatifs, mais ce serait un gros effort pour un maigre résultat, car, nous le verrons plus loin, cela ne donnerait pas un estimateur efficace au sens de Fisher.

## I.4. Comparaison avec les méthodes usuelles.

Pour comparer la qualité de précision de notre estimateur avec les estimateurs usuels des moindres carrés, il faut se placer dans des conditions où cette comparaison est possible, soit sous l'hypothèse :

(COMP) : $d=1$, $EX_0=0$, $s_k=0$ pour $k<0$.

On choisit alors la normalisation $\mathbb{U}$ pour l'estimateur, et on compare donc avec $\hat{\hat{b}}_n$ l'estimateur $\hat{d}_n$ obtenu par minimisation de la pseudovraisemblance gaussienne $-1/2(n.\mathrm{Log}2\pi\sigma^2+{}^tY(n).[T_n(2\pi h)]Y^{-1}(n))$ où : $Y(n)=(Y_k), 1\leqslant k\leqslant n$, $h$ est la densité spectrale de $Y$, et $T_n$ est la matrice de Toeplitz des coefficients de Fourier de $2\pi h$.

Dans les applications, les estimateurs sont calculés par la méthode "aller-retour" due à Box et Jenkins.(voir [1] et [5])

On a : $n^{d/2}(\hat{d}_n-b) \xrightarrow{\lambda} N(0,\Gamma)$ (voir [3])

$n^{d/2}(\hat{\hat{b}}_n-b) \xrightarrow{\lambda} N(0,\Sigma^*)$ (voir th.2)

où $\Gamma = T^{-1}$ et $\Sigma^* = \dfrac{E\psi^2(X_0)}{EX_0^2.(E\psi'^2(X_0))}.T^{-1}$

Pour comparer la précision des deux estimateurs, on compare leur variance asymptotique, et il s'ensuit que :

Si le contraste linéaire employé est tel que $v(\psi,X)<1$, la méthode "contraste linéaire" est meilleure que "moindres-carrés". Le résultat est inversé si $V(\psi,X)>1$.

On remarque aussi que, puisque l'on a toujours $V(\psi,X)\geqslant \dfrac{1}{\mathrm{Var}X.I(X)-1}$ $V(\psi,X)>1$ dès que $\mathrm{Var}X.I(X)<2$, ce qui est le cas si $X$ est "proche" d'une loi gaussienne, où $\hat{d}_n$ est efficace.

Ces considérations présentent notre méthode par objectifs comme une alternative aux méthodes usuelles du second ordre dans les cas classiques d'estimation de paramètres d'un processus auto-régressif causal : utiliser un objectif linéaire quand on est loin de la situation gaussienne, et les moindres carrés lorsque l'on en est proche.

Simulations.

Les simulations portent sur des AR1, c'est à dire pour $b=(1,\alpha)$. On utilise comme objectif le cumulant standardisé d'ordre 4 :

$$J(c) = \left| \frac{E(X_0(c)-EX_0(c))^4}{(\mathrm{Var}X_0(c))^2} - 3 \right| .$$

On choisit celui-ci pour des facilités de simulation, car les hypothèses (H3) sont alors vérifiées dès que la loi de X est symétrique par rapport à l'origine. Pour une loi asymétrique, le choix du cumulant d'ordre 3 serait plus judicieux.

Numériquement, le calcul de $\hat{\alpha}_n$ se ramène alors à la recherche des racines d'un polynome (on s'en aperçoit facilement en calculant le gradient de $J_n$). Les calculs ont été effectués sur un Vax 750, en utilisant le programme IMSL de recherche de racines d'un polynome. Pour chaque loi, on se place à n=1000, et on en réalise 100 simulations. On regarde alors la loi de $\hat{\alpha}_{1000}$ à partir du 100-échantillon que l'on a.

Avec les mêmes données, on calcule l'estimateur par moindres carrés et sa loi avec la méthode de Box et Jenkins implantée sur Genstat. On compare alors les variances théoriques et numériques des deux méthodes.

Les variances théoriques sont :

$\Lambda=1-\alpha^2$ et $\Sigma^*=(1-\alpha^2).R$ où $R=\dfrac{EX^6EX^2-(EX^4)^2}{(3(EX^2)^2-EX^4)^2}$ lorsque X est centré.

On s'intéresse donc au rapport des deux variances R, d'autant plus petit que la nouvelle méthode est efficace.

On remarque que R est nul si et seulement si $X^3= X$ en loi (ceci se voit par l'inégalité de Schwarz appliquée à $X^3$et X), donc si et seulement si X suit une loi de Bernoulli, soit une loi de support {-1;1}. Dans ce cas particulier, il est facile de voir que toutes les hypothèses de consistance et d'asymptotique sont vérifiées, donc que $n^{d/2}(\hat{\alpha}_n - \alpha) \xrightarrow{\lambda} N(0,0)$, c'est à dire que $n^{d/2}(\hat{\alpha}_n - \alpha)$ tend en probabilité vers 0, ce qui indique que la vitesse est supérieure à $n^{d/2}$.(On n' est bien évidemment pas dans le cas où l'information de Fisher du modèle existe et où la borne pour $V(\psi,X)$ est calculée).

On s'intéresse à des lois ayant pour densité $p_m$ (voir ci-dessous), qui vérifient toutes les hypothèses ; on a choisi ces lois car elles sont particulièrement simples à simuler et qu'en faisant varier le paramètre m, on s'écarte plus ou moins de la gaussienne de façon intuitive :

$$p_m(x) = 1/2\sqrt{2\pi}\ .[\exp-\frac{(x+m)^2}{2} + \exp-\frac{(x-m)^2}{2}]$$

On a :

$$R = \frac{4m^6+18m^4+9m^2+6}{4m^8}$$

Donc en particulier :

. $\lim_{m\longrightarrow 0} R = +\infty$ : on s'approche de la loi gaussienne où la méthode n'est pas définie.

. $\lim_{m \to \infty} R = 0$ : plus les sommets sont éloignés, plus on est loin de la gaussienne.

Les simulations ont été effectuées pour différentes valeurs de $\alpha$. Nous donnons ici les résultats relatifs à $\alpha = -0.7$ .

Tableau des résultats.

| m | 50 | 10 | 1 | 0.5 |
|---|---|---|---|---|
| R numérique | $3.\ 10^{-3}$ | $1.8\ 10^{-2}$ | 7.05 | 4300 |
| R théorique | $4.\ 10^{-4}$ | $10^{-2}$ | 9.25 | 604 |

Pour les valeurs extrêmes de m, on voit que l'asymptotique n'est pas atteinte à n=1000, mais R numérique donne néanmoins une bonne vision de l'efficacité relative des deux méthodes.

Histogrammes des lois de $(\hat{\alpha}_n - \alpha)$ : voir en annexe. Ils illustrent le tableau précédent.

## I.5. Problèmes d'efficacité.

Lorsque d=1 et avec la normalisation $\hat{\theta} \in \mathbb{U}$, on a un modèle paramétré de paramètre $\theta=(c_1,\ldots,c_p)$. On calcule l'information de Fisher du modèle lorsqu'elle existe :

$I = \lim_{\infty} 1/n\ I_n(\theta)$, où $I_n(\theta)=E(\partial L_n/\partial c_i . \partial L_n/\partial c_j . 1/L_n^2\ )$,

$L_n$ étant la vraisemblance du processus X . On peut en effet montrer que le résultat est le même que si l'on calcule $I_n$ avec la vraisemblance de Y, il s'agit d'un changement de paramètre biunivoque.
On rappelle qu'alors, quel que soit l'estimateur du paramètre, sa variance asymptotique $\Sigma$ quand elle existe au sens décrit en I.3

vérifie:

$\Sigma - I^{-1}$ est s.d.p. Si $\Sigma = I^{-1}$, on dit que l'estimateur est efficace.

Sous des conditions de régularité et d'intégrations de la fonction caractéristique de la loi de X, on obtient (voir le calcul et les conditions plus loin) :

**THEOREME** 7 : $I_{i,j} = \mathrm{Var}X.I(X) \sum_{k\neq 0} s_{k-i}s_{k-j} + \sum_{k\neq 0} s_{k-i}s_{-k-j}$

$+ \{E[(X-EX)^2 f'^2/f^2]-1\}\, s_{-i}s_{-j} + (EX)^2 Ef'^2/f^2 (\Sigma f_k)^2$

$+ [EX.E(Xf'^2/f^2) - (EX)^2 Ef'^2/f^2](s_{-i}+s_{-j})\Sigma\, s_k$

$f$ étant la densité de la loi de $X_0$ .

Démonstration:

On calcule les vraisemblances à l'aide de leurs fonctions caractéristiques. Si $\varphi_{c,n}$ est la fonction caractéristique de la loi jointe de $(X_1(c),\ldots,X_n(c))$ , on a :

$\varphi_{c,n}(t_1,\ldots,t_n) = \Pi_{l=-\infty}^{+\infty} \varphi(\Sigma_{k=0}^{n}(c*s)_{k-l}t_k)$, où $\varphi$ est la fonction caractéristique de la loi de X.

Les hypothèses sont du type suivant :

.Régularités, dérivabilités de $\varphi$

.Intégrabilité de fonctions de $\varphi$ et de ses dérivées et majoration par des fonctions intégrables (Ceci pour échanger dérivation et intégration).

Ce sont celles qui permettent d'écrire le formulaire suivant :

$f(x) = \int \varphi(t)\exp -itx \;.1/2\pi$ , $\int \varphi'(t)\exp -itx = 2\pi i x f(x)$ ,

$\int t\varphi(t)\exp -itx = 2\pi i f'(x)$ , $\int t\varphi'(t)\exp -itx = 2\pi.-(f(x)+xf'(x))$ ,

$\int t^2\varphi(t)\exp -itx = -2\pi f''(x)$ , $\int t^2\varphi'(t)\exp -itx = 2\pi.-i(f'(x)+xf''(x))$ ,

$\int \varphi''(t)\exp -itx = 2\pi.-x^2 f(x)$ , $\int t\varphi''(t)\exp -itx = -2\pi i.(2xf(x)+x^2 f'(x))$

$$\int t^2\varphi''(t)\exp -itx = 2\pi.-(2f(x)+4xf'(x)+x^2f''(x))$$

On écrit alors :

$$\frac{\partial\varphi_{c,n}}{\partial c_i}=\Sigma_l\ (\Sigma_{k=0}^{n}s_{k-l-i}t_k)\varphi'(\Sigma_{k=0}^{n}(c*s)_{k-l}t_k)\ \prod_{m\neq l}\varphi(\Sigma_{k=0}^{n}(c*s)_{k-m}t_k)$$

$$\frac{\partial^2\varphi_{c,n}}{\partial c_i\partial c_j} = \Sigma_l(\Sigma_{k=0}^{n}s_{k-l-i}t_k)(\Sigma_{k=0}^{n}s_{k-l-j}t_k)\varphi''(\Sigma_{k=0}^{n}(c*s)_{k-l}t_k)\ .$$

$$\prod_{m\neq l}\varphi(\Sigma_{k=0}^{n}(c*s)_{k-m}t_k)\ +\ \Sigma_l\Sigma_{m\neq l}(\Sigma_{k=0}^{n}s_{k-l-i}t_k)(\Sigma_{k=0}^{n}s_{k-m-j}t_k).$$

$$\varphi'(\Sigma_{k=0}^{n}(c*s)_{k-l}t_k)\varphi'(\Sigma_{k=0}^{n}(c*s)_{k-m}t_k).\prod_{p\neq m\,et\,l}\varphi(\Sigma_{k=0}^{n}s_{k-p}t_k).$$

