

THÈSES D'ORSAY

MARC LAVIELLE

**Déconvolution 2-D et détection de ruptures : applications
en géophysique / Marc Lavielle**

Thèses d'Orsay, 1990

http://www.numdam.org/item?id=BJHTUP11_1990__0269__P0_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.



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/>*

63724

ORSAY
n° d'ordre :

UNIVERSITE PARIS SUD

CENTRE d'ORSAY

T H E S E

présentée pour obtenir

Le TITRE de DOCTEUR EN SCIENCES

SPECIALITE : MATHEMATIQUES

par

Marc LAVIELLE

Sujet : DECONVOLUTION 2-D ET DETECTION DE RUPTURES.
APPLICATIONS EN GEOPHYSIQUE.

Soutenu le 25 Juin 1990 devant le Jury composé de :

DACUNHA CASTELLE	D.	Président
COMETS	F.	
DELPINO	E.	
DEMOMENT	G.	
GASSIAT	E.	
GUYON	X.	
LEON	J.	

2-D DECONVOLUTION AND FAILURES' DETECTION. APPLICATIONS IN GEOPHYSICS.

Abstract:

We study the application in geophysics of some technics used for image restoration or signal processing. In the first part, we consider the deconvolution problem when the signal is assumed to belong to a mixture of Gaussian variables. Because of its numerous applications, the particular case of Bernoulli-Gaussian signals is treated. The model is parametric and a stochastic algorithm of estimation is used for the model's identification. The deconvolution technics we present optimizes some natural criteria of good recovery, related to the observations and wrong detections. To recover the reflectivity sequence of the earth, the deconvolution of real seismic traces is then studied. To take into account the layers' structure existing in the earth, we consider this problem as a 2-D deconvolution problem. The layers are not always parallel, so we need to define a non-homogeneous neighborhood system. Examples arising from real data are presented. In the second part, we deal with the problem of sequential failures' detection in the distribution of a sequence of multidimensional random variables. We first consider variables which distribution is parametric, in most of these cases, failures affect the mean and the variance. Some applications are proposed, to detect abrupt failures of lithology in the earth and, on the other hand, to identify stationnary sea-states. If the data are dependent, the changes can affect the spectral distribution of the process. We use the empirical spectral estimators to build some test statistics which are asymptotically normal and independent. Their use changes the original problem to the detection of parameter changes in a sequence of multivariate independent Gaussian random variables.

Key words : Deconvolution. Image processing. Gibbs Sampler. Parametric estimation. Bernoulli-Gaussian signals. Seismic traces. Failures' detection. Spectral estimation.

L'ensemble de ce travail a été effectué au Vénézuéla, à Caracas. Ce furent pour moi trois années merveilleuses . . . Je veux profiter de ces quelques lignes pour remercier du fond du cœur tous ceux qui m'ont permis de réaliser cette thèse dans des conditions aussi exceptionnelles.

Je tiens tout d'abord à exprimer toute ma gratitude et mon amitié à Didier Dacunha-Castelle pour son aide, son soutien, et le constant intérêt qu'il a pu porter à mon travail.

Un abrazo fuerte a José Rafael León qui m'a accueilli si chaleureusement à l'Université Centrale du Vénézuéla et m'a rendu un peu "venezolano" . . .

Je remercie Xavier Guyon pour avoir accepté d'être rapporteur de cette thèse, ses conseils ont toujours été bienvenus, d'un coté de l'océan comme de l'autre. Merci aussi à Elizabeth Gassiat, Francis Comets et Guy Demoment pour faire partie du jury.

Ce fut pour moi un très grand plaisir de travailler au sein de l'équipe de géophysique de l'INTEVEP, j'en remercie tous ses membres pour leur gentillesse et leur disponibilité. Muchas gracias a Eulogio del Pino qui a suivi attentivement mon travail et m'a donné tous les moyens dont il pouvait disposer pour m'aider.

Merci à Fabrice Gamboa, à Paul Doukhan, enfin bref, à toute l'équipe d'Orsay, et bien videmment à celle de Caracas.

A Elena,

A Renaud,

(Y a Cristóbal Colón, por supuesto !!!)

La géophysique représente un domaine privilégié d'applications pour un statisticien, les problèmes posés sont nombreux, variés et le nombre de données disponibles est souvent considérable. Dans le but d'identifier la composition du sous-sol (ou lithologie), la recherche pétrolière a largement contribué au développement de nouvelles techniques de prospection et à l'emploi de méthodes mathématiques diverses, depuis le traitement de données traditionnel (clustering, analyse factorielle, etc . . .) à des techniques plus sophistiquées de traitement de signal.

Nous disposerons essentiellement de deux types de données :

** Des mesures effectuées dans des puits permettent une reconstruction du sous-sol de très haute résolution au voisinage de ces puits. Plusieurs enregistrements sont effectués, mesurant des paramètres qui caractérisent le type de lithologie en présence: impédance, radio-activité, résistivité, etc . . . La réalisation d'un puit et de ces enregistrements représente un coût extrêmement élevé, les puits sont donc peu nombreux et espacés de plusieurs kilomètres.*

** Des données beaucoup plus nombreuses car moins coûteuses, mais plus pauvres en information, sont les traces sismiques : une onde est émise à la surface et se propage dans le sous-sol, à chaque changement de lithologie une partie du signal est réfléchié alors qu'une autre est transmise. Le signal mesuré à la surface par le sismogramme est constitué de la somme des signaux réfléchis. Cette information fournira au géophysicien la possibilité de décrire l'agencement des différentes couches ainsi que leur composition. En effet, une section sismique contient plusieurs centaines de traces, espacées chacune de quelques dizaines de mètres, cela offre donc une vision en coupe du sous-sol où la structure de couches est évidente. Il faut remarquer que dans certaines zones d'intérêt particulier, le terrain est quadrillé, les traces obtenues fournissent alors une vision tridimensionnelle du sous-sol.*

Par leurs natures, ces deux types de données demandent une approche différente. Le choix arbitraire d'une modélisation nous a conduit à voir l'interprétation des données sismiques comme un problème de déconvolution, le champ inconnu que l'on cherche à reconstruire est constitué des coefficients de réflexion. D'un autre côté, les enregistrements provenant de puits seront considérés comme des processus aléatoires dans lesquels des ruptures sont détectées, une rupture indiquant un changement brusque de lithologie.

Cette thèse est présentée en deux parties : la première partie est consacrée principalement à la déconvolution de traces sismiques alors que la seconde partie traite du problème de détection de ruptures dans un cadre général, une des applications présentées concernant les enregistrements décrits ci-dessus.

PARTIE I

DECONVOLUTION

INTRODUCTION

Considérons le modèle de convolution suivant:

$$\begin{cases} \mathbf{Z} = \mathbf{f} * \mathbf{X} + \boldsymbol{\varepsilon} \\ \mathbf{Y} = (\mathbf{Q}, \mathbf{X}) \end{cases} \quad (*)$$

$\mathbf{Z} = \{Z(k)\}_{1 \leq k \leq n}$ est une série d'observations, \mathbf{f} le filtre de convolution, $\mathbf{Y} = \{Y(k)\}_{1 \leq k \leq n}$ une suite de variables aléatoires indépendantes et $\boldsymbol{\varepsilon}$ un bruit blanc gaussien de variance $\sigma_{\boldsymbol{\varepsilon}}^2$.

Pour tout i , la loi de $Y(i) = (Q(i), X(i))$ est définie par:

- $P(Q(i) = k) = \lambda_k$ pour $1 \leq k \leq m$ ($\sum_{k=1}^m \lambda_k = 1$)
- Conditionnellement à $Q(i) = k$, $X(i)$ suit une loi P_k

on suppose P_k paramétré par un paramètre α_k et ayant une densité $p(\cdot, \alpha_k)$ par rapport à une certaine mesure $d\mu$, nous noterons θ_Y l'ensemble des paramètres caractérisant la loi de \mathbf{Y} ,

$$\theta_Y = (\lambda_1, \alpha_1, \lambda_2, \alpha_2, \dots, \lambda_m, \alpha_m)$$

et θ l'ensemble des paramètres définissant le modèle (*),

$$\theta = (\theta_Y, \mathbf{f}, \sigma_{\boldsymbol{\varepsilon}}^2).$$

Notre problème est de reconstruire \mathbf{Y} lorsque seul \mathbf{Z} est observé, toutefois, cette reconstruction ne pourra être réalisée que si le modèle est identifié. Le modèle étant ici paramétrique, son identification revient à estimer le paramètre θ .

Dans la suite, les lois P_k seront supposées gaussiennes, nous traiterons alors deux cas particulier du modèle (*):

- Modèle M1 : le filtre f est réduit à l'identité et $\varepsilon = 0$:

$$\begin{cases} Z = X \\ Y = (Q, X) \end{cases}$$

Nous sommes face à un problème classique de mélange de populations: les paramètres du mélange doivent d'une part être estimés et d'autre part, chaque observation doit être attribuée à une population. En effet, l'estimation de Y se réduit ici à l'estimation de Q .

- Modèle M2 : Q est une variable de Bernoulli de paramètre λ (donc $m=2$) et les lois P_0 et P_1 sont des lois normales centrées de variances σ_0^2 et σ_1^2 (on supposera $\sigma_0^2 < \sigma_1^2$).

Un cas particulier consiste à choisir $\sigma_0^2 = 0$, Y est alors une suite de variables Bernoulli-Gaussiennes. Estimer Y revient à détecter les signaux (estimer Q) et estimer leur amplitude (estimer X). Ce type de modèle est fréquemment employé en géophysique, échographie, signal, etc ... [12], [13], [18], [21].

Ainsi, en géophysique, ce modèle est généralement utilisé pour représenter une trace sismique: Z est la série des valeurs observées du sismogramme, X celle des coefficients de réflexion, le filtre f est une onde émise à la surface et ε un bruit de mesure. La modélisation B-G de Y signifie ici qu'un réflecteur n'est présent que lorsque $Q=1$.

Ce modèle n'est pas très réaliste puisque l'on suppose les couches parfaitement uniformes. Nous préférons utiliser le modèle proposé dans [22] où σ_0^2 est non nul, les réflecteurs sont maintenant séparés en deux groupes: des réflecteurs principaux marquent les changements de couches ($Q=1$) et des réflecteurs secondaires, de plus faible amplitude, traduisent les irrégularités existantes à l'intérieur des couches ($Q=0$).

Comme nous l'avons dit, le modèle (*) est paramétré par θ , notons alors P_θ la loi du couple (Y, Z) et $L_\theta(Y, Z)$ sa vraisemblance sous P_θ . Une approche Bayésienne pour estimer θ et reconstruire Y consiste à utiliser la distribution à postériori de Y , $P_\theta(Y/Z)$. Généralement, maximiser la vraisemblance conditionnelle $L_\theta(Y/Z)$ à la fois par rapport au paramètre θ et par rapport à Y conduit à des solutions dégénérées si on n'impose aucune restriction sur l'ensemble des solutions. On trouvera dans [8] plusieurs exemples illustrant ce comportement de la vraisemblance.

Dans [14] et [22], Mendel propose une procédure itérative afin de maximiser cette vraisemblance conditionnelle dans le cas du modèle M2: $L_\theta(Y/Z)$ est maximisé alternativement en Y et en θ . L'estimateur

obtenu de θ est alors biaisé, c'est-à-dire que le vrai paramètre que nous noterons θ^* n'est pas un point fixe de l'algorithme. On dira que θ^* est un point fixe pour un algorithme itératif d'estimation si il vérifie la propriété suivante:

Soit $\theta^{(k)}$, l'estimateur de θ à l'itération k . Si $\theta^{(k)} = \theta^*$ alors $E(\theta^{(k+1)}) = \theta^*$.

Considérons par exemple, le cas simple où le filtre est réduit à l'identité et où le bruit de mesure est nul. Notons f_1 (resp. f_0) la densité d'une normale centrée de variance σ_1^2 (resp. σ_0^2). Pour tout i , la distribution à postériori de $Q^{(i)}$ est donnée par:

$$P(Q^{(i)} = 1 / Z) = \frac{\lambda f_1(Z^{(i)})}{\lambda f_1(Z^{(i)}) + (1-\lambda)f_0(Z^{(i)})}$$

Dans ce cas, $L_\theta(Y/Z)$ est maximum pour $\lambda=1$ et $Q=1$ ou bien pour $\lambda=0$ et $Q=0$, ce qui correspond aux solutions dégénérées décrites plus haut.

Supposons maintenant $\theta^* = (\lambda, \sigma_0^2, \sigma_1^2)$ connu. Soit Q^* maximisant $L_{\theta^*}(Q/Z)$, prédire Y par $Y^* = (Q^*, Z)$ conduit à adopter la règle de décision suivante:

$$Q^{*(i)} = 1 \text{ si } |Z^{(i)}| > \tilde{z},$$

où \tilde{z} est tel que $\lambda f_1(\tilde{z}) = (1-\lambda)f_0(\tilde{z})$.

Soit maintenant $\hat{\theta} = (\hat{\lambda}, \hat{\sigma}_0^2, \hat{\sigma}_1^2)$ l'estimateur de θ maximisant $L_\theta(Y^*/Z)$:

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n Q^{*(i)}, \quad \hat{\sigma}_1^2 = \frac{1}{n\hat{\lambda}} \sum_{i=1}^n Q^{*(i)} Z^{2(i)}, \quad \hat{\sigma}_0^2 = \frac{1}{n(1-\hat{\lambda})} \sum_{i=1}^n (1-Q^{*(i)}) Z^{2(i)}.$$

On voit facilement que $E(\hat{\theta}) \neq \theta^*$ (Supposons par exemple que $\frac{\lambda}{\sigma_1^2} > \frac{1-\lambda}{\sigma_0^2}$, on décidera alors $Q^{*(i)} = 1$ pour tout i , et donc $\hat{\lambda} = 1$).

Une telle procédure ne pouvant être utilisée, nous allons procéder en deux étapes:

- 1) Identification du modèle.
- 2) Estimation de Y .

Le Chapitre 1 est consacré à l'estimation des paramètres. Dans le cas des modèles M1 et M2, la vraisemblance de Z est particulièrement complexe et il n'existe pas de forme explicite de l'estimateur du maximum de vraisemblance de θ . Au contraire, lorsque Y est connu, la vraisemblance du couple (Z, Y) peut facilement être maximisée pour estimer θ . Nous allons donc mettre en place un algorithme itératif où

à chaque itération, un échantillon \mathbf{Y} est simulé et un estimateur de θ est calculé. \mathbf{Y} étant simulé avec la loi à postérieure $P_{\theta}(\mathbf{Y}/\mathbf{Z})$, le vrai paramètre θ^* est un point fixe pour cet algorithme. Cet algorithme peut s'interpréter comme une version stochastique de l'algorithme EM [7], on trouvera des procédures d'estimation semblables dans [4] et [32] dans le cas de traitement d'images et dans [2] dans le cas d'un mélange de populations.

Dans le Chapitre 2, nous nous intéressons à la reconstruction de la suite \mathbf{Y} dans le cas des modèles M1 et M2 lorsque les paramètres sont connus. Nous demanderons à la solution cherchée d'optimiser certains critères naturels de bonne reconstruction concernant les erreurs d'affectation ou la qualité d'ajustement avec les observations. Nous verrons tout particulièrement comment des techniques Bayésiennes, classiques en traitement d'images, peuvent être adaptées au problème de déconvolution de signaux Bernoulli-Gaussiens.

Le Chapitre 3 est consacré à la déconvolution de traces sismiques. Face à des données réelles, les techniques développées sur des modèles théoriques et au moyen de simulations demandent à être adaptées. Alors que jusqu'à présent, et comme c'est généralement le cas dans la littérature ([14], [18], [20], [21], [24], [24], [27]) nous avons étudié chaque trace séparément, nous aborderons maintenant ce problème comme un problème bidimensionnel. En effet, une continuité latérale, due à la structure de couches, est évidente dans le sous-sol et il est naturel de chercher à retrouver cette continuité dans la solution, de plus, nous devons tenir compte du fait que les couches ne sont pas toujours horizontales. Un autre problème délicat à considérer vient de la non-stationnarité du champ \mathbf{Z} , en effet, lorsque l'onde se propage dans le sous-sol ses caractéristiques peuvent être modifiées: l'onde s'amortit et des hautes fréquences disparaissent. D'autre part, il n'est pas raisonnable de considérer le champ \mathbf{Y} comme stationnaire: dans une même section, des zones peuvent être extrêmement accidentées et d'autres, au contraire, très régulières. Nous serons alors amenés à traiter la section par morceaux, l'hypothèse de stationnarité étant faite à l'intérieur de chaque fenêtre.

CHAPITRE 1

ALGORITHME D'ESTIMATION

INTRODUCTION

Soit (Y, Z) un couple de variables aléatoires à valeurs dans un ensemble E et dont la distribution Q appartient à un ensemble \mathcal{R} de lois sur E .

Soit R un élément de \mathcal{R} , on notera R^Z la distribution marginale de Z et $R^{Y/Z}$ la distribution de Y conditionnellement à Z .

Soit \mathcal{P} un sous-ensemble de \mathcal{R} , \mathcal{P} est une famille de lois paramétrées telle que:

- Les éléments de \mathcal{P} sont en bijection avec ceux de Θ , l'espace des paramètres.
- Le modèle est dominé par une mesure μ et tout élément P_θ de \mathcal{P} a une densité p_θ .
- L'hypothèse suivante d'identifiabilité est vérifiée:

$$\forall (P_{\theta_1}, P_{\theta_2}) \in \mathcal{P}, \quad P_{\theta_1}^Z = P_{\theta_2}^Z \Rightarrow \theta_1 = \theta_2$$

Soit \mathcal{Q} une famille de lois de \mathcal{R} définie par:

$$\mathcal{Q} = \{ R \in \mathcal{R}, R^Z = Q^Z \}$$

Faisant l'hypothèse que Q appartient à la famille \mathcal{P} , nous nous proposons d'estimer la loi de (Y, Z) lorsque seule la composante Z est observée, c'est-à-dire chercher un élément de \mathcal{P} appartenant également à \mathcal{Q} . D'après les définitions des ensembles \mathcal{P} et \mathcal{Q} , la solution de ce problème est unique puisque $\mathcal{P} \cap \mathcal{Q} = \{ Q \}$.

Soit θ^* l'élément de Θ tel que $Q = P_{\theta^*}$, nous sommes alors ramenés au problème d'estimation du paramètre θ^* . Nous présentons dans un premier temps un algorithme stochastique d'estimation pour résoudre ce problème, nous comparerons ensuite le comportement de cet algorithme avec une version déterministe: l'algorithme EM. Finalement, des exemples sont proposés pour estimer les paramètres d'un mélange de populations et ceux d'un modèle de convolution.

Cet algorithme est utilisé avec des données réelles dans le Chapitre 3.

1.1 ALGORITHME STOCHASTIQUE D'ESTIMATION

1.1.1 INTRODUCTION

Sans aucun type de restriction, nous supposons Y et Z tous les deux de dimension n . Notre problème est alors d'estimer le paramètre θ au moyen des observations $Z = \{Z(i)\}_{1 \leq i \leq n}$.

Dans de nombreux exemples, et particulièrement dans ceux que nous étudions plus loin (mélange de populations et modèle de convolution), la vraisemblance de Z a une expression complexe, il est alors difficile de trouver un estimateur de θ ayant de bonnes propriétés asymptotiques comme seule fonction de cette variable. Au contraire, lorsque Y et Z sont connus, des estimateurs ayant de bonnes propriétés asymptotiques peuvent être utilisés: estimateur du maximum de vraisemblance, du pseudo-maximum de vraisemblance, des moindres carrés, etc . . .

Partant de cette remarque, nous allons construire un algorithme itératif où Y est simulé à chaque itération pour permettre l'estimation de θ .

Notons $L_{\theta}(Y, Z)$ la vraisemblance de (Y, Z) sous P_{θ} , nous ferons les deux hypothèses suivantes:

- Lorsque (Y, Z) est connu, on peut calculer $\hat{\theta}_n$, l'estimateur du maximum de vraisemblance de θ : $\hat{\theta}_n = \underset{\theta \in \Theta}{\text{ArgMax}} L_{\theta}(Y, Z)$.
- Pour un paramètre θ donné, on peut simuler Y avec la loi conditionnelle $P_{\theta}^{y/z}$. Le couple (Y, Z) suit alors la loi $P_{\theta}^{y/z} P_{\theta^*}^z$.

1.1.2 ALGORITHME

L'algorithme que nous proposons est un algorithme itératif qui requiert le choix d'un paramètre initial et qui, à chaque itération, simule Y et estime θ :

-
- On initialise le paramètre à $\theta(0)$.
 - Soit $\theta^{(i-1)}$, l'estimateur de θ à l'étape $i-1$. A l'étape i :
 - a) On simule $Y^{(i)}$ avec la loi $P_{\theta^{(i-1)}}^{y/z}$
 - b) On calcule $\theta^{(i)}$, le nouvel estimateur de θ ,

$$\theta^{(i)} = \underset{\theta \in \Theta}{\text{ArgMax}} L_{\theta}(Y^{(i)}, Z)$$
 - On arrête la procédure quand l'algorithme a convergé: $\theta^{(i)} \approx \theta^{(i-1)}$.
-

L'idée de cet algorithme est d'utiliser les observations, qui sont une réalisation de la vraie loi P_{θ^*} , comme une force de rappel face aux données simulées sous une fausse loi, et ce jusqu'à une cohérence entre ces deux distributions.

On peut facilement voir que sous les hypothèses d'identifiabilité définies plus haut, le vrai paramètre θ^* est le seul point fixe de cet algorithme: si $\theta^{(i-1)} = \theta^*$ alors $E(\theta^{(i)}) = \theta^*$.

Nous avons présenté ici un algorithme d'estimation dans un cadre très général, toutefois, l'application qui nous a motivé pour mettre en place un tel algorithme est l'identification d'un modèle de convolution. On trouvera dans la littérature d'autres champs d'application où le même type d'algorithme est utilisé:

Dans [2], Celeux et Diebolt considèrent le cas particulier d'un mélange de populations et utilisent ce même algorithme sous le nom de SEM (Stochastic - Expectation - Maximisation), en effet, nous verrons que cet algorithme est une version stochastique de l'algorithme EM (Expectation - Maximisation) [15]. Nous reviendrons plus en détail sur cette application puisqu'elle correspond au cas particulier du modèle de convolution lorsque le filtre est réduit à l'identité.

Chalmond estime les paramètres d'un champ de Gibbs dans le cas de traitement d'images en utilisant une procédure semblable sous le nom de Gibbsian EM [4]. Les paramètres estimés sont d'une part ceux liés à la distribution du bruit et d'autre part ceux liés à celle du champ X mesurant les niveaux de gris de l'image originale. Le nombre d'états du champ X est fini puisqu'on suppose qu'en chaque pixel, m niveaux de gris sont possibles (image m -aire). Dans le cas d'un champ binaire (X prend la valeur 0 ou 1 en chaque pixel), la distribution de X est totalement définie par l'ensemble des $P_{1,k}$, probabilité que $X=1$ en un site sachant que k de ses voisins valent 1. Dans son article, Chalmond estime les $P_{1,k}$ dans le cas d'un système à quatre voisins.

Dans [26], toujours le même type de procédure est employée sous le nom d'algorithme IP (Imputation Posterior). Tanner et Wong s'intéressent essentiellement à des échantillons de petite taille dont certaines données sont manquantes. Ils adoptent un point de vue totalement Bayésien puisque le paramètre θ qui les intéresse, lié à la loi de l'échantillon, est considéré comme une variable aléatoire dont ils estiment la loi à postériori $P(\theta/Z)$, Z étant l'ensemble des données observées. Dans un des exemples proposés, les données sont douze Gaussiennes bidimensionnelles indépendantes, Tanner et Wong s'intéressent alors à la corrélation entre composantes lorsque pour huit de ces douze observations, une des composantes est manquante. Les autres exemples sont consacrés à des tables de contingence incomplètes.

Dans la suite, pour éviter la multiplicité des appellations pour un même algorithme, nous reprendrons le nom de SEM.

1.1.3 CONVERGENCE

Appelons θ_n^{emv} l'estimateur de θ maximisant la vraisemblance de Z . Pour n fixé, la suite $\{\theta^{(i)}\}$ ne converge pas vers un point de Θ , mais vers une variable aléatoire ζ_n . Nous aimerions bien sûr pouvoir définir un ensemble de conditions suffisantes telles que $\sqrt{n}(\zeta_n - \theta_n^{emv})$ converge vers une variable aléatoire Gaussienne centrée.

Ce problème est délicat et requiert une étude particulière à chaque application. A ce sujet, on trouve dans [2] quelques résultats dans le cas d'estimation des paramètres d'un mélange de populations. Nous résumons ici les résultats obtenus par Celeux et Diebolt:

- La suite $\{\theta^{(i)}\}$ constitue une chaîne de Markov homogène ergodique qui converge en loi vers son unique loi stationnaire Ψ_n .
- Sous les hypothèses suivantes: i) Les proportions de chaque population du mélange sont strictement positives ii) La matrice de Fisher au point θ est définie positive, et quelques hypothèses techniques supplémentaires, nous avons le théorème suivant:

Soit ξ_n une variable aléatoire sur Θ de loi Ψ_n . Alors $\sqrt{n}(\xi_n - \theta_n^{emv})$ converge en loi, quand n tend vers l'infini, vers une variable aléatoire Gaussienne centrée de variance Σ . Σ est une matrice qui peut être exprimée en fonction de θ .

Nous n'allons pas essayer de généraliser ce résultat, mais étudier la version déterministe de cet algorithme et tenter d'expliquer, de manière heuristique, son comportement.

1.2 ALGORITHME D'ESTIMATION E.M

1.2.1 INTRODUCTION

Soient \mathcal{R}_1 et \mathcal{R}_2 deux sous-ensembles de \mathcal{R} , \tilde{R}_1 et \tilde{R}_2 les éléments de \mathcal{R}_1 et \mathcal{R}_2 minimisant la distance entre ces deux ensembles:

$$(\tilde{R}_1, \tilde{R}_2) = \underset{R_1 \in \mathcal{R}_1, R_2 \in \mathcal{R}_2}{\text{Argmin}} K(R_1, R_2)$$

$K(R_1, R_2)$ représente la distance de Kullback entre ces deux distributions:

$$K(R_1, R_2) = \begin{cases} \int \text{Log} \frac{dR_1}{dR_2} dR_1 & \text{si } R_1 \ll R_2 \\ +\infty & \text{sinon} \end{cases}$$

Dans [6] Csiszár et Tusnády construisent deux suites $\{R_{1(i)}\}$ et $\{R_{2(i)}\}$ qui convergent vers \tilde{R}_1 et \tilde{R}_2 sous certaines hypothèses. Nous allons reprendre ici cette construction et voir sous quelles conditions les suites convergent vers les distributions recherchées.

On définit les applications suivantes:

$$A_P : \mathcal{P} \rightarrow \mathcal{Q}, \quad A_P(P^{(i)}) = Q^{(i)} \quad \text{tel que} \quad Q^{(i) \frac{y}{z}} = P^{(i) \frac{y}{z}} \quad (\text{et } Q^{(i) z} = Q^z)$$

$$A_Q : \mathcal{Q} \rightarrow \mathcal{P}, \quad A_Q(Q^{(i)}) = P^{(i+1)} \quad \text{tel que} \quad P^{(i+1)} = \underset{P \in \mathcal{P}}{\text{Argmin}} K(Q^{(i)}, P)$$

Pour une initialisation donnée $P^{(0)}$, on a donc construit deux suites $\{P^{(i)}\}$ et $\{Q^{(i)}\}$. Les éléments de \mathcal{P} étant paramétrés, générer une suite $\{P^{(i)}\}$ revient à générer une suite de paramètres $\{\theta^{(i)}\}$, et réciproquement.

Nous allons voir que si \mathcal{P} est un ensemble convexe, alors les suites convergent vers $P_{\theta^*} = Q$.

1.2.2 REPRESENTATION GRAPHIQUE

Pour tout élément R de \mathcal{R} tel que $Q \ll R$, nous avons les relations:

$$K(Q, R) = \int \text{Log} \frac{dQ}{dR} dQ = \int \text{Log} \frac{dQ^z}{dR^z} dQ^z + \iint \text{Log} \frac{dQ^{y/z}}{dR^{y/z}} dQ^{y/z} dQ^z$$

$$K(Q, R) = K(Q^z, P^z) + E_{Q^z}(K(Q^{y/z}, P^{y/z}))$$

Puisque dans la suite les distances sont toujours calculées par rapport à Q , on pose:

$$d(P) = K(Q, P) \quad , \quad d_1(P) = K(Q^z, P^z) \quad \text{et} \quad d_2(P) = E_{Q^z}(K(Q^{y/z}, P^{y/z})).$$

On a bien sûr $d(P) = d_1(P) + d_2(P)$ et on peut représenter un élément R de \mathcal{R} par ses coordonnées $d_1(P)$ et $d_2(P)$. On a alors en particulier $Q = (0, 0)$ et $Q(i) = (0, d_2(Q(i)))$ pour tout i .

Nous présentons en Figure1 une représentation graphique du comportement des suites $\{P(i)\}$ et $\{Q(i)\}$.

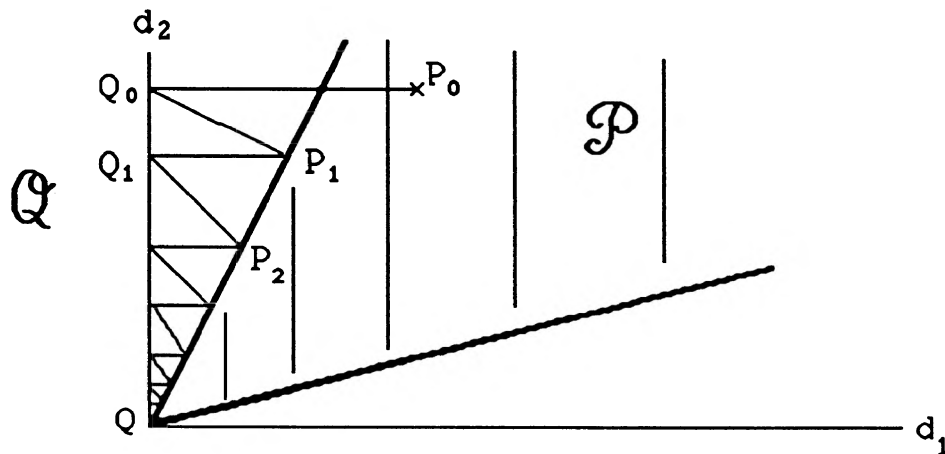


Fig.1 Comportement des suites $\{P(i)\}$ et $\{Q(i)\}$.