En écrivant ces relations en c=b, en les multipliant par $\exp -i\Sigma t_k x_k$ et en les intégrant à l'aide du formulaire précédent, on obtient

$$\frac{1}{f_{n,c}}.\frac{\partial f_{n,c}}{\partial c_i}\Bigg|_{c=b}\quad \text{et}\quad \frac{1}{f_{n,c}^2}.\frac{\partial f_{n,c}}{\partial c_i}.\frac{\partial f_{n,c}}{\partial c_j}-\frac{1}{f_{n,c}}.\frac{\partial^2 f_{n,c}}{\partial c_i\,\partial c_j}\Bigg|_{c=b}$$

puis on en prend l'espérance pour obtenir :

$$((I_n)_{i,j} = -\ (n+1)s_{-i}s_{-j} + \Sigma^n_{k\neq 0,=-n}\ s_{k-i}s_{-k-j}.(n-|k|+1) + (n+1)s_{-i}s_{-j}EX^2f'^2/f^2$$

$$EX^2.Ef'^2/f^2.\Sigma_{l=0}^{n}\Sigma^n_{k\neq l,=0}\ s_{k-l-i}s_{k-l-j}$$

$$+(EX)^2Ef'^2/f^2.(\Sigma_{l_1,l_2}\Sigma^n_{k\neq l_1 et l_2,=0}\ s_{k-l_1-i}s_{k-l_2-j}\ -\ \Sigma_l\Sigma^n_{k\neq l,=0}\ s_{k-l-i}s_{k-l-j})$$

$$+EX.EXf'^2/f^2.[\Sigma_{l_1}\Sigma^n_{l_2\neq l_1,=0}(s_{-i}s_{l_2-l_1-j}+s_{-j}s_{l_2-l_1-i})].$$

La limite $I_{i,j}$ s'en déduit aussitôt.

Lorsque X est centré et que le filtre s est causal:

$$I^{-1}=\ \frac{1}{VarX.I(X)}\ T^{-1}\quad .$$

On s'apercoit alors que la méthode d'estimation par objectif de type linéaire n'est jamais efficace, et que sous les hypothèses

(COMP), la méthode Box-Jenkins est efficace uniquement dans le cas gaussien.

Réflexion sur un estimateur adaptatif.

Dans le cas où s est causal, il est connu que l'estimateur du maximum de vraisemblance est efficace.

Le problème adaptatif est le suivant : est-il possible, lorsque F est inconnue, de faire aussi bien que si F est connue ?
Un tel problème est très largement étudié par Bickel ([3]) dans le cas d'observations indépendantes : il indique comment et dans quels cas cela est possible. Transposant ses idées au cas d'observations corrélées, on regarde la condition de Stein ([26])de décorrélation du paramètre estimé et de l'estimateur de densité de cette façon :
Supposons que la loi de $X_0$ est centrée et de la forme :
$f=(1-t)g+th$. Calculons l'information de Fisher asymptotique du modèle paramétré par (b,t). Avec un calcul analogue à celui développé précédemment, on obtient facilement que le terme croisé en $b_i$,t est proportionnel à $s_{-i}$. Pour que l'adaptation soit possible, ce terme devrait être nul. Ceci implique alors : $\forall i=1...p$, $s_{-i}=0$, et puisque s a un inverse de longueur p, ceci implique : $\forall i>0$, $s_{-i}=0$.
Ceci suggère qu'une condition nécessaire pour mener un calcul adaptatif serait que le filtre s soit causal. Nous faisons alors la conjecture :

conjecture : Un estimateur adaptatif efficace existe seulement si s est causal.

Si ce résultat est vrai, il s'agit ensuite de trouver sous quelles conditions sur la loi F on peut construire un estimateur adaptatif efficace et d'en proposer une construction.

Les calculs menés par Bickel utilisent de manière constante l'indépendance des observations ; dans la situation qui nous intéresse, il faudrait étudier d'autres calculs s'adaptant au cas mélangeant, ce qui nécessite certainement un important travail que nous n'avons pas mené ici.

## I.6. Identification = estimation de l'ordre p.

L'idée est d'utiliser pour estimer p une méthode analogue à la méthode de vraisemblance compensée. (Voir [1]).

Si d>1, on supposera $p_1=\ldots=p_d=p$.
On introduit le paramètre supplémentaire q paramétrisant l'ordre, et l'on pose :

$A_n(q,c) = n^{d\alpha} J_n(q,c) - \lambda q$ , où $J_n(q,c)$ est la fonction d'objectif lorsque définie lorsque le modèle est d'ordre q.
On choisit alors comme estimateur $(q_n,c_n)$ le couple (q,c) qui maximise $A_n(q,c)$ pour $q \in \{0,\ldots,Q\}$ et c vérifiant la condition de normalisation choisie.

Q peut être arbitrairement grand, mais il doit être fixé.
On a alors :

**THEOREME** 8 : Pour $0<\alpha<1$ et $\lambda>0$, sous les conditions (C1) et (C2), et (A1)....(A3) :

$$(q_n,c_n) \xrightarrow{P} (p,b)$$

et l'asymptotique pour $c_n$ est la même que lorsque l'on n'estime pas p.

<u>Démonstration</u> :

1). La 1ère partie du théorème découle du fait que, si $\alpha\in]0;1[$, et si $q>p$, $n^{d\alpha}[J_n(q,c_{*n})-J_n(q,b_*)]\xrightarrow{P}0$ où $c_{*n}$ maximise $J_n(q,c)$ à $q$ fixé et où $b_*=(b,0..0)$.

En effet, à $q>p$ fixé, tous les résultats précédents sont valables (le modèle peut être considéré comme étant d'ordre $q$ et de paramètre $b_*$), et en particulier :

$$J_n(q,c_{*n})-J_n(q,b) = 1/2.{}^t(c_{*n}-b_*)\ [D^2J_n(c_{*n})+ O_p(1)]\ (c_{*n}-b_*)$$

et le résultat est alors immédiat sachant que $n^{d/2}(c_{*n}-b_*)$ est tendu. On en déduit immédiatement :

Si $q>p$, $A_n(q,c_{*n})-A_n(p,b_{*n})\xrightarrow{P} -\lambda\ (q-p) < 0$ .

Maintenant :

$\{q_n>p\} \subset \{\ \exists\ q\in\{p+1...Q\}/\ A_n(q,c_{*n})-A_n(p,b_{*n})\geqslant 0\ \}$

donc $P(q_n>p) \leqslant \Sigma_{q=p+1}^{Q}P(A_n(q,c_{*n})-A_n(p,b_{*n})\geqslant 0)$, et $P(q_n>p)\xrightarrow[n\to\infty]{}0$.

D'autre part, on a de la même façon :

$$P(q_n<p) \leqslant \Sigma_{q=1}^{p-1}\ P(A_n(q,c_{*n})-A_n(p,b_{*n})\geqslant 0)$$

$$\leqslant \Sigma_{q=1}^{p-1}\ P(J_n(q,c_{*n})\geqslant J_n(p,b_{*n})-\lambda p/n^{d\alpha}\}.$$

Il est immédiat, du fait de l'optimalité stricte de $J(b)$ et du fait que l'optimum est atteint que l'on a, si la normalisation choisie est $\mathbb{T}$ ou $\mathbb{U}$ :

$$\sup_c J(q,c) < \sup_b J(p,b).$$

Il est de même, grâce à cela et à la convergence uniforme de $J_n(q,c)$ vers $J(q,c)$, facile de montrer que, pour $q<p$ :

$\forall\varepsilon>0$, $P(J_n(q,c_{*n})\geqslant J_n(p,b_{*n})-\varepsilon) \longrightarrow 0$ quand $n\to\infty$.

On en déduit que $P(q_n<p)\longrightarrow 0$ quand $n\to\infty$, et donc que $q_n\xrightarrow{P} p$ .

La convergence p.s. de $q_n$ n'est pas étudiée ici, elle nécessite des calculs plus fins.

2). La 2ème partie du théorème découle du fait que $q_n \longrightarrow p$ en probabilité et que sur $\{q_n=p\}$, l'asymptotique est la même que celle étudiée précédemment.

## I.7. Robustesse.

Robustesse à une sous-paramétrisation (cas d=1).

On se place dans la situation où l'on a sous estimé p, c'est à dire :

$b = (b_1,b_2)$, $b_1 \in \mathbb{R}^{p-q}$ et $b_2 \in \mathbb{R}^q$, $\|b_2\|=\varepsilon_q$ est petit, et l'on cherche b dans $\mathbb{R}^{p-q}$, c'est à dire que l'on maximise $J_n(p-q,c)$ en c.

Soit $\omega$ le module de continuité de $J(p,c)$ en b et $\omega_{-1}$ le module de continuité de $J^{-1}$(fonction inverse de $J(p,c)$) en b.

On a le résultat suivant :

**THEOREME** 9 : $\forall\ \delta>0$, $P[\|c_n-b\| \leqslant \omega_{-1}(\delta+\omega(\varepsilon_q))] \longrightarrow 0$ quand $n \longrightarrow \infty$ où $c_n$ désigne l'estimateur complété par q 0.

Ce théorème garantit une certaine robustesse lorsque l'objectif J est assez régulier, ainsi que son inverse, autour de b.

Démonstration.

On reprend les notations de la démonstration de consistance.

Sur $E_n(\delta/2)$, on a :

$J(c_n)+\delta/2 \geqslant J_n(c_n) \geqslant J_n(b_1,0) \geqslant J(b_1,0)-\delta/2 \geqslant J(b)-\delta/2-\omega(\varepsilon_q)$.

Donc $|J(c_n)-J(b)| \leqslant \delta+\omega(\varepsilon_q)$, et $\|c_n-b\| \leqslant \omega_{-1}(\delta+\omega(\varepsilon_q))$.

Donc $P[\|c_n-b\| \leqslant \omega_{-1}(\delta+\omega(\varepsilon_q))] \leqslant P(E_n(\delta/2))$.

Robustesse au bruit.

On suppose maintenant que le signal sortant est bruité :

$Y = s*X + \varepsilon$, où $(\varepsilon)$ est un bruit indépendant de X.

La question est la suivante : L'estimateur de maximum d'objectif conserve-t-il de bonnes propriétés ?

Nous étudierons seulement un cas particulier du problème, mais relativement représentatif, le cas où J est un cumulant standardisé et où les $\varepsilon_t$ sont i.i.d de loi gaussienne.

On a alors le résultat suivant :

**THEOREME** 6 : Dans le cadre décrit ci-dessus, on a :

$\forall\ \delta > 0,\ \exists\ \alpha > 0$ ($\alpha$ ne dépendant pas de X) tel que :

$$|\mathrm{Var}\varepsilon_0/\mathrm{Var}X_0| < \alpha \Longrightarrow P(\limsup_n \|b_n - b\| \leqslant \delta) = 1.$$

Ce résultat garantit une bonne robustesse au bruit lorsque le rapport signal sur bruit est suffisamment élevé.

Démonstration.

Par un calcul facile on obtient:

$$J(c) = \frac{\sum (c*s)_k^m}{\left(\sum (c*s)_k^2 + \sum\left(\begin{smallmatrix} p; & ;p \\ 0, & ,0 \end{smallmatrix}\right) c_k^2 . \mathrm{Var}\varepsilon_0/\mathrm{Var}X_0\right)^{m/2}} \cdot h(X_0)$$

$$= g(c, \mathrm{Var}\varepsilon_0/\mathrm{Var}X_0) . h(X_0)$$

$h(X_0)$ étant le cumulant d'ordre m standardisé de $X_0$.

On choisit l'estimateur $b_n$ comme réalisant le maximum global de $J_n(c)$ sur $\mathbb{T}$ (il peut y avoir plusieurs choix possibles).

Soit $c(\eta)$ une suite sur $\mathbb{T}$ de maximisateurs de $J(c)$ lorsque $\mathrm{Var}\varepsilon_0$ tend vers 0. Etant sur un compact, elle a au moins une valeur d'adhérence $c_*$. Par passage à la limite, $g(c_*,0)=0$. Or, pour

$\mathrm{Var}\varepsilon_0=0$, on sait que J a un maximisateur unique b. Donc $c(\eta)$ a une seule valeur d'adérence b qui est ainsi limite de la suite. Ceci s'écrit :

(1) $\forall\delta>0, \exists\alpha>0$ tel que $|\mathrm{Var}\varepsilon_0/\mathrm{Var}X_0|<\alpha \Rightarrow \|c(\eta)-b\|<\delta$, $\eta=\mathrm{Var}\varepsilon_0$.

Il n'y a, pour tout $\eta$, qu'un nombre fini de valeurs de $c(\eta)$ (g est une équation polynomiale). $\alpha$ dans (1) peut donc être choisi indépendamment de $c(\eta)$.

D'autre part, rappelons qu'en étudiant $J_n(c)-J_n(c)$, il est facile de montrer que $J_n(c)$ converge presque surement vers $J(C)$ uniformément en c. Soit $\Omega = \{\omega/\lim_n \sup_{\mathbb{T}} |J_n(c)-J(c)|\}$, $P(\Omega)=1$.
Soit $\omega$ dans $\Omega$, et soit $b_*$ une valeur d'adhérence de $b_n(\omega)$. Il est facile de voir que $b_*$ est un $c(\eta)$ (=maximisateur de $J(c)$), et donc que $\|b_*- b\|<\delta$ dès que $|\mathrm{Var}\varepsilon_0/\mathrm{Var}X_0|< \alpha$.
On en déduit aisément le théorème.

## I.8. Moments infinis et objectifs cumulants.

Tout au long de cette étude, on a supposé l'existence de moments finis sur $X_0$. Ceci n'est certainement pas une condition nécessaire. Intuitivement, lorsque les moments de $X_0$ ne sont pas bornés, on peut s'attendre à ce que la méthode fonctionne encore "mieux" du fait que la fonction est plus pointue autour de son optimum.
Des essais numériques effectués sur un AR1 où X suit une loi de Cauchy montrent que le procédé fonctionne encore et même mieux.
Les résultats théoriques suivants confirment ce résultat dans le cas d' objectifs de type "cumulants" et pour d=1.
Dans tout le paragraphe, on se placera donc dans le cadre suivant :
$J_n(c)=|$cumulant standardisé empirique d'ordre m de $X_0(c)|$

On considère les hypothèses suivantes :

(H1) : .Les $X_t$ sont i.i.d.

.$\exists\alpha>0$ tel que : $\forall\varepsilon>0$, $X_t\in L^{\alpha-\varepsilon}$, $EX_t=0$ si $\alpha>1$,

$$\text{et } \forall\varepsilon>0,\ \frac{\sum X_t^p}{n^{p/\alpha-\varepsilon}} \xrightarrow{P} +\infty \text{ pour } p=2 \text{ et } p=m$$

.$X_t^m$ est positif.

(H2) : Le filtre s est causal, i.e. $s_k=0$ pour $k<0$.

<u>Remarque</u> :

L'hypothèse suivante implique (H1) :

(H') : Les $X_t$ sont i.i.d., dans le domaine d'attraction d'une loi stable d'index $\alpha \in ]0,2[$, et :

a). Si $\alpha\leqslant1$, les $X_t$ sont positifs

b). Si $\alpha>1$, les $X_t$ sont centrés ($EX_t=0$) et m (ordre du cumulant utilisé) est pair.

Le fait que (H′) $\Longrightarrow$ (H1) est un résultat de Miller (voir [20]).

On a alors les résultats suivants :

**THEOREME** 11 : Si (H1), alors $J_n(b) \longrightarrow +\infty$ p.s.

$$\text{et} \quad \frac{J_n(c)}{J_n(b)} \longrightarrow h(c) = \left| \frac{\Sigma(c*s)_k}{(\Sigma(c*s)_k^2)^{1/2}} \right|$$

en probabilité, et la convergence est uniforme si $c$ varie dans un compact. La convergence n'a pas lieu p.s.

**THEOREME** 12 : Si (H1), alors $\hat{b}_n \xrightarrow{P} b/\|b\|$ et $\hat{\hat{b}}_n \xrightarrow{P} b/b_0$.

Ce théorème est une application immédiate du théorème 1 et du théorème 11.

On a ensuite le résultat suivant, indiquant une vitesse de convergence lorsque le filtre est causal :

**THEOREME** 13 : Si (H1) et (H2), alors $\forall\ \delta>0$ :

.Si $\alpha \geqslant 1$, $n^{1/(\alpha+\delta)}(\hat{\hat{b}}_n - b) \xrightarrow{P} 0$

.Si $\alpha \leqslant 1$, $n^{1-\delta}(\hat{\hat{b}}_n - b) \xrightarrow{P} 0$.

Ce théorème montre bien que sous les hypothèses considérées, la vitesse de convergence est plus grande que lorsque l'on a l'existence de moments finis.

On utilisera constamment les résultats suivants, dus respectivement à Chaterjii et Kanter ([6] et [14]) :

(C) : Si $Z_1,\dots,Z_n,Z$ sont des variables aléatoires vérifiant :

ou bien $Z \in L^p$, $0<p<2$, $p \neq 1$ et $P(|Z_n| \geqslant x) \leqslant P(|Z| \geqslant x)$ $\forall x$,

ou bien $Z \in L^1$ et $P(|Z_n| \geqslant x \ /Z_1,\dots,Z_{n-1}) \leqslant P(|Z| \geqslant x \ /Z_1,\dots,Z_{n-1})$

Alors $\dfrac{\Sigma_1^n(Z_k-\alpha_k)}{n^{1/p}} \longrightarrow 0$ p.s. avec :

$\alpha_k=0$ si $0<p<1$ et $\alpha_k=E(Z_k/Z_1,\ldots,Z_{k-1})$ si $1\leqslant p<2$.

En particulier, les hypothèses de ce théorème sont vérifiées dès que $Z_k$ est un processus stationnaire.

(K) : Si $X=a*Z$ et $Y=b*Z$ où $Z_n$ est i.i.d. positif avec $EZ_n=+\infty$, et où $\Sigma|a_k|<\infty$ et $\Sigma|b_k|<\infty$, alors :

$\dfrac{\Sigma X_n}{\Sigma Y_n} \xrightarrow{P} \dfrac{\Sigma a_k}{\Sigma b_k}$ . La convergence n'a pas lieu p.s.

Si $C_m$ est le cumulant de X d'ordre m, on a :

$C_m= \Sigma_{k=0}^{p}\lambda_k(EX^m)^{a_{k,m}}\ldots\ldots(EX)^{a_{k,1}}$

où : $\Sigma_{j=1}^{m} j.a_{k,j}= m$, $a_{0,m}=1$ et $a_{0,j}=0$ pour $j\neq m$, $a_{k,m}=0$ pour $k>0$, et donc $\Sigma a_{k,j}>1$ pour $k\neq 0$ (i.e. $\geqslant 2$)

<u>Démonstration du théorème 11</u>.

Notons $S_k(c) = \Sigma_{t=1}^{n}(X_t(c))^k$, $S_k(b) = S_k$. On a alors :

$$J_n(b) = \left|\frac{\lambda_0.1/n.S_m+\Sigma_{k=1}^{p}\lambda_k.\Pi_{j=1}^{m}(S_j/n)^{a_{k,j}}}{(1/n.S_2-(1/n.S_1)^2)^{m/2}}\right|$$

On choisit $\varepsilon$ vérifiant :

$$0<\varepsilon<\frac{m/2-1}{m/2+1} \quad \text{et} \quad 0<\varepsilon< \inf_{k>0}\frac{\Sigma_{j=1}^{m-1}a_{k,j}-1}{\Sigma_{j=1}^{m-1}a_{k,j+1}}$$

On écrit alors de la façon suivante :

$$J_n(b) = n^{m/2-1-\varepsilon(m/2+1)}.\left|\frac{N_n}{D_n}\right| \quad \text{avec :}$$

$$N_n = \lambda_0 \cdot \frac{S_m}{n^{m/\alpha-\varepsilon}} + \Sigma_{k=1}^{m} \lambda_k \cdot \frac{1}{n^{\Sigma a_{k,j}-1-\varepsilon(\Sigma a_{k,j}+1)}} \cdot \Pi_{j=1}^{m-1} \left[\frac{S_j}{n^{j/\alpha+\varepsilon}}\right]^{a_{k,j}}$$

$$D_n = \frac{1}{n^{2/\alpha+\varepsilon}} \cdot S_2 - 1/n \cdot \left(\frac{1}{n^{1/\alpha+\varepsilon/2}} \cdot S_1\right)^2$$

On a alors :

$$\frac{S_m}{n^{m/\alpha-\varepsilon}} \longrightarrow +\infty \text{ p.s. par (H1),}$$

et $\forall j$, $\dfrac{S_j}{n^{j/\alpha+\varepsilon}} \longrightarrow 0$ p.s. par (C), car $X_t^j$ est i.i.d. dans $L^{\alpha/j-\varepsilon}$

et si $\alpha>1$ et $j=1$, $E(X_t/X_1,\ldots,X_{t-1})=EX_t=0$

Il s'ensuit que $D_n \longrightarrow 0$, $N_n \longrightarrow +\infty$, et l'on conclut facilement que $J_n(b)$ tend presque surement vers plus l'infini.

Démontrons maintenant que $C_m^n(c)/C_m^n(b) \longrightarrow \Sigma_k (c*s)_k^m$ en probabilité.

$$\left|\frac{C_m^n(c)}{C_m^n(b)}\right| = \left[\frac{S_m(c)}{S_m(b)} + \Sigma_{k=1}^{p} \mu_k \cdot \frac{1}{n^{\Sigma a_{k,j}-1}} \cdot \Pi_{j=1}^{m-1} (S_j(c))^{a_{k,j}} \frac{1}{S_m(b)}\right] \cdot \left[\frac{S_m(b)}{S_m+\Sigma\mu_k \cdots}\right]$$

où $\mu_k=\lambda_k/\lambda_0$.

Il est immédiat, en regardant les calculs faits précédemment, que le 2ème terme du produit tend p.s. vers 1.

D'autre part, $\forall k\neq 0$ :

$$\frac{1}{n^{\Sigma a_{k,j}-1}} \cdot \Pi_{j=1} (S_j(c))^{a_{k,j}} \frac{1}{S_m(b)} = \frac{1}{n^{\gamma-\varepsilon.\beta}} \cdot \Pi \left[\frac{S_j(c)}{n^{j/\alpha+\varepsilon}}\right] \cdot \left[\frac{S_m(b)}{n^{m/\alpha-\varepsilon}}\right]^{-1}$$

où $\beta=\Sigma a_{k,j}+1$ et $\gamma$ est précisé plus loin.