Nous allons maintenant voir sous quelles conditions le type de schéma présenté ci-dessus représente bien le comportement des suites $\{P(i)\}$ et $\{Q(i)\}$, c'est-à-dire quand ces suites convergent vers Q .

1.2.3 CONVERGENCE DES SUITES

En utilisant les résultats de [6], nous allons montrer le résultat suivant:

Théorème

|| Si \mathcal{P} est un ensemble convexe, alors les suites $\{P(i)\}$ et $\{Q(i)\}$ convergent vers Q .

Pour démontrer ce théorème, nous allons en fait démontrer les deux lemmes suivants:

Lemme 1

|| Si \mathcal{P} est un ensemble convexe, alors $d(P^{(i+1)}) \leq d(Q^{(i)})$

Pour tout t de $[0, 1]$, on définit $P_t = (1 - t)Q + tP^{(i+1)}$. Comme \mathcal{P} est convexe, $P_t \in \mathcal{P}$ et par définition,

$$K(Q^{(i)}, P^{(i+1)}) \leq K(Q^{(i)}, P_t).$$

On a donc:

$$\frac{K(Q^{(i)}, P^{(i+1)}) - K(Q^{(i)}, P_t)}{1 - t} \leq 0, \text{ pour } t \neq 1.$$

En faisant tendre t vers 1, on obtient que $\frac{d}{dt} K(Q^{(i)}, P_t) \leq 0$ en $t=1$.

D'autre part, comme $K(Q^{(i)}, P_t) = - \int \text{Log} \frac{dP_t}{dQ^{(i)}} dQ^{(i)}$, alors:

$$\frac{d}{dt} K(Q^{(i)}, P_t) = \int \frac{dQ - dP^{(i+1)}}{dP^{(i+1)}} dQ^{(i)} \text{ en } t=1.$$

On en déduit que $\int \frac{dQ}{dP^{(i+1)}} dQ^{(i)} \leq 1$, et donc que $\int \text{Log} \frac{dQ^{(i)}}{dP^{(i+1)}} dQ \leq 0$.

Finalement, on obtient de la dernière inégalité que $K(Q, P^{(i+1)}) \leq K(Q, Q^{(i)})$.

Lemme 2

|| Si $d(P^{(i+1)}) \leq d(Q^{(i)})$, alors $P^{(i)}$ converge vers Q .

Par définition, nous avons $d(Q^{(i+1)}) = d_2(P^{(i+1)}) = d(P^{(i+1)}) - d_1(P^{(i+1)})$. Donc,

$$d(Q^{(i)}) - d(Q^{(i+1)}) = d(Q^{(i)}) - d(P^{(i+1)}) + d_1(P^{(i+1)}).$$

Par hypothèse, $d(P^{(i+1)}) \leq d(Q^{(i)})$, donc $d(Q^{(i)}) - d(Q^{(i+1)}) \geq d_1(P^{(i+1)}) \geq 0$. Et,

$$\sum_{n_1}^{n_2} d_1(P(i)) \leq d(Q(n_1)) - d(Q(n_2)) < +\infty,$$

en faisant tendre n_2 vers l'infini, on déduit $\sum_{n_1}^{+\infty} d_1(P(i)) < +\infty$, donc $d_1(P(i)) \rightarrow 0$ et par conséquent $d(P(i)) \rightarrow 0$.

1.2.4 ALGORITHME EM

L'algorithme EM est décrit dans [2], [6], [7] et [30]. Cet algorithme est un cas particulier de la procédure décrite ci-dessus.

On suppose que tous les éléments de \mathcal{R} ont une densité, soient p_i et q_i les densités des lois $P(i)$ et $Q(i)$. La procédure est alors la suivante:

- On initialise le paramètre à $\theta(0)$.
- Soit $\theta_{(i-1)}$, l'estimateur de θ à l'itération $i-1$. A l'itération i :
 - a) On détermine $Q(i)$ tel que $q_i(\mathbf{Y}, \mathbf{Z}) = q_i(\mathbf{Z}) p_{i-1}(\mathbf{Y} / \mathbf{Z})$: étape E.
 - b) On détermine $P(i)$ en minimisant $K(Q(i), \mathcal{P})$,
(il est alors équivalent de maximiser $E_{P(i-1)}(\text{Log } p(\mathbf{Y}, \mathbf{Z}) / \mathbf{Z})$) : étape M.

Si l'hypothèse de convexité n'est pas vérifiée, les résultats concernant la convergence de cet algorithme restent très locaux. On trouve dans [30] les résultats suivants:

Sans hypothèses préalables, on ne peut rien dire sur la convergence de la suite $\{\theta(i)\}$, celle-ci dépend du point initial $\theta(0)$. Il existe toujours le risque de converger vers un point-selle ou un minimum local de la vraisemblance. Si cette vraisemblance est unimodale et n'admet qu'un point stationnaire, $\{\theta(i)\}$ converge vers l'estimateur du maximum de vraisemblance.

Dans le cas de mélange de populations, Redner et Walker montrent le résultat suivant [25]:

Si la matrice de Fisher calculée au point θ^* (vrai paramètre) est définie positive et si les proportions des populations sont toutes strictement positives, alors, pour n assez grand, l'estimateur du maximum de vraisemblance θ_n^{emv} est bien défini et la suite $\{\theta(i)\}$ converge linéairement vers θ_n^{emv} si le point initial en est suffisamment proche.

1.3 REMARQUES

1.3.1 COMPARAISON ENTRE VERSION STOCHASTIQUE ET VERSION DETERMINISTE

Soient $\{P(i)\}$ et $\{Q(i)\}$ (resp. $\{P_n(i)\}$ et $\{Q_n(i)\}$) les suites définies dans la version déterministe (resp. stochastique). $P(i)$ et $P_n(i)$ sont paramétrées par $\theta(i)$ et $\theta_n(i)$.

Soit $\varepsilon_n(i) = d(Q_n(i)) - d(Q(i))$, nous avons alors $\varepsilon_n(i) = d_2(P_n(i)) - d_2(P(i))$. D'autre part, $d_2(P(i))$ (resp. $d_2(P_n(i))$) est une fonction de $\theta(i)$ (resp. $\theta_n(i)$), posons alors:

$$d_2(P(i)) = g(\theta(i)) \quad , \quad d_2(P_n(i)) = g(\theta_n(i)).$$

$\theta_n(i)$ est l'estimateur du maximum de vraisemblance de $\theta(i)$, donc $g(\theta_n(i))$ est l'estimateur du maximum de vraisemblance de $g(\theta(i))$. D'après les propriétés d'un tel estimateur, $\sqrt{n} (g(\theta_n(i)) - g(\theta(i)))$ converge vers une variable Gaussienne centrée, de variance $\Sigma(i)$, où $\Sigma(i)$ dépend des paramètres du modèle.

$\varepsilon_n(i)$ est la perturbation aléatoire qui différencie la trajectoire déterministe de la trajectoire stochastique. Asymptotiquement, les deux trajectoires coïncident.

1.3.2 AVANTAGE DE LA VERSION STOCHASTIQUE

Pour un paramètre θ donné, la loi conditionnelle $P_\theta^{y/z}$ ne peut être calculée explicitement que dans certains cas. Lorsque les données sont dépendantes, on n'aura pas toujours moyen de résoudre les équations de vraisemblance et effectuer les étapes E et M de l'algorithme EM. Au contraire, cela n'est pas un obstacle pour la mise en place l'algorithme SEM, il suffit en effet de pouvoir simuler la loi conditionnelle $P_\theta^{y/z}$, ce qui sera souvent possible en utilisant un échantillonneur de Gibbs, par exemple. Nous verrons plus loin une telle application où on estime les paramètres d'un modèle de convolution.

1.3.3 \mathcal{P} NON CONVEXE

Dans de très nombreux cas, la famille de loi \mathcal{P} n'est pas convexe. Les algorithmes de type déterministe comme l'EM, visant à maximiser la vraisemblance de Z , convergent vers des solutions qui dépendent du point initial, du fait de la présence de maxima locaux ou de points selle dans la fonction de vraisemblance. Si dans un voisinage de P_{θ^*} , la distribution recherchée, \mathcal{P} est localement convexe, les résultats obtenus en 1.3.2 restent valables si l'initialisation est faite dans ce voisinage.

Dans le cas du SEM, la suite des estimateurs est une chaîne de Markov homogène, on peut alors espérer éviter certains de ces maxima locaux. Maintenant, si le nombre de données est important, nous

avons vu que les algorithmes stochastique et déterministe ont des trajectoires très proches. Les résultats obtenus avec le SEM dépendent alors du point initial. D'un autre côté, les perturbations aléatoires seront importantes si il y a peu de données, les maxima locaux seront évités, mais l'estimation sera aussi très erratique.

Un algorithme idéal serait d'utiliser d'une part la nature aléatoire de la suite des estimateurs pour ne pas rester dans le voisinage de mauvaises solutions, et d'autre part rendre l'algorithme déterministe lorsqu'on est dans un voisinage convexe de la solution. Celeux et Diebolt proposent sous le nom de SAEM (Simulating Annealing EM) une version de type recuit simulé [3]:

Soient $\theta_{(i)}^{em}$ l'estimateur de θ obtenu par l'algorithme EM à l'itération i et $\theta_{(i)}^{sem}$ celui obtenu par l'algorithme SEM. On pose alors:

$$\theta_{(i)}^{saem} = \theta_{(i)}^{em} + \gamma_{(i)} (\theta_{(i)}^{sem} - \theta_{(i)}^{em}).$$

La suite $\{\gamma_{(i)}\}$ décroît lentement vers 0 quand i tend vers l'infini en partant de $\gamma_{(0)}=1$. Ce paramètre joue le même rôle que la température dans les algorithmes classiques de refroidissement. L'algorithme est utilisé pour l'identification de mélanges de densités de probabilité, lorsque la probabilité à postériori $P_{\theta_{(i)}}^{y/z}$ peut être calculée explicitement.

Dans un cadre plus général, on peut simuler plusieurs réalisations de la loi $P_{\theta_{(i)}}^{y/z}$ et diminuer ainsi la variance de l'estimateur. En reprenant les notations utilisées en 1.3.1, la perturbation $\epsilon_{n(i)}$ aura une variance divisée par $m(i)$ si on effectue $m(i)$ simulations indépendantes à l'itération i au lieu d'une seule. On a alors un nouvel algorithme: $d(Q_{n(i)}) = d(Q_{(i)}) + \frac{1}{\sqrt{m(i)}} \epsilon_{n(i)}$. Le comportement de cet algorithme est similaire à celui du SAEM si ce nombre de simulations croît à chaque itération, c'est-à-dire si $\{m(i)\}$ tend vers l'infini en partant de 1. Le problème, qui semble très difficile, consiste alors à trouver la suite $\{m(i)\}$ qui optimise la convergence de la suite $\{\theta_{n(i)}\}$ vers le bon paramètre θ^* .

Il est également difficile de calculer explicitement la variance de $\theta_{n(i)}$. Néanmoins, si la règle d'arrêt est fournie par un critère de variance (on arrête les itérations lorsque par exemple $E(\theta_{n(i+1)} - \theta_{n(i)})^2 \leq \delta$, où δ est un seuil), on peut estimer la suite des variances empiriquement en effectuant plusieurs simulations.

1.3.4 ERREURS DE MODELISATION

Dans le cas d'une application, face à des données réelles, dire que Q appartient à une famille \mathcal{P} de lois paramétrées est une approximation, on utilise un modèle que l'on espère le plus proche possible de la réalité.

Lorsque \mathcal{P} est convexe, les suites de lois $\{P(i)\}$ et $\{Q(i)\}$ décrites en 1.3.2 ne convergent alors plus vers la vraie distribution Q mais vers des lois \tilde{P} et \tilde{Q} telles que:

$$(\tilde{P}, \tilde{Q}) = \underset{P \in \mathcal{P} \quad Q \in \mathcal{Q}}{\text{Argmin}} K(P, Q)$$

\tilde{Q} vérifie alors $\tilde{Q}^z = Q^z$ et $\tilde{Q}^{y/z} = \tilde{P}^{y/z}$

(La démonstration est identique à celle du théorème 1.3.2: on montre que $d(P(i+1)) - d(\tilde{P}) \leq d(Q(i))$ et que $d(P(i+1)) - d(\tilde{P}) \rightarrow 0$ en prenant cette fois-ci \tilde{Q} comme origine (Figure 2)).

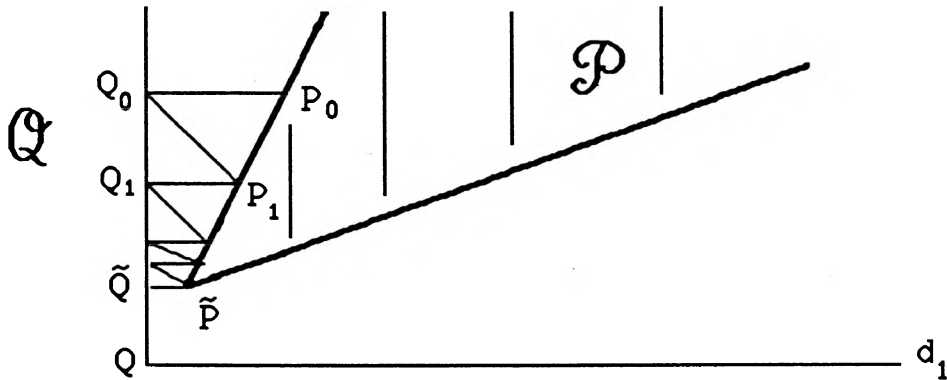


Fig.2 Comportement des suites $\{P(i)\}$ et $\{Q(i)\}$ lorsque $Q \notin \mathcal{P}$.

1.4 EXEMPLES

1.4.1 MELANGES DE POPULATIONS

1.4.1.a Mise en place de l'algorithme

Soit un n -échantillon d'une variable aléatoire (X, Q) à valeurs dans $\mathbb{R}^p \times \mathbb{N}$ définie par:

- $P(Q = k) = \lambda_k \quad 1 \leq k \leq m$
- Conditionnellement à $Q = k$, X suit une loi P_k . (On supposera P_k paramétré par un paramètre α_k et ayant une densité $p(\cdot, \alpha_k)$)

Dans les exemples proposés, X est une variable Gaussienne, donc $\alpha_k = (m_k, \sigma_k^2)$. On cherche alors à estimer $\theta = (\lambda_1, m_1, \sigma_1^2, \dots, \lambda_m, m_m, \sigma_m^2)$ lorsque seul $X = \{X(i)\}_{1 \leq i \leq n}$ est observé.

La condition d'identifiabilité est vérifiée si les α_k sont tous différents et si les λ_k sont tous strictement positifs.

La mise en place de l'algorithme est très simple dans ce cas, puisque les données sont indépendantes:

-
- On initialise θ à $\theta^{(0)}$.
 - A l'étape i ,
 - a) les $Q(j)$ sont simulés indépendamment avec la loi à postériori:

$$P(Q(j) = k / X(j) = x(j)) = \frac{\lambda_{k^{(i-1)}} p(x(j), \alpha_{k^{(i-1)}})}{\sum_{p=1}^m \lambda_{p^{(i-1)}} p(x(j), \alpha_{p^{(i-1)}})}$$

On pose alors $t_{k^{(j)}} = 1$ si $Q(j) = k$, 0 sinon.

- b) On estime θ par l'estimateur du maximum de vraisemblance:

$$\lambda_{k^{(i)}} = \frac{1}{n} \sum_{j=1}^n t_{k^{(j)}} \quad , \quad m_{k^{(i)}} = \frac{1}{\sum t_{k^{(j)}}} \sum_{j=1}^n t_{k^{(j)}} x(j) \quad , \quad \sigma_{k^{(i)}}^2 = \frac{1}{\sum t_{k^{(j)}}} \sum_{j=1}^n t_{k^{(j)}} (x(j) - m_{k^{(i)}})^2.$$

1.4.1.b Simulations

Exemple 1

Nous avons simulé 2 populations avec les paramètres suivants:

$$\begin{array}{llll} n_1 = 3000 & \lambda_1 = 0.6 & m_1 = 0 & \sigma_1^2 = 1 \\ n_2 = 2000 & \lambda_2 = 0.4 & m_2 = 1 & \sigma_2^2 = 2 \end{array}$$

La suite $\theta^{(i)}$ est représentée en Figure 3.

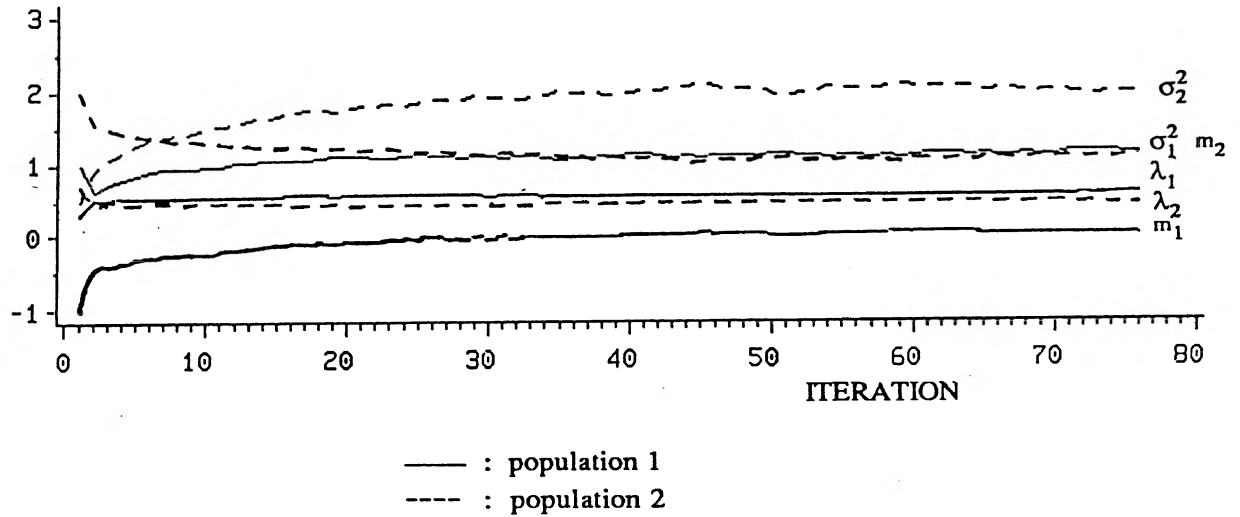


Fig.3 Convergence de l'estimateur $\theta(i)$ (2 populations).

Exemple 2

3 populations ont été simulées avec les paramètres suivants:

$n_1 = 2000$	$\lambda_1 = 0.27$	$m_1 = -1$	$\sigma_1^2 = 0.2$
$n_2 = 2500$	$\lambda_2 = 0.33$	$m_2 = +1$	$\sigma_2^2 = 0.2$
$n_3 = 3000$	$\lambda_3 = 0.40$	$m_3 = 0$	$\sigma_3^2 = 1$

Nous avons représenté en Figure 4 les erreurs quadratiques $e_\lambda(i)$, $e_m(i)$ et $e_\sigma(i)$.

$$e_\lambda(i) = 100 \sum_{k=1}^3 (\lambda_{k(i)} - \lambda_k)^2, \quad e_m(i) = \sum_{k=1}^3 (m_{k(i)} - m_k)^2, \quad e_\sigma(i) = \sum_{k=1}^3 (\sigma_k^2(i) - \sigma_k^2)^2$$

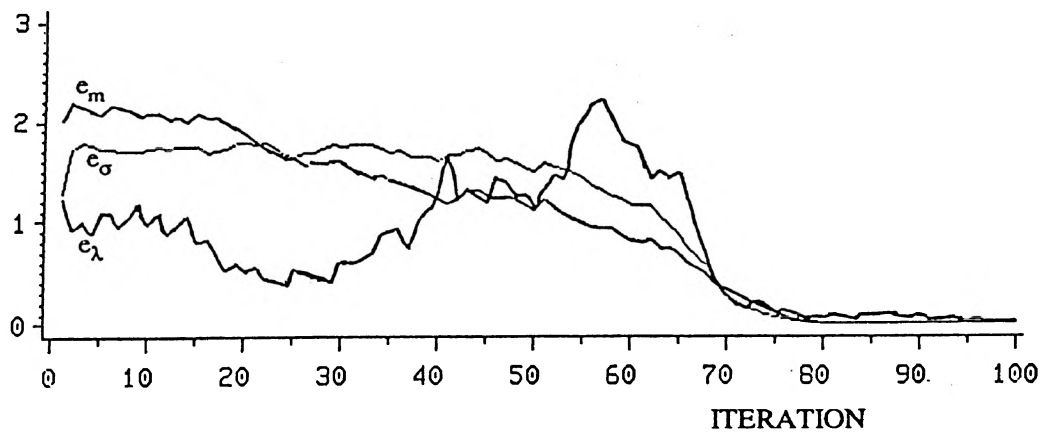


Fig.4 Erreurs d'estimation (3 populations).

Comme nous l'avons dit, Celeux et Diebolt étudient cette application très en détail dans [2] et [3]. On trouvera dans leurs travaux de nombreux exemples et également une technique permettant d'estimer m , le nombre de populations:

On commence par faire tourner l'algorithme avec un nombre de populations que l'on peut raisonnablement considérer comme un majorant de m . Si au cours des itérations, un des λ_k estimés est plus petit qu'un seuil donné, on peut penser que le nombre de populations a été surestimé: on recommence alors avec une population de moins.

1.4.2 MODELE DE CONVOLUTION

1.4.2.a Description du modèle

Comme nous l'avons dit en introduction, le modèle de convolution que nous proposons ici est utilisé pour modéliser des traces sismiques (Chapitre 3), nous allons le traiter ici dans un cadre général. Le modèle est le suivant:

$$\begin{cases} Z(k) = \sum_{i=p}^p f_i X(k-i) + \varepsilon(k) \\ Y(k) = (Q(k), X(k)) \end{cases}$$

$Z = \{Z(k)\}_{1 \leq k \leq n}$ sont les observations, $Y = \{Y(k)\}_{1 \leq k \leq n}$ une suite de variables aléatoires indépendantes et ε un bruit blanc Gaussien de variance σ_ε^2 .

Q est une suite de variables de Bernoulli de paramètre λ et X un mélange de 2 Gaussiennes:

- Conditionnellement à $Q = 0$, X suit une loi $\mathcal{N}(0, \sigma_0^2)$
- Conditionnellement à $Q = 1$, X suit une loi $\mathcal{N}(0, \sigma_1^2)$

Le champ (Z, Y) est supposé stationnaire.

Les paramètres déterminant la distribution de Y sont alors $(\sigma_0^2, \sigma_1^2, \lambda)$ et l'ensemble des paramètres à estimer est:

$$\theta = (f, \sigma_\varepsilon^2, \sigma_0^2, \sigma_1^2, \lambda).$$

1.4.2.b Identifiabilité du modèle

Dans les simulations présentées dans ce paragraphe, le filtre f est celui utilisé en géophysique (Fig. 5-b). Nous pouvons voir dans la Fig. 5 un exemple de non-unicité de la décomposition $(\sigma_0^2, \sigma_1^2, \lambda)$, en effet la troisième séquence présentée en Fig.5-a appartient au noyau de f : la convolution de toute combinaison linéaire de cette séquence avec f donne 0. Nous pouvons donc trouver une infinité de suites Y ayant des distributions différentes et produisant la même convolution. Une contrainte doit alors être fixée arbitrairement, nous avons choisi ici de fixer la valeur de λ (nous verrons dans le Chapitre 3 que cela revient à imposer une densité de réflecteurs).

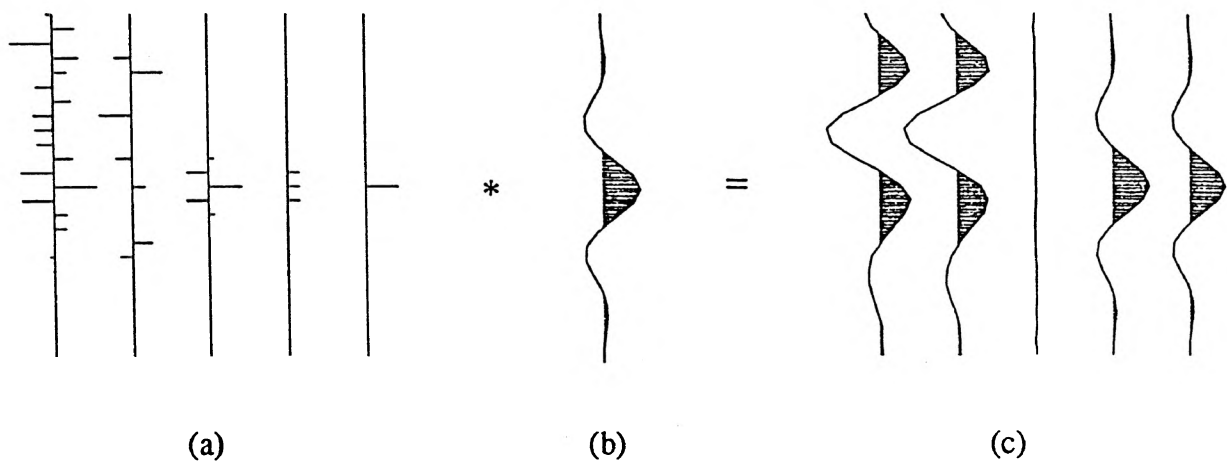


Fig. 5 Exemple de convolution
(a) Réflecteurs (b) Filtre (c) Convolution

Nous avons les égalités suivantes: $(af) * (\frac{X}{a}) = (-f) * (-X) = f * X$. Pour avoir unicité du filtre, il faut donc:

- imposer sa phase, en fixant le signe d'un des f_i .
- choisir une échelle en fixant l'énergie du filtre $\sum_{i=-p}^p f_i^2$ ou la variance de X .

Finalement, nous allons estimer $\theta = (f, \sigma_\epsilon^2, \sigma_0^2, \sigma_1^2)$ avec les contraintes $f_0 > 0$ et $\sum f_i^2 = E_f$ où E_f est choisi arbitrairement.

1.4.2.c Mise en place de l'algorithme

Comme nous l'avons vu, l'implémentation de l'algorithme d'estimation requiert la résolution de deux problèmes:

Simulation de Y

La simulation de Y est réalisée en utilisant un échantillonneur de Gibbs, en effet, si les $Y(i)$ ont des distributions à priori indépendantes, à postériori, ils sont dépendants et ne peuvent être simulés séparément.

Soit θ un paramètre donné et soit P_θ la loi de (Y, Z) . Nous avons les relations:

$$P_\theta(Y / Z) = \frac{P_\theta(Z / Y) P_\theta(Y)}{P_\theta(Z)} \quad (1)$$

$$P_\theta(Y / Z) = \frac{P_\theta(Z / X) P_\theta(X / Q) P_\theta(Q)}{P_\theta(Z)} \quad (2)$$

Avec,

$$P_\theta(Z / X) = \prod_{k=1}^n \left(\frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(\frac{-1}{2\sigma_\epsilon^2} (Z(k) - f \cdot X(k))^2\right) \right) \quad (3)$$

$$P_\theta(X / Q) = \prod_{k=1}^n \left(\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(\frac{-X(k)^2}{2\sigma_1^2}\right) \right)^{Q(k)} \left(\frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(\frac{-X(k)^2}{2\sigma_0^2}\right) \right)^{1-Q(k)} \quad (4)$$

$$P_\theta(Q) = \prod_{k=1}^n \lambda^{Q(k)} (1 - \lambda)^{1-Q(k)} \quad (5)$$

$P_\theta(Z)$ est une constante puisque Z est l'observation, la loi à postériori de Y est donc connue et peut être simulée (toute la procédure est décrite en annexe).

Estimation de θ

Lorsque le couple (Y, Z) est connu, on peut calculer $\hat{\theta}$ l'estimateur du maximum de vraisemblance de θ . Soit $L_\theta(Y, Z)$ la vraisemblance de (Y, Z) en θ ,

$$\hat{\theta} = \underset{\theta \in \Theta}{\text{Argmax}} L_\theta(Y, Z).$$

D'après (3), (4) et (5) le calcul de $\hat{\theta}$ peut être divisé en 3 parties:

* \hat{f} et $\hat{\sigma}_\varepsilon^2$ sont obtenus en maximisant $P_\theta(Z/X)$.

Comme ε est Gaussien, il est équivalent de calculer \hat{f} en maximisant $P_\theta(Z/X)$ ou en minimisant $\|Z - f * X\|^2$. L'estimateur du maximum de vraisemblance et celui des moindres carrés coïncident:

$$\hat{f} = (X'X)^{-1} X'Z$$

Z est le vecteur colonne $(Z(1), Z(2), \dots, Z(n))^t$

X est la matrice $n \times (2p+1)$

$$\begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ X(i-p-1) & X(i-p) & \cdot & \cdot & X(i+p-1) \\ X(i-p) & X(i-p+1) & \cdot & \cdot & X(i+p) \\ X(i-p+1) & X(i-p+2) & \cdot & \cdot & X(i+p+1) \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

La variance du bruit est estimée par:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n} \sum_{k=1}^n (Z(k) - \hat{f} * X(k))^2$$

* $\hat{\sigma}_0^2$ et $\hat{\sigma}_1^2$ sont obtenus en maximisant $P_\theta(X/Q)$.

$$\hat{\sigma}_0^2 = \frac{1}{n_0} \sum_{k=1}^n (1 - Q(k)) X(k)^2 \quad \hat{\sigma}_1^2 = \frac{1}{n_1} \sum_{k=1}^n Q(k) X(k)^2$$

n_0 et n_1 sont définis par : $n_0 = \sum_{k=1}^n (1 - Q(k))$ et $n_1 = n - n_0 = \sum_{k=1}^n Q(k)$

* $\hat{\lambda}$ est obtenu en maximisant $P_\theta(Q)$.

$$\hat{\lambda} = \frac{n_1}{n}$$

1.4.2.d Simulations

Nous avons simulé une suite $Z = \{Z(k)\}_{1 \leq k \leq n}$ avec $n=1500$ (cela représente un nombre moyen de données dans une trace sismique).

Les paramètres utilisés sont les suivants: $\sigma_0^2 = 0.1$, $\sigma_1^2 = 1$ et $\lambda = 0.1$. Le rapport signal/bruit ($\sigma_z^2/\sigma_\varepsilon^2$) est de 50dB et le filtre est présenté en Figure 6.

La procédure d'estimation contient 2 étapes:

- On utilise l'algorithme d'estimation pour estimer f , σ_ε^2 , σ_0^2 , σ_1^2 et λ , seuls \hat{f} et $\hat{\sigma}_\varepsilon^2$ sont retenus. L'estimation du filtre est présentée en Fig.6-b et celle du rapport signal/bruit a donné 45dB.

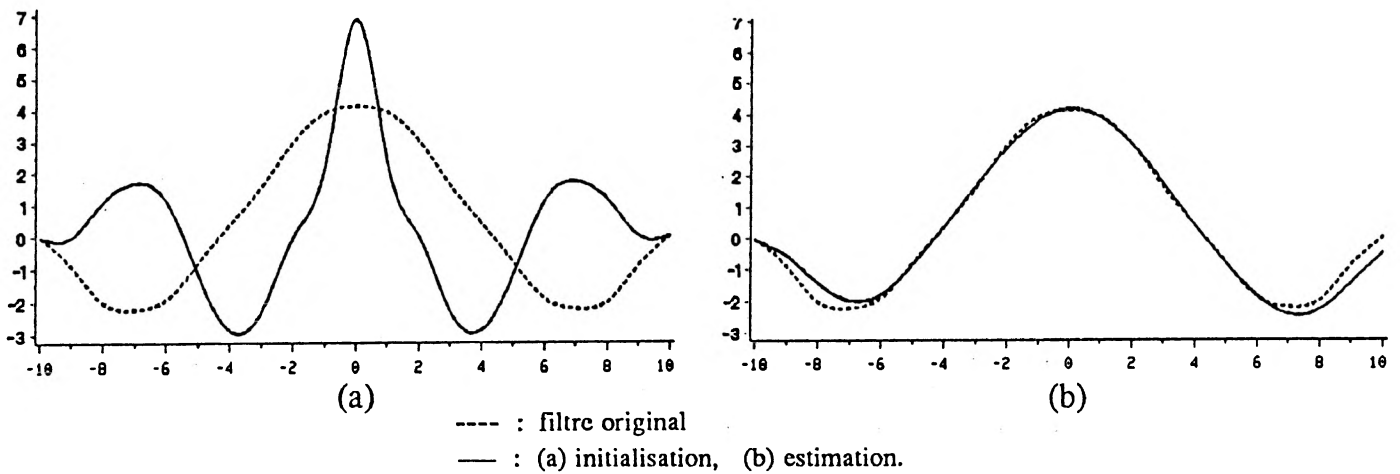


Fig. 6 Estimation du filtre.

- f et σ_ε^2 sont fixés à leurs valeurs estimées et λ à sa vraie valeur. On estime alors σ_0^2 et σ_1^2 . Les résultats obtenus avec différentes initialisations sont en Figure 7.

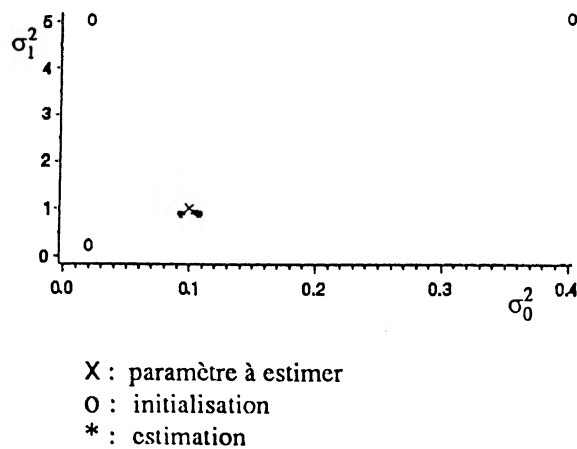


Fig. 7 Estimation des variances σ_0^2 et σ_1^2 .

ANNEXE

L'échantillonneur de Gibbs

Nous voulons simuler $\mathbf{Y} = \{Y(k)\}_{1 \leq k \leq n}$ avec la distribution à postérieure $P(\mathbf{Y}/\mathbf{Z})$ où $\mathbf{Z} = \{Z(k)\}_{1 \leq k \leq n}$ est l'observation. La procédure est la suivante:

- On initialise $\mathbf{Y} = (y_1, y_2, \dots, y_n)$
- A l'itération k :
 - a) On choisit un site i de $[1, n]$ (de manière cyclique ou aléatoire).
 - b) On simule $Y(i)$ avec la loi $P(Y(i)/\mathbf{Z}, Y(j)=y_j, j \neq i)$. Pour $j \neq i$, les $Y(j)$ restent inchangés.

Si on passe par tous les sites un grand nombre de fois, la distribution de \mathbf{Y} converge alors vers la distribution à postérieure $P(\mathbf{Y}/\mathbf{Z})$. On trouvera dans [32] une étude détaillée de cet échantillonneur et de ses propriétés ergodiques.

Calcul de la loi conditionnelle $P(Y(i)/\mathbf{Z}, Y(j), j \neq i)$

D'après Bayes, nous avons:

$$P(Y(i)/\mathbf{Z}, Y(j), j \neq i) = \frac{P(\mathbf{Z}/\mathbf{Y}) P(Y(i)/Y(j), j \neq i)}{P(\mathbf{Z} / Y(j), j \neq i)} \quad (6)$$

Comme les $Y(i)$ sont indépendants,

$$P(Y(i)/\mathbf{Z}, Y(j), j \neq i) = \frac{P(\mathbf{Z}/\mathbf{Y}) P(X(i)/Q(i))P(Q(i))}{P(\mathbf{Z} / Y(j), j \neq i)} \quad (7)$$

Dans notre cas, la loi conditionnelle de $Y(i)$ est un mélange de deux lois normales. En effet, d'après les équations (3), (4), (5) et (7), on déduit :

$$P(Y(i)=(x(i),q(i))/ Z, Y(j)) = \left(\frac{\lambda_i}{\sqrt{2\pi V_{1,i}}} \exp\left(\frac{-1}{2V_{1,i}} (x(i)-m_{1,i})^2\right) \right)^{q(i)} \left(\frac{1-\lambda_i}{\sqrt{2\pi V_{0,i}}} \exp\left(\frac{-1}{2V_{0,i}} (x(i)-m_{0,i})^2\right) \right)^{1-q(i)}$$

Les paramètres λ_i , $m_{0,i}$, $V_{0,i}$, $m_{1,i}$ et $V_{1,i}$ sont calculés par identification:

Soit μ_i la prédiction au sens des moindres carrés de $X(i)$:

$$\mu_i = \underset{x(i)}{\text{ArgMin}} \| Z - f * X \|^2$$

En posant $U_{i,k} = Z_{i+k} - \sum_{j \neq k} f_j x_{i+k-j}$ et $E_f = \sum_{j=-p}^{+p} f_j^2$ on obtient,

$$\mu_i = \frac{1}{E_f} \sum_{k=-p}^{+p} f_k U_{i,k}$$

On en déduit alors,

$$V_{0,i} = V_0 = \left(\frac{E_f}{\sigma_\varepsilon^2} + \frac{1}{\sigma_0^2} \right)^{-1}, \quad V_{1,i} = V_1 = \left(\frac{E_f}{\sigma_\varepsilon^2} + \frac{1}{\sigma_1^2} \right)^{-1}$$

$$m_{0,i} = \frac{E_f}{\sigma_\varepsilon^2} V_0 \mu_i, \quad m_{1,i} = \frac{E_f}{\sigma_\varepsilon^2} V_1 \mu_i$$

$$\lambda_i = \left[1 + \frac{1-\lambda}{\lambda} \sqrt{\frac{V_0 \sigma_1^2}{V_1 \sigma_0^2}} \exp \left\{ \frac{1}{2} \left(\frac{m_{0,i}^2}{V_0} - \frac{m_{1,i}^2}{V_1} \right) \right\} \right]^{-1}$$

CHAPITRE 2

TECHNIQUES BAYESIENNES DE DÉCONVOLUTION. APPLICATION A DES SIGNAUX BERNOULLI-GAUSSIENS.

INTRODUCTION

Comme dans le chapitre précédent consacré à l'estimation de paramètres, nous nous intéressons au modèle de convolution:

$$\begin{cases} \mathbf{Z} = \mathbf{f} * \mathbf{X} + \boldsymbol{\varepsilon} \\ \mathbf{Y} = (\mathbf{Q}, \mathbf{X}) \end{cases}$$

Dans ce chapitre, nous nous consacrerons à la reconstruction de la suite \mathbf{Y} lorsque le modèle est identifié et lorsque \mathbf{Z} est observé. Nous traitons ici uniquement le cas unidimensionnel, le Chapitre 3 sera consacré à la déconvolution bidimensionnelle de traces sismiques.

\mathbf{Y} est une suite de variables aléatoires indépendantes et équidistribuées où \mathbf{Q} est une suite de variables de Bernoulli et \mathbf{X} un mélange de 2 populations Gaussiennes centrées, de variances différentes. Nous nous intéresserons tout particulièrement au cas où une des variances est nulle: \mathbf{Y} est alors une suite de variables Bernoulli-Gaussiennes.

Pour illustrer les éventuelles incohérences qui peuvent exister entre une fonctionnelle à maximiser et des critères de bonne reconstruction que l'on se donne, nous consacrerons la première partie à un problème simple: la prédiction dans le cas d'un mélange de populations (un cas particulier est celui de la déconvolution lorsque le filtre est réduit à l'identité).

La seconde partie traite du modèle de convolution et des différents algorithmes Bayésiens de reconstruction. Ces algorithmes sont généralement utilisés en traitement d'image [31] et peuvent parfaitement être adaptés à notre problème. Néanmoins, nous verrons que les probabilités à postériori (globale ou marginales) ne peuvent être utilisées comme fonctionnelles à maximiser sans quelques précautions préalables: nous serons conduits à travailler sous des fausses lois pour optimiser certains critères naturels de bonne reconstruction.

La troisième partie est consacrée exclusivement au modèle Bernoulli-Gaussien, ce type de modèle a été largement étudié puisqu'on rencontre de nombreuses applications (échographie, géophysique, signal, etc . . .) [18], [12], [13], [14], [21]. Nous verrons que quelques modifications apportées aux algorithmes Bayésiens traditionnels permettent d'améliorer notablement les résultats.

Dans le cas du modèle Bernoulli-Gaussien, le problème de déconvolution peut être décomposé en deux sous-problèmes: un problème de détection (localiser les signaux non nuls) et un problème d'estimation (prédire leurs valeurs). Ces deux problèmes sont indissociables, en effet, quelque-soit l'algorithme utilisé, un signal est détecté si et seulement si sa valeur prédite est supérieure en valeur absolue à un certain seuil. Ce seuil est déterminé au moyen de critères de bonne reconstruction concernant le nombre de mauvaises détections, le nombre d'oublis et la qualité d'ajustement avec les observations.

Différentes reconstructions d'une suite B-G, issues de différents algorithmes, sont présentées à la fin de ce chapitre. Le temps de calcul n'a pas été ici notre préoccupation majeure, comme c'est le cas dans [12], toutefois, nous verrons qu'un algorithme type ICM conduit très rapidement à une excellente solution. Alors que dans l'algorithme ICM traditionnel, le seuil de détection définit ci-dessus est constant, nous allons ici le rendre variable pour éviter de converger vers une mauvaise configuration. Des algorithmes type MAP et MPM sont également proposés.

2.1 PREDICTION DANS LE CAS D'UN MELANGE DE POPULATIONS

Soit Ω , un ensemble de variables aléatoires indépendantes composé de 2 populations Ω_0 et Ω_1 . On note λ la proportion de la population Ω_1 , P_i la loi des éléments de Ω_i et p_i la densité associée pour $i=0,1$.

Soit (Q, X) un élément de Ω :

- Q indique à quelle population appartient Y :

$$P(Q = 1) = \lambda \quad P(Q = 0) = 1 - \lambda$$

- X est la valeur de Y , on note:

$$P(X = x / Q = q) = p_q(x) \quad , \quad q = 0,1$$

$$P(X = x) = \lambda p_1(x) + (1 - \lambda) p_0(x)$$

La loi de Y est alors donnée par:

$$P(Y = (q, x)) = P(X=x / Q=q) P(Q=q) = \lambda^q (1-\lambda)^{1-q} p_q(x)$$

Nous nous plaçons dans le cas où le paramètre λ et les lois P_q sont connus, nous nous intéressons alors au double problème de décision-estimation:

- Décider à quelle population appartient Y (prédire Q).
- Estimer sa valeur (prédire X)

Plusieurs cas sont à envisager:

2.1.1 PROBLEME DE DECISION : X est connu et Q est inconnu

Nous pouvons calculer la distribution conditionnelle de Q:

$$\tilde{\lambda} = P(Q=1 / X=x) = \frac{\lambda p_1(x)}{\lambda p_1(x) + (1 - \lambda) p_0(x)}$$

La règle de décision naturelle est la suivante: on décide $Q=1$ si $\tilde{\lambda} \geq \alpha$. Une fonction de coût doit être définie pour choisir le seuil α et il est naturel de construire celle-ci à partir des erreurs de classement:

Pour $\tilde{\lambda}$ donné, on appelle J_1 la probabilité de décider $Q=1$ lorsque $Q=0$ et J_0 celle de décider $Q=0$ quand $Q=1$.

$$J_1 = (1 - \tilde{\lambda}) \mathbf{1}_{\tilde{\lambda} \geq \alpha} \quad , \quad J_0 = \tilde{\lambda} \mathbf{1}_{\tilde{\lambda} \leq \alpha}$$

Lorsqu'on dispose d'un n-échantillon comme observation, on s'intéresse au nombre de mauvais classements, on cherche à minimiser les deux types d'erreurs $E_1 = E(J_1)$ et $E_0 = E(J_0)$, où l'espérance est calculée sous la loi de $\tilde{\lambda}$.

Soit t appartenant à $[0, 1]$, on se donne comme fonction de coût:

$$J_D = t E_1 + (1 - t) E_0$$

Théorème

|| La fonction de coût J_D est minimisée en choisissant comme seuil $\alpha = t$, et cela indépendamment de la loi de $\tilde{\lambda}$.

Démonstration

Nous avons,

$$J_D = t E [(1 - \tilde{\lambda}) \mathbf{1}_{\tilde{\lambda} \geq \alpha}] + (1 - t) E [\tilde{\lambda} \mathbf{1}_{\tilde{\lambda} \leq \alpha}]$$

Soit $P(\tilde{\lambda})$ la loi de $\tilde{\lambda}$,

$$J_D = t \int_{\alpha}^1 (1 - \tilde{\lambda}) dP(\tilde{\lambda}) + (1 - t) \int_0^{\alpha} \tilde{\lambda} dP(\tilde{\lambda})$$

On peut facilement vérifier que la dérivée partielle de J_D par rapport à α s'annule pour $\alpha = t$.

Remarque

Ce résultat peut s'appliquer dans le cas général où $\lambda = P(Q=1)$ est une variable aléatoire de loi inconnue, étant fonction des observations, par exemple. Le problème est alors d'estimer Q lorsque λ est connu.

2.1.2 PROBLEME D'ESTIMATION : X est inconnu et Q est connu

Connaître Q veut dire connaître la loi de X , puisque conditionnellement à $Q=q$, X suit une loi P_q . La fonction de coût raisonnable et fréquemment utilisée pour estimer X est l'erreur de prédiction $\|X - \hat{X}\|_p$ calculée sous la loi P_q , où \hat{X} est la prédiction de X .

Dans les exemples que nous traitons plus loin, ainsi que dans les applications du Chapitre 3, les variables sont des mélanges de Gaussiennes. Nous estimerons donc toujours X en minimisant l'erreur quadratique $J_E = E_{P_q}(X - \hat{X})^2$, c'est-à-dire en maximisant la vraisemblance de X :

$$\hat{X} = \underset{x}{\text{Argmax}} p_q(x)$$

\hat{X} est alors le mode de P_q : $\hat{X} = E_{P_q}(X)$.

2.1.3 PROBLEME DE DECISION-ESTIMATION : X et Q sont inconnus

Les deux fonctions de coût définies plus haut, J_D et J_E associées aux erreurs de détection et aux erreurs d'estimation doivent servir à prédire $Y = (Q, X)$. En effet, alors que la loi de Y se décompose en une partie discrète $P(Q)$ et une partie continue $P(X/Q)$, la fonction de coût se décompose en un terme lié aux erreurs de classement et un terme lié à l'erreur d'estimation. Par conséquent, on est amené à estimer Y en deux temps, de façon hiérarchique:

- On estime Q en minimisant J_D .
- On estime X en minimisant J_E .

Si les deux types d'erreur de classement définis en 2.1.1 ont le même poids ($t=1/2$) et si la minimisation de J_E est obtenue en maximisant la vraisemblance de X , alors \hat{Y} , la prédiction de Y , est défini par:

$$\hat{Q} = \underset{q}{\text{Argmax}} \lambda^q (1-\lambda)^{1-q}$$

$$\hat{X} = \underset{x}{\text{Argmax}} p_Q(x)$$

Soit m_q le mode de la loi P_q , la fonctionnelle à maximiser est alors donnée par:

$$\hat{Y} = \underset{x, q}{\text{Argmax}} \lambda^q (1-\lambda)^{1-q} \frac{p_q(x)}{p_q(m_q)}$$

Remarque

Choisir d'estimer Y en maximisant la vraisemblance $p(Y)$ peut conduire à de très mauvais résultats, supposons que $p(Y)$ atteint son maximum en $\tilde{Y} = (q, x)$:

Pour tout q et tout x , $\lambda^{\tilde{q}} (1-\lambda)^{1-\tilde{q}} p_{\tilde{q}}(x) \geq \lambda^q (1-\lambda)^{1-q} p_q(x)$. La probabilité d'effectuer un mauvais classement est donnée par $P(Q \neq \tilde{q}) = (1-\lambda)^{\tilde{q}} \lambda^{1-\tilde{q}}$, dès que $p_{\tilde{q}}(x) \gg p_q(x)$, cette probabilité peut être très grande.

Considérons par exemple le cas de deux Gaussiennes: $P(Q=1) = 1 - P(Q=0) = \lambda$, conditionnellement à $Q=q$, $X \sim \mathcal{N}(m_q, \sigma_q^2)$:

Maximiser $P(Y)$ conduit à estimer (Q, X) par $(1, m_1)$ dès que $\lambda \sqrt{\sigma_0^2} > (1-\lambda) \sqrt{\sigma_1^2}$. Si σ_1^2/σ_0^2 est très petit, la probabilité $1-\lambda$ d'effectuer un mauvais classement peut être très grande.

Dans ce cas, comme nous l'avons vu plus haut, la fonctionnelle à maximiser que nous proposons pour estimer Y est donnée par:

$$\hat{Y} = \underset{x, q}{\text{Argmax}} \lambda^q (1-\lambda)^{1-q} \exp \left\{ -\frac{1}{2\sigma_q^2} (x-m_q)^2 \right\}$$

Il est équivalent de maximiser $P(Y)$ si les variances σ_1^2 et σ_0^2 sont égales.

2.2 DECONVOLUTION D'UN MELANGE DE DEUX GAUSSIENNES

2.2.1 MODELE

Nous rappelons brièvement le modèle utilisé:

$$\begin{cases} Z(i) = \sum_{k=p}^p f_k X(i-k) + \varepsilon(i) \\ Y(i) = (Q(i), X(i)) \end{cases}$$

$Z = \{Z(k)\}_{1 \leq k \leq n}$ sont les observations, $Y = \{Y(k)\}_{1 \leq k \leq n}$ une suite de variables aléatoires indépendantes et ε un bruit blanc Gaussien de variance σ_ε^2 .

Q est une suite de variables de Bernoulli de paramètre λ et X un mélange de 2 Gaussiennes:

- Conditionnellement à $Q = 0$, X suit une loi $\mathcal{N}(0, \sigma_0^2)$
- Conditionnellement à $Q = 1$, X suit une loi $\mathcal{N}(0, \sigma_1^2)$

(On suppose $\sigma_1^2 > \sigma_0^2$)

Notre but est ici de reconstruire la suite $Y = \{Y(k)\}_{1 \leq k \leq n}$ lorsque tous les paramètres du modèle sont connus.

(On peut remarquer qu'en l'absence de filtre, on est ramené pour chaque $Y(k)$ au problème précédent, dans le cas d'un mélange de deux Gaussiennes.)

2.2.2 ALGORITHMES BAYESIENS DE RECONSTRUCTION

Plusieurs techniques utilisées en traitement d'images utilisent la distribution à postériori de la suite Y pour sa reconstruction [31]:

2.2.2.a MPM (Marginal Probability Mode)

Le MPM consiste à prédire Y en maximisant chaque marginale $P(Y^{(i)} / Z)$.

La distribution à postériori de $Y^{(i)}$, $P(Y^{(i)}/Z)$ ne peut être calculée explicitement, on a alors recours à des techniques de Monte-Carlo pour l'estimer: on simule de façon indépendante plusieurs réalisations de la loi $P(Y/Z)$ au moyen d'un échantillonneur de Gibbs (voir l'annexe du Chapitre 1) puis on estime empiriquement chaque distribution $P(Y^{(i)} / Z)$.

Pour tout i et conditionnellement à Z , $Q^{(i)}$ est une variable de Bernoulli de paramètre $\tilde{\lambda}_i$, $X^{(i)}$ la combinaison d'une $\mathcal{N}(\tilde{m}_{0,i}, \tilde{V}_{0,i})$ et d'une $\mathcal{N}(\tilde{m}_{1,i}, \tilde{V}_{1,i})$. Notons $\hat{\lambda}_i$, $\hat{m}_{0,i}$, $\hat{V}_{0,i}$, $\hat{m}_{1,i}$ et $\hat{V}_{1,i}$ les estimateurs empiriques des différents paramètres.

Chacune des distributions $P(Y^{(i)}/Z)$ est estimée indépendamment des autres, et chaque $Y^{(i)}$ sera estimé indépendamment des autres. On est donc ramené au problème décrit dans le paragraphe 2.1.3, concernant le mélange de deux populations. Comme nous l'avons vu, plutôt que de maximiser $P(Y^{(i)}/Z)$, nous allons nous donner une fonction de coût à minimiser:

On définit $J_D = t N_1 + (1 - t) N_0$, N_1 représente le nombre de fois où on décide $Q=1$ alors que $Q=0$ et N_0 le nombre de fois où on décide $Q=0$ quand $Q=1$. On estime la suite $\{Q^{(i)}\}$ en appliquant la règle suivante de décision:

$$\hat{Q}^{(i)} = \mathbf{1}_{\hat{\lambda}_i \geq t}$$

La suite $\{X^{(i)}\}$ est estimée ensuite en minimisant l'erreur quadratique de prédiction:

$$\hat{X}^{(i)} = \hat{m}_{0,i} \mathbf{1}_{\hat{Q}^{(i)}=0} + \hat{m}_{1,i} \mathbf{1}_{\hat{Q}^{(i)}=1}$$

2.2.2.b MAP (Maximum Posterior Probability)

L'estimateur du MAP consiste à maximiser la distribution à postériori $P(Y/Z)$.

En notant $K = \text{Log} \left\{ \frac{\sigma_0^2}{\sigma_1^2} \left(\frac{\lambda}{1-\lambda} \right)^2 \right\}$, la distribution à postériori de Y est donnée par:

$$P(\mathbf{Y}=\mathbf{y} / \mathbf{Z}) = Cte \times \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left[\frac{(Z^{(i)} - f \cdot X^{(i)})^2}{\sigma_\varepsilon^2} + X^{(i)2} \left(\frac{(1-q^{(i)})}{\sigma_0^2} + \frac{q^{(i)}}{\sigma_1^2} \right) - Kq^{(i)} \right] \right\}$$

Supposons que $\tilde{\mathbf{y}} = (\tilde{\mathbf{q}}, \tilde{\mathbf{x}})$ maximise $p(\mathbf{Y}/\mathbf{Z})$ et soit $\tilde{y}^{(i)}$ un élément de cette suite, nous avons la relation:

$$P(\mathbf{Y}=\tilde{\mathbf{y}}/\mathbf{Z}) = P(Y^{(i)}=\tilde{y}^{(i)} / \mathbf{Z}, Y^{(j)}=\tilde{y}^{(j)}, j \neq i) P(Y^{(j)}=\tilde{y}^{(j)} j \neq i / \mathbf{Z}),$$

pour une suite $\{\tilde{y}^{(j)}\}_{j \neq i}$ donnée, $P(Y^{(i)} / \mathbf{Z}, Y^{(j)}=\tilde{y}^{(j)})$ est donc maximum en $Y^{(i)}=\tilde{y}^{(i)}$.

Conditionnellement à \mathbf{Z} et aux $Y^{(j)}, j \neq i$, $Q^{(i)}$ est une variable de Bernoulli de paramètre λ_i , $X^{(i)}$ la combinaison d'une $\mathcal{N}(m_{0,i}, V_0)$ et d'une $\mathcal{N}(m_{1,i}, V_1)$ (Les différents paramètres sont calculés en annexe du Chapitre 1).

Si $Q^{(i)} = 0$, $P(Y^{(i)} / \mathbf{Z}, Y^{(j)}=\tilde{y}^{(j)})$ atteint son maximum en $X^{(i)} = m_{0,i}$ et vaut $\frac{1-\lambda_i}{\sqrt{2\pi V_0}}$

Si $Q^{(i)} = 1$, $P(Y^{(i)} / \mathbf{Z}, Y^{(j)}=\tilde{y}^{(j)})$ atteint son maximum en $X^{(i)} = m_{1,i}$ et vaut $\frac{\lambda_i}{\sqrt{2\pi V_1}}$

D'où:

$$\tilde{y}^{(i)} = \begin{cases} (0, m_{0,i}) & \text{si } (1-\lambda_i) \sqrt{V_1} > \lambda_i \sqrt{V_0} \\ (1, m_{1,i}) & \text{sinon} \end{cases}$$

Nous avons la relation (Annexe Chap.1):

$$\lambda_i = \left[1 + \frac{1-\lambda}{\lambda} \sqrt{\frac{V_0 \sigma_1^2}{V_1 \sigma_0^2}} \exp \left\{ \frac{1}{2} \left(\frac{m_{0,i}^2}{V_0} - \frac{m_{1,i}^2}{V_1} \right) \right\} \right]^{-1}$$

En utilisant les définitions de $m_{1,i}$ et $m_{0,i}$ on déduit:

$$\tilde{q}^{(i)} = 1 \text{ si } m_{1,i}^2 \geq \frac{V_1^2}{V_1 - V_0} \text{Log} \left[\frac{\sigma_1^2}{\sigma_0^2} \left(\frac{1-\lambda}{\lambda} \right)^2 \right]$$

On retrouve bien le résultat que l'on attendait (obtenu en 2.1.3): la probabilité que $Q=1$ croît avec λ et σ_0^2 . D'autre part, la vraisemblance à postériori est maximisée lorsque la règle suivante est vérifiée:

$Q(i) = 1$ si et seulement si $m_{1,i}$, la prédiction de $X(i)$ lorsque $Q(i)=1$, est supérieure en valeur absolue à un seuil X_{\min} défini ci-dessus.

Comme nous l'avons vu dans le paragraphe 2.1.3, la probabilité d'erreur peut être grande si $V_0 \ll V_1$, on décidera en effet $Q \equiv 0$.

Si on veut donner le même poids aux deux types d'erreurs de classification, on préférera décider $Q(i)=1$ si $\lambda_i > 1 - \lambda_i$, cela revient alors à utiliser la règle de décision suivante:

$$\hat{Q}(i) = 1 \text{ si } m_{1,i}^2 \geq \frac{V_1^2}{V_1 - V_0} \text{Log} \left[\frac{V_0 \sigma_1^2}{V_1 \sigma_0^2} \left(\frac{1 - \lambda}{\lambda} \right)^2 \right]$$

On peut remarquer qu'on a seulement déplacé le seuil X_{\min} , la prédiction de $X(i)$ reste inchangée quand à elle.

La fonctionnelle à maximiser dans ce cas est la vraisemblance $P'(Y/Z)$ calculée sous une loi P' telle que P' et P ne diffèrent que sur la distribution de Q : λ' le paramètre de Bernoulli associé à la loi P' est défini par:

$$\lambda' = \frac{\lambda \sqrt{V_1}}{\lambda \sqrt{V_1} + (1 - \lambda) \sqrt{V_0}}$$

Finalement, on fera comme si Y suivait la "fausse loi" P' : on cherche alors à maximiser sa distribution à postériori.

Des techniques de recuit simulé permettent d'optimiser la maximisation de cette fonctionnelle et de réduire les temps de calcul [19], toutefois, ceux-ci restent importants si le nombre de données est grand. Cette procédure est décrite en annexe du chapitre 3.

2.2.2.c ICM (Iterative Conditional Mode)

L'ICM est un algorithme rapide, qui procède par itérations et aboutit à un maxima local de la vraisemblance $P(Y/Z)$:

A l'étape k , on change Y en un seul site i , choisi aléatoirement ou cycliquement, tel que $Y(i)$ maximise la distribution conditionnelle $P(Y(i)/Z, Y(j)_{j \neq i})$.

Cette distribution conditionnelle a été étudiée ci-dessus:

$$P(Q(i) = 1 / Z, Y(j)_{j \neq i}) = 1 - P(Q(i) = 0 / Z, Y(j)_{j \neq i}) = \lambda_i$$

$$P(X^{(i)} = x / Z , Y^{(j)}_{j \neq i} , Q^{(i)} = q) = \frac{1}{\sqrt{2\pi V_q}} \exp \left\{ -\frac{1}{2V_q} (x - m_{q,i})^2 \right\}$$

Si à l'étape k , on veut estimer $Y^{(i)}$ par $\hat{y}^{(i)} = (\hat{q}^{(i)}, \hat{x}^{(i)})$ en utilisant la règle suivante:

$$\hat{q}^{(i)} = \underset{q}{\text{Argmax}} P(Q^{(i)} = q / Z , Y^{(j)}_{j \neq i})$$

$$\hat{x}^{(i)} = \underset{x}{\text{Argmax}} P(X^{(i)} = x / Z , Y^{(j)}_{j \neq i} , Q^{(i)} = q),$$

on se placera dans le "faux" modèle défini ci-dessus pour le MAP et pour lequel un nouveau paramètre de Bernoulli λ' est défini.