$X(c)_t$ est stationnaire, et $X(c)_t \in L^{\alpha/j-\varepsilon}$ $\forall\varepsilon>0$.

Donc si $\alpha<1$, on peut appliquer (C), et $S_j(c)/n^{j/\alpha+\varepsilon} \longrightarrow 0$ p.s.

Ici on a $\gamma=\Sigma a_{k,j+1}$. On choisit $\varepsilon<\operatorname{Inf}_{k>0} \dfrac{\Sigma a_{k,j}-1}{\Sigma a_{k,j}+1}$, et en utilisant (H1), on obtient la convergence presque sûre du 2ème terme de la somme vers 0.

Pour $\alpha\geqslant 1$, on met $S_1(c)/n$ à la place de $S_1(c)/n^{1/\alpha+\varepsilon}$ et on a aussi $S_j(c)/n^{j/\alpha+\varepsilon} \longrightarrow 0$ p.s. par (C) pour $j>1$ et par la loi des grands nombres pour $j=1$ puisque $E(X_t)=0$.

Ici $\gamma=\Sigma a_{k,j}-2+1/\alpha$, et puisque $\Sigma a_{k,j}\geqslant 2$ dès que $j\neq m$, on peut choisir $0<\varepsilon<\mathrm{Inf}_k \dfrac{1/\alpha}{\Sigma a_{k,j}+1}$ et l'on conclut comme précédemment.

Il ne reste donc plus qu'à étudier $S_m(c)/S_m(b)$.

Notons $\gamma_k= (c*s)_k$.

On a : $S_m(c)=\Sigma_{t=1}^n(\Sigma_k\gamma_k X_{t-k})^m$

$$=\Sigma_{t=1}^n(\Sigma_k\gamma_k^m X_{t-k}^m) + \Sigma_1^n \Sigma_{\beta\in B} C(\beta).Z_{\beta,t}$$

où $B= \{\beta= (\beta_1,\ldots,\beta_m)/ \beta_j$ entier, $\beta_j\geqslant 0\ \forall j$,

$\exists j\neq k$ tels que $\beta_j\neq 0$ et $\beta_k\neq 0$, et $\beta_1+\ldots+\beta_m=m\}$.

(B contient un nombre fini d'éléments)

et où $Z_{\beta,t}= \Sigma_{k_1\ldots k_m\neq 2\text{à}2} \gamma_{k_1}^{\beta_1}\ldots\gamma_{k_m}^{\beta_m}.X_{t-k_1}^{\beta_1}\ldots X_{t-k_m}^{\beta_m}$.

On va montrer d'abord que $\forall\beta\in B$ $\Sigma_{t=1}^n Z_{\beta,t} / \Sigma_{t=1}^n X_t^m \longrightarrow 0$ p.s.

Il est quasi immédiat que $\forall\varepsilon>0$, $Z_{\beta,t}\in L^{\alpha/q-\varepsilon}$ où $q=\max(2,\max \beta_j)$, $2\leqslant q<m$ car deux $\beta_j$ au moins sont non nuls; $\dfrac{\alpha}{q}<1$, $Z_{\beta,t}$ est stationnaire, on peut donc appliquer (C).

On écrit alors :

$$\Sigma_{t=1}^n Z_{\beta,t} / \Sigma_{t=1}^n X_t^m = n^{(q-m)/\alpha+\varepsilon}.\frac{1}{n^{q/\alpha+\varepsilon}}.\Sigma_1^n Z_{\beta,t}.\left[\frac{1}{n^{m/\alpha-\varepsilon}}\Sigma_1^n X_t^m\right]^{-1}$$

qui tend p.s. vers 0 en choisissant bien $\varepsilon$.

Enfin, $\dfrac{\Sigma_{t=1}^n(\Sigma_k\gamma_k^m X_{t-k}^m)}{\Sigma_{t=1}^n X_t^m} \xrightarrow{P} \Sigma_k \gamma_k^m$ grâce à (K), et il n'y a pas de convergence p.s.

Résultat et démonstration sont les mêmes pour $C_2^n(c)/C_2^n(b)$, et l'on a alors $J_n(c)/J_n(b)\longrightarrow h(c)$ en probabilité.

Pour achever la démonstration, reste à prouver l'uniformité de la

convergence sur un compact. Il suffit de le faire pour $C_m^n(c)/C_m^n(b)$ et pour $C_2^n(c)/C_2^n(b)$. Nous ferons la démonstration pour le premier terme, valable aussi pour le deuxième.

En regardant le théorème de Prokhorov, on s'aperçoit qu'il est suffisant de montrer :

$\left| \dfrac{C_m^n(c) - C_m^n(d)_n}{C_m^n(b)} \right| \leqslant \omega_n(\|c-d\|).\Delta_n$, où $\omega_n$ est une fonction déterministe convergeant uniformément vers $\omega$ continue en 0 avec $\omega(0)=0$, et où $\Delta_n$ ne dépend pas de c et d et converge vers $\Delta$ borné.

Avec les mêmes notations que précédemment, on écrit :

$$\frac{C_m^n(c) - C_m^n(d)}{C_m^n(b)} = \left[ \frac{\Sigma_{t=1}^n \Sigma_k [(c*s)_k^m - (d*s)_k^m] X_{t-k}^m}{\Sigma_{t=1}^n X_t^m} + \frac{\Sigma_{t=1}^n \Sigma_{\beta\in B} C(\beta)(Z_{\beta,t}(c) - Z_{\beta,t}(d))}{\Sigma_t^m X_t} \right.$$

$$(\alpha) \qquad \left. + \Sigma_{k=1}^n \mu_k . \frac{1}{{}_n\Sigma_j a_{k,j} - 1} . \Pi_{j=1}^{m-1} (S_j(c))^{a_{k,j}} \frac{1}{S_m(b)} + \text{id. avec d} \right]$$

$$. \frac{S_m(b)}{S_m(b) + \Sigma_k \mu_k \cdot \cdot \cdot \cdot \cdot}$$

Comme précédemment, le deuxième terme du produit tend vers 1.

$\forall j,\ |S_j(c)| = |\Sigma_{t=1}^n (\Sigma_k (c*s)_k X_{t-k})^j| \leqslant \Sigma_{t=1}^n (\Sigma_k g_k |X_{t-k}|)^j$,

avec $g_k = \Sigma_{j=0}^p |s_{k-j}|$.

On a la même majoration pour les $S_j(d)$. Ceci montre, utilisant encore les mêmes arguments, que tous les termes du premier facteur du produit ($\alpha$) sauf le premier sont bornés indépendamment de c et d par une quantité ayant une limite finie (par (K)).

Regardons maintenant le premier terme :

$$\omega_n(c,d) = \frac{\Sigma_{t=1}^n \Sigma_k [(c*s)_k^m - (d*s)_k^m]\ X_{t-k}^m}{\Sigma_t X_t^m}$$

$$= \frac{\Sigma_{t=1}^n \Sigma_k ((c-d)*s)_k . P((c*s)_k ; (d*s)_k) . X_{t-k}^m}{\Sigma_t X_t^m}$$

où P est un polynome homogène de degré m-1 à coefficients positifs. On peut le borner par $P(|g_k|,|g_k|)$.

$|((c-d)*s)_k| \leqslant \|c-d\|.\lambda$ où $\lambda$ est une constante car les $s_k$ sont bornés. Alors :

$$|\omega_n(c,d)| \leqslant \lambda.\|c-d\|.\frac{\Sigma_t(\Sigma_k\eta_k X^m_{t-k})}{\Sigma_t X^m_{t-k}}\ ,\ \eta_k=P(|g_k|;|g_k|),$$ qui a la forme

cherchée puisque, par (K), $\dfrac{\Sigma^n_{t=1}(\Sigma_k\eta_k X^m_{t-k})}{\Sigma^n_{t=1}X^m_{t-k}} \xrightarrow{P} \Sigma_k\eta_k$.

<u>Démonstration du théorème 13</u>.

On choisit ici la normalisation $b_0=1$ qui donne encore un estimateur $\hat{\hat{b}}_n$ consistant.

On écrit le développement de Taylor :

$$-\text{Grad } J_n(b) = {}^t(\hat{\hat{b}}_n-b).\int_0^1 D^2J_n(b+u(\hat{\hat{b}}_n-b))\ du\ .$$

On a :

$$J_n(c) = \frac{N_n(c)}{(D_n(c)^{m/2}} \text{ où } N_n= C_m(c) \text{ et } D_n= C_2(c).$$

$$\frac{\partial N_n}{\partial c_i} = [1/n.\Sigma(X_t^{m-1}(c)Y_{t-i}$$
$$+\Sigma^p_{k=1}\lambda_k\frac{1}{\Sigma a_{k,j}}\Sigma_j j a_{k,j}S_j(c)^{a_{k,j}-1}\prod_{l\neq j}S_l(c)^{a_{k,l}}(\Sigma(X_t^{j-1}(c)Y_{t-i})].$$

$$\frac{\partial D_n}{\partial c_i} = 2/n.\Sigma X(c)_t\ Y_{t-i}-2/n^2(\Sigma Y_{t-i}).(\Sigma X(c)_t)$$

$$\frac{\partial^2 D_n}{\partial c_i\partial c_j} = 2/n.\Sigma Y_{t-i}Y_{t-j}-2/n^2.\Sigma Y_{t-i}\Sigma Y_{t-j}$$

$$\frac{\partial^2 N_n}{\partial c_i\partial c_j} = \Sigma_k\frac{\lambda_k}{\Sigma a_{k,j}}.\Sigma_j j a_{k,j}(a_{k,j}-1)(S_j(c))^{a_{k,j}-2}.(\Sigma(X(c)_t^{j-1}).Y_{t-i}).$$
$$(\Sigma(X(c)_t^{j-1}).Y_{t-j}).\Pi_{l\neq j}(S_l(c)^{a_{k,l}})$$

$$+\Sigma_k\frac{\lambda_k^n}{{}_n\Sigma a_{k,j}}.\Sigma_{j,l}j a_{k,j}l a_{k,l}(S_j(c))^{a_{k,j}-1}.(S_l(c))^{a_{k,l}-1}.\prod_{p\neq j \text{ et } l}(S_p(c))^{a_{k,p}}$$
$$.(\Sigma(X(c)_t^{j-1})\ Y_{t-i}).(\Sigma(X(c)_t^{l-1}).Y_{t-j})$$

On a ensuite :

$$\frac{\partial J_n(c)}{\partial c_i} = \frac{\partial N_n(c)}{\partial c_i}.\frac{1}{D_n^{m/2}} - m/2.\frac{\partial D_n}{\partial c_i}.\frac{N_n}{D_n^{m/2+1}}$$

$$\frac{\partial^2 J_n(c)}{\partial c_i \partial c_j} = \frac{\partial^2 N_n}{\partial c_i \partial c_j}.D_n^{-m/2} - m/2.\frac{\partial N_n}{\partial c_j}.\frac{\partial D_n}{\partial c_i}.D_n^{-m/2-1} - m/2.\frac{\partial N_n}{\partial c_i}.\frac{\partial D_n}{\partial c_j}.D_n^{-m/2-1}$$

$$+ m/2(m/2+1).N_n.D_n^{-m/2-2}.\frac{\partial D_n}{\partial c_i}.\frac{\partial D_n}{\partial c_j} - m/2.N_n D_n^{-m/2-1}.\frac{\partial^2 D_n}{\partial c_i \partial c_j}$$