2.3 DECONVOLUTION DE SIGNAUX BERNOULLI-GAUSSIENS

2.3.1 MODELE BERNOULLI GAUSSIEN

On rencontre dans la littérature différentes approches pour aborder le problème de la déconvolution dans le cas particulier de signaux Bernoulli-Gaussiens: la vraisemblance à postériori est maximisée dans [18], une déconvolution récursive utilisant un filtre de Kalman est proposée dans [12], [13], [14] ou [21]. Nous nous proposons ici d'adapter des techniques Bayésiennes utilisées en traitement d'image, mais surtout, comme nous l'avons vu plus haut, de présenter des techniques qui permettent d'optimiser un ensemble de critères.

$$\text{Revenons au modèle de convolution initial: } \begin{cases} Z^{(i)} = \sum_{k=p}^p f_k X^{(i-k)} + \varepsilon(i) \\ Y^{(i)} = (Q^{(i)}, X^{(i)}) \end{cases}$$

Q est une variable aléatoire de Bernoulli de paramètre λ . Plaçons-nous maintenant dans le cas particulier où la distribution de X est la combinaison d'une δ -Dirac en 0 et d'une distribution normale centrée de variance σ^2 :

- Si $Q=0$, $X=0$
- Si $Q=1$, $X \sim \mathcal{N}(0, \sigma^2)$

La δ -Dirac en 0 est définie ici comme la distribution limite d'une $\mathcal{N}(0, \sigma_0^2)$ lorsque σ_0^2 tend vers 0. Les résultats obtenus plus haut peuvent alors être utilisés en faisant tendre σ_0^2 vers 0.

De manière générale, nous dirons que $Y = (Q, X)$ suit une loi B-G(λ, μ, σ^2) si:

$$Q \sim B(\lambda)$$

$$X \sim \mathcal{N}(\mu, \sigma^2) \mathbf{1}_{Q=1} + \delta_0 \mathbf{1}_{Q=0}$$

Nous avons alors:

* $P(Y(i))$, la distribution à priori de $Y(i)$, est une B-G($\lambda, 0, \sigma^2$)

* $P(Y(i)/Z, Y(j) \text{ } j \neq i)$, la distribution conditionnelle de $Y(i)$, est une B-G(λ_i, m_i, V), les paramètres λ_i, m_i et V ont été calculés en annexe du chapitre 1:

$$V = \left(\frac{E_f}{\sigma_\varepsilon^2} + \frac{1}{\sigma^2} \right)^{-1}, \quad m_i = \frac{E_f}{\sigma_\varepsilon^2} V \mu_i$$

$$\lambda_i = \left[1 + \frac{1-\lambda}{\lambda} \sqrt{\frac{\sigma^2}{V}} \exp\left\{ -\frac{m_i^2}{2V} \right\} \right]^{-1}$$

$$\text{Où } E_f = \sum_{j=-p}^{+p} f_j^2 \quad \text{et} \quad \mu_i = \frac{1}{E_f} \sum_{k=-p}^{+p} f_k [Z_{i+k} - \sum_{j \neq k} f_j x_{i+k-j}].$$

* $P(Y(i)/Z)$, la distribution à postérieure de $Y(i)$ est une B-G($\tilde{\lambda}_i, \tilde{m}_i, \tilde{\sigma}_i^2$), les paramètres $\tilde{\lambda}_i, \tilde{m}_i, \tilde{\sigma}_i^2$ sont estimés par Monte-Carlo.

Nous allons maintenant revenir sur les algorithmes Bayésiens décrits en 2.1:

2.3.2 IMPLEMENTATION DES ALGORITHMES BAYESIENS DE DECONVOLUTION

2.3.2.a MPM

Comme nous l'avons dit, $P(Y(i)/Z)$, la distribution à postérieure de $Y(i)$, est une B-G($\tilde{\lambda}_i, \tilde{m}_i, \tilde{\sigma}_i^2$), les paramètres $\tilde{\lambda}_i, \tilde{m}_i, \tilde{\sigma}_i^2$ sont estimés empiriquement en simulant plusieurs réalisations de la loi $P(Y/Z)$.

On applique ici exactement la même règle qu'en 2.2.2.a pour estimer $Y(i)$ si on cherche à minimiser la fonction de coût $J_D = \alpha N_1 + (1 - \alpha) N_0$:

$$\hat{q}^{(i)} = \mathbf{1}_{\hat{\lambda}_i \geq \alpha}$$

$$\hat{x}^{(i)} = \hat{m}_i \mathbf{1}_{\hat{q}^{(i)}=1}$$

2.3.2.b MAP

Il faut tout d'abord remarquer que la vraisemblance d'une δ -Dirac n'est pas définie: dans le cas d'une $\mathcal{N}(0, \sigma_0^2)$, la vraisemblance en 0 vaut $(2\pi\sigma_0^2)^{-\frac{1}{2}}$ et lorsque la variance σ_0^2 tend vers 0, la vraisemblance tend vers l'infini.

On rencontre dans la littérature diverses conventions pour définir la vraisemblance d'une δ -Dirac (en la considérant égale à 1, par exemple). Nous allons voir que cette convention est liée à un critère très simple: on décide que $Q=1$ (c'est-à-dire $X \neq 0$) si et seulement si la valeur prédite de X est supérieure en valeur absolue à un certain seuil.

Par convention, nous allons poser ici que la vraisemblance d'une δ -Dirac en 0 est celle d'une $\mathcal{N}(0, \sigma_0^2)$ pour un certain σ_0^2 donné. Cette vraisemblance atteint son maximum en 0 où elle vaut $(2\pi\sigma_0^2)^{-\frac{1}{2}}$.

On peut remarquer que ce modèle fabriqué artificiellement est équivalent au modèle suivant:

$$Z(i) = \sum_{k=p}^p f_k Q(i-k)X(i-k) + \varepsilon(i),$$

Si $Q = 0$, X suit une loi $\mathcal{N}(0, \sigma_0^2)$ et si $Q = 1$, X suit une loi $\mathcal{N}(0, \sigma^2)$.

Supposons que $\tilde{y} = (\tilde{q}, \tilde{x})$ maximise $p(Y/Z)$, la vraisemblance à postériori de Y . Soit $\tilde{y}(i)$ un élément de cette suite, $\tilde{y}(i)$ doit vérifier:

$$\tilde{y}(i) = \begin{cases} (0, 0) & \text{si } |m_i| \leq X_{\min} \\ (1, m_i) & \text{si } |m_i| > X_{\min} \end{cases}$$

$$\text{où } X_{\min}^2 = V \text{Log} \left[\frac{\sigma^2}{\sigma_0^2} \left(\frac{1-\lambda}{\lambda} \right)^2 \right]$$

Définir conventionnellement la vraisemblance d'une δ -Dirac, en se donnant σ_0^2 revient donc à se fixer un seuil minimum X_{\min} pour X non nul et donc la proportion d'évènements $Q=1$.

Le critère de reconstruction utilisé ici est donc: trouvons la séquence de signaux Bernoulli-Gaussiens la plus vraisemblable pour produire la série Z observée et telle que ses éléments non nuls aient des valeurs dans $\mathbb{R} - [-X_{\min}, X_{\min}]$, où le seuil X_{\min} est choisi.

2.3.2.c ICM

A une itération quelconque, on calcule $\hat{y}^{(i)} = (\hat{q}^{(i)}, \hat{x}^{(i)})$ la nouvelle prédiction de $Y^{(i)}$ telle que $P(Y^{(i)}/Z, Y^{(j)}, j \neq i)$ soit maximisé. Pour cela, on calcule m_i la prédiction de $X^{(i)}$ lorsque $Q^{(i)} = 1$ et on applique la règle suivante de décision:

- si $|m_i| \geq X_{\min}$, $\hat{y}^{(i)} = (1, m_i)$
- si $|m_i| < X_{\min}$, $\hat{y}^{(i)} = (0, 0)$.

Lorsque cet algorithme est utilisé sous cette forme, on se cale très vite sur une mauvaise solution, du fait de la structure fortement multimodale de la fonction de vraisemblance à postériori de Y . Nous proposons alors une variante de cet algorithme qui consiste à faire varier le seuil X_{\min} au cours des itérations.

Donnons-nous deux seuils δ_{\min} et δ_{\max} tels que les éléments non nuls de la suite $\{\hat{x}^{(i)}\}_{1 \leq i \leq n}$ que l'on cherche comme solution, appartiennent à $\mathbb{R} -]-\delta_{\min}, \delta_{\min}[$ et tels que $\delta_{\max} \geq \delta_{\min}$.

Le seuil $X_{\min}(k)$ varie de la manière suivante au cours des itérations:

- $X_{\min}(0) = 0$: on accepte tous les évènements.
- $X_{\min}(k)$ croît jusqu'à δ_{\max} : on ne garde que les plus importants.
- $X_{\min}(k)$ décroît jusqu'à δ_{\min} : on en récupère une partie.

Le choix optimal de δ_{\max} et de la fonction $X_{\min}(k)$ reste un problème ouvert. Dans les exemples proposés plus loin, $X_{\min}(k)$ est une fonction linéaire, plusieurs solutions correspondant à différentes valeurs de δ_{\max} sont proposées.

2.4 SIMULATIONS

Nous avons simulé une série de 600 variables Bernoulli-Gaussiennes telles que $\lambda = 0.1$ et $\sigma^2 = 1$, le filtre de convolution est celui présenté dans le Chapitre 1 (Fig. 6) et le rapport signal/bruit est de 10dB. Nous présentons plusieurs reconstructions, issues d'algorithmes différents.

Dans les figures 1-3, sont représentées: $\left\{ \begin{array}{l} | : \text{série originale de Bernoulli-Gaussiennes.} \\ * : \text{série estimée.} \end{array} \right.$

1) MPM

Nous avons simulé 100 réalisations de la loi $P(Y/Z)$ et choisi comme seuil de détection $\alpha = 40\%$.

Comme nous l'avons vu dans le chapitre précédent, différentes suites Y peuvent produire des convolutions très proches, les paramètres $\tilde{\lambda}_i$, \tilde{m}_i et $\tilde{\sigma}_i^2$ ne sont pas uniques. Si on réalise N simulations indépendantes (en réinitialisant l'échantillonneur de Gibbs entre chaque simulation), les réalisations obtenues peuvent être très différentes et ne permettent donc pas une bonne estimation des paramètres, on a tendance à détecter les signaux en paquets (Figure 1-a).

Au contraire, si ces réalisations sont dépendantes (en ne réinitialisant pas l'échantillonneur de Gibbs), l'estimation des paramètres a un sens puisqu'on reste dans un voisinage d'une même solution (Fig. 1-b).

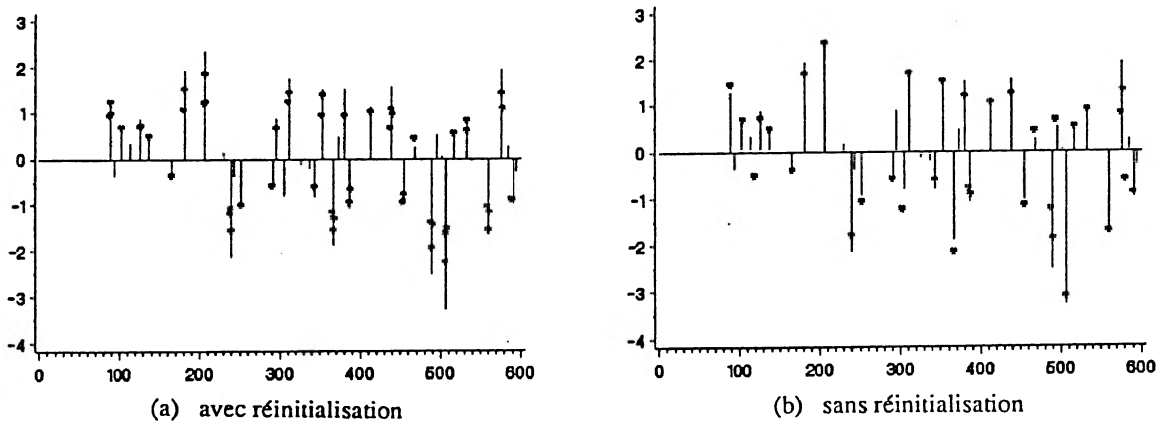
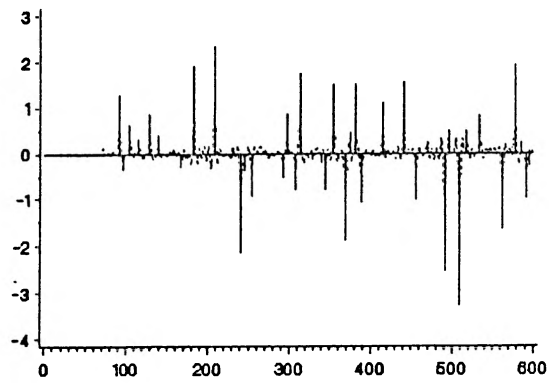
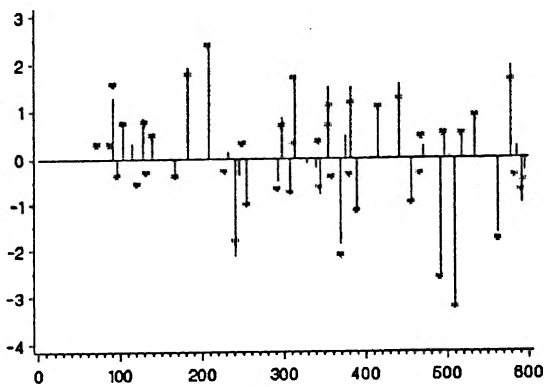
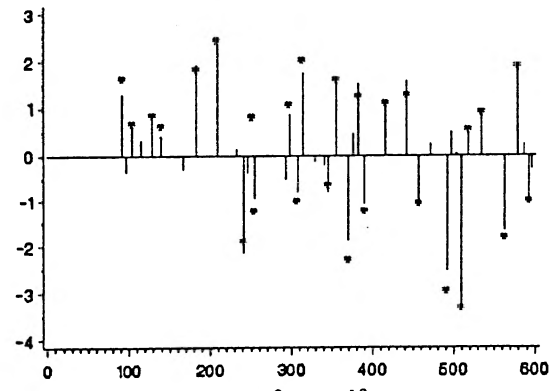


Fig.1 MPM

2) MAP

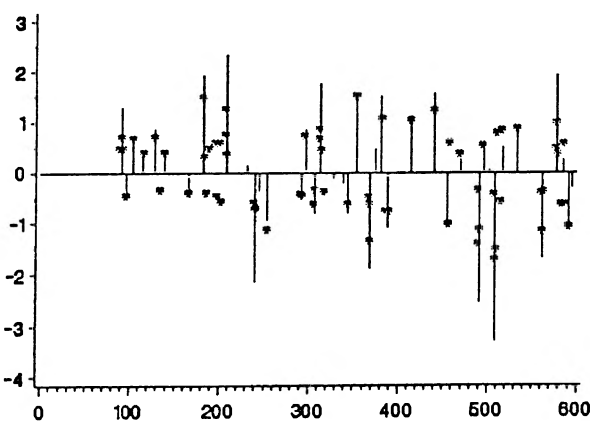
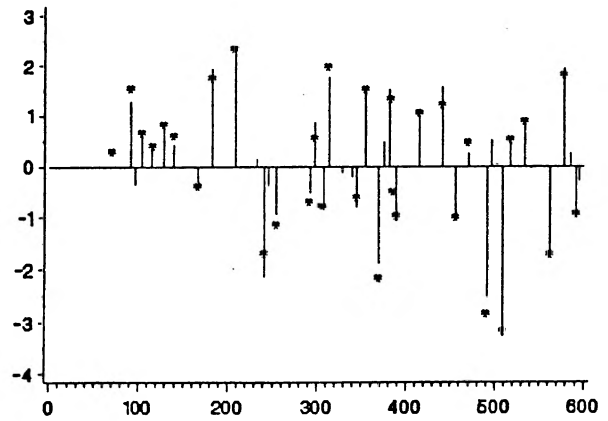
Nous avons choisi différentes valeurs de σ_0^2 correspondant à des seuils X_{\min} différents:

En Figure 2-a, on a choisi $X_{\min} = 0$: on décide toujours $Q=1$ quelque-soit la valeur de \hat{X} . En Figure 2-b et 2-c, on a fixé respectivement $\sigma_0^2 = 1$ et $\sigma_0^2 = 10^{-10}$.

(a) $X_{\min} = 0$ (b) $\sigma_0^2 = 1$ (c) $\sigma_0^2 = 10^{-10}$ **Fig.2** MAP

3) ICM

Nous avons utilisé ici différents couples $(\delta_{\min}, \delta_{\max})$ en effectuant chaque fois 100 itérations: un algorithme ICM traditionnel est utilisé en Figure 3-a, puisque $\delta_{\min} = \delta_{\max} = 0.3$. Nous avons ensuite utilisé un seuil variable tel que $\delta_{\min} = 0.3$ et $\delta_{\max} = 1.5$ (Figure 3-b).

(a) $\delta_{\max} = \delta_{\min} = 0.3$ (b) $\delta_{\max} = 1.5, \delta_{\min} = 0.3$ **Fig.3** ICM

En comparant les divers résultats obtenus, l'ICM à seuil variable semble l'algorithme le plus approprié pour détecter ce type de signaux tout en étant très rapide en comparaison avec le MAP ou le MPM.

La non-unicité des solutions, due au filtre utilisé, rend l'utilisation des autres algorithmes très délicate, en effet, le terme $P(Z/Y)$ n'a pas un maxima global unique, $P(Y/Z)$ est fortement multimodale et seul le terme $P(Y)$ permet d'en différencier ses divers mode. La maximisation de $P(Y/Z)$ est alors rendue particulièrement difficile et requiert un temps de calcul très important, en effet, si la décroissance de la température est trop rapide, on risque de converger vers un maxima local de la vraisemblance, donc une mauvaise configuration.

CHAPITRE 3

DECONVOLUTION 2-D DE TRACES SISMIQUES

INTRODUCTION

L'interprétation des traces sismiques est un problème abondamment traité en géophysique. Notamment, de nombreuses techniques ont été développées pour retrouver la série des coefficients de réflexion associés aux changements de lithologie, utilisant le sismogramme comme seule observation [1], [14], [17], [20], [21], [22], [23], [24], [27], [28].

Le modèle mathématique généralement admis pour modéliser une trace sismique consiste à voir celle-ci comme une version filtrée de la série des réflecteurs, c'est-à-dire comme la convolution de cette séquence avec l'onde émise à la surface. Un bruit de mesure est également présent.

Nous reprendrons ici ce modèle et présentons un algorithme de déconvolution dont la tâche consiste à estimer les coefficients de réflexion. Les paramètres identifiant le modèle sont ceux liés à la distribution des réflecteurs et à celle du bruit ainsi que l'onde (filtre de convolution). Ils sont estimés dans un premier temps en utilisant l'algorithme stochastique d'estimation présenté dans le Chapitre 1.

La technique de déconvolution présentée ici diffère essentiellement en deux aspects des autres techniques utilisées en géophysique:

- * Dans la plupart des travaux traitant ce sujet, chaque trace sismique est déconvoluée séparément. Aucune continuité latérale n'est alors demandée et d'éventuelles incohérences dans les résultats sont possibles.

Nous n'allons pas utiliser le sismogramme comme seule information, mais également cette continuité latérale des réflecteurs que nous recherchons. Cela veut dire que toutes les traces de la section

sismique seront déconvoluées simultanément. La prédiction en un point donné dépendra de celle effectuée dans un voisinage bidimensionnel de ce point.

Toutes les techniques appliquées ici sont largement inspirées de celles développées en traitement d'images. En effet, une section sismique peut être vue comme une image bidimensionnelle que l'on cherche à restaurer: l'image originale, que l'on veut retrouver, est constituée des réflecteurs, cette image a été dégradée par l'effet d'un filtre de convolution, puis bruitée. Nous sommes alors dans le cadre typique utilisé en traitement d'images par Geman et Geman dans [11].

* Les méthodes de déconvolution qui consistent à maximiser (ou minimiser) une fonctionnelle donnée, permettent d'obtenir des solutions optimisant un critère de nature mathématique. Ainsi, Mendel propose dans [14] et [22] une déconvolution par maximum de vraisemblance (MLD). La déconvolution par minimum d'entropie (MED) vise, en géophysique, à maximiser le kurtosis [28]. Nous proposons ici une approche différente: nous allons nous donner des critères de bonne reconstitution et voir ensuite quels outils mathématiques utiliser pour obtenir un résultat optimisant ces critères.

On demandera aux réflecteurs estimés de vérifier certaines propriétés, parmi les propriétés suivantes:

- (1) Présenter un bon ajustement avec les données observées du sismogramme.
- (2) Montrer une continuité latérale.
- (3) Faire apparaître essentiellement les réflecteurs les plus importants.

Pour une combinaison de critères donnée, la solution cherchée maximise une vraisemblance à postériori calculée sous une "fausse" loi en accord avec ces critères.

Les exemples proposés proviennent de données réelles fournies par l'INTEVEP (Vénézuéla). En effet, l'idée de ce travail est de présenter un algorithme qui donne de bons résultats avec des données réelles, données pour lesquelles le modèle mathématique proposé reste une modélisation, une approximation que l'on espère la plus proche possible de la réalité. Nous verrons que les résultats obtenus sont bien en accord avec les critères que l'on s'est imposé, l'algorithme semble robuste aux erreurs de modélisation commises.

Dans la 1^{ère} partie, la modélisation des traces et des réflecteurs est décrite. Pour ceux-ci, nous reprenons en partie le modèle utilisé par Mendel [22] qui consiste à séparer les réflecteurs en deux groupes: des réflecteurs principaux indiquant les séparations entre les différentes couches et des

réflecteurs secondaires qui traduisent les irrégularités existantes à l'intérieur des couches, ils sont de faible amplitude et seront considérés comme du bruit.

Les réflecteurs principaux forment des lignes, non nécessairement horizontales, leur inclinaison dépend de l'agencement des couches du sous-sol. Nous définirons donc une variable annexe qui mesurera la pente du sous-sol en chaque point.

La 2^{ème} partie est consacrée à l'estimation des différents paramètres du modèle. L'algorithme d'estimation présenté dans le Chapitre 1 est particulièrement bien adapté à ce problème puisque nous sommes dans un cas typique de données incomplètes: on ne dispose que des traces sismiques, la série des coefficients de réflexion n'étant pas observée.

Les paramètres estimés varient avec la profondeur: plus on s'éloigne de la surface, et plus l'onde s'amortit et perd des hautes fréquences alors que le niveau de bruit augmente. D'autre part, les caractéristiques du sous-sol peuvent varier d'un endroit à un autre d'une même section: des zones sont extrêmement accidentées et d'autres très régulières. Néanmoins, nous ferons l'approximation suivante: à l'intérieur d'une fenêtre de taille raisonnable, une trace sismique est un processus stationnaire et par conséquent, les paramètres sont considérés comme constants.

Dans la 3^{ème} partie, l'algorithme de déconvolution est présenté. Dans un premier temps, un système de voisinage est défini afin de vérifier la propriété suivante: l'estimation en un site d'un réflecteur dépend seulement de ses voisins et non de tous les autres réflecteurs.

Les caractéristiques locales de la distribution des réflecteurs ainsi que des critères de bonne reconstruction nous permettent ensuite de construire une vraisemblance telle que la solution proposée la maximise. Nous allons donc pouvoir:

- Discriminer les réflecteurs en deux populations (principaux et secondaires).
- Estimer les valeurs des coefficients de réflexion en chaque site.

Dans la 4^{ème} partie nous verrons une application de cet algorithme sur une section sismique. Nous présentons des solutions correspondant à différents critères de reconstruction. L'impédance acoustique peut être calculée à partir de la série des coefficients de réflexion, une représentation graphique de cette variable permet de mettre en évidence les couches du sous-sol.

3.1 MODELISATION

3.1.1. MODELE DE CONVOLUTION

3.1.1.a Modélisation d'une trace sismique

La modélisation mathématique d'une trace sismique que nous présentons plus loin peut se justifier de la manière suivante:

Le sous-sol est constitué de différentes couches superposées. A chaque changement de lithologie est lié un coefficient de réflexion dont la valeur dépend des impédances des deux milieux. L'impédance est une fonction de la densité ρ du milieu et de la vitesse de propagation V d'une onde dans ce milieu, cette grandeur physique caractérise donc le type de lithologie. La connaissance du sous-sol peut alors être directement résumée de celle des coefficients de réflexion.

Les données dont nous disposons sont discrètes: un intervalle de temps constant sépare chaque mesure et les traces sont espacées d'une distance constante. Par conséquent, nous considérerons le sous-sol comme discrétisé.

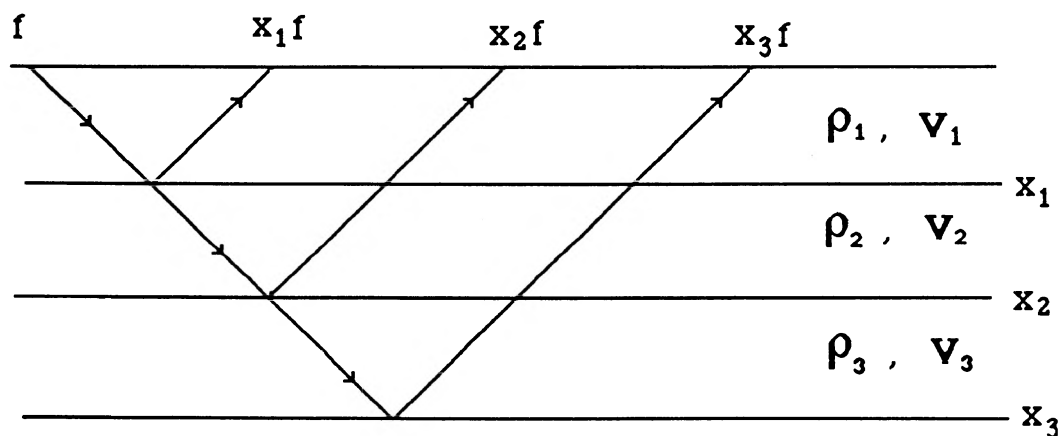


Fig.1 Propagation d'une onde dans le sous-sol.

Une onde est émise à la surface (vibration ou explosion) et se transmet dans le sous-sol (Figure 1). A chaque interface, une partie du signal est réfléchi tandis qu'une autre est transmise. La trace sismique Z est l'enregistrement à la surface de tout ce qui a été réfléchi.

Soit τ l'intervalle de temps entre deux mesures et δ , la distance parcourue par l'onde pendant le temps $\tau/2$. ($\tau = 2 \cdot 10^{-3} \text{ s}$, $\delta \approx 1 \text{ m}$)

En première approximation, nous pouvons dire que ce qui est mesuré à la surface au temps $k\tau$ est la somme de plusieurs signaux réfléchis et d'un bruit de mesure: la partie du signal émis au temps $(k-i)\tau$ qui se réfléchit à une profondeur $i\delta$ arrive à la surface au temps $k\tau$ pour $1 \leq i \leq k-1$.

Soit $X(k\delta)$, la valeur du coefficient de réflexion à la profondeur $k\delta$, $f(k\tau)$ l'amplitude du signal émis au temps $k\tau$, $Z(k\tau)$ l'amplitude du signal mesuré au temps $k\tau$ et ϵ un bruit de mesure,

$$Z(k\tau) = f(\tau) X(k\delta) + f(2\tau) X((k-1)\delta) + \dots + f((k-1)\tau) X(\delta) + \epsilon(k\tau).$$

Les valeurs de X étant petites devant 1, on a négligé les réflexions d'ordre supérieur (qui seraient de l'ordre de X^3) et considéré les coefficients de transmission égaux à 1, c'est-à-dire que le signal n'est pas altéré lors d'un changement de couche.

Notons $Z(k) = Z(k\tau)$, $X(k) = X(k\delta)$, $\epsilon(k) = \epsilon(k\tau)$ et $f_k = f(k\tau)$. On obtient alors l'équation de convolution:

$$Z(k) = \sum_{i=1}^{k-1} f_i X(k-i) + \epsilon(k)$$

Dans la suite, nous considérerons le signal émis f comme le filtre de convolution. Différents types de filtre peuvent être utilisés en géophysique, dans notre cas, l'onde émise a la forme suivante:

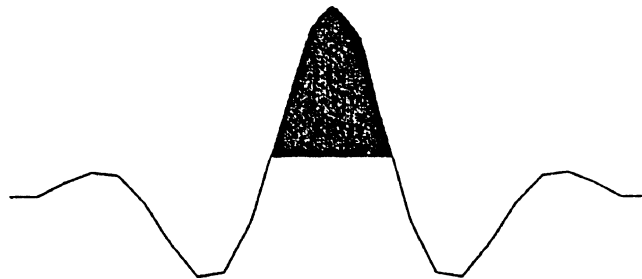


Fig.2 Filtre de convolution

Cette onde est celle utilisée théoriquement, mais du fait de perturbations et d'effets d'amortissement qui la déforment, nous devons la réestimer.

- Par convention, le filtre est translaté tel que $f_0 = \sup f_i$

- Presque toute l'énergie du signal est concentrée dans sa partie centrale, une approximation raisonnable consiste alors à tronquer le filtre tel que $f_k = 0$ pour $|k| > p$: nous utiliserons donc dans la suite un filtre de longueur $(2p+1)$, dans les exemples présentés, p a été choisi égal à 10.

3.1.1.b Modélisation d'une section sismique

L'ensemble des données à traiter est en fait constitué de plusieurs traces sismiques mesurées sur une ligne droite, chaque trace étant espacée d'environ 20m l'une de l'autre.

- Soit nt le nombre de traces sismiques contenues dans la section étudiée et n le nombre de données par trace. (Dans le cas présent, $nt=300$ et $n=1500$).
- Soit S l'ensemble des points de la section sismique (on dira de manière équivalente point ou site).

$$S = \{ (i, j) \quad i = 1 \dots n, \quad j = 1 \dots nt \}$$

On notera N le nombre total de sites, $N = |S| = nxnt$.