Il est alors facile de voir, avec des techniques identiques à celles développées pour les théorèmes précédents que l'on a, en notant $\gamma_k=(c*s)_k$ et pour un $\delta>0$ :

$$\frac{\partial N_n}{\partial c_i} = 1/n.\Sigma X_t^m.m\Sigma\gamma_k^{m-1}.s_{k-i}.(1+n^{-\delta}o_p(1))$$

$$\frac{\partial D_n}{\partial c_i} = 1/n.\Sigma X_t^2.2\Sigma\gamma_k s_{k-i}.(1+n^{-\delta}o_p(1))$$

$$\frac{\partial^2 D_n}{\partial c_i \partial c_j} = 1/n.\Sigma X_t^2.2.\Sigma s_{k-i}s_{k-j}.(1+n^{-\delta}o_p(1))$$

$$\frac{\partial^2 N_n}{\partial c_i \partial c_j} = 1/n.\Sigma X_t^m.m.(m-1).\Sigma\gamma_k^{m-2}.s_{k-i}s_{k-j}.(1+n^{-\delta}o_p(1))$$

$$N_n = 1/n.\Sigma X_t^m.\Sigma\gamma_k^m.(1+n^{-\delta}o_p(1))$$

$$D_n = 1/n.\Sigma X_t^2.\Sigma\gamma_k^2.(1+n^{-\delta}o_p(1))$$

où les $o_p(1)$ le sont uniformément en c.

On obtient alors :

$$\frac{\partial J_n}{\partial c_i}(b) = n^{m/2-1+\delta-\min(1,1/\alpha)}.o_p(1)$$

$$\frac{\partial^2 J_n}{\partial c_i \partial c_j} = \frac{1/n.S_m(b)}{(1/nS_2(^b))^{m/2}}.[\frac{m(m-1)\Sigma\gamma_k^{m-2}s_{k-i}s_{k-j}}{(\Sigma\gamma_k^2)^{m/2}} - m.m/2.\frac{(\Sigma\gamma_k^{m-1}s_{k-i}).(2\Sigma\gamma_k\ s_{k-j})}{(\Sigma\gamma_k^2)^{m/2+1}}$$

$$- m.m/2.\frac{(\Sigma\gamma_k^{m-1}s_{k-j}).(\Sigma\gamma_k\ s_{k-i})}{(\Sigma\gamma_k^2)^{m/2+1}}$$

$$+ m/2(m/2+1).\frac{\Sigma\gamma_k\ s_{k-i}\Sigma\gamma_k\ s_{k-j}}{\Sigma\gamma_k^m}$$

$$- m.\frac{\Sigma\gamma_k^m}{(\Sigma\gamma_k^2)^{m/2}}.\Sigma s_{k-i}s_{k-j}\ ].(1+n\ o_p^{-\delta}(1))$$

On sait (comme pour le théorème 11 par (C) et (H1)) :

$$\frac{S_m(b)/n}{(S_2(b)^{m/2}} = n^{m/2-1-\varepsilon} \cdot u_n \quad \text{où } u_n \longrightarrow \infty$$

En regardant le développement de Taylor et compte-tenu des résultats précédents, on obtient, pour un certain $\delta>0$ :

$$n^{\min(1/\alpha;1)-\delta} \cdot {}^t(\hat{b}_n-b) \cdot A_n(c_n) = o_P(1) \quad , \quad c_n \longrightarrow b$$

$A_n$ est la matrice intervenant dans le hessien de $J_n$ écrit ci-dessus. Il est clair que $A_n \longrightarrow A$ où $A_{i,j} = \Sigma_k s_{k-i} s_{k-j}$ car $\gamma_k=0$ pour $k\neq 0$ en $c=b$ et $s_{-i}=0$ pour $i>0$.

$A$ est inversible, car $c \in \mathrm{Ker}A \Rightarrow {}^tcAc=0$, soit $\Sigma_k (c*s)_k^2 = 0$, soit $c*s=0$, soit $c=0$.

Le résultat final s'en déduit immédiatement.

## II. RECHERCHE DE CONTRASTES ET ALGORITHME STOCHASTIQUE POUR L'IDENTIFICATION DE S.

Dans ce chapitre, nous étudions les résultats et la méthode développés par A. Benveniste, M. Goursat et G. Ruget dans leur article : Robust identification of a nonminimum phase system : Blind adjustment of a linear equalizer in data communications. ([2])

Le problème étudié est celui présenté en introduction où l'on ne suppose pas que l'inverse de s est fini-dimensionnel. Dans ce cadre les méthodes proposées dans le chapitre précédent ne s'appliquent que si l'on impose à s l'appartenance à un compact fixé, condition non imposée dans le travail de Ruget et al.

On cherche à construire un système linéaire de filtre $\theta$ de telle sorte que le système global $r=\theta * s$ soit l'identité avec un retard éventuel. Le moyen proposé est l'utilisation d'une fonctionnelle $V(\theta) = J(r)$ construite à partir de la loi des variables $U_t$ issues du système global : $U_t = \Sigma_k r_k X_{t-k} = \Sigma_k \theta_k Y_{t-k}$, cette fonctionnelle présentant en $\theta = s^{-1}$ un minimum global.

On reconnait la méthode de type "contraste", identique à celle utilisée dans le chapitre précédent.

Le point de vue adopté ensuite est celui d'un algorithme stochastique, présentant des avantages de calcul importants par rapport aux méthodes d'estimation par minimum de contraste.

Les auteurs restreignent leur étude au cas où la loi de X est symétrique et, bien entendu, toujours non gaussienne.
Ils étudient un ensemble particulier de contrastes, ceux qui sont totalement linéaires, soit de la forme :
$V(\theta) = J(r) = E\psi(U_t)$, où $\psi$ est une fonction paire pour donner le même poids à un point et à son symétrique puisque la loi initiale est paire ; ces contrastes conduisent à un algorithme stochastique de la forme :
$\theta_k^{t+1} = \theta_k^t - \tau Y_{t-k}\varphi(c_t)$, pour $k=-N,\ldots,N$ (en pratique, on tronque les suites à distance finie), avec $c_t = \sum_{k=-N}^{+N} \theta_k^t Y_{t-k}$ et où $\varphi$ est la dérivée de $\psi$.
Nous allons maintenant présenter les diverses étapes étudiées justifiant la méthode, et nous attacherons ensuite plus longuement à la recherche des fonctionnelles pouvant servir de contraste.

## II.1. Etude de la fonction V(θ).

L'algorithme stochastique présenté est un algorithme de gradient. Il ne suffit donc pas de s'assurer que $\theta=s^{-1}$ est un minimum strict de la fonction V, il faut aussi étudier les lignes de descente et leurs attracteurs.
On suppose que s est d'énergie finie, i.e. que $\sum_k s_k^2<\infty$, ce qui permet de travailler avec pour espace de paramètre l'espace des suites

de carré sommable. ( $\theta$ est d'énergie finie si et seulement si r l'est et l'on peut donc travailler sur V ou sur J.)

On étudie tout d'abord la fonction J(r) sur la sphère unité ( i.e. pour $\Sigma r_k^2=1$); avec des hypothèses suffisantes (dont nous reparlerons plus loin) sur la fonction $\psi$ et sur la loi initiale F des $X_t$, J(r) a pour seuls minima +Id et -Id à un retard près (travaillant avec des lois paires et avec $\psi$ paire, on ne peut distinguer une loi de sa symétrique) ; on montre de plus que +Id et -Id sont les seuls attracteurs (à un retard près) des lignes de descente.

On généralise ensuite le résultat à $l^2$ avec une condition supplémentaire sur la dérivée $\varphi$ de $\psi$ :
$\varphi(x) = -\gamma.\mathrm{sign}(x) + \eta(x)$ ,$\eta$ étant la partie dérivable de $\varphi$ et vérifiant $EX_0\eta(X_0)=\gamma E|X_0|$, et $\gamma>0$.
Ceci est une condition analogue à une condition d'échelle : il est nécessaire de connaître l'échelle de la loi pour que, sans condition de normalisation du paramètre, les minima locaux soient bien $\pm$Id à un retard près.
Il faut ensuite revenir au véritable paramètre $\theta$. Il est immédiat que, sous les mêmes conditions que les résultats précédents, $V(\theta)$ a pour seuls minima $\pm s^{-1}$ à un retard près. On obtient encore que ce sont les seuls attracteurs des lignes de descente.
En remarquant que $\dfrac{dr}{du} = s*\dfrac{d\theta}{du}$ , on voit facilement que si $\theta$ suit une ligne de descente de V, r suit une ligne de descente de J lorsque $l^2$ est muni de la métrique :
$\|r\|_\Lambda = < r \ , \ \Lambda^{-1}r >$ où $\Lambda_{i,j}= \Sigma_k s_k s_{k+i-j}$.

Ceci montre que si $\Lambda$ est mal conditionnée ( si ses valeurs propres

sont loin d'être voisines) alors l'efficacité de l'algorithme est moindre sur certaines directions.

## II.2. Algorithme avec contrainte.

Pour pallier au problème de conditionnement de $\Lambda$, qui est la matrice de covariance de la sortie Y divisée par la variance de l'entrée $X_0$, on effectue le même traitement sur une sortie "blanchie", i.e. de matrice de covariance proportionnelle à l'identité. Précisément : On filtre les $Y_t$ linéairement pour obtenir un processus $Z=(Z_t)$ qui vérifie : $EZ_tZ_s= \delta_{t-s}.EX_0^2$ . Un tel filtrage est classique. Nous appellerons w le filtre blanchissant. Les raisonnements du II.1 montrent que lorsque $\theta$ suit une ligne de descente de V, le système global r (maintenant $r = s*w*\theta$ ) suit une ligne de descente de J pour $l^2$ munie de la métrique usuelle multipliée par $EX_0^2$ : il n'y a plus de direction sur laquelle l'efficacité est moindre.

De plus, on a alors $\|\theta\| = \|r\|.(EX_0^2)$, et l'on peut donc suivre les lignes de descente des fonctionnelles J et V en même temps sur la sphère $\{ r \ / \ \Sigma r_k^2=1 \}$ . Ceci conduit à l'algorithme avec contrainte suivant :

$\theta^{t+1}_k= \lambda_t.(\theta^t_k - Z_{t-k}.\varphi(c_t))$ $k = -N,...,N$ , avec $c_t = \Sigma_{-N}^{+N}\theta^t_k.Z_{t-k}$

et $\lambda_t$ étant choisi de telle sorte que $\|\theta^{t+1}\|=1$.

Il ne reste alors plus qu'à ajuster l'échelle.

Remarque : Dans les situations étudiées, c'est à dire pour une loi initiale paire et avec des fonctions $\psi$ paires, il reste encore à déterminer un point initial assurant la convergence vers $s^{-1}$et non

vers $-s^{-1}$, ce qui nécessite sur le filtre une connaissance à priori supplémentaire.

Aspects techniques.

La mise en oeuvre de la méthode ne peut se faire, comme indiqué sur les formules définissant l'algorithme, qu'avec des paramètres finidimensionnels. Il faut alors mesurer l'effet de la troncature sur la localisation des minima de $V(\theta)$, et imposer des conditions supplémentaires pour que ces minima soient dans un voisinage des minima obtenus sans troncature et pour que les lignes de crêtes correspondant aux points stationnaires soient peu stables.

## II.3. Recherche des contrastes J(r).

Le problème de recherche de conditions suffisantes, portant sur $\psi$, et relativement générales, sous lesquelles $J$ est un contraste, est étudié ici. Nous en proposerons une interprétation.

Nous nous intéresserons à la restriction de $J(r)$ à la sphère, l'extension à $l^2$ tout entier est ensuite simplement un problème d'échelle.

Il est alors naturel d'utiliser des coordonnées sphériques : $(i,j)$ étant fixé, on considère le système de coordonnées pour $r$ : $R,\alpha$, et $(r_k), k\neq i$ et $k\neq j$, où $R$ et $\alpha$ sont donnés par $R^2 = 1 - \Sigma_{k\neq i,j} r_k^2$ , et $r_i = R.\cos\alpha$, $r_j = R.\sin\alpha$.

On note $\mu$ la loi de $\Sigma_{k\neq i,j} r_k X_{-k}$ et $\varphi(x) = \int \varphi(x-y)\mu(dy)$.

On a alors le premier résultat suivant :

**LEMME** : Si $\psi$ et F sont paires, alors :

$$\frac{\partial J}{\partial \alpha} = 2R.\int_0^\infty\int_0^\infty (y\psi^\mu(Rx) - x\psi^\mu(Ry)).P_\alpha(dx,dy)$$

où $P_\alpha$ est la transformée de $F\otimes F$ par la rotation d'angle $\alpha$.

Compte-tenu de la symétrie de F, ce lemme montre que $\frac{\partial J}{\partial \alpha}$ est nul pour $\alpha = k\pi/4$, et donc que r est un point stationnaire de J dès que toutes ses coordonnées $r_k$ non nulles ont même module.
Ceci n'est pas suffisant pour assurer que Id et -Id sont les seuls minima de J. Nous allons énoncer le résultat obtenu par Benveniste, Goursat et Ruget (la démonstration en est assez simple par passage aux coordonnées polaires) et nous analyserons ensuite le sens de ce résultat.

Le théorème porte sur deux classes de lois, sous gaussiennes et supergaussiennes, lois ayant une densité de la forme :
$F(dx) = K.\exp(-g(x))dx$, et on a les définitions suivantes :

1). F est sous-gaussienne si g est paire et si $g(x)$ et $g'(x)/x$ sont strictement croissantes sur $\mathbb{R}^+$, ou bien si F est uniforme sur un segment $[-d;+d]$.

2). F est super-gaussienne si g est paire, strictement croissante sur $\mathbb{R}$ et $g'(x)/x$ strictement décroissante sur $\mathbb{R}^+$.

Par exemple, la loi de densité $C.\exp-|x|^\gamma$ est sous-gaussienne si $\gamma>2$ et super-gaussienne si $\gamma<2$.
On a alors :

**THEOREME** : Si F est sous-gaussienne (resp. super-gaussienne), et si $\varphi$ est deux fois dérivable sauf peut être à l'origine avec $\varphi(0+)\leqslant 0$ (resp. $\varphi(0+)\geqslant 0$ ) et $\varphi''\geqslant 0$ sur $\mathbb{R}^+$ (resp. $\varphi''\leqslant 0$ ), l'une des deux inégalités étant stricte, alors :
$\frac{\partial J}{\partial \alpha} > 0$ pour $\alpha \in ]0,\pi/4[$, et Id et -Id à un retard près sont les seuls minima de J sur la sphère unité.

Ceci conduit à des fonctions $\psi$ de la forme :
$\psi(x) = \alpha x^2 + \varepsilon.\gamma|x| + \beta - \delta.\mathrm{Log}\ h(x)$, où $\gamma\geqslant 0$, $\delta\geqslant 0$, $h(x)=K.\exp(-g(x))$, g étant paire et trois fois dérivable sauf peut-être à l'origine, K normalisant h au sens $l_1$, et avec :

1). Si F est sous-gaussienne, $\varepsilon=-1$ et $g'''\geqslant 0$ , l'inégalité étant stricte si $\gamma=0$ et $g'(0+)=0$.

2). Si F est super-gaussienne, $\varepsilon=+1$ et $g'''\leqslant 0$, l'inégalité étant stricte si $\gamma=0$ et $g'(0+)=0$.
Le contraste est alors donné par :
$J(r) = \alpha.E(\Sigma r_k X_{-k})^2 + \beta + \varepsilon.\gamma\ E|\Sigma r_k X_{-k}| - E(\mathrm{Log}\ h(\Sigma_k r_k X_{-k}))$.
Le terme de variance est constant sur la sphère unité. A un terme constant près, les contrastes considérés sont des combinaisons de $\pm E|Y|$ et $-E(\mathrm{Log}\ h(Y))$.
On remarque tout d'abord que si J est un contraste pour une loi initiale sous-gaussienne, alors -J est un contraste pour une loi initiale super-gaussienne.

Si $F_r$ désigne la loi de la variable aléatoire $\Sigma_k r_k X_{-k}$, et si H désigne la loi de densité h, et si $H_0$ désigne la loi super-gaussienne de densité $C.\exp(-|x|)$, on a :

$J(r) = \delta.(K(F_r, H) - K(F_r,G)) + \varepsilon\gamma.(K(F_r;H_0) - K(F_r;G))$, où G est la loi gaussienne centrée de même variance que $X_0$ et K désigne la distance de Kullback.

Ceci appelle l'interprétation suivante :

Soient les ensembles de loi suivants :

U = { F, loi de densité $h(x) = K.\exp(-g(x))$ où g est paire, strictement croissante sur $\mathbb{R}^+$, trois fois dérivable sauf peut-être à l'origine, et $g'''>0$ sur $R_*^+$ }.

S = { F, loi de densité $h(x) = K.\exp(-g(x))$ où g est paire, strictement croissante sur $\mathbb{R}^+$, trois fois dérivable sauf peut-être à l'origine, et $g'''<0$ sur $R_*^+$ }.

U est un sous ensemble des lois sous-gaussiennes et S un sous ensemble des lois super-gaussiennes.

On a alors :

Si la loi initiale F est dans U (resp. dans S), alors si H est une loi quelconque de U (resp.S), $K(F_r,H) - K(F_r,G)$ est un contraste.

On peut de même choisir $J(T) = \underset{H\in U \text{ (resp.} H\in S)}{\text{Min}} K(F_r,H) - K(F_r,G)$.

Le fait que de tels J soient des contrastes peut maintenant s'énoncer de la façon suivante :

> Filtrer la loi initiale linéairement écarte plus de la classe de lois considérées que de se placer en un point quelconque de la classe, ou encore : la "distance" entre la loi filtrée et la classe initiale est plus grande que la plus grande largeur de la classe, tous les termes d'éloignement étant pris au sens distance de Kullback compensée par l'écart à la gaussienne.

D'autre part, on a vu que l'opposé d'un contraste pour une loi initiale dans une classe était un contraste pour une loi initiale dans

l'autre classe. On peut alors donner une autre formulation :

La distance entre les deux classes de lois est plus grande que la distance d'une loi filtrée linéaire (la loi initiale étant dans une des deux classes) à une quelconque des deux classes.

Le résultat du théorème est donc, formulé ainsi, une caractérisation de deux classes de lois par rapport à une certaine géométrie et de leur comportement conjoint par filtration linéaire.

Il serait intéressant de trouver une caractérisation portant plus sur les fonctionnelles que sur les lois, et d'obtenir des contrastes opérant sur des classes de lois plus larges. En regardant des contrastes particuliers obtenus par cette méthode, on s'aperçoit qu'ils opèrent pour des classes de lois initiales beaucoup plus vastes; on retrouve en fait des contrastes cités dans le chapitre précédent :

a). En prenant pour H la fonction initiale, on a :
$J(r) = K(F_r,F) - K(F_r,G)$. On reconnait le contraste optimal au sens variance asymptotique pour les estimateurs d'objectif de type linéaire étudiés dans le chapitre précédent.

b). En regardant la famille de lois sous- et super-gaussiennes de densité $C.\exp-|x|^{\gamma}$, on a $J(r) = E|\Sigma_k r_k X_{-k}|^{\gamma}$.
Pour $\gamma=4$, on obtient :

$E(\Sigma r_k X_{-k})^4 = (\Sigma r_k^4)(EX_0^4-3(EX_0^2)^2) + 3(\Sigma r_k^2)^2(EX_0^2)^2$. On voit donc que $J(r) = E(\Sigma t_k X_{-k})^4$ est un contraste pour des lois initiales ayant un moment d'ordre 4 et vérifiant $EX_0^4<3(EX_0^2)^2$, et que $-E(\Sigma r_k X_{-k})^4$ est un contraste pour des lois dont les moments d'ordre 4 et 2 véri-

fient l'inégalité contraire.

On retrouve en fait ici le cumulant d'ordre 4.

Rappelons enfin que la valeur absolue des cumulants standardisés est un contraste pour toutes les lois initiales possédant un tel cumulant non nul, mais que ce contraste n'est pas (sauf pour l'ordre 4) de la forme très simple étudiée ici $E\psi(Y)$.

## III. DECONVOLUTION AVEUGLE PAR ESTIMATION SPECTRALE.

Ce chapitre étudie le point de vue spectral, qui est celui adopté par Rosenblatt et Lii pour résoudre le problème de déconvolution aveugle ([17]).

Par cette méthode, on estime la fonction spectrale :

$\alpha(\exp -i\lambda) = \Sigma_k s_k \exp -i\lambda$

en estimant le carré de son module, la densité spectrale, par des méthodes connues (périodogramme lissé) et sa phase à l'aide de la densité spectrale d'un cumulant d'ordre k choisi.

Par cette méthode, on ne suppose pas nécessairement que le paramètre est fini-dimensionnel; on suppose simplement la sommabilité de série des carrés $s_k^2$.

En contrepartie, la vitesse de convergence n'est plus $n^{d/2}$.

Sous des hypothèses un peu plus contraignantes sur s et X ( existence de tous les moments de $X_0$, sommabilité de certaines séries liées aux cumulants,...) et sous l'hypothèse de mélange fort pour Y, Rosenblatt obtient une vitesse de convergence de l'ordre de $n^{-2/5}$.

En contraignant de plus en plus la décroissance à l'infini des $s_k$, il est probable que, de même que pour l'estimation de densités, on approche de plus en plus la vitesse $\sqrt{n}$ (on approche le cas où s a un inverse de longueur finie et où $\alpha$ est $C^\infty$).

Nous présenterons la méthode pour un processus indexé dans $\mathbb{Z}$ pour plus de lisibilité. Mais celle-ci s'étend à d quelconque.

<u>Cumulants, densité spectrale cumulante d'ordre k et estimation de la phase de</u> $\alpha(\exp -i\lambda)$.