Nous reprenons ici le modèle traditionnel de convolution unidimensionnel:

$$Z = X * f + \varepsilon$$

$$Z_{i,j} = \sum_{k=p}^p f_k X_{i-k,j} + \varepsilon_{i,j}$$

$Z = \{Z_s\}_{s \in S}$ sont les valeurs observées du sismogramme, $X = \{X_s\}_{s \in S}$ les valeurs recherchées des coefficients de réflexion, $f = (f_{-p}, \dots, f_0, \dots, f_p)$ l'onde émise à la surface et $\varepsilon = \{\varepsilon_s\}_{s \in S}$ une série de bruits de mesure. On fera l'hypothèse que ε est un bruit blanc Gaussien de variance σ_ε^2 .

On dispose donc, comme observation, d'une juxtaposition de nt traces, chacune d'elles étant le résultat bruité de la convolution d'une série de réflecteurs avec un filtre.

3.1.2 MODELISATION DES REFLECTEURS

3.1.2.a Introduction

Il est clair que la modélisation des coefficients de réflexion doit être motivée par une bonne connaissance de l'aspect physique du problème, à savoir le type de géologie que l'on rencontre, l'agencement des différentes couches du sous-sol.

Pour justifier notre modélisation, nous ferons les hypothèses suivantes:

- * Le sous-sol est constitué d'une superposition de couches distinctes.

- * Ces couches ne sont pas nécessairement horizontales.
- * Ces couches sont continues et relativement uniformes, même si à l'intérieur de chacune d'elles, existent de petites variations.

Une telle modélisation implique, pour les coefficients de réflexion, d'ajouter une information supplémentaire à leur simple valeur. Nous allons donc introduire des variables annexes pour caractériser totalement un réflecteur.

Nous noterons $\mathbf{Y} = \{Y_s\}_{s \in S}$ l'ensemble de ces variables: le réflecteur au site s sera noté Y_s , où Y_s est multidimensionnel. On suppose dans ce qui suit que le champ \mathbf{Y} est stationnaire.

Dans un premier temps, nous définirons l'ensemble de ces variables et étudierons les distributions marginales des réflecteurs. Ensuite, pour tenir compte des dépendances existantes dans le champ \mathbf{Y} , nous introduirons un système de voisinage et définirons les caractéristiques locales de la distribution de \mathbf{Y} .

3.1.2.b Distribution marginale des réflecteurs

Type de réflecteur

On peut classer les réflecteurs en deux classes:

- Les réflecteurs principaux marquent les séparations entre les différentes couches.
- Les réflecteurs secondaires indiquent de petites variations, de petites irrégularités à l'intérieur des couches.

On définit alors une variable Q qui détermine le type de réflecteur: pour tout $s \in S$, si $Q_s = 0$, Y_s est un réflecteur secondaire, si $Q_s = 1$, Y_s est un réflecteur principal.

Q_s est une variable aléatoire de Bernoulli, de paramètre λ :

$$P(Q_s = 1) = 1 - P(Q_s = 0) = \lambda$$

(λ indique la proportion de réflecteurs principaux)

Valeur du coefficient de réflexion

Chaque coefficient de réflexion prend une valeur dans \mathbb{R} . Soit X_s la valeur du réflecteur Y_s , on supposera X_s Gaussien:

$$\text{si } Q_s = 0 \quad , \quad X_s \sim \mathcal{N}(0, \sigma_0^2)$$

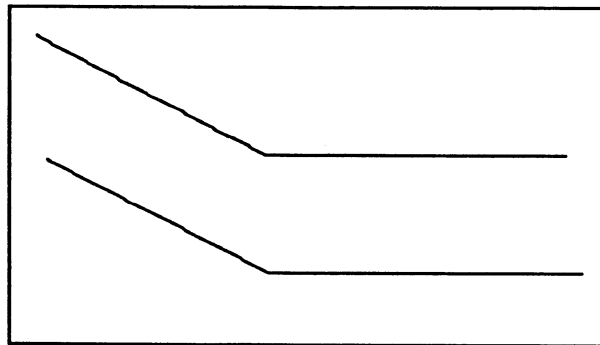
$$\text{si } Q_s = 1 \quad , \quad X_s \sim \mathcal{N}(0, \sigma_1^2)$$

Les valeurs des réflecteurs principaux étant plus grandes que celles des secondaires, on a $\sigma_1^2 > \sigma_0^2$ (la modélisation d'un coefficient de réflexion à valeur dans $[-1, 1]$ par une variable Gaussienne à valeur dans \mathbb{R} se justifie pour des valeurs très petites de σ_1^2 et σ_0^2).

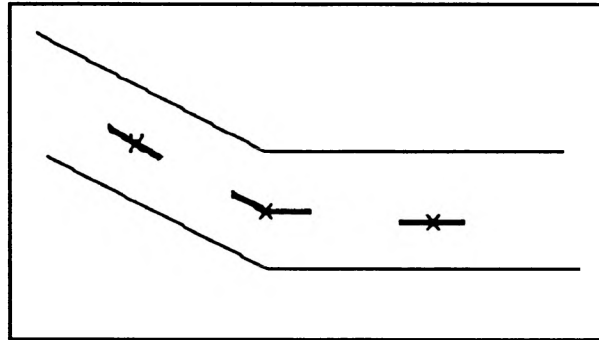
Cette modélisation des variables Q et X est celle utilisée par Mendel [14], [22]. Celui-ci étudie chaque trace sismique isolément, il suppose alors les réflecteurs indépendants. Dans le cas bidimensionnel qui nous intéresse, nous ne ferons pas cette hypothèse d'indépendance puisque la continuité latérale du sous-sol implique une dépendance entre réflecteurs voisins: nous supposerons le sous-sol constitué de lignes de réflecteurs indépendantes. Pour estimer Y en un point, l'information apportée par les réflecteurs voisins doit être d'une aide précieuse et il va nous être alors nécessaire de définir un voisinage en chaque site.

3.1.2.c Système de voisinage

Considérons une configuration simple du sous-sol:



Il est naturel, d'une part, de considérer les éléments d'une même ligne de réflecteurs comme voisins, c'est-à-dire que le système de voisinage dépendra de la configuration du sous-sol. D'autre part, les couches ne sont pas toujours horizontales, une notion de pente doit alors être introduite. Pour tenir compte d'éventuelles ruptures de pente, nous définirons conventionnellement en chaque site une pente à droite et une pente à gauche:



Définissons une variable C qui mesure cette pente, nous proposons le codage suivant:

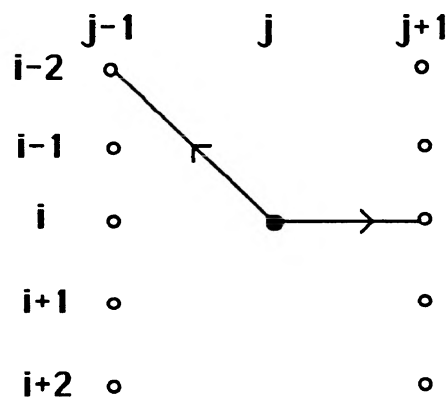
Soit $C_{i,j} = (CG_{i,j}, CD_{i,j})$ la variable C au point (i, j) : $CG_{i,j}$ indique la pente à gauche de (i, j) et $CD_{i,j}$ la pente à droite. Par définition, on pose $CG_{i,j}$ (ou $CD_{i,j}$) positif si la pente est montante de gauche à droite au point (i, j) , négatif si la pente est descendante, et nul si elle est horizontale. Du fait de la discrétisation du sous-sol, $CG_{i,j}$ et $CD_{i,j}$ sont à valeurs entières.

Soit CM la valeur de la plus grande pente possible, on a donc:

$$-CM \leq CG_{i,j} \leq CM \quad , \quad -CM \leq CD_{i,j} \leq CM.$$

Au point (i, j) sont associés deux voisins:

- Un voisin à gauche $VG_{i,j} = (i + CG_{i,j}, j-1)$
- Un voisin à droite $VD_{i,j} = (i - CD_{i,j}, j+1)$



$$C_{i,j} = (-2, 0)$$

$$VG_{i,j} = (i-2, j-1) \quad , \quad VD_{i,j} = (i, j+1)$$

Fig.3 Exemple de voisinage

La suite des C_{ij} définit un système de voisinage sur S , et un graphe orienté a été créé:

- Les sommets sont les éléments de S .
- Les arêtes sont les segments liant chaque point à ses deux voisins.

Le système de voisinage que nous désirons doit vérifier la propriété de symétrie suivante [11]:

$$s \text{ voisin de } t \iff t \text{ voisin de } s$$

Cette propriété n'étant pas toujours vérifiée sur le graphe tel qu'il a été créé, nous allons le symétriser de la façon suivante:

Une arête existe entre s et t si et seulement si la condition de symétrie définie ci-dessus est vérifiée. Cette symétrisation revient à supprimer une partie des arêtes du graphe initial. Après cette transformation, les sommets du nouveau graphe ont 0, 1 ou 2 voisins.

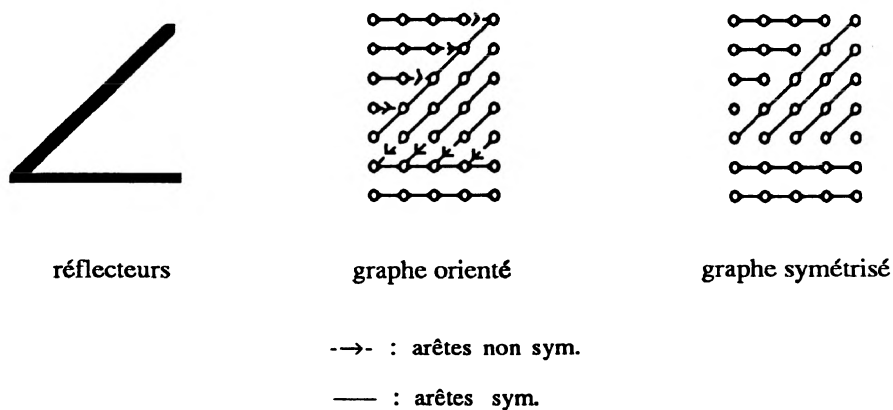


Fig.4 Exemple de symétrisation du graphe

Il est évident que certaines configurations ne sont pas acceptables: on ne peut imaginer, par exemple, que des lignes de réflecteurs se croisent. Nous verrons dans le paragraphe 3.3.1 consacré à l'estimation de C (donc à la construction du graphe) que de telles configurations ne pourront être retenues.

3.1.2.d Caractéristiques locales de la distribution des réflecteurs

Finalement, un réflecteur au site s est défini par trois variables X_s , Q_s et C_s :

- Le champ $C = \{C_s\}$ génère un système de voisinage $V = \{V_s\}$, où V_s est le voisinage du site s .
- Nous noterons désormais Y le champ $\{Y_s\}$ où $Y_s = (X_s, Q_s)$.

Comme nous l'avons déjà dit, les réflecteurs ne sont pas supposés indépendants. Alors que nous avons étudié plus haut les distributions marginales des Y_s , nous allons maintenant étudier les caractéristiques locales de la distribution de Y conditionnellement à V , c'est-à-dire quand le système de voisinage est connu.

Hypothèse de Markovianité

Nous ferons l'hypothèse de Markovianité suivante:

$$P(Y_s / Y_t, t \neq s, t \in S) = P(Y_s / Y_t, t \in V_s)$$

Cela signifie qu'il n'est pas nécessaire de connaître tout le reste du champ Y pour faire une prédiction au point s , mais seulement ses deux au plus voisins.

Distribution du champ Q

Le nombre de configurations possibles du champ Q est fini. Soit $P(Q = q)$ la probabilité d'avoir la configuration $q = \{q_s\}$, on pose:

$$P(Q = q) = \frac{1}{K(\alpha, \beta)} \exp \left\{ -\alpha \sum_{s \in S} q_s - \beta \sum_{\substack{(s,t) \in S \\ s,t \text{ vois.}}} 1_{q_s \neq q_t} \right\}$$

$K(\alpha, \beta)$ est une constante de normalisation telle que $P(Q)$ soit une mesure de probabilité.

Dans le cas présent, on suppose que α et β sont positifs.

- Plus α est grand, et plus la proportion de réflecteurs principaux diminue.
- Plus β est grand, et plus grande est la régularité du champ Q .

(On rappelle que $Q_s=1$ si le réflecteur Y_s est un réflecteur principal et que $Q_s=0$ sinon)

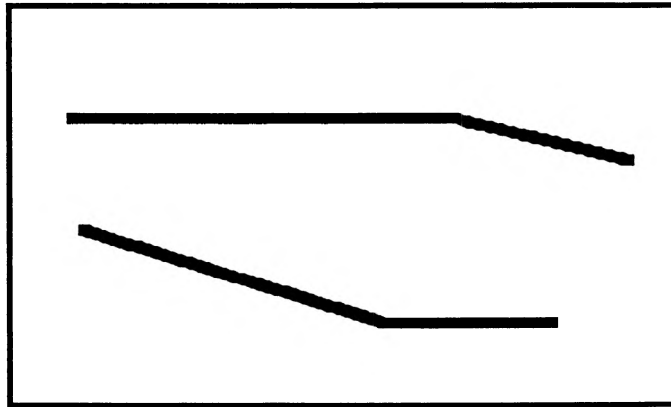
La loi conditionnelle de Q_s est alors donnée par:

$$P(Q_s = q_s / Q_t, t \neq s) = \frac{e^{-\alpha q_s - \beta \sum_{V_s} 1_{q_s \neq q_t}}}{e^{-\beta \sum_{V_s} q_t} + e^{-\alpha - \beta \sum_{V_s} (1 - q_t)}}$$

Cette loi conditionnelle varie donc suivant le nombre d'éléments de V_s .

Distribution conditionnelle du champ X

Une fois Q fixé, le graphe est constitué de lignes de réflecteurs principaux et de réflecteurs secondaires:



(on a marqué les lignes de réflecteurs principaux en traits pleins)

i) Les réflecteurs secondaires sont des variables aléatoires indépendantes, identiquement distribuées, Gaussiennes, centrées et de variance σ_0^2 .

ii) Les lignes de réflecteurs principaux sont indépendantes. Les réflecteurs principaux appartenant à une même ligne sont corrélés, nous noterons ρ la corrélation entre deux réflecteurs principaux voisins (ρ est positif et supposé constant).

Considérons une ligne de réflecteurs principaux de longueur m et notons $X = (X_1, X_2, \dots, X_m)$ les valeurs des coefficients de réflexion de cette ligne, le problème devient alors unidimensionnel:

Soit Γ la matrice de covariance du vecteur X , on suppose Γ de la forme:

$$\Gamma = \sigma_1^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{m-1} \\ \rho & 1 & \rho & \dots & \rho^{m-2} \\ \rho^2 & \rho & 1 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{m-1} & \dots & \dots & \rho^2 & \rho & 1 \end{pmatrix}$$

La loi de \mathbf{X} est alors donnée par:

$$P(\mathbf{X}) = (2\pi\sigma_1^2)^{-\frac{m}{2}} (1-\rho^2)^{-\frac{m-1}{2}} \exp \left\{ \frac{-1}{2(1-\rho^2)\sigma_1^2} \left[(1-\rho^2)X_1^2 + \sum_{i=1}^{m-1} (X_{i+1} - \rho X_i)^2 \right] \right\}$$

On a les représentations AR suivantes:

$$\begin{aligned} X_i &= \rho X_{i+1} + \eta_i^+ & \eta_i^+ &\sim \mathcal{N}(0, (1-\rho^2)\sigma_1^2) & 1 \leq i \leq m-1 \\ X_i &= \rho X_{i-1} + \eta_i^- & \eta_i^- &\sim \mathcal{N}(0, (1-\rho^2)\sigma_1^2) & 2 \leq i \leq m \end{aligned}$$

On a également la représentation markovienne:

$$X_i = \frac{\rho}{1+\rho^2} (X_{i-1} + X_{i+1}) + e_i \quad e_i \sim \mathcal{N}\left(0, \frac{1-\rho^2}{1+\rho^2} \sigma_1^2\right) \quad 2 \leq i \leq m-1$$

Ces diverses représentations déterminent totalement les distributions conditionnelles des réflecteurs principaux.

Alors que la distribution marginale d'un réflecteur principal X_s est une $\mathcal{N}(0, \sigma_1^2)$, la distribution de X_s conditionnellement à V_s varie suivant le nombre de voisins qui sont aussi des réflecteurs principaux:

- Tous les éléments de V_s sont des réflecteurs secondaires (le réflecteur est isolé):

$$X_s \sim \mathcal{N}(0, \sigma_1^2)$$

- Un seul élément de V_s est un réflecteur principal (on se trouve à une extrémité d'une ligne de réflecteurs), notons x_t la valeur de ce réflecteur:

$$X_s \sim \mathcal{N}(\rho x_t, (1-\rho^2)\sigma_1^2)$$

- Les deux éléments de V_s sont des réflecteurs principaux (on se trouve au milieu d'une ligne de réflecteurs), notons x_t et x_r les valeurs de ces deux réflecteurs:

$$X_s \sim \mathcal{N}\left(\frac{\rho}{1+\rho^2}(x_t+x_r), \frac{1-\rho^2}{1+\rho^2}\sigma_1^2\right)$$

3.2 IDENTIFICATION DU MODELE

3.2.1 INTRODUCTION

3.2.1.a Paramètres à estimer

L'ensemble des paramètres caractérisant la distribution de Y est:

$$\theta_Y = (\alpha, \beta, \rho, \sigma_0^2, \sigma_1^2)$$

L'ensemble des paramètres caractérisant le modèle de convolution est:

$$\theta = (\theta_Y, f, \sigma_\varepsilon^2)$$

Nous verrons dans la 3^{ème} partie, consacrée à la déconvolution, que α , β et ρ jouent les rôles de paramètres de régularisation, ils sont fixés arbitrairement suivant le type de solution désirée: il faut donc estimer f , σ_ε^2 , σ_0^2 et σ_1^2 . Ces paramètres sont estimables sur chaque trace et l'algorithme proposé dans le Chapitre 1 peut être utilisé.

3.2.1.b Variabilité des paramètres dans la section

On ne peut pas considérer les paramètres comme constants dans toute la section. En particulier, le niveau de bruit augmente avec la profondeur et des changements apparaissent dans le spectre de l'onde au cours de sa propagation: les hautes fréquences s'atténuent. D'autre part, en des endroits différents d'une même section les mesures sont effectuées indépendamment et les caractéristiques du sous-sol peuvent varier considérablement. Pour tenir compte de ces éventuelles variations, nous ne traiterons pas la section complète dans son ensemble mais par morceaux, en supposant les différents paramètres du modèle

constants à l'intérieur d'une même fenêtre si celle-ci est bien choisie. Le problème du choix optimal de la fenêtre n'est pas traité ici, nous nous remettons à la connaissance pratique du géophysicien qui est habitué à ce type de données et sait effectuer un découpage raisonnable de la section.

Les fenêtres utilisées pour estimer les paramètres sont constituées de 10 traces, chacune d'elles contenant 500 données (cela correspond à 1s de propagation)

3.2.2 ALGORITHME

Comme nous l'avons vu, la décomposition de θ_Y n'est pas unique: différentes séquences de réflecteurs associées à des paramètres θ_Y différents peuvent générer des traces identiques (voir Fig.1 du Chapitre 1). Il faut alors fixer λ , la proportion de réflecteurs principaux, et estimer σ_0^2 et σ_1^2 , les variances des coefficients de réflexion.

Soient Z^1, Z^2, \dots, Z^m les m traces de la fenêtre employée pour l'estimation de $\theta = (f, \sigma_\varepsilon^2, \sigma_0^2, \sigma_1^2)$. Nous allons employer exactement la même procédure décrite dans le paragraphe 1.4.2 du Chapitre 1 pour chacune des m traces, mais à chaque itération, l'estimateur de θ sera calculé comme la moyenne des m estimateurs obtenus indépendamment sur les m traces:

- On initialise le paramètre θ à $\theta(0)$.
- Soit $\theta^{(i)}$ l'estimateur de θ à l'itération i . A l'itération $i+1$:

Pour la trace j :

- a) On simule $Y^{j(i+1)}$ avec la distribution $P_{\theta^{(i)}}(Y/Z^j)$ au moyen de l'échantillonneur de Gibbs décrit en annexe du Chapitre 1.
- b) On calcule $\theta^{j(i+1)}$, l'estimateur du maximum de vraisemblance de θ sur la trace j :

$$\theta^{j(i+1)} = \underset{\theta \in \Theta}{\text{Argmax}} L_\theta(Y^{j(i+1)}, Z^j).$$

On répète cette procédure pour les m traces de la fenêtre et on calcule $\theta^{(i+1)}$:

$$\theta^{(i+1)} = \frac{1}{m} \sum_{j=1}^m \theta^{j(i+1)}$$

- On arrête les itérations lorsque l'algorithme a convergé: $\theta^{(i+1)} \approx \theta^{(i)}$

3.2.3 RESULTATS

L'estimation de l'onde est montrée en Fig.5 pour deux fenêtres de temps différentes. La fenêtre A est proche de la surface: données 100 à 600 (entre 0.2 et 1.2s). La fenêtre B en est éloignée: données 900 à 1400 (entre 1.8 et 2.8s). On voit en Fig5.b les différences existantes entre les deux spectres respectifs, l'onde est plus pauvre en hautes fréquences à mesure qu'elle se propage dans le sous-sol.

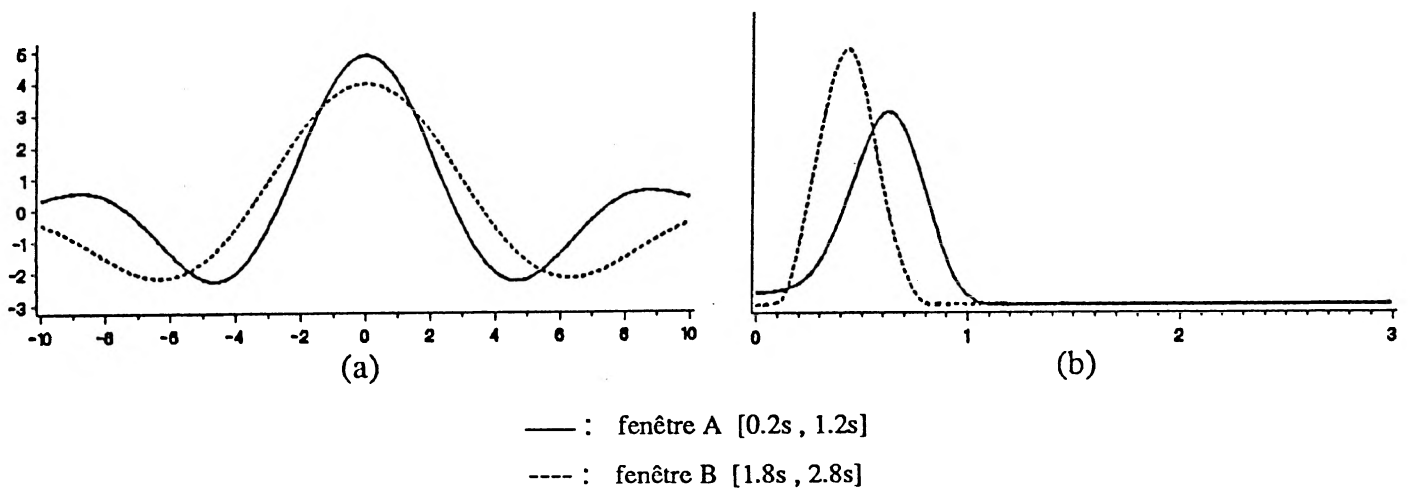


Fig.5 (a) Estimations de l'onde dans les fenêtres A et B.
(b) Spectres des ondes estimés.

L'estimation du niveau de bruit donne le résultat suivant: le rapport signal/bruit dans la fenêtre A (1000dB) est plus élevé que dans la fenêtre B (100dB).

Pour différentes valeurs de λ fixées, nous avons estimé les variances σ_0^2 et σ_1^2 (Figure 6):

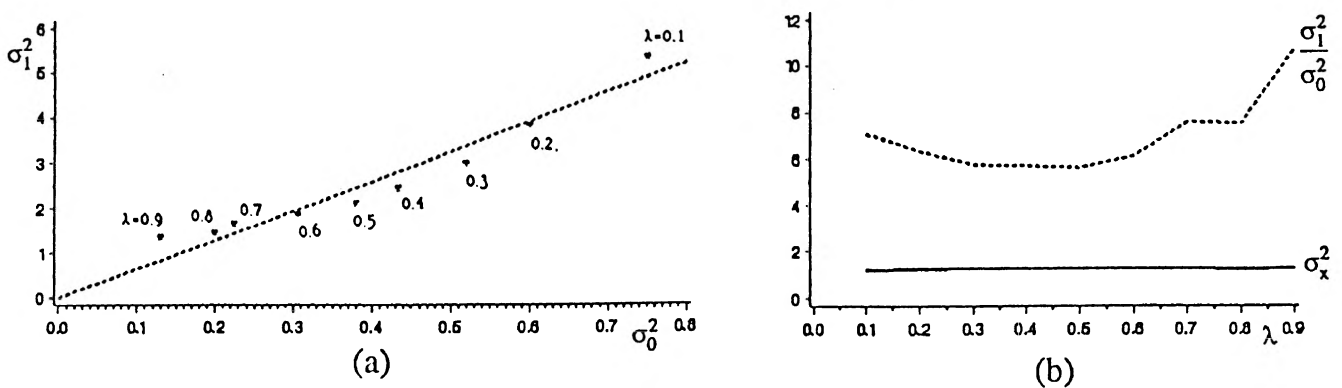


Fig.6 (a) Estimation de σ_1^2 et σ_0^2 pour différentes valeurs fixées de λ .
(b) Estimation du rapport $\frac{\sigma_1^2}{\sigma_0^2}$ et de σ_x^2 en fonction de λ .

Deux remarques sont intéressantes:

- L'estimation du rapport σ_1^2/σ_0^2 est une fonction de λ que l'on peut approximer à une constante pour λ compris entre 0.1 et 0.8 : $\sigma_1^2/\sigma_0^2 \approx 6.5$.
- Pour toute valeur de λ , la variance des coefficients de réflexion $\sigma_x^2 = \lambda\sigma_1^2 + (1-\lambda)\sigma_0^2$ reste constante: $\sigma_x^2 \approx 1.2$ (Fig. 6-b).

Comme nous l'avons remarquer dans le Chapitre 1 (§ 1.4.2.b), à cause du filtre utilisé, la distribution de \mathbf{Y} n'est pas estimable puisque la décomposition $(\lambda, \sigma_0^2, \sigma_1^2)$ n'est pas unique. Pour illustrer ceci, nous avons utiliser l'algorithme en ne fixant plus λ cette fois-ci. Nous présentons en Fig.7 les estimations de λ et σ_1^2/σ_0^2 obtenues avec différentes initialisations:

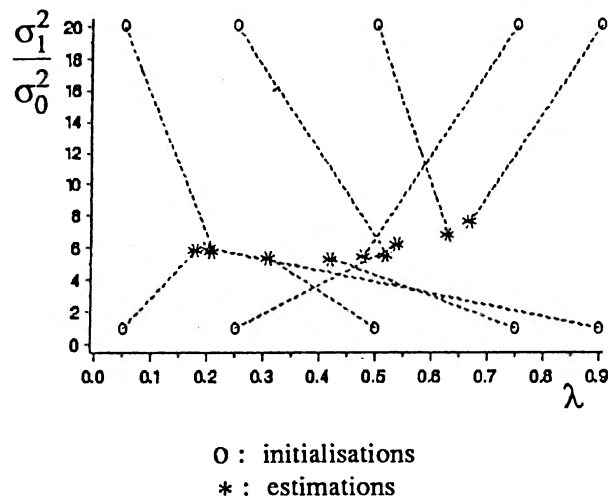


Fig.7 Estimation de λ et σ_1^2/σ_0^2 .

Les résultats confirment bien ce que l'on pouvait attendre:

- Quelque-soit l'initialisation, σ_1^2/σ_0^2 converge vers une valeur à peu près constante (≈ 6.5).
- Suivant l'initialisation, l'estimation de λ varie entre 0.2 and 0.7. Ce paramètre n'est pas estimable.

3.3 ESTIMATION DES REFLECTEURS

Le problème d'estimation du couple $(\mathbf{X}, \mathbf{Q}) = \{X_s, Q_s\}_{s \in S}$ et de $\mathbf{C} = \{C_s\}_{s \in S}$ va être scindé en deux parties:

- Estimation de C , directement d'après les traces sismiques. On en déduit alors $V = \{V_s\}$, le système de voisinage sur S .
- Estimation de (X, Q) en utilisant le système de voisinage précédemment construit.

3.3.1 CONSTRUCTION DU SYSTEME DE VOISINAGE

3.3.1.a Algorithme

En observant une section sismique, on voit de façon évidente la similitude qui existe entre deux traces conjointes. Cela transcrit la continuité qui existe dans le sous-sol, les couches sont régulières et d'une longueur importante devant la distance séparant deux traces. Toutefois, il arrive que ces couches ne soient pas horizontales. Dans ce cas, des shifts apparaissent entre des traces voisines et les valeurs de ces shifts mesurent les inclinaisons des pentes.

Le champ Z est une version filtrée du champ X . A cause de la forme du filtre (proche d'une δ -Dirac), nous allons utiliser l'information apportée par Z pour déterminer le système de voisinage du champ X .

Plaçons-nous sur la trace j , au point (i, j) . Estimer $C_{i,j}$ veut dire faire une estimation locale de la pente du sous-sol en (i, j) :

Soit $\mathcal{Z}_{i,j}$ le vecteur $(Z_{i-m,j}, \dots, Z_{i,j}, \dots, Z_{i+m,j})$, on pose $s = (i, j)$ et $s' = (i', j')$, on appelle alors $\rho_{s,s'}$ la corrélation empirique calculée entre \mathcal{Z}_s et $\mathcal{Z}_{s'}$:

$$\rho_{s,s'} = \frac{\sum_{k=-m}^{+m} (Z_{i+k,j} - \bar{\mathcal{Z}}_s)(Z_{i'+k,j'} - \bar{\mathcal{Z}}_{s'})}{\sqrt{\sum_{k=-m}^{+m} (Z_{i+k,j} - \bar{\mathcal{Z}}_s)^2 \sum_{k=-m}^{+m} (Z_{i'+k,j'} - \bar{\mathcal{Z}}_{s'})^2}}$$

$$\text{où } \bar{\mathcal{Z}}_s = \frac{1}{2m+1} \sum_{k=-m}^{+m} Z_{i+k,j} .$$

On rappelle que $VD(s)$ est le voisin de droite du site s et $VG(s)$, celui de gauche. Supposons que les sites s , $VD(s)$ et $VG(s)$ appartiennent à une même ligne de réflecteurs principaux, d'après les représentations AR du champ X décrites en 3.1.2.d, nous avons:

$$X_s = \rho X_{VG(s)} + \eta_s^+ \quad , \quad X_s = \rho X_{VD(s)} + \eta_s^-$$

où η^+ et η^- sont des bruits blancs Gaussiens.