Si $X = (X_1,\dots,X_k)$ est un vecteur aléatoire, soit $\varphi_X(t_1,\dots,t_k)$ sa fonction caractéristique. Si tous les moments de X jusqu'à l'ordre n existent, on définit les cumulants de X d'ordre $\nu=(\nu_1,\dots,\nu_k)$ noté $C_X^\nu$ pour $\nu_1+\dots+\nu_k\leqslant n$ comme étant les coefficients du développement en série de Taylor :

$$\operatorname{Log} \varphi_X(t_1,\dots,t_k) = \sum_{\nu_1+\dots+\nu_k\leqslant n} \frac{i^{\nu_1+\dots+\nu_k}}{\nu_1!\dots\nu_k!} \cdot C_X^\nu . t_1^{\nu_1}\dots t_k^{\nu_k} + o(|t|^n).$$

On obtient facilement, en développant $\varphi_X(t)$, la relation entre cumulants et moments. En particulier, on a :

$$C^{(1,\ ,1)}_{(Y_1,\ ,Y_k)} = \operatorname{cum}(Y_1,\dots,Y_k) = \sum_\nu (-1)^{p-1}.(p-1)!\ E(\prod_{j\in\nu_1} Y_j)\dots E(\prod_{j\in\nu_p} Y_j)$$

où $\nu = (\nu_1,\dots,\nu_p)$ est une partition de $\{1,\dots,k\}$.

On voit facilement que $\operatorname{cum}(Y_1,\dots,Y_k)$ est une forme k-linéaire, et que si on peut partitionner $\{Y_1,\dots,Y_k\}$ en deux ensembles tels que chaque variable d'un ensemble est indépendante de l'autre ensemble, alors $\operatorname{cum}(Y_1,\dots,Y_k) = 0$.

Maintenant, pour un processus stationnaire, on définit la densité spectrale cumulante d'ordre k comme étant la série de Fourier des cumulants d'ordre 1 des k-uples $(Y_t, Y_{t+j_1},\dots,Y_{t+j_{k-1}})$ :

$$f_k(\lambda_1,\dots,\lambda_{k-1})=(2\pi)^{-k+1}\sum_j \operatorname{cum}(Y_t, Y_{t+j_1},\dots,Y_{t+j_{k-1}})e^{-ij_1\lambda_1}\dots e^{-ij_{k-1}\lambda_{k-1}}$$

Lorsque Y est un processus linéaire s'écrivant :

$Y_t = \Sigma_k s_k X_{t-k}$ où les $X_t$ sont indépendants centrés de variance 1 et identiquement distribués, alors :

$f_k(\lambda) = \gamma_k(2\pi)^{-k+1}\alpha(e^{-i\lambda_1})\ldots\alpha(e^{-i\lambda_{k-1}}).\alpha(e^{i(\lambda_1+\ldots+\lambda_{k-1})})$ , où $\gamma_k$ est le cumulant d'ordre k de $X_0$.

Pour k=2, on retrouve la densité spectrale $f(\lambda)=|\alpha(e^{-i\lambda})|^2/2\pi$.

On a alors le résultat suivant :

**THEOREME** : Si $\sum_k |k||s_k|<\infty$, et si $X_0$ possède des moments finis jusqu'à l'ordre $k>2$ où $\gamma_k\neq 0$, alors $\alpha(\exp -i\lambda)$ peut être identifié à partir des observations $Y_t$ à un facteur $\pm \exp ia\lambda$ près, $a\in\mathbb{N}$.

Posons $h(\lambda) = \arg \{ \alpha(\exp -i\lambda).\alpha(1)/|\alpha(1)| \}$.

Au signe près, on a :

$\alpha(\exp -i\lambda) = \sqrt{2\pi f(\lambda)}.\exp ih(\lambda)$.

(Le signe ne peut pas être déterminé car en multipliant le filtre s par -1 et l'entrée X par -1, on obtient la même sortie Y).

Sachant que l'on peut construire des estimateurs consistants de la densité spectrale $f(\lambda)$, il ne reste plus qu'à pouvoir estimer $h(\lambda)$.

Par ailleurs, en décalant l'indexation des $X_t$ de $a\in\mathbb{N}$ et celle des $s_k$ de $-a$, on obtient la même sortie Y. $h(\lambda)$ ne peut donc être déterminée qu'à un facteur $a\lambda$ près, $a\in\mathbb{N}$.

On a :

$(\alpha(1)/|\alpha(1)|)^k.\gamma_k = (2\pi)^{k/2-1}.\ f_k(0,\ldots,0)(f(0))^{-k/2}$, et

$$h(\lambda_1)+..+h(\lambda_{k-1})-h(\lambda_1+..+\lambda_{k-1})=\arg[(\alpha(1)/|\alpha(1)|^k)\gamma_k f_k^{-1}(\lambda_1,..,\lambda_{k-1})]$$

$$=\arg f_k(\lambda_1,..,\lambda_{k-1}) + \arg f_k(0,..,0). \qquad (1)$$

D'autre part :

$$h'(0)-h'(\lambda)= \lim_{\Delta\longrightarrow 0} \frac{h(\lambda)+(k-2)h(\Delta)-h(\lambda+(k-2)\Delta)}{(k-2)\Delta} . \qquad (2)$$

Sachant estimer $f_k(\lambda_1,..,\lambda_{k-1})$ (ce qui est possible, nous le ver-

rons plus loin), (1) et (2) permettent d'estimer $h'(0)-h'(\lambda)$ par :

$$\frac{\arg \hat{f}_k(\lambda,\Delta,\ldots,\Delta) + \arg \hat{f}_k(0,\ldots,0)}{(k-2)\Delta}$$ en prenant $\Delta$ assez petit.

Il est alors possible d'estimer $h_1(\lambda) = \int_0^\lambda (h'(u)-h'(0))du$.

Or, $h_1(\lambda) = h(\lambda) - \lambda.h'(0)$.

Pour atteindre le but fixé, il ne s'agit plus que de pouvoir estimer $h'(0)$ à une constante entière additive près.

$h_1(\pi) = h(\pi) - \pi h'(0)$. Les $s_k$ étant réels, $h(\pi)=p\pi$ où $p\in\mathbb{N}$. Donc :

$h'(0) = p - \frac{h_1(\pi)}{\pi}$. A une constante entière additive près, on estimera $h'(0)$ par $-h_1(\pi)/\pi$.

Nous allons maintenant revenir brièvement sur les estimateurs de la densité spectrale et des densités spectrales cumulantes d'ordre k, puis nous détaillerons procédure d'estimation de $h_1(\lambda)$ et vitesse de convergence lorsque le cumulant d'ordre 3 de $X_0$ est non nul.

<u>Estimation de la densité spectrale</u>.

On construit le périodogramme :

$I_N(\lambda) = 1/2\pi N.|\Sigma_{k=0}^{N-1} Y_k \exp -ik\lambda|^2$.

Cet estimateur n'est pas très bon ponctuellement car sa variance ne tend en général pas vers 0. Mais pour des $\lambda_j$ distincts, les $I_N(\lambda_j)$ sont indépendants, ce qui suggère de prendre comme estimateur de $f(\lambda)$ une "moyenne" du périodogramme ; on choisit ainsi :

$f_N(\lambda) = \int W_N(\lambda-\mu)\ I_N(\mu)\ d\mu$, où $W_N$ est un noyau régularisant qui tend vers $\delta_0$ quand N tend vers l'infini.

On peut prendre $W_N(x) = \frac{1}{b(N)}.W(\frac{x}{b(N)})$

On mesure la vitesse de convergence par la racine carrée de l'écart

quadratique moyen de $f_N(\lambda)$ à $f(\lambda)$, lequel est la somme de deux quantités : le carré du biais $|f(\lambda)-Ef_N(\lambda)|^2$ et la variance $\mathrm{var}(f_N(\lambda))$.

Avec des hypothèses suffisantes sur le noyau W, $N^{-1}=o(b(N))$, et si l'on suppose $\Sigma_k k^2|r_k|<\infty$, où $r_k=\mathrm{cov}(Y_t,Y_{t+k})$, le carré du biais est de l'ordre de $b(N)^4$ et la variance de l'ordre de $(Nb(N))^{-1}$. On obtient donc une vitesse optimale de $N^{-2/5}$ en prenant $b(N)=N^{-1/5}$.

<u>Estimation de la densité spectrale cumulante d'ordre</u> k.

On généralise la construction faite précédemment de la façon suivante :

posons $d_N(\lambda) = \Sigma_{k=0}^{N-1} Y_k \exp -ik\lambda$.

On définit le périodogramme d'ordre k par :

$$I_N^k(\lambda_1,\ldots,\lambda_{k-1}) = (2\pi)^{-k+1}\ N^{-1}(\Pi_{j=1}^{k-1} d_N(\lambda_j)).d_N(-\Sigma_{j=1}^{k-1}\lambda_j).$$

Pour k=2, on reconnait $I_N(\lambda)$, le périodogramme classique.

On choisit comme estimateur de la densité spectrale cumulante d'ordre k la somme de Riemann du périodogramme d'ordre k lissé :

$$f_k^N(\lambda_1,\ldots,\lambda_{k-1}) = (2\pi)^{k-1}N^{-k+1}\Sigma_{ki=-\infty}^{+\infty} W_N(\lambda_1-2\pi k_1/N;\ldots;\lambda_{k-1}-2\pi k_{k-1}/N; -\Sigma\ \lambda_j+2\pi\ \Sigma k_j/N)$$
$$.\ I_N^k(2\pi k_1/N,\ldots,2\pi k_{k-1}/N)$$

où $W_N$ vérifie : $W_N(u_1,..,u_k) = b(N)^{-k+1}W(B(N)^{-1}u_1;\ldots;b(N)^{-1}u_k)$.

Moyennant des conditions de régularité sur W (voir [25]), cette somme de Riemann est égale à la convolution du périodogramme d'ordre k et de $W_N(u_1,..,u_{k-1},-u_1+..+u_{k-1})$ plus $o(N^{-k+1})$.

Maintenant, si l'on suppose que X a tous ses moments finis, et que $\sum_{v_1,..,v_{k-1}}(1+|v_j|)|\mathrm{cum}(Y_t,Y_{t+v_1},\ldots,Y_{t+v_{k-1}})|<\infty$ pour $j=1\ldots k-1$, ainsi que certaines hypothèses techniques sur le noyau W, on obtient lorsque $N^{-1}=o(b(N)^{k-1})$, que la variance est de l'ordre de $(Nb(N)^{k-1})^{-1}$ et le carré du biais de l'ordre de $b(N)^2$. La vitesse

de convergence est alors au mieux de l'ordre de $N^{-1/k+1}$.

<u>Etude du cas</u> $k=3$ ; <u>estimation de</u> $h_1(\lambda)$.

Dans cette partie, nous décrivons l'estimation de $h_1(\lambda)$ basée sur la densité spectrale cumulante d'ordre 3. On supposera donc que $X_0$ possède un cumulant d'ordre 3 non nul.

Nous supposerons que $f_3(0,0)\geqslant 0$.

On estimera $h_1$ aux points $\lambda=k\Delta$, où $\Delta=\Delta(N)\longrightarrow 0$.

On remarque tout d'abord que $h_1(\lambda)=h(\lambda)-h'(0)\lambda$ est à peu près égal à $h(k\Delta)-k\Delta/\Delta.h(\Delta)$, qui est égal à :

$\Sigma_{j=1}^{k-1}\{h(j\Delta)+h(\Delta)-h((j+1)\Delta)\} = -\Sigma_{j=1}^{k-1}\arg f_3(j\Delta,\Delta)$ grace à (1).