Nous ferons l'approximation suivante: les représentations AR employées pour le champ X peuvent être étendues au champ Z , notons alors,

$$E_{\text{cor}} = \frac{1}{N} \sum_{s \in S} \left[\| \mathcal{Z}_s - \rho_{s, \text{VD}(s)} \mathcal{Z}_{\text{VD}(s)} \|^2 + \| \mathcal{Z}_s - \rho_{s, \text{VG}(s)} \mathcal{Z}_{\text{VG}(s)} \|^2 \right]$$

On voit facilement que choisir CD_s (resp. CG_s) en chaque s de S qui minimise E_{cor} revient à maximiser la corrélation entre \mathcal{Z}_s et $\mathcal{Z}_{\text{VD}(s)}$ (resp. $\mathcal{Z}_{\text{VG}(s)}$).

Ne sachant pas encore où se trouvent les réflecteurs principaux, il faut estimer C en chaque site. Supposons maintenant qu'un réflecteur secondaire se trouve en $s=(i, j)$, X_s est indépendant de ses voisins, mais du fait de l'effet de convolution, Z_s sera corrélé à ses voisins si une ligne de réflecteurs principaux est proche de s (à une distance inférieure à p). Dans le cas contraire, Z_s est indépendant des Z appartenant aux traces voisines, un critère autre que la corrélation est donc nécessaire pour estimer C .

D'autre part, des problèmes apparaissent si on estime C en minimisant E_{cor} : lorsque le graphe est symétrisé, de nombreuses arêtes disparaissent et nombreux sont les sommets isolés ou à un voisin. Or, plus le graphe possède d'arêtes (plus de sommets ont deux voisins) et meilleure est la continuité dans la solution obtenue par la suite. Nous allons donc introduire un terme de régularisation pour améliorer le graphe dans ce sens et pouvoir estimer C sur tout S . Soient,

$$N_1 = \frac{1}{N} \sum_{s \in S} |CD_s - CG_s|$$

$$N_2 = \frac{1}{N} \sum_{(i,j) \in S} (|CD_{i,j} - CD_{i+1,j}| + |CG_{i,j} - CG_{i+1,j}|)$$

$$N_3 = \frac{1}{N} \sum_{s \in S} (|CD_s - CG_{\text{VD}(s)}| + |CG_s - CD_{\text{VG}(s)}|)$$

N_1 mesure les changements d'angles en chaque site, N_2 les différences de pente entre deux sites voisins d'une même verticale, et N_3 la symétrie du graphe (différences entre arêtes incidentes et arêtes sortantes en chaque sommet).

Il est clair que si $C_s = (h, h)$ pour tout s de S , alors $N_1 = N_2 = N_3 = 0$ et le graphe est complet (tous les sommets ont deux voisins).

Nous allons estimer C en minimisant:

$$J = a_0 E_{\text{cor}} + a_1 N_1 + a_2 N_2 + a_3 N_3$$

Un choix arbitraire de m , a_0 , a_1 , a_2 et a_3 doit être fait. Le paramètre m détermine la taille de la fenêtre sur laquelle sont calculées les covariances, si m est pris trop grand, des détails vont disparaître, et si m est trop petit, on aura une estimation erratique des corrélations, m a été choisi ici égal à 5.

Le choix des autres coefficients s'est fait de manière empirique, en utilisant les deux règles suivantes:

- * Soit \tilde{E}_{cor} la valeur de E_{cor} calculée avec $a_1 = a_2 = a_3 = 0$, c'est-à-dire lorsque le terme de régularisation ne joue aucun rôle. On impose alors à E_{cor} de rester proche de cette valeur optimale:

$$E_{\text{cor}} < 105\% \tilde{E}_{\text{cor}}$$

- * Soit M , le nombre de sommets du graphe ayant deux voisins, on demande à M d'être proche de N :

$$M > 90\% N$$

Il faut également remarquer que les termes de régularisation que nous avons introduits ont permis d'éliminer d'éventuelles configurations inacceptables (lignes de réflecteurs qui se croisent, par exemple)

3.3.1.b Résultats

Le graphe utilisé ici est de taille 12×100 ($N=1\ 000$). Pour ce graphe, $\tilde{E}_{\text{cor}} = 1594$. Nous avons fixé $a_1 = a_2 = 1$ et $a_3 = 3$. Nous donnons dans le tableau ci-dessous les résultats obtenus pour différentes valeurs de a_0 :

a_0	E_{cor}	% \tilde{E}_{cor}	Nb de sites isolés	% N	Nb de sites à 1 voisin	% N	Nb de sites à 2 voisins	% N
0	2661	170	0	0	0	0	1000	100
0.2	2133	134	0	0	21	2.1	979	97.9
1	1705	107	0	0	40	4.0	960	96.0
10	1608	101	1	0.1	53	5.3	946	94.6
$+\infty$	1594	100	10	1	129	12.9	861	86.1

Finalement, nous avons choisi $a_0 = 10$, $a_1 = a_2 = 1$ et $a_3 = 3$ puisque la corrélation reste très proche de la corrélation optimale alors que le nombre de sites ayant deux voisins croît considérablement.

J a été minimisé en utilisant un algorithme de recuit. Nous présentons en Figure 8 une portion du graphe construit, en superposition du sismogramme original:

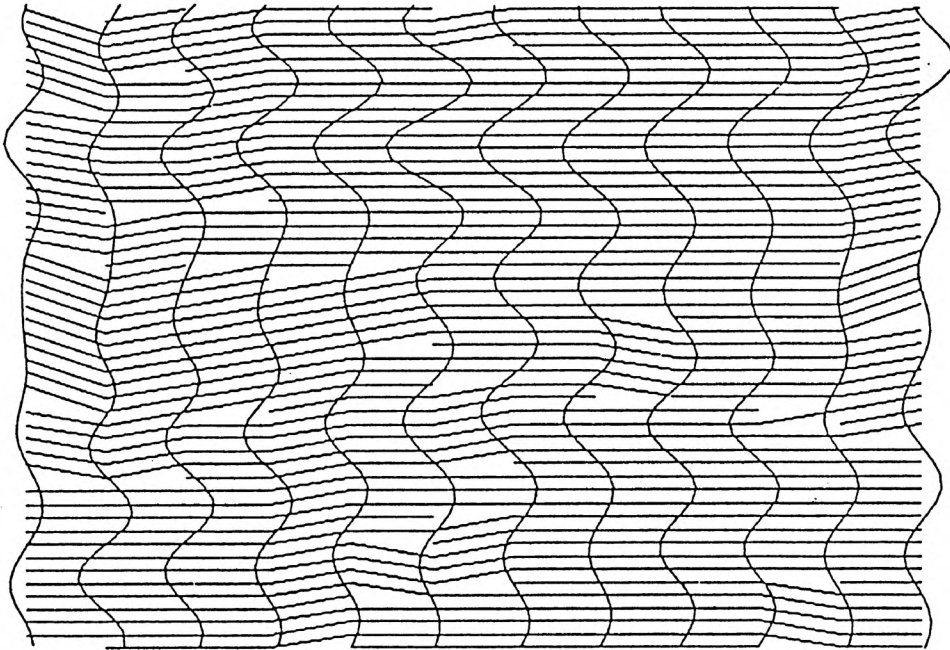


Fig.8 Traces sismiques et graphe

3.3.2 ESTIMATION DE Y

3.3.2.a Introduction

Revenons au modèle de convolution initial:

$$\begin{cases} \mathbf{Z} = \mathbf{f} * \mathbf{X} + \boldsymbol{\varepsilon} \\ \mathbf{Y} = (\mathbf{Q}, \mathbf{X}) \end{cases}$$

Le système de voisinage ayant été fixé, nous disposons des traces sismiques comme seule information pour estimer $\mathbf{Y} = (\mathbf{X}, \mathbf{Q})$. Le filtre \mathbf{f} et les paramètres statistiques α , β , σ_0^2 , σ_1^2 , σ_ε^2 et ρ sont supposés connus.

Techniques de déconvolution :

Nous nous trouvons dans un cadre classique de données incomplètes: dans le triplet $(\mathbf{Z}, \mathbf{X}, \mathbf{Q})$, seul \mathbf{Z} est observé. Quelles sont les approches possibles pour reconstruire le couple (\mathbf{X}, \mathbf{Q}) ?

a) Critères de bon ajustement avec les observations

La distribution de Z , d'après le modèle de convolution, est étroitement liée à celle de X . Si seule cette composante nous intéresse, une première idée paraît naturelle pour calculer \hat{X} , l'estimation de X :

- Travailler dans l'espace du temps et choisir \hat{X} qui ajuste au mieux les données, c'est-à-dire trouver la combinaison de réflecteurs qui produirait les traces sismiques les plus proches possible de celles observées. Une question se pose alors: que signifie " le plus proche possible " ?

* La technique des moindres carrés, consiste à minimiser l'erreur quadratique [1], [29]:

$$E_2 = \| Z - f * X \|^2 = \sum_{s \in S} (Z_s - f * X_s)^2$$

$$\hat{X} = \underset{X}{\text{Argmin}} E_2$$

lorsque le bruit est Gaussien, cela revient à maximiser la distribution conditionnelle de Z , $P(Z / X)$.

Cette technique offre une solution relativement facile à calculer, mais a le désavantage d'être très instable devant le bruit, ou devant une variation du filtre. Nous avons vu en effet qu'une infinité de séquences peuvent produire les mêmes traces.

* Cette instabilité diminue si, au lieu de minimiser la norme L^2 , on choisit de minimiser la norme L^1 de l'erreur [5], [27]:

$$E_1 = \| Z - f * X \|^1 = \sum_{s \in S} |Z_s - f * X_s|$$

$$\hat{X} = \underset{X}{\text{Argmin}} E_1$$

(La difficulté est alors d'atteindre le minima global de cette expression)

- Une autre alternative consiste à chercher un filtre inverse en passant dans le domaine des fréquences:

Partant de l'équation de convolution en l'absence de bruit $Z = f * X$, on en déduit la relation $\hat{Z} = \hat{f} \hat{X}$ (\hat{Z} est la transformée de Fourier de Z). On calcule alors $\hat{X} = \hat{f}^{-1} \hat{Z}$, il ne reste plus ensuite qu'à calculer la transformée de Fourier inverse de \hat{X} pour retrouver X . Cette solution offre un inconvénient majeur: le filtre f est à bande limitée (Fig.5), \hat{f} s'annule sur un intervalle et rend impossible le calcul de \hat{X}

sur cet intervalle. D'autre part, la présence d'un bruit de mesure rend le problème plus complexe. On trouvera dans [20], [23] différentes techniques utilisées pour calculer un filtre inverse approché \hat{f}^{-1} .

Quelque-soit la norme choisie ou le filtre inverse utilisé, la configuration du sous-sol obtenue peut être totalement irréaliste, en désaccord avec une connaissance à priori du problème, puisqu'aucune contrainte sur X n'a été introduite.

b) Utilisation d'une information à priori

Une seconde approche consiste alors à utiliser une information à priori sur les réflecteurs:

- La déconvolution par minimum d'entropie (MED) cherche une solution constituée de pics éparses, on demande donc aux réflecteurs d'avoir une distribution la plus éloignée possible de la Gaussienne [9], [28]. L'expression " minimum d'entropie " vient du fait que, à variance donnée, la distribution Gaussienne est d'entropie maximum. Plusieurs algorithmes existent, chacun d'eux visant à maximiser une fonctionnelle qui mesure une distance à la distribution Gaussienne: kurtosis, négentropie . . .

- Les méthodes bayésiennes, quant à elles, supposent connue la distribution à priori de Y , on choisit alors \hat{Y} le plus probable "à postérieur". On tient alors compte des données observées, mais également de cette distribution à priori: observant les traces sismiques, quelle est, parmi toutes les séquences de réflecteurs supposées possibles, la plus probable pour générer de telles traces ?

Plusieurs algorithmes de reconstruction existent [31], nous les avons étudiés en détail dans le chapitre précédent dans le cas unidimensionnel. Dans le cas de déconvolution sismique, l'estimateur le plus couramment employé est celui du MAP [22]. Cet estimateur est obtenu en maximisant la distribution à postérieur $P(Y/Z)$, la méthode que nous proposons est une modification de cet algorithme.

3.3.2.b Algorithme de déconvolution

Soit P_θ la distribution du couple (Y, Z) . Nous avons les relations suivantes:

$$P_\theta(Y/Z) = P_\theta(X, Q/Z)$$

$$P_\theta(Y/Z) = \frac{P_\theta(Z/X) P_\theta(X/Q) P_\theta(Q)}{P_\theta(Z)}$$

$P_\theta(Z/X)$ représente une information à postérieur apportée par les observations. Maximiser ce terme signifie chercher X qui rende ces observations les plus probables possibles.

$P_\theta(\mathbf{X}, \mathbf{Q})$ représente l'information à priori que l'on met. Le rôle de ce terme est d'aboutir à une solution le plus en accord possible avec cet à priori lorsqu'il est maximisé, d'éviter d'obtenir des solutions peu probables. Dans notre cas, ce terme doit tendre à régulariser la solution obtenue, mais quel poids donner à ce terme et quel poids donner aux observations pour aboutir à une solution optimale [8] ?

En effet, la solution qui nous paraîtra la meilleure est celle qui optimisera un critère, ou plus exactement, une combinaison de plusieurs critères. Rien ne nous dit que maximiser la probabilité à postériori $P_\theta(\mathbf{Y}/\mathbf{Z})$ aboutit toujours à cette solution.

Donnons-nous donc plutôt des critères de bonne restauration, et voyons quelle est la fonction de coût associée à minimiser, ou de manière équivalente, quelle est la fonctionnelle à maximiser, s'il s'agit d'une probabilité.

Nous ne nous proposons pas de retrouver la série originale des coefficients de réflexion, mais de présenter un algorithme qui, pour un ensemble de critères donnés, aboutit à la solution optimale. Des critères qui semblent raisonnables sont les suivants:

- (a) Un bon ajustement avec les données doit être respecté.
- (b) Les réflecteurs les plus importants doivent constituer des lignes, pour mettre en évidence la structure de couches.
- (c) Une continuité latérale dans les valeurs des réflecteurs principaux doit apparaître.

Voyons maintenant comment chaque critère peut être traduit en termes mathématiques.

Critère de bon ajustement avec les données: Si on cherche comme solution des séquences de réflecteurs qui produisent des traces les plus semblables (au sens des moindres carrés) à celles observées, on cherche \mathbf{X} qui minimise:

$$E_2 = \|\mathbf{Z} - \mathbf{f} * \mathbf{X}\|^2 = \sum_{s \in S} (Z_s - \mathbf{f} * X_s)^2$$

Le bruit ε étant Gaussien, cela équivaut à maximiser $P_\theta(\mathbf{Z}/\mathbf{X})$:

$$P_\theta(\mathbf{Z}/\mathbf{X}) = (2\pi \sigma_\varepsilon^2)^{-N} \exp \left\{ \frac{-1}{2\sigma_\varepsilon^2} \sum_{s \in S} (Z_s - \mathbf{f} * X_s)^2 \right\}$$

On notera dans la suite $\hat{\sigma}_\varepsilon^2$ la variance estimée de ε et $\hat{\mathbf{f}}$ le filtre estimé.

Critère de bon agencement des réflecteurs: Le critère (b) est directement lié aux caractéristiques locales du champ Q :

- Le paramètre α détermine la proportion de réflecteurs principaux: plus α est grand et plus cette proportion est faible.

- Le paramètre β détermine la continuité des lignes de réflecteurs (β est toujours supposé positif): plus β est grand et plus on aura tendance à supprimer les réflecteurs principaux isolés et les trous dans les lignes de réflecteurs.

On se fixera donc un jeu de paramètres $(\tilde{\alpha}, \tilde{\beta})$ en accord avec le type de solution recherchée.

Critère de régularité dans les valeurs des réflecteurs : Ce critère est directement lié à la loi conditionnelle de $X_s, P_\theta(X_s / Q_s, Y_t, t \in V_s)$.

Si $Q_s=0$, Y_s est un réflecteur secondaire, $P_\theta(X_s / Q_s=0)$ est maximum lorsque $X_s=0$. Cela correspond à la régularité optimale.

Si $Q_s=1$, Y_s est un réflecteur principal, X_s est corrélé à ses voisins si ils sont aussi des réflecteurs principaux. Plus ρ est grand, et plus les valeurs des coefficients de réflexion sont proches sur une même ligne de réflecteurs.

Les variances σ_0^2 et σ_1^2 sont estimées, notons $\hat{\sigma}_0^2$ et $\hat{\sigma}_1^2$ les estimateurs obtenus. La corrélation ρ est fixée arbitrairement à une valeur $\tilde{\rho}$, suivant la régularité demandée à la solution.

Finalement, une combinaison des trois critères cités plus haut nous mène à maximiser une fonctionnelle qui tient compte des trois termes $P_\theta(Z/X)$, $P_\theta(X/Q)$ et $P_\theta(Q)$.

L'estimateur du MAP est obtenu lorsqu'on donne le même poids à chacun de ces termes puisqu'il consiste à maximiser le produit $P_\theta(Z/X)P_\theta(X,Q)$. On peut alors proposer un estimateur plus général qui maximise:

$$H_r(Y) = P_\theta(Z/X)^r P_\theta(X, Q)$$

r indique le poids que l'on donne à la partie "ajustement" face à la partie "à priori".

Le vrai paramètre θ est bien sûr inconnu, on utilise en fait les paramètres estimés $(\hat{f}, \hat{\sigma}_\varepsilon^2, \hat{\sigma}_0^2, \hat{\sigma}_1^2)$ et les paramètres que l'on s'est fixé comme paramètres de régularisation $(\tilde{\alpha}, \tilde{\beta}, \tilde{\rho})$. Notons maintenant

$\hat{\sigma}_\varepsilon^2 = \frac{1}{r} \hat{\sigma}_\varepsilon^2$ et $\tilde{\theta} = (\hat{f}, \hat{\sigma}_\varepsilon^2, \hat{\sigma}_0^2, \hat{\sigma}_1^2, \hat{\alpha}, \hat{\beta}, \hat{\rho})$, maximiser $H_r(\mathbf{Y})$ revient alors à maximiser $P_{\tilde{\theta}}(\mathbf{Y}/\mathbf{Z})$, c'est-à-dire choisir comme estimateur de \mathbf{Y} l'estimateur du MAP dans un modèle paramétré par $\tilde{\theta}$.

Finalement, nous avons construit en deux temps une fonctionnelle à maximiser:

- Dans la partie d'estimation des paramètres du modèle, du fait de la non unicité de la solution, nous avons été amenés à estimer un ensemble de distributions à priori des réflecteurs. En accord avec le type de solution cherchée, on choisit un élément particulier de cet ensemble en se fixant α , β et ρ et en estimant σ_0^2 et σ_1^2 . On estime également le filtre f et le niveau de bruit σ_ε^2 .

- Suivant l'importance que l'on donne dans la reconstruction à la partie ajustement, on se fixe artificiellement un niveau de bruit $\tilde{\sigma}_\varepsilon^2 = \frac{1}{r} \hat{\sigma}_\varepsilon^2$.

La solution cherchée $\hat{\mathbf{Y}}$ maximise alors $P_{\tilde{\theta}}(\mathbf{Y}/\mathbf{Z})$:

$$\hat{\mathbf{Y}} = \underset{\mathbf{Y}}{\text{ArgMax}} P_{\tilde{\theta}}(\mathbf{Y}/\mathbf{Z})$$

La maximisation de $P_{\tilde{\theta}}(\mathbf{Y}/\mathbf{Z})$ est réalisée en utilisant un algorithme de recuit simulé [11], [16], [19], cette procédure est décrite en annexe.

3.4 RESULTATS

Les résultats présentés ici proviennent de données réelles fournies par l'INTEVEP (Vénézuéla). De 12 traces de la section sismique complète, nous avons sélectionné une petite fenêtre de 100 données, proche de la surface (entre 0.5 et 0.7 sec). Les données ont été normalisées afin que la variance de \mathbf{Z} soit égale à 1.

La variance estimée du bruit est égale à 10^{-3} et le filtre estimé est celui présenté en trait plein en Fig.5.

Les différentes solutions proposées correspondent à des valeurs distinctes de $\tilde{\theta}$ telles que:

$$\hat{\mathbf{Y}} = \underset{\mathbf{Y}}{\text{ArgMax}} P_{\tilde{\theta}}(\mathbf{Y}/\mathbf{Z})$$

Comme pour chaque valeur de $\tilde{\theta}$, on se fixe r , α , β et ρ , pour plus de simplicité, nous noterons dans la suite $\tilde{\theta} = (r, \alpha, \beta, \rho)$.

4 jeux distincts de paramètres ont été utilisés:

$$(a) \quad \tilde{\theta}_1 = (1 , 2.00 , 0.00 , 0.00)$$

$$(b) \quad \tilde{\theta}_2 = (0.1 , 0.10 , 0.20 , 0.90)$$

$$(c) \quad \tilde{\theta}_3 = (0.01 , 0.05 , 0.30 , 0.99)$$

$$(d) \quad \tilde{\theta}_4 = (0.01 , 2.00 , 4.00 , 0.99)$$

a) Pour $\tilde{\theta}_1$, nous avons choisi $\beta = \rho = 0$, cela signifie qu'il n'y a aucune dépendance latérale pour les variables Q et X, chaque trace est alors déconvoluée séparément. α est tel que λ , la proportion à priori de réflecteurs principaux, soit proche de 0.1.

b) Les valeurs de $\tilde{\theta}_2$ sont telles que, à priori, les réflecteurs principaux soient peu nombreux et bien corrélés. Le rapport signal/bruit estimé étant très grand (10^3), nous avons choisi une valeur de σ_ε^2 plus élevée que celle estimée pour diminuer le rôle de l'ajustement dans notre reconstruction en posant $r=0.1$.

c) Pour $\tilde{\theta}_3$, α a été pris très petit et ρ très grand, cela correspond d'une part à une proportion à priori importante de réflecteurs principaux et d'autre part à une très forte corrélation entre ceux-ci. Là aussi, le rôle de l'ajustement a été considérablement atténué.

d) Pour $\tilde{\theta}_4$, on a choisi cette fois-ci une grande valeur de α pour avoir une faible proportion de réflecteurs principaux.

Les différents résultats présentés dans les figures 9-12 sont les suivants:

- Figure 9 : estimation des réflecteurs principaux et secondaires.
- Figure 10 : estimation des réflecteurs principaux.
- Figure 11 : convolution des coefficients de réflexion de la Fig.9 avec le filtre estimé.
- Figure 12 : estimation de l'impédance acoustique, calculée à partir des coefficients de réflexion de la Fig.9.

Les coefficients de réflexion $\{X(k)\}$ et l'impédance acoustique $\{I(k)\}$ sont liés par les relations:

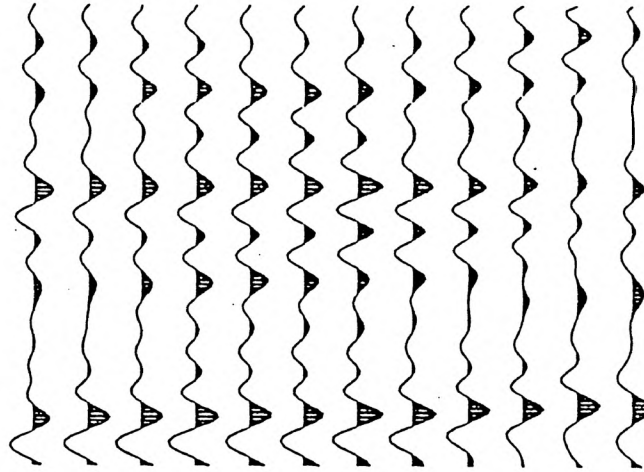
$$I(k+1) = I(k) \frac{1 + X(k)}{1 - X(k)} \quad , \quad X(k) = \frac{I(k+1) - I(k)}{I(k+1) + I(k)}$$

Dans chaque figure, nous présentons également les traces originales utilisées.

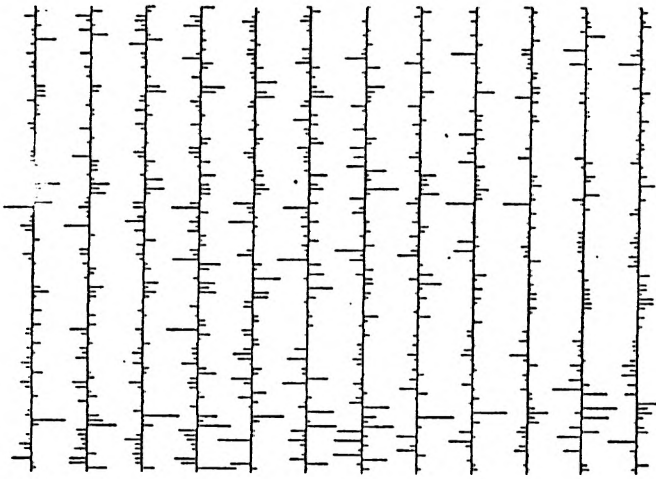
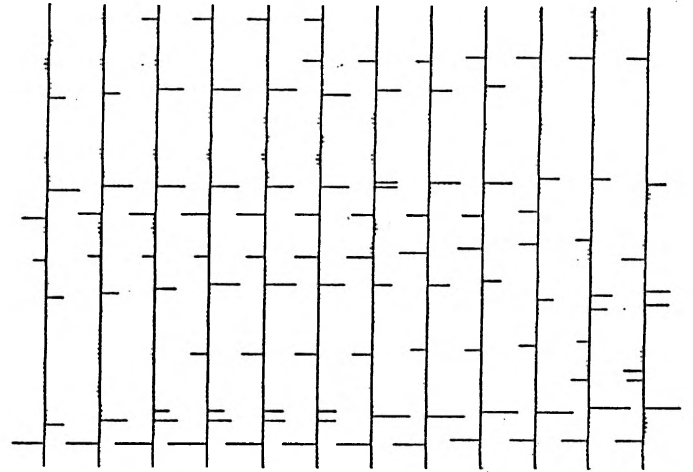
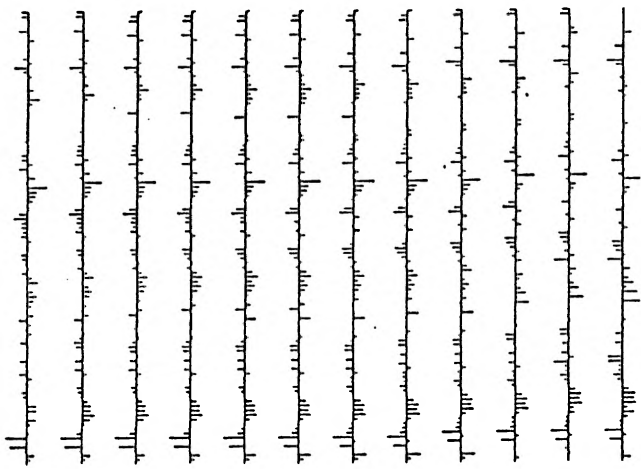
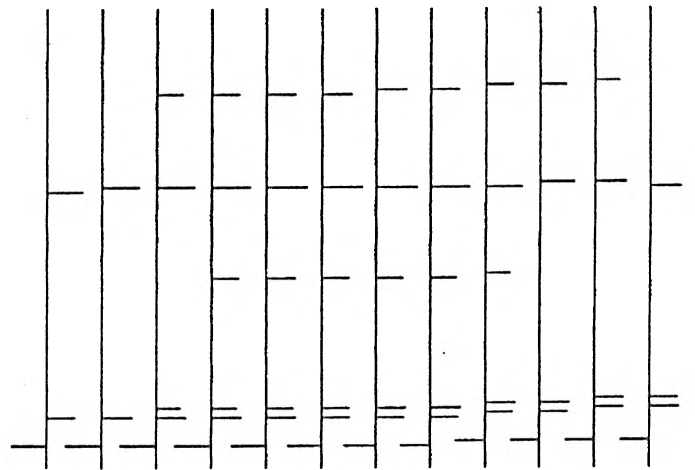
Nous donnons ci-dessous, l'erreur quadratique résiduelle pour chaque choix de $\tilde{\theta}$:

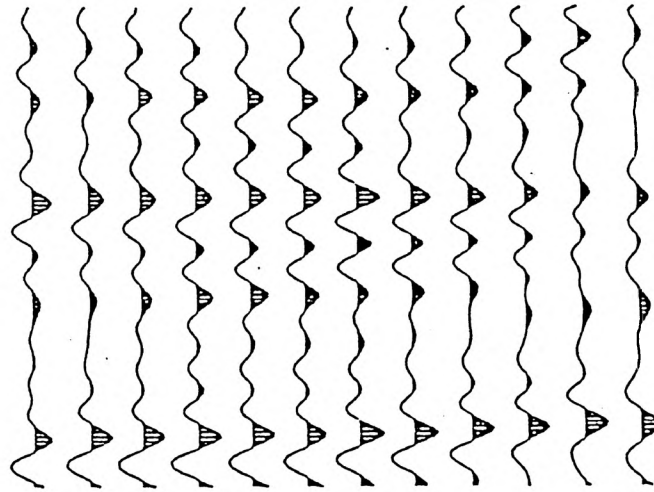
$$e_2 = \frac{1}{N} \|Z - f*\hat{X}\|^2 = \frac{1}{N} \sum_{s \in S} (Z_s - f*\hat{X}_s)^2$$

$\tilde{\theta}$	$\tilde{\theta}_1$	$\tilde{\theta}_2$	$\tilde{\theta}_3$	$\tilde{\theta}_4$
e_2	$3 \cdot 10^{-4}$	$4.5 \cdot 10^{-3}$	10^{-2}	$1.5 \cdot 10^{-1}$