On va donc choisir comme estimateur de $h_1(k\Delta)$ la quantité :

$H_n(k\Delta) = -\Sigma_{j=1}^{k-1}\arg f_N^3(j\Delta,\Delta)$, où $f_N^3(\lambda_1,\lambda_2)$ est l'estimateur de $f_3(\lambda_1,\lambda_2)$ défini au paragraphe précédent.

Le résultat de convergence obtenu est alors le suivant :

**<u>THEOREME</u>** : Si tous les moments de $X_0$ existent, si le noyau $W$ vérifie les conditions usuelles et si $f_3$ est deux fois dérivable alors $H_n(\lambda)-h_1(\lambda)$ est asymptotiquement normal ($\lambda=\lim k\Delta$), avec une variance de l'ordre de $(\Delta^3N)^{-1}$.

L'écart quadratique moyen de $H_n(\lambda)$ à $h_1(\lambda)$ est borné par $C_1\Delta^2+ C_2(\Delta^3N)^{-1}$.

La vitesse optimale de convergence est donc $N^{-1/5}$. Si l'on suppose $f_3$ trois fois continument dérivable, cette borne devient : $C_1\Delta^4+C_2(\Delta^3N)^{-1}$, et la vitesse optimale de convergence est $N^{-2/7}$.

On a supposé $f_3(0,0)\geqslant 0$. On estimera le signe de $f_3(0,0)$ par celui de la partie réelle de $f_3^N(0,0)$, et si ce dernier est négatif on estimera $h_1(\lambda)$ par

$$-\Sigma_{j=1}^{k-1}\arg(-f_3^N(j\Delta,\Delta)).$$

## REFERENCES BIBLIOGRAPHIQUES

[1]. Azencott R., Dacunha-Castelle D. : Séries d'observations irrégulières, Masson, 1984.

[2]. Benveniste A., Goursat M., Ruget G. : Robust identification of a non minimum phase system : blind adjustment of a linear equalizer in data communications, IEEE Trans. aut. cont., vol A.C. 25, 3, 1980.

[3]. Bickel P.J. : The 1980 wald memorial lectures on adaptative estimation, Ann. stat., 10, 3, 647-671, 1982.

[4]. Billingsley P. : Convergence of probability measures, Wiley, 1968.

[5]. Box G.E., Jenkins G.M. : Time series analysis, forecasting and control, Holden-day, 1976.

[6]. Chaterjii S. : Notes on an $L^p$convergence theorem, Ann. math. stat., 40, 3, 1068-1070, 1969.

[7]. Crump N.D. : A Kalman filter approach to the deconvolution of seismic signals, Geophysics, 39, 1-13, 1974.

[8]. Dacunha-Castelle D., Duflo M. : Probabilités et statistiques, tome 2, Masson, 1983.

[9]. Donoho D. : On minimum entropy deconvolution, Proc. of 2nd appl. time series symposium, Tulsa, D.F. Findley, 565-608, 1980.

[10]. Gorodetski V.V. : On the strong mixing property for linear sequences, Theory prob. appl., 22, 411-413, 1977.

[11]. Goussart Y., Demoment G. : Recursive deconvolution of Bernoulli-Gaussian processes using a MA representation, LSS internal report number GPI 87/03, 1987.

[12]. Guyon X. : Champs stationnaires sur $\mathbb{Z}^2$ : modèles, statistiques et simulations, prepub. Orsay, 1985.

[13]. Kailath T. : Lectures on Wiener and Kalman filtering, Springer Verlag, 1981.

[14]. Kanter M. : On quotients of moving average processes with infinite mean, Proc. amer. math. soc., 56, 281-287, 1976.

[15]. Lai T.L., Wei C.Z. : Asymptotic properties of general autoregressive models and strong consistency of least square estimates of their parameters, J. of mult. an., 13, 1-23, 1983.

[16]. Levy S., Fullagar P.K. : Reconstruction of a sparse spike train from a portion of its spectrum and application to high resolution deconvolution, Geophysics, 46(9), 1235-1243, 1981.

[17]. Lii K.S., Rosenblatt M. : Deconvolution and estimation of transfer function phase and coefficients for non gaussian linear processes, Ann. stat., 10, 4, 1195-1208, 1982.

[18]. Mac-Leish : Invariance principle for dependant variables, Z. wahrsch., 32, 165-178, 1975.

[19]. Mahalanabis A.K., Prasad S., Mohandas K.P. : A fast optimal deconvolution algorithm for real seismic data using Kalman predictor model, IEEE trans. geoscience, GE19, 216-221, 1981.

[20]. Miller H.D. : A note on sums of independant random variables with infinite first moment, Ann. math. stat., 38, 751-758, 1967.

[21]. Mokkadem A. : Entropie de processus et erreur de prédiction, C.R.A.S, 298, série I, 493-496, 1984.

[22]. Mokkadem A. : Sur le mélange d'un processus ARMA vectoriel, C.R.A.S., 303, série I, 11, 519-521, 1986.

[23]. Oldenburg D.W., Scheuer T., Levy S. : Recovery of the acoustic impedance from reflection seismograms, Geophysics, 48, 10, 1318-1337.

[24]. Pinsker : Information and information stability of random variables and processes, 1964.

[25]. Rosenblatt M. : stationary sequences and random fields, Birckhauser Boston, 1985.

[26]. Stein C. : Efficient non parametric testing and estimation, Proc. third Berkeley symp. math. stat. prob. 1, 187-196, University of California press, 1956.

[27]. Taylor H.L., Banks S.C., Mc Coy J.F. : Deconvolution with the l1 norm, Geophysics, 44, 1, 39-52, 1979.

[28]. Whittle P. : Gaussian estimation in stationary time series, Bull. I.S.I. 39, 105-129.

[29]. Zolotarev : ideal metrics in problems of probability theory and mathematical statistics, Austral. J. stat. 21, 193-203, 1979.

$m = 0.5$

HISTOGRAM OF $\Gamma$

```
        - 0.630      0
0.630 - 0.645        1 *
0.645 - 0.660        3 ***
0.660 - 0.675       10 **********
0.675 - 0.690       21 *********************
0.690 - 0.705       31 ******************************
0.705 - 0.720       19 *******************
0.720 - 0.735        8 ********
0.735 - 0.750        6 ******
0.750 -              1 *

MISSING VALUES       0

SCALE: 1 ASTERISK REPRESENTS  1.00 UNIT(S)
```

HISTOGRAM OF $\Sigma^*$

```
    - -5       1 *
-5 - -4        1 *
-4 - -3        2 **
-3 - -2        5 *****
-2 - -1       23 ***********************
-1 -  0       32 ********************************
 0 -  1       16 ****************
 1 -  2       18 ******************
 2 -  3        0
 3 -           2 **

MISSING VALUES       0

SCALE: 1 ASTERISK REPRESENTS  1.00 UNIT(S)
```

$m = 1$

HISTOGRAM OF $\Gamma$

```
        - 0.648      1 *
0.648 - 0.660        3 ***
0.660 - 0.672       10 **********
0.672 - 0.684       13 *************
0.684 - 0.696       20 ********************
0.696 - 0.708       22 **********************
0.708 - 0.720       17 *****************
0.720 - 0.732        6 ******
0.732 - 0.744        5 *****
0.744 -              3 ***

MISSING VALUES       0
```

SCALE: 1 ASTERISK REPRESENTS 1.00 UNIT(S)

HISTOGRAM OF $\Sigma^*$

```
        --0.92       1 *
-0.92 --0.88         0
-0.88 --0.84         0
-0.84 --0.80         6 ******
-0.80 --0.76         9 *********
-0.76 --0.72        16 ****************
-0.72 --0.68        31 *******************************
-0.68 --0.64        27 ***************************
-0.64 --0.60         8 ********
-0.60 -              2 **

MISSING VALUES       0
```

SCALE: 1 ASTERISK REPRESENTS 1.00 UNIT(S)

$m = 10$

HISTOGRAM OF $\Gamma$

```
          - 0.660      1 *
0.660 - 0.672      8 ********
0.672 - 0.684     17 *****************
0.684 - 0.696     19 *******************
0.696 - 0.708     28 ****************************
0.708 - 0.720     11 ***********
0.720 - 0.732      9 *********
0.732 - 0.744      4 ****
0.744 - 0.756      1 *
0.756 -            2 **

MISSING VALUES     0
```

SCALE: 1 ASTERISK REPRESENTS 1.00 UNIT(S)

HISTOGRAM OF $\Sigma^*$

```
          - -0.7050      3 ***
-0.7050 - -0.7035      7 *******
-0.7035 - -0.7020     16 ****************
-0.7020 - -0.7005     23 ***********************
-0.7005 - -0.6990     15 ***************
-0.6990 - -0.6975     16 ****************
-0.6975 - -0.6960     13 *************
-0.6960 - -0.6945      2 **
-0.6945 - -0.6930      4 ****
-0.6930 -              1 *

MISSING VALUES         0
```

SCALE. 1 ASTERISK REPRESENTS 1.00 UNIT(S)

$m = 50$

```
HISTOGRAM OF        Γ

            - 0.660      1 *
    0.660 - 0.672        6 ******
    0.672 - 0.684       19 *******************
    0.684 - 0.696       23 ***********************
    0.696 - 0.708       23 ***********************
    0.708 - 0.720       13 *************
    0.720 - 0.732        8 ********
    0.732 - 0.744        4 ****
    0.744 - 0.756        2 **
    0.756 -              1 *

  MISSING VALUES         0

SCALE: 1 ASTERISK REPRESENTS  1.00 UNIT(S)
```

```
HISTOGRAM OF        Σ*

            ---0.7008       19 *******************
  -0.7008 ---0.7000         42 ******************************************
  -0.7000 ---0.6992         23 ***********************
  -0.6992 ---0.6984          7 *******
  -0.6984 ---0.6976          5 *****
  -0.6976 ---0.6968          0
  -0.6968 ---0.6960          2 **
  -0.6960 ---0.6952          1 *
  -0.6952 ---0.6944          1 *
  -0.6944 -                  0

  MISSING VALUES             0

SCALE: 1 ASTERISK REPRESENTS  1.00 UNIT(S)
```

# TABLE DES MATIERES


| | | pages |
|---|---|---|
| O. | INTRODUCTION | 4 |
| I. | DECONVOLUTION AVEUGLE, OU ESTIMATION DES PARAMETRES D'UN PROCESSUS AUTO-REGRESSIF DE LOI INCONNUE NON GAUSSIENNE | 16 |
| I.1. | Description de la méthode, consistance de l'estimateur | 16 |
| I.2. | Exemples | 20 |
| I.3. | Comportement asymptotique | 25 |
| I.4. | Comparaison avec les méthodes usuelles | 38 |
| I.5. | Problèmes d'efficacité | 41 |
| I.6. | Identification : Estimation de l'ordre p | 45 |
| I.7. | Robustesse | 47 |
| I.8. | Moments infinis et objectifs cumulants | 50 |
| II. | RECHERCHE DE CONTRASTES ET ALGORITHMES STOCHASTIQUES POUR L'IDENTIFICATION DE S | 59 |
| II.1. | Etude de la fonction $V(\theta)$ | 60 |
| II.2. | Algorithme avec contrainte | 62 |
| II.3. | Recherche des contraintes $J(r)$ | 63 |
| III. | DECONVOLUTION AVEUGLE PAR ESTIMATION SPECTRALE | 69 |
| Références bibliographiques | | 76 |
| Annexes | | 78 |