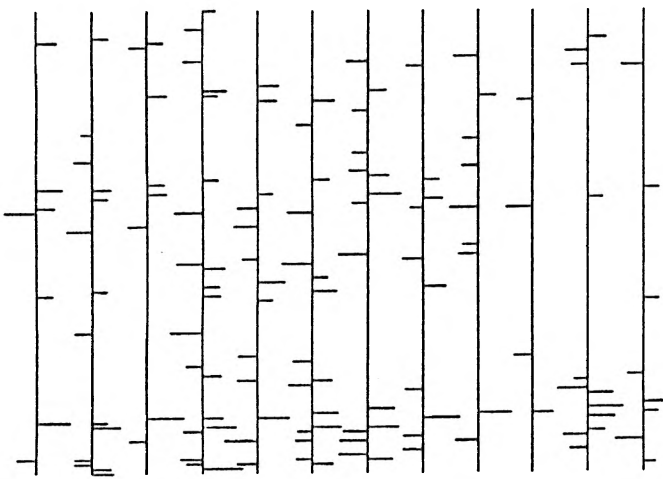
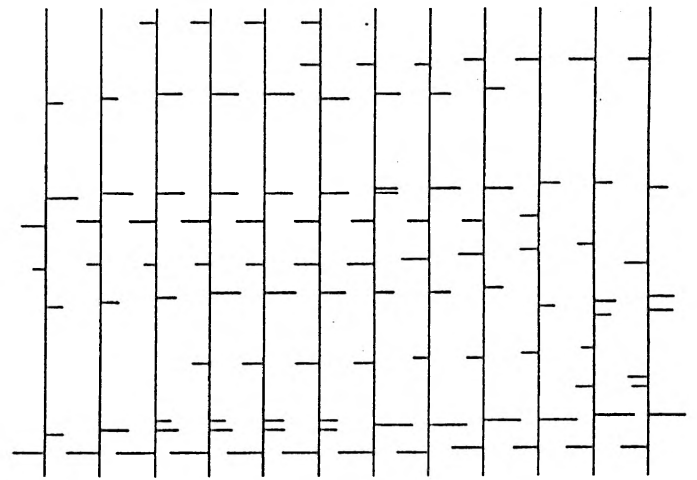
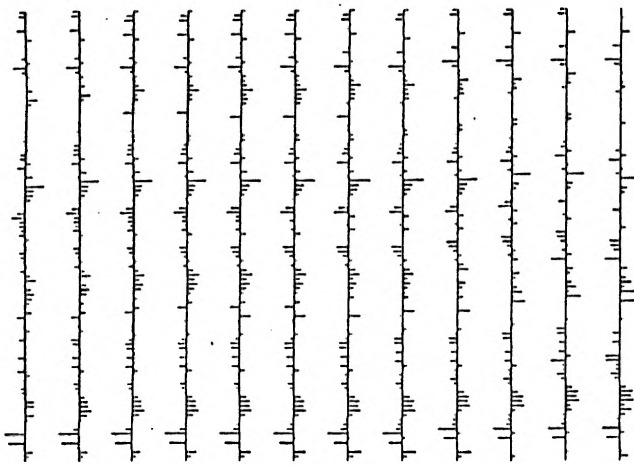
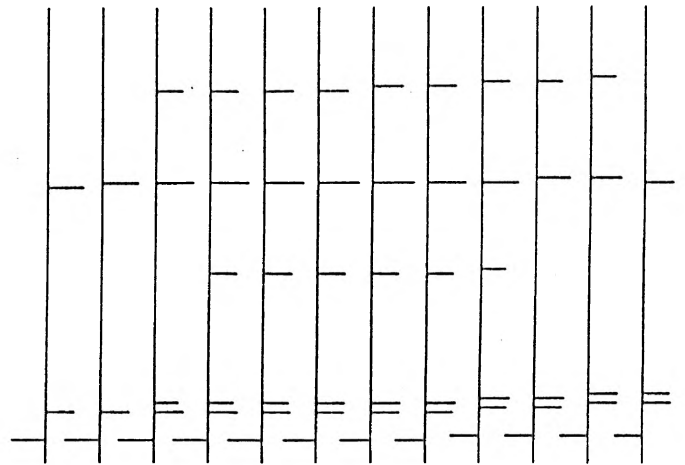


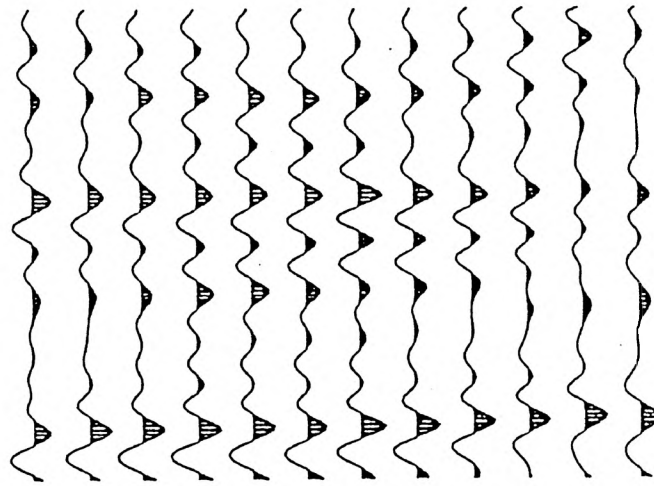
Traces originales

(a) $\tilde{\theta}_1 = (1; 2; 0; 0)$ (b) $\tilde{\theta}_2 = (0.1; 0.1; 0.2; 0.9)$ (c) $\tilde{\theta}_3 = (0.01; 0.05; 0.3; 0.9)$ (d) $\tilde{\theta}_4 = (0.01; 2; 4; 0.9)$ **Fig.9** Estimation des coefficients de réflexion principaux et secondaires.

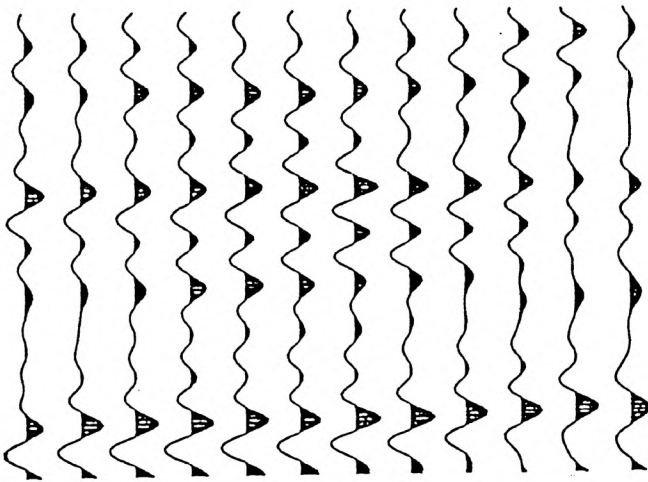
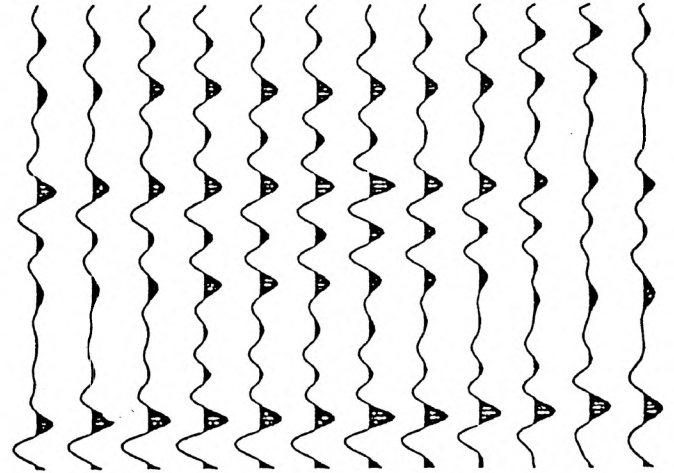
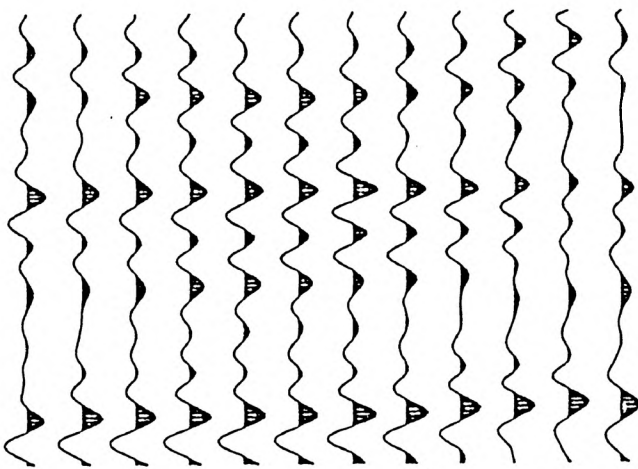
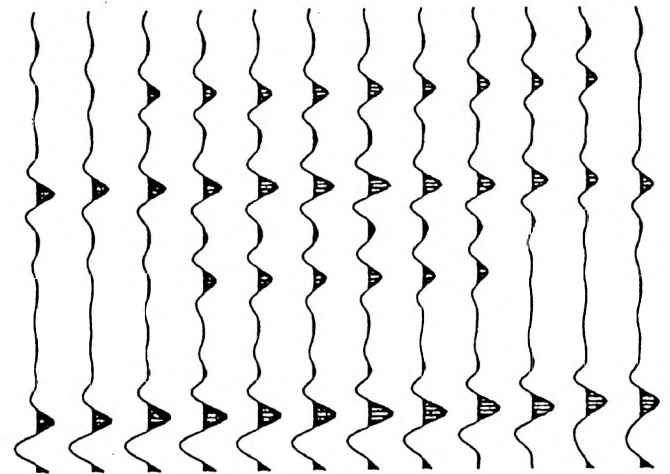


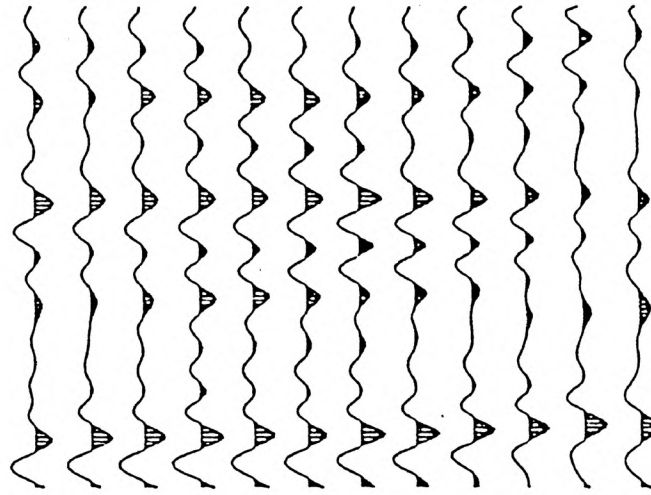
Traces originales

(a) $\tilde{\theta}_1 = (1; 2; 0; 0)$ (b) $\tilde{\theta}_2 = (0.1; 0.1; 0.2; 0.9)$ (c) $\tilde{\theta}_3 = (0.01; 0.05; 0.3; 0.9)$ (d) $\tilde{\theta}_4 = (0.01; 2; 4; 0.9)$ **Fig.10** Estimation des coefficients de réflexion principaux.

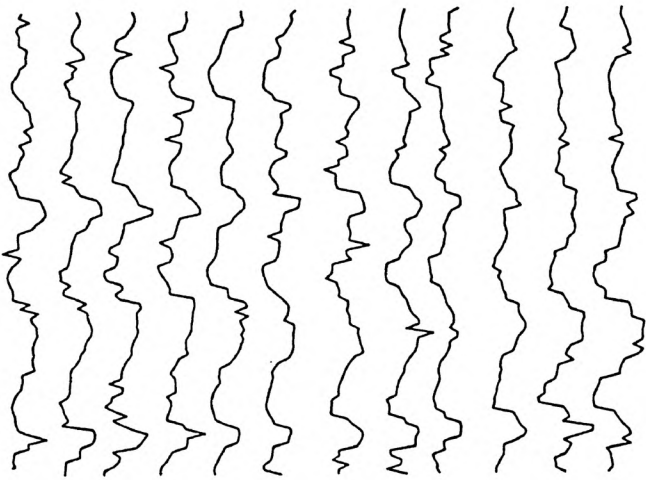
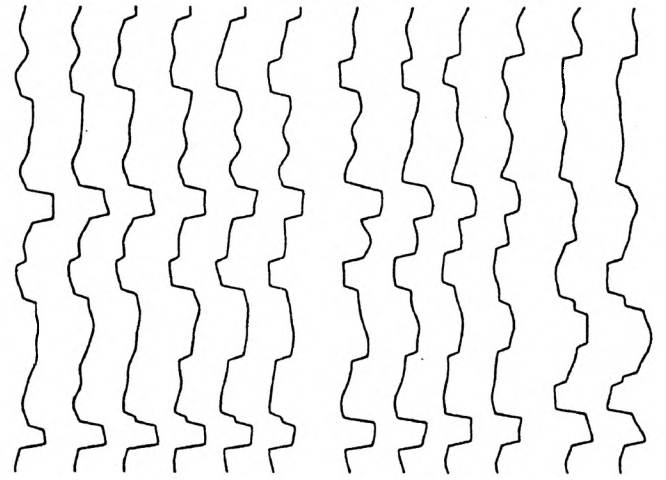
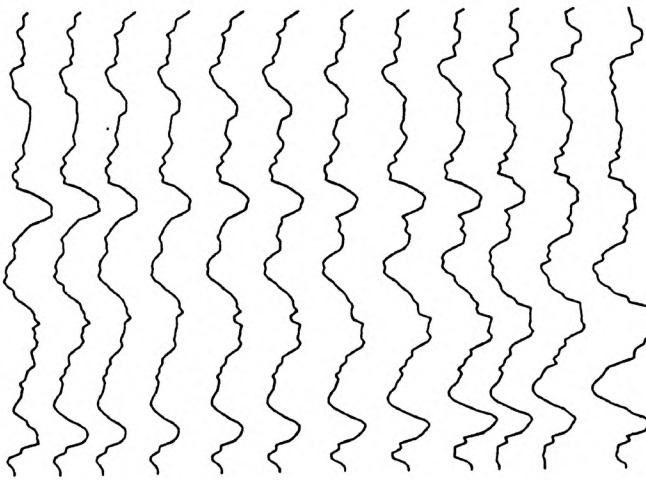
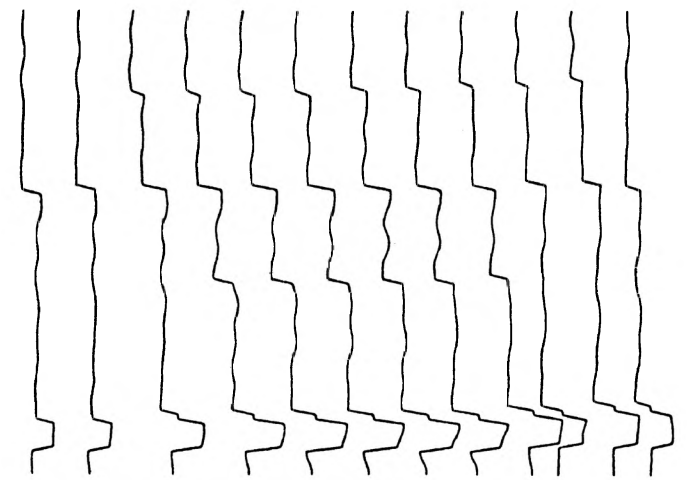


Traces originales

(a) $\tilde{\theta}_1 = (1; 2; 0; 0)$ (b) $\tilde{\theta}_2 = (0.1; 0.1; 0.2; 0.9)$ (c) $\tilde{\theta}_3 = (0.01; 0.05; 0.3; 0.9)$ (d) $\tilde{\theta}_4 = (0.01; 2; 4; 0.9)$ **Fig.11** Convolution des coefficients de réflexion avec le filtre estimé.



Traces originales

(a) $\tilde{\theta}_1 = (1; 2; 0; 0)$ (b) $\tilde{\theta}_2 = (0.1; 0.1; 0.2; 0.9)$ (c) $\tilde{\theta}_3 = (0.01; 0.05; 0.3; 0.9)$ (d) $\tilde{\theta}_4 = (0.01; 2; 4; 0.9)$ **Fig.12** Estimation de l'impédance acoustique.

Au vu de ces résultats, quelques remarques peuvent être faites :

Lorsque chaque trace est déconvoluée séparément (Fig.9-a et 10-a), un très bon ajustement avec les données peut être atteint (Fig.11-a), mais la continuité latérale des réflecteurs n'apparaît pas. Au contraire, cette continuité est évidente dans les autres exemples. En effet, lorsque seul un critère de bon ajustement est requis, on peut aboutir à des solutions distinctes, qui produisent des traces semblables une fois convoluées. D'un autre côté, plus une connaissance à priori du modèle est bonne, et plus les contraintes sur les solutions sont fortes, le nombre de solutions possibles diminue donc considérablement.

La comparaison entre les solutions obtenues en Figure 9 avec $\tilde{\theta}_3$ et $\tilde{\theta}_4$ est intéressante, puisque ce sont deux solutions extrêmes, quand à la proportion de réflecteurs principaux. Dans (d), seuls apparaissent les événements les plus importants de (c), et sous une forme "condensée": plusieurs réflecteurs voisins de faible amplitude dans (c) sont réduits à un réflecteur unique d'amplitude plus grande dans (d).

Nous présentons en Figure 13 un exemple de convolution où un réflecteur unique et un paquet de petits réflecteurs produisent des traces semblables.

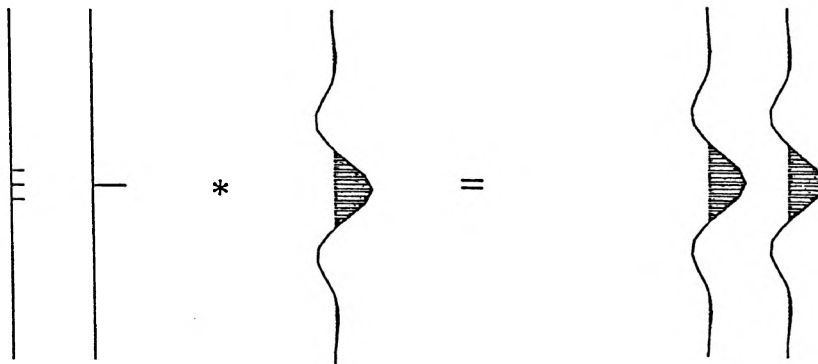


Fig.13 Exemple de non unicité dans la déconvolution.

Alors que la continuité latérale des réflecteurs n'apparaît pas lorsque chaque trace est déconvoluée séparément (Fig.9-a), une certaine continuité est évidente dans l'impédance acoustique (Fig.12-a). Les diverses estimations de l'impédance, issues d'autres jeu de paramètres (Fig.12-b, 12-c et 12-d), présentent également cette continuité spatiale tout en étant plus régulières. De plus chacune de ces représentations montrent les mêmes tendances, plus ou moins filtrées.

L'estimation de l'impédance est robuste au choix des paramètres utilisés dans l'algorithme de déconvolution. Ce résultat est particulièrement important puisque pour une analyse géologique du sous-sol, le géophysicien utilisera plus volontiers l'impédance acoustique que la série des coefficients de réflexion.

CONCLUSION

Notre but n'est pas de reconstruire la série des coefficients de réflexion originale. En effet, ce problème ne peut être résolu si on ne dispose que des traces sismiques comme information, plusieurs solutions peuvent alors être proposées mais qu'est-ce qui permet de dire qu'une solution est meilleure qu'une autre ?

Une connaissance à priori du problème devrait permettre de réduire le nombre de ces solutions. Un spécialiste (un géophysicien dans notre cas) dispose là d'un outil qui lui permet de construire la solution optimale en accord avec les contraintes qu'il se donne. Si de plus, une information extérieure (enregistrements provenant de forages et mesurant en certains endroits de la section l'impédance du milieu, la résistivité, la radio-activité, etc . . .) est à sa disposition, une automatisation du choix des paramètres à employer peut être envisagée.

En particulier, la mesure de l'impédance acoustique dans un puit doit permettre de calibrer les résultats obtenus à partir de données sismiques puisque cette même variable peut être directement estimée d'après la série des coefficients de réflexion.

ANNEXE

Calcul de la loi conditionnelle $P(Y_s / Z, Y_t, t \neq s)$

Une des composantes de Y est discrète alors que l'autre est continue. La loi conditionnelle de Y_s sera alors naturellement décomposée en deux parties :

- $P(Q_s / Z, Y_t = y_t, t \neq s)$, nous allons montrer que cette loi est une loi de Bernoulli, de paramètre λ_s .
- $P(X_s / Z, Q_s = q_s, Y_t = y_t, t \neq s)$, nous allons montrer que cette loi est une loi normale de paramètre $(m_{1,s}, V_{1,s})$ si $Q_s = 1$ et $(m_{0,s}, V_{0,s})$ si $Q_s = 0$.

Calculons les paramètres $m_{1,s}, V_{1,s}, m_{0,s}, V_{0,s}$ et λ_s . Nous avons la relation :

$$P(Y_s / Z, Y_t = y_t, t \neq s) = \frac{P(Z/Y) P(Y_s / Y_t = y_t, t \neq s)}{P(Z / Y_t = y_t, t \neq s)}$$

Contrairement au cas unidimensionnel, les Y_s ne sont pas indépendants,

$$P(Y_s / Z, Y_t = y_t, t \neq s) = \frac{P(Z/Y) P(X_s / Q_s = q_s, Y_t = y_t) P(Q_s / Y_t = y_t)}{P(Z / Y_t = y_t, t \neq s)}$$

Pour toute variable U continue et ayant une densité par rapport à la mesure de Lebesgue, nous écrirons dans la suite de façon abrégée $P(U=u)$ à la place de $\frac{P(U \in [u, u+du])}{du}$, c'est-à-dire la densité calculée au point u .

* Conditionnellement à Q_s et aux Y_t , X_s est Gaussien:

$$P(X_s = x_s / Q_s = q_s, Y_t = y_t) = \frac{1}{\sqrt{2\pi V_L}} \exp\left\{-\frac{1}{2V_L} (x_s - \mu_L)^2\right\}$$

μ_L est la prédiction "latérale" de X_s , dépendant du voisinage V_s et de q_s , V_L est la variance associée à cette prédiction.

Nous allons noter dans la suite $(\mu_L, V_L) = \begin{cases} (\mu_{L0}, V_{L0}) & \text{si } q_s = 0 \\ (\mu_{L1}, V_{L1}) & \text{si } q_s = 1 \end{cases}$

* Du fait de l'équation de convolution $Z = X * f + \varepsilon$, Z est seulement fonction de X , donc,

$$P(Z/Y) = P(Z/X_s, X_t \neq s).$$

Cette distribution est Gaussienne et a une densité qui peut s'écrire sous la forme

$$P(Z=z/Y) = h(z, X_t \neq s) \frac{1}{\sqrt{V_V}} \exp\left\{ -\frac{1}{2V_V} (x_s - \mu_V)^2 \right\}$$

où h est une fonction qui ne dépend pas de Y_s .

μ_V peut s'interpréter comme la prédiction "verticale" de X_s , c'est-à-dire calculée en utilisant une information verticale. V_V est la variance associée à cette prédiction.

* Conditionnellement aux Y_t , Q_s est une variable de Bernoulli de paramètre λ_L :

$$P(Q_s=1/Y_t=y_t) = 1 - P(Q_s=0/Y_t=y_t) = \lambda_L$$

Nous avons les relations :

$$P((Q_s, X_s) = (1, x_s) / Z, Y_t) = Cte \times \frac{\lambda_L}{\sqrt{V_{L1}}} \exp\left\{ -\frac{1}{2V_V} (x_s - \mu_V)^2 - \frac{1}{2V_{L1}} (x_s - \mu_{L1})^2 \right\}$$

$$P((Q_s, X_s) = (0, x_s) / Z, Y_t) = Cte \times \frac{(1-\lambda_L)}{\sqrt{V_{L0}}} \exp\left\{ -\frac{1}{2V_V} (x_s - \mu_V)^2 - \frac{1}{2V_{L0}} (x_s - \mu_{L0})^2 \right\}$$

$$P((Q_s, X_s) = (q_s, x_s)) \text{ signifie ici : } P(Q_s = q_s) \frac{P(X_s \in [x_s, x_s + dx_s] / Q_s = q_s)}{dx_s}$$

Par identification, on en déduit :

$$V_{0,s} = \left(\frac{1}{V_V} + \frac{1}{V_{L0}} \right)^{-1}, \quad m_{0,s} = V_{0,s} \left(\frac{\mu_V}{V_V} + \frac{\mu_{L0}}{V_{L0}} \right)$$

$$V_{1,s} = \left(\frac{1}{V_V} + \frac{1}{V_{L1}} \right)^{-1}, \quad m_{1,s} = V_{1,s} \left(\frac{\mu_V}{V_V} + \frac{\mu_{L1}}{V_{L1}} \right)$$

$$\lambda_s = \left[1 + \frac{1-\lambda_L}{\lambda_L} \sqrt{\frac{V_{0,s} V_{L1}}{V_{1,s} V_{L0}}} \exp \left\{ \frac{1}{2} \left(\frac{m_{0,s}^2}{V_{0,s}} - \frac{m_{1,s}^2}{V_{1,s}} + \frac{\mu_{L1}^2}{V_{L1}} - \frac{\mu_{L0}^2}{V_{L0}} \right) \right\} \right]^{-1}$$

Restent alors à calculer λ_L , μ_{L0} , V_{L0} , μ_{L1} , V_{L1} , μ_V et V_V .

1 CALCUL DE : (μ_V, V_V)

$$\mu_V = \underset{X_s}{\text{ArgMin}} \| Z - f * X \|^2$$

Posons $s=(i, j)$, la prédiction verticale de X_s ne dépend alors que des Y et des Z de la trace j . On est donc ramené à un problème unidimensionnel, les calculs sont faits en annexe du Chapitre 1. En reprenant les notations déjà utilisées:

$$U_{i,k} = Z_{i+k} - \sum_{j \neq k} f_j x_{i+k-j}, \quad E_f = \sum_{j=-p}^{+p} f_j^2,$$

on obtient,

$$\mu_V = \frac{1}{E_f} \sum_{k=-p}^{+p} f_k U_{i,k}, \quad V_V = \frac{\sigma_\varepsilon^2}{E_f}$$

2 CALCUL DE : (μ_{L0}, V_{L0})

Par hypothèse, lorsque $Q_s = 0$, $X_s \sim \mathcal{N}(0, \sigma_0^2)$, donc :

$$\mu_{L0} = 0, \quad V_{L0} = \sigma_0^2$$

3 CALCUL DE : ($\mu_{L1}, V_{L1}, \lambda_L$)

Appelons t et r les sites voisins de s . Nous avons les relations :

$$\lambda_L = P(Q_s = 1 / (X_t, Q_t) = (x_t, q_t), (X_r, Q_r) = (x_r, q_r))$$

$$\lambda_L = \frac{P(X_t=x_t, X_r=x_r / Q_s=1, Q_t=q_t, Q_r=q_r) P(Q_s=1 / Q_t=q_t, Q_r=q_r)}{P(X_t=x_t, X_r=x_r / Q_t=q_t, Q_r=q_r)}$$

Notons p_0 , p_1 et p_2 les probabilités conditionnelles de Q_s :

$$p_0 = P(Q_s=1 / Q_t=0, Q_r=0) = \frac{e^{-\alpha-2\beta}}{1+e^{-\alpha-2\beta}}$$

$$p_1 = P(Q_s=1 / Q_t=1, Q_r=0) = \frac{e^{-\alpha}}{1+e^{-\alpha}}$$

$$p_2 = P(Q_s=1 / Q_t=1, Q_r=1) = \frac{e^{-\alpha+2\beta}}{1+e^{-\alpha+2\beta}}$$

3.a Etude du cas: $q_t = q_r = 0$

Y_s, Y_t et Y_r sont indépendants, on a donc :

$$P(X_t=x_t, X_r=x_r / Q_s=1, Q_t=Q_r=0) = P(X_t=x_t / Q_t=0)P(X_r=x_r / Q_r=0)$$

donc,

$$\lambda_L = P(Q_s=1 / Q_t=0, Q_r=0)$$

De plus, par hypothèse, $X_s \sim \mathcal{N}(0, \sigma_1^2)$, donc :

$$\lambda_L = p_0, \quad \mu_{L1} = 0, \quad v_{L1} = \sigma_1^2$$

3.b Etude du cas : $q_t = 0, q_r = 1$

$$P(X_t=x_t, X_r=x_r / Q_s=1, Q_t=0, Q_r=1) = P(X_t=x_t / Q_t=0) P(X_r=x_r / Q_s=Q_r=1)$$

Or,

$$P(X_r=x_r / Q_s=Q_r=1) = P(X_r=x_r / Q_r=1)$$

Donc,

$$\lambda_L = P(Q_s=1 / Q_t=0, Q_r=1)$$

La représentation Markovienne est donnée par :

$$X_s = \rho X_r + \eta_s, \quad \eta_s \sim \mathcal{N}(0, (1-\rho^2)\sigma_1^2)$$

Donc,

$$\lambda_L = p_1, \quad \mu_{L1} = \rho x_r, \quad v_{L1} = (1-\rho^2)\sigma_1^2$$

3.c Etude du cas: $q_t = 1, q_r = 1$

La loi conditionnelle de (X_t, X_r) est donnée par :

$$P(X_t=x_t, X_r=x_r / Q_s=1, Q_t=Q_r=1) = \frac{1}{\sqrt{(2\pi\sigma_1^2)^2(1-\rho^4)}} \exp\left\{ \frac{-1}{2(1-\rho^4)\sigma_1^2} (x_t^2 + x_r^2 - 2\rho^2 x_t x_r) \right\}$$

$$P(X_t=x_t, X_r=x_r / Q_s=0, Q_t=Q_r=1) = \frac{1}{\sqrt{(2\pi\sigma_1^2)^2}} \exp\left\{ \frac{-1}{2\sigma_1^2} (x_t^2 + x_r^2) \right\}$$

Donc,

$$\lambda_L = Cte \times \frac{p_2}{\sqrt{(1-\rho^4)}} \exp\left\{ \frac{-1}{2(1-\rho^4)\sigma_1^2} (x_t^2 + x_r^2 - 2\rho^2 x_t x_r) \right\}$$

$$1 - \lambda_L = Cte \times (1-p_2) \exp\left\{ \frac{-1}{2\sigma_1^2} (x_t^2 + x_r^2) \right\}$$

D'où :

$$\lambda_L = \left[1 + \frac{1-p_2}{p_2} \sqrt{1-\rho^4} \exp\left\{ \frac{1}{2(1-\rho^4)\sigma_1^2} (\rho^4(x_t^2 + x_r^2) - 2\rho^2 x_t x_r) \right\} \right]^{-1}$$

On a la représentation Markovienne :

$$X_s = \frac{\rho}{1+\rho^2} (X_t + X_r) + e_s, \quad e_s \sim \mathcal{N}\left(0, \frac{1-\rho^2}{1+\rho^2} \sigma_1^2\right)$$

Donc :

$$\mu_{L1} = \frac{\rho}{1+\rho^2} (x_t + x_s), \quad v_{L1} = \frac{1-\rho^2}{1+\rho^2} \sigma_1^2$$

Des calculs effectués ci-dessus, on déduit que $V_{1,s}$ et $V_{0,s}$ sont constantes, ne dépendant pas du site s . On notera alors V_1 et V_0 ces deux variances.

L'algorithme de recuit simulé

Nous voulons trouver la suite $\mathbf{Y} = \{Y_s\}_{s \in S}$ qui maximise la distribution à postériori $P(\mathbf{Y}/\mathbf{Z})$ où $\mathbf{Z} = \{Z_s\}_{s \in S}$ est l'observation.

$$P(\mathbf{Y}=\mathbf{y} / \mathbf{Z}) = Cte \times \exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} \sum (Z_s - f * x_s)^2 - \alpha \sum q_s - \beta \sum 1_{q_s \neq q_t} - \frac{1}{2} \sum (1 - q_s) \left(\frac{x_s^2}{\sigma_0^2} + \text{Log}(2\pi\sigma_0^2) \right) + \sum L(X) \right\}$$

$L(X)$ est la Log-vraisemblance de $X = (X_1, X_2, \dots, X_m)$, les valeurs des coefficients de réflexion sur une ligne de réflecteurs principaux de longueur m (voir 3.1.2.d). On a alors:

$$L(X) = -\frac{m}{2} \text{Log}(2\pi\sigma_1^2) - \frac{m-1}{2} \text{Log}(1-\rho^2) - \frac{1}{2(1-\rho^2)\sigma_1^2} \left[(1-\rho^2)X_1^2 + \sum_{i=1}^{m-1} (X_{i+1} - \rho X_i)^2 \right]$$

La procédure est la suivante :

- On initialise $\mathbf{Y} = (y_1, y_2, \dots)$ et $T = T_0$
- A l'itération k :
 - a) On choisit un site s de S (de manière cyclique ou aléatoire)
 - b) On simule Y_s avec la loi $P_T(Y_s / \mathbf{Z}, Y_t = y_t \ t \neq s)$ définie plus bas. Pour $t \neq s$, les Y_t restent inchangés.

Si tous les sites sont visités un grand nombre de fois et si à chaque itération, on diminue la température T suffisamment lentement, on converge vers le maxima global de $P(\mathbf{Y} / \mathbf{Z})$.

Théoriquement, la température doit être de la forme $T(k) = T_0 / \text{Log}(k)$, dans la pratique, nous avons utilisé une décroissance plus rapide: $T(k) = \alpha T(k-1)$ où α est proche de 1.

La loi à postériori de \mathbf{Y} , $P(\mathbf{Y}/\mathbf{Z})$ peut s'écrire sous la forme :

$$P(\mathbf{Y}=\mathbf{y} / \mathbf{Z}) = \frac{1}{K} \exp\{ -U(\mathbf{y}) \}$$

On note alors

$$P_T(\mathbf{Y}=\mathbf{y} / \mathbf{Z}) = \frac{1}{K(T)} \exp\left\{ -\frac{U(\mathbf{y})}{T} \right\}$$

avec $K(T) = \int \exp\left\{ -\frac{U(\mathbf{y})}{T} \right\} d\mathbf{y}$.

Nous avons vu que :

$$P(Y_s/Z, Y_t=y_t, t \neq s) = \left(\frac{\lambda_s}{\sqrt{2\pi V_1}} \exp\left(\frac{-1}{2V_1} (x_s - m_{1,s})^2\right) \right)^{q_s} \left(\frac{1-\lambda_s}{\sqrt{2\pi V_0}} \exp\left(\frac{-1}{2V_0} (x_s - m_{0,s})^2\right) \right)^{1-q_s}$$

Nous pouvons alors en déduire

$$P_T(Y_s/Z, Y_t=y_t, t \neq s) = \left(\frac{\lambda_{s(T)}}{\sqrt{2\pi V_{1(T)}}} \exp\left(\frac{-1}{2V_{1(T)}} (x_s - m_{1,s})^2\right) \right)^{q_s} \left(\frac{1-\lambda_{s(T)}}{\sqrt{2\pi V_{0(T)}}} \exp\left(\frac{-1}{2V_{0(T)}} (x_s - m_{0,s})^2\right) \right)^{1-q_s}$$

Les calculs donnent

$$V_{1(T)} = TV_1, \quad V_{0(T)} = TV_0$$

$$\lambda_{s(T)} = \left[1 + \left(\sqrt{\frac{V_1}{V_0}} \right)^{\frac{1}{T}-1} \left(\frac{1-\lambda_s}{\lambda_s} \right)^{\frac{1}{T}} \right]^{-1}$$

BIBLIOGRAPHIE

- [1] A.J. Berkout, 1977 *"Least squares inverse filtering and wavelet deconvolution"* Geophysics Vol 42 p 1369-1383.
- [2] G. Celeux, J. Diebolt, 1987 *"A random imputation principle: The stochastic EM Algorithm"* Rapport de Recherche INRIA.
- [3] G. Celeux, J. Diebolt, 1990 *"Une version de type recuit simulé de l'algorithme SEM"* C.R.A.S t.310 Série I p 119-124.
- [4] B. Chalmond, 1989 *"An iterative Gibbsian technique for reconstruction of m-ary images "* Pattern recognition.
- [5] J.F. Clearbout, F. Muir, 1973 *"Robust modeling with erratic data"* Geophysics Vol 38 p 826-844.
- [6] C. Csiszár, G. Tusnády, 1984 *"Information Geometry and Alternating Minimization Procedures"* Statistics & Decisions Supp. Issue N° 1 p 205-237.
- [7] A. Demster, N. Laird, D. Rubin, 1977 *"Maximum likelihood estimation from incomplete data via the EM algorithm "* Journal of the Royal Statistical Society vol 39 p 1-38.
- [8] J.M. Dinten, X. Guyon, J. F. Yao, 1988 *"On the choice of the regularization parameter: the case of binary restoration framework"* Preprint Orsay.
- [9] D. Donoho, 1980 *"On minimum entropy deconvolution"* Proc of 2nd appl. time series symposium, Tulsa, D.F. Findley p 565-608.
- [10] E. Gassiat, 1988 *"Deconvolution aveugle"* Thèse de docteur en Sciences. Université de Paris-Sud.
- [11] D. Geman, S. Geman, 1984 *"Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images"* I.E.E.E. Trans. Pattern Analysis Mch. Intell. Vol 6 p 721-741.
- [12] Y. Goussard, G. Demoment, 1987 *"Recursive deconvolution of Bernoulli-Gaussian processes using a MA representation"* Traitement du Signal Vol 4 n° 5 .
- [13] Y. Goussard, G. Demoment, 1988 *"Détection-Estimation récursive rapide de séquences Bernoulli-Gaussiennes"* Rapport interne LSS n° GPI 88/05.
- [14] J. Goutsias, J.F. Mendel, 1986 *"Maximum-likelihood deconvolution : an optimization theory perspective"* Geophysics Vol 51 n° 6 p 1206-1220.
- [15] X. Guyon, 1985 *"Champs stationnaires sur \mathbb{Z}^2 , Modèles statistiques et simulations"* Preprint Orsay.
- [16] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, 1983 *"Optimization by simulated annealing"* Science 220 p 671-680.
- [17] S.D. Kollias, C.C. Halkias, 1985 *"An instrumental variable approach to minimum-variance seismic deconvolution"* IEEE Trans. Geoscience & Remote Sensing GE-23 p 778-788.
- [18] J. Kormylo, J.F. Mendel, 1982 *"Maximum-likelihood detection and estimation of Bernoulli-Gaussian processes"* IEEE Trans. Inf. Theory IT-28 p 482-488.

- [19] P.J.M. van Laarhoven, E.H.L. Aarts, 1987 *"Simulated annealing : Theory and applications"* Mathematics and its Applications D. Reidel Publishing Company.
- [20] S. Levy, P.K. Fullagar, 1981 *"Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution"* Geophysics Vol 46 n° 9 p 1235-1243.
- [21] A.K. Mahalanabis, S. Prasad, K.P. Mohandas, 1982 *"Recursive decision directed estimation of reflection coefficients for seismic data deconvolution"* Automatica 18 p 721-726.
- [22] J. F. Mendel, 1983 *"Optimal seismic deconvolution"* Academic Press.
- [23] D.W. Oldenburg, 1980 *"A comprehensive solution to the linear deconvolution problem"* Geophys. J. Roy. Astr. Soc. Vol 65 p 331-357.
- [24] D.W. Oldenburg, T. Scheuer, S. Levy 1983 *"Recovery of the acoustic impedance from reflection seismograms"* Geophysics Vol 48 n° 10 p 1318-1337.
- [25] R.A. Redner, H.F. Walker, 1984 *"Mixture densities, maximum likelihood and the EM algorithm"* SIAM Rev. Vol 26 p 195-239.
- [26] M.A. Tanner, W.H. Wong, 1987 *"The Calculation of Posterior Distributions by Data Augmentation"* JASA Theory and Methods Vol 82 N° 398 p 527-551.
- [27] H.L. Taylor, S.C. Banks, J.F. Mc Coy, 1979 *"Deconvolution with the l_1 -norm "* Geophysics Vol 49 p 39-52.
- [28] R. Wiggins, 1978 *"Minimum entropy deconvolution"* Geexplor. Vol 16 p 21-35.
- [29] L.C. Wood, R.C. Heiser, S. Treitel, P.L. Riley, 1978 *"The debubbling of marine source signature"* Geophysics Vol 43 p 715-729.
- [30] C.F. Jeff Wu, 1983 *"On the convergence properties of the EM Algorithm"* The Annals of Statistics Vol 11 N° 1 p 95-103.
- [31] J.F. Yao, 1988 *"Methodes bayésiennes en segmentation d'images"* Bayesian Statistics. Proc. 8th Franco-Belgian Meeting of statisticians (CORE , Université Catholique de Louvain , Belgique)
- [32] L. Younes, 1988 *"Problèmes d'estimation paramétrique pour des champs de Gibbs Markoviens. Applications en traitement d'images"* Thèse de docteur en Sciences. Université de Paris-Sud.

PARTIE II

**DETECTION
DE
RUPTURES**

INTRODUCTION

Dans un grand nombre d'applications, la détection de ruptures est un problème important, en effet, de nombreux systèmes physiques sont soumis aléatoirement à des changements brusques. On cherche bien sûr à détecter ces changements, mais aussi à les localiser et à identifier leur nature. Nous ne considérerons ici que le cas de systèmes uni-indexés (les systèmes indexés par le temps étant les plus courants). On supposera qu'un processus réel, de dimension fini, représente ce système.

Des changements de différentes natures peuvent affecter les caractéristiques du processus : rupture de la moyenne, de la variance, changements dans son spectre, etc . . . Quelque-soit le type de rupture, nous nous plaçons dans un cadre le plus général possible, en supposant le processus multidimensionnel d'une part, et en supposant d'autre part le nombre de changements inconnu, les données doivent alors être traitées séquentiellement.

Le Chapitre 4 est essentiellement consacré à la présentation des techniques de base utilisées pour la détection de ruptures dans une suite de variables, lorsque la distribution de ces variables est paramétrée. C'est par conséquent le paramètre caractérisant cette distribution qui se trouve affecté par les changements (les sauts de moyenne étant l'exemple le plus courant). Une variante de ces algorithmes classiques de détection est proposée lorsque les lois sont inconnues. Cet algorithme a été utilisé pour identifier d'une part, des états de mer stationnaires et détecter d'autre part, des changements de lithologie dans le sous-sol.

Dans le Chapitre 5, on cherche à détecter séquentiellement des ruptures dans le spectre d'un processus multidimensionnel de loi non paramétrique. Les propriétés asymptotiques des estimateurs empiriques des fonctions de distribution spectrale vont nous permettre de construire des statistiques de test asymptotiquement normales et indépendantes pour lesquelles l'algorithme de détection présenté dans le Chapitre 4 est parfaitement bien adapté. Plusieurs simulations sont proposées, offrant de bons résultats aussi bien dans le cas Gaussien que non Gaussien.

CHAPITRE 4

DETECTION SEQUENTIELLE DE RUPTURES DANS LA DISTRIBUTION D'UNE SUITE DE VARIABLES

INTRODUCTION

Nous considérons une suite $\{X(k)\}_{1 \leq k \leq n}$ de variables aléatoires, des changements brusques affectent la distribution de probabilité de ces variables. On suppose que la loi de cette suite est paramétrée par un paramètre θ de dimension fini. C'est alors θ qui se trouve affecté par ces ruptures.

Plusieurs types de problèmes peuvent se présenter, nous allons alors procéder par étapes, partant du problème le plus simple et terminer avec le plus difficile:

Tout d'abord, nous considèrerons le cas d'une rupture unique à détecter dans cette suite de variables, lorsque les lois sont connues puis inconnues. On est alors ramené à un test de rapport de vraisemblance entre deux hypothèses: H_0 "il n'y a pas rupture" et H_1 "il y a rupture".

Nous aborderons ensuite ce problème de manière séquentielle, c'est-à-dire lorsque le nombre de ruptures est inconnu et lorsque celles-ci doivent être détectées au fur et à mesure que les données apparaissent. Nous présentons l'algorithme de détection généralement employé et décrit dans [1], [2], [3], [4], [10], [12], [16]. Comme précédemment, la statistique de test est construite à partir du rapport de vraisemblance entre les hypothèses H_0 et H_1 , une rupture est détectée chaque fois que le détecteur est supérieur à un seuil donné, ce seuil déterminant le retard à la détection et la probabilité de fausse alarme. Lorsque les lois sont inconnues, nous proposons un nouvel algorithme qui utilise un deuxième détecteur afin d'évaluer si les changements de distribution sont significatifs ou non. Ce détecteur est défini comme la distance de Kullback entre les distributions estimées avant et après l'instant de rupture.

Finalement, deux exemples d'application sont proposés. Le premier consiste à détecter des ruptures dans le processus de houle, on cherche ainsi à identifier des états de mer stationnaires. Les données utilisées sont une suite de mesures de hauteurs de vagues, les changements affectent la moyenne de cette suite. Dans le second exemple, nous tâcherons de détecter les changements de lithologies qui existent dans le sous-sol. Les données proviennent d'enregistrements effectués dans un puit et mesurant les caractéristiques du milieu. Là aussi, la moyenne de cette suite de données est affectée par les ruptures.

4.1 DETECTION D'UNE RUPTURE UNIOUE

4.1.1 CAS DE LOIS CONNUES

TEST DE RAPPORT DE VRAISEMBLANCE

On se place ici dans un cadre assez simple: on se donne deux lois connues P_0 et P_1 de densités respectives p_0 et p_1 . On observe une suite de variables $\{X(k)\}_{1 \leq k \leq n}$ où $X(k)$ suit la loi P_0 jusqu'à un instant $r-1$ et la loi P_1 à partir de r . L'instant de rupture r est bien sûr inconnu et on cherche à l'estimer, on est naturellement tenté de chercher l'instant \hat{r} tel que les distributions avant et après cet instant soient les plus différentes possibles.

La traduction mathématique de ce test consiste à utiliser un test de rapport de vraisemblance.

Soient les deux hypothèses:

H_0 : il n'y a pas de rupture, $X(k)$ suit une loi P_0 pour $1 \leq k \leq n$.

H_1 : il y a rupture en $k = t$, $X(k)$ suit une loi P_0 pour $1 \leq k \leq t-1$ et une loi P_1 pour $t \leq k \leq n$.

Soit $V_i^j(P_0)$ (resp. $V_i^j(P_1)$) la vraisemblance de $(X(i), X(i+1), \dots, X(j))$ sous P_0 (resp. P_1) et:

$$R_i^j(P_0, P_1) = \frac{V_i^j(P_1)}{V_i^j(P_0)}$$

On estimera alors l'instant de rupture r par:

$$\hat{r} = \underset{1 \leq t \leq n}{\text{Argmax}} R_t^n(P_0, P_1)$$

et on acceptera l'hypothèse H_1 si $R_{\hat{r}}^n(P_0, P_1) > T$. T est un seuil fixé qui détermine ici les probabilités de fausse alarme et de non détection:

Pour des niveaux donnés α et β , T est tel que:

$$P_0(\text{Max}_{1 \leq t \leq n} R_t^n(P_0, P_1) > T) < \alpha.$$

$$P_1(\text{Max}_{1 \leq t \leq n} R_t^n(P_0, P_1) < T) < \beta.$$

4.1.2 CAS DE LOIS INCONNUES

TEST GENERALISE DE RAPPORT DE VRAISEMBLANCE

Le problème est maintenant plus délicat puisque les lois P_0 et P_1 sont inconnues. On supposera toutefois que ces lois appartiennent à un ensemble \mathcal{P} de lois paramétrées. On notera P_θ un élément de \mathcal{P} paramétré par θ , $\theta \in \Theta$, P_0 et P_1 seront donc notés P_{θ_0} et P_{θ_1} . On supposera de plus que tout élément de \mathcal{P} est absolument continu par rapport à une certaine mesure de référence et a une densité p_θ .

Notre problème consiste toujours à estimer l'instant de rupture r mais aussi les lois P_{θ_0} et P_{θ_1} , c'est-à-dire les paramètres θ_0 et θ_1 .

En notant $R_t^n(\theta_0, \theta_1) = R_t^n(P_{\theta_0}, P_{\theta_1})$, on estimera r , θ_0 et θ_1 par:

$$\hat{r} = \text{Argmax}_{1 \leq t \leq n} R_t^n(\hat{\theta}_0, \hat{\theta}_1)$$

où:

$$\hat{\theta}_0 = \text{Argmax}_{\theta} V_1^{t-1}(\theta) \quad \text{et} \quad \hat{\theta}_1 = \text{Argmax}_{\theta} V_t^n(\theta).$$

La règle de décision est la même que celle définie précédemment: on accepte H_1 si $R_{\hat{r}}^n(\hat{\theta}_0, \hat{\theta}_1) > T$.

4.2 DETECTION SEQUENTIELLE DE RUPTURES

Lorsque le nombre de ruptures est inconnu, le problème se complique car les hypothèses sont multiples. On ne peut pas raisonnablement construire une unique fonctionnelle à maximiser en bloc pour estimer les différents instants de ruptures. Nous allons aborder le problème de façon séquentielle, c'est-à-dire que nous traiterons les données successivement, dans leur ordre d'arrivée. A chaque instant, nous testerons la présence ou non de rupture sans remettre en question les décisions prises avant cet instant.

Nous continuerons d'utiliser $R_t^n(P_0, P_1)$ comme statistique de test, en considérant ce processus comme un processus bi-indexé (t et n). Deux instants nous intéressent maintenant: l'instant où la rupture

s'est produite et celui où elle est détectée. Le temps séparant ces deux instants est appelé retard à la détection et sa minimisation est parfois un problème important (dans le cas de détection de panne, par exemple).

4.2.1 CAS DE LOIS CONNUES

Le test utilisé pour détecter une rupture unique peut être facilement adapté dans ce cas, en utilisant la règle d'arrêt de Page-Hinkley [1], [2], [3], [4], [8], [12], [16]. Pour n fixé, notons:

$$G_n = \text{Max}_{1 \leq t \leq n} R_t^n(P_0, P_1)$$

$$v = \min\{n, G_n > T\}$$

$$\hat{r} = \text{Argmax}_{1 \leq t \leq v} R_t^v(P_0, P_1)$$

v est l'instant où la rupture est détectée et \hat{r} celui où elle est localisée. Lorsque une rupture est détectée à un instant \hat{r} , le passé est oublié et l'algorithme est réinitialisé en considérant \hat{r} comme nouvelle origine du temps.

L'idée de ce test est d'attendre d'avoir un nombre suffisant de données qui suivent le plus vraisemblablement la loi P_1 pour accepter H_1 . Ce seuil de confiance est déterminé par T , ce paramètre fixe également le retard à la détection:

Plus T est grand, plus la probabilité de fausse alarme diminue mais plus le retard à la détection augmente. Le choix de ce paramètre demande alors un compromis entre ces deux critères et dépend du type de problème traité.

Shiryayev a montré que ce test est optimal: pour un temps moyen entre fausses alarmes donné, le retard à la détection est minimum [15].

Willsky se place dans le domaine de l'automatique [16]. Il considère un modèle Gaussien et utilise un filtre de Kalman pour calculer à chaque instant l'erreur de prédiction qui devrait être un bruit blanc Gaussien en l'absence de changement.

4.2.2 CAS DE LOIS INCONNUES

La combinaison des algorithmes proposés en 4.1.2 et 4.2.1 est bien sûr envisageable: pour n fixé, on estime θ_0 , θ_1 et r comme précédemment, puis on fait varier n jusqu'à ce que G_n soit supérieur au seuil T fixé. Lorsque les lois sont connues, la statistique G_n peut être réactualisée pour calculer G_{n+1} ,

malheureusement, ce n'est plus généralement le cas lorsque ces lois sont inconnues (toutefois, dans le cas de modèles Gaussiens, Willsky propose un calcul récursif du rapport de vraisemblance [16]).

Nikiforov étudie également ce problème, et tout particulièrement dans le cas de modèle ARMA [12]. Il considère le cas de lois connues, mais aussi celui de lois inconnues dans deux cas de figure:

- $\theta = \theta^* + \lambda C$ où $\|C\| = 1$ et $\dim(C) = \dim(\theta)$. Les hypothèses H_0 et H_1 sont alors définies par:
 $H_0 : \lambda < 0$ et $H_1 : \lambda \geq 0$.
- Avant la rupture, le paramètre θ est connu: $\theta = \theta_0$. L'hypothèse H_1 est définie par: après la rupture, $(\theta - \theta_0)^t F(\theta_0) (\theta - \theta_0) > \delta$ ($F(\theta)$ est la matrice d'information de Fisher).

Nous proposons une variante de l'algorithme décrit plus haut, où une valeur de plus petit changement détectable intervient et permet de simplifier la procédure tout en restant efficace. Nous allons tout d'abord considérer le cas simple d'un paramètre unidimensionnel:

4.2.2.a Cas : $\theta \in \mathbb{R}$

Supposons que $X(i)$ suit une loi P_θ jusqu'à l'instant $k-1$, le paramètre étant unidimensionnel, deux types de changement sont alors possibles en k : θ croît ou θ décroît.

Soit μ , la plus petite variation du paramètre θ que l'on cherche à détecter. A l'instant k , les hypothèses à tester sont donc:

$$H_0 : X(k) \text{ suit une loi } P_\theta.$$

$$H_1^+ : X(k) \text{ suit une loi } P_{\theta^+}, \theta^+ > \theta + \mu.$$

$$H_1^- : X(k) \text{ suit une loi } P_{\theta^-}, \theta^- < \theta - \mu.$$

On peut alors utiliser deux détecteurs simultanément, un premier testera H_1^+ contre H_0 et l'autre H_1^- contre H_0 .

Si $\hat{\theta}_k$ est un estimateur de θ , calculé comme fonction de $X(1), X(2), \dots, X(k-1)$, on est dans le cas de loi connues en posant d'une part $\theta_0 = \hat{\theta}_k$ et d'autre part $\theta_1 = \hat{\theta}_k + \mu$ (resp $\hat{\theta}_k - \mu$) pour tester H_1^+ (resp H_1^-) contre H_0 .

Nous verrons plus loin, dans le cadre de deux applications, la mise en place d'un tel algorithme pour détecter des ruptures dans la moyenne d'un processus Gaussien unidimensionnel.

4.2.2.b Cas : $\theta \in \mathbb{R}^p$

Dès que p est différent de 1, les types de ruptures possibles sont nombreux puisque les directions que peut prendre le paramètre θ sont diverses. Il n'est alors pas raisonnable de mettre en place l'algorithme présenté ci-dessus dans le cas $\theta \in \mathbb{R}$ pour se retrouver dans un cadre de détection de ruptures lorsque les lois sont connues. A chaque instant, la direction prise par θ sera d'une part estimée, on testera d'autre part si le saut est significatif, mais pour cela, une distance entre paramètres doit être définie. L'ensemble des paramètres étant en bijection avec un ensemble de distributions, nous proposons ici l'utilisation de la distance de Kullback pour décider si ce changement est significatif.

Algorithme

Soient θ_k^- et θ_k^+ de Θ tels que $X^{(k-1)}$ suive la loi $P_{\theta_k^-}$ et $X^{(k+1)}$ la loi $P_{\theta_k^+}$.

L'algorithme consistera alors:

- * A tester si θ_k^- et θ_k^+ sont différents.
- * A décider si $X^{(k)} \sim P_{\theta_k^-}$ ou si $X^{(k)} \sim P_{\theta_k^+}$.

... , θ_k^- et θ_k^+ sont inconnus et doivent être estimés. Soient $(\hat{\theta}_k^-, \hat{\theta}_k^+)$ un estimateur de (θ_k^-, θ_k^+) :

$\hat{\theta}_k^-$ est un estimateur du passé, et dépend de $(X^{(k-1)}, X^{(k-2)}, \dots)$

$\hat{\theta}_k^+$ est un estimateur du futur, et dépend de $(X^{(k+1)}, X^{(k+2)}, \dots)$

Soit $K(k) = K(\hat{\theta}_k^-, \hat{\theta}_k^+)$ la distance de Kullback entre $P_{\hat{\theta}_k^-}$ et $P_{\hat{\theta}_k^+}$:

$$K(\hat{\theta}_k^-, \hat{\theta}_k^+) = \int_{\Omega} \text{Log} \left[\frac{dP_{\hat{\theta}_k^-(x)}}{dP_{\hat{\theta}_k^+(x)}} \right] dP_{\hat{\theta}_k^-(x)}$$

Soit $S_k^m = (X^{(k)}, X^{(k+1)}, \dots, X^{(k+m)})$, nous noterons $V_{\theta}(S_k^m)$ la vraisemblance de S_k^m sous P_{θ} et $R(S_k^m)$ le rapport de vraisemblance défini par:

$$R(S_k^m) = \frac{L_{\hat{\theta}_k^+}(S_k^m)}{L_{\hat{\theta}_k^-}(S_k^m)}$$

On décidera au temps $k+j$ qu'il y a rupture à l'instant k si les 2 conditions suivantes sont vérifiées simultanément:

$$c_1 : \text{pour tout } i, 0 \leq i \leq j, \quad K(\hat{\theta}_{k+i}^-, \hat{\theta}_{k+i}^+) > \delta$$

$$c_2 : \text{pour tout } i, 0 \leq i < j, \quad 0 < R(S_k^i) < T \text{ et } R(S_k^j) > T$$

δ et T sont 2 seuils fixés arbitrairement:

- δ représente le plus petit changement que l'on cherche à détecter.
- Comme précédemment, T détermine la probabilité de fausse alarme et d'oubli ainsi que le retard à la détection.

Cas de variables Gaussiennes indépendantes

Si les variables sont indépendantes, nous aurons $R(S_k^m) = \prod_{j=0}^m R(X(k+j))$, le détecteur utilisé est donc le suivant:

$$Z(k) = [Z(k-1) + \text{Log } R(X(k))]^+$$

Une rupture est détectée lorsque simultanément $K > \delta$ et $Z > \text{Log}(T)$. Elle est localisée après le dernier instant où $Z = 0$.

Dans le cas Gaussien, $\theta_0 = (m_0, \Gamma_0)$ et $\theta_1 = (m_1, \Gamma_1)$, nous avons les relations:

$$\text{Log } R(X(k)) = \text{Log} \frac{L_{\theta_1}(X(k))}{L_{\theta_0}(X(k))} \quad \text{et} \quad K(\theta_0, \theta_1) = -E_{\theta_0} \text{Log } R(X(k)).$$

On obtient alors:

$$\text{Log } R(X(k)) = \frac{1}{2} \left[\text{Log det}(\Gamma_0 \Gamma_1^{-1}) + (X(k) - m_0)^t \Gamma_0^{-1} (X(k) - m_0) - (X(k) - m_1)^t \Gamma_1^{-1} (X(k) - m_1) \right]$$

$$K(\theta_0, \theta_1) = -\frac{1}{2} \left[\text{Log det}(\Gamma_0 \Gamma_1^{-1}) + p - \text{tr}(\Gamma_0 \Gamma_1^{-1}) - (m_0 - m_1)^t \Gamma_1^{-1} (m_0 - m_1) \right]$$

On vérifie que $K(\theta_0, \theta_1) \geq 0$ avec $K(\theta_0, \theta_1) = 0$ si et seulement si $\theta_0 = \theta_1$.

Estimation des paramètres

Soit $X(1)$ la première observation après la dernière rupture détectée et k tel que $Z(k-1) = 0$. Cela signifie donc que $X(1), X(2) \dots X(k-1)$ sont supposés avoir la même distribution.

$\hat{\theta}_k^+$ est une fonction de $(X(k+1), X(k+2) \dots X(k+h))$ où h est un entier donné.

$\hat{\theta}_k^-$ est une fonction de $(X(1), X(2) \dots X(k-1))$

* Si $Z(k) > 0$, $\hat{\theta}_{k+1}^- = \hat{\theta}_k^-$

* Si $Z(k) = 0$, on calcule un nouvel estimateur $\hat{\theta}_{k+1}^-$ comme fonction de $(X(1), \dots X(k))$.

Dans le cas Gaussien les paramètres à estimer sont la moyenne et la matrice de covariance, on utilise les estimateurs empiriques usuels.

4.3 APPLICATIONS

4.3.1 IDENTIFICATION D'ETATS DE MER

Notre objectif a été d'identifier des états de mer considérés comme stationnaires, c'est-à-dire des intervalles de temps durant lesquels les caractéristiques de la houle (hauteur et période des vagues, vent, etc . . .) restent constantes.

La hauteur significative $H_{1/3}$ représente la moyenne du tiers supérieur des vagues et peut être considérée comme un résumé exhaustif du phénomène. Nous disposons comme données de l'estimation de ce paramètre chaque 20mn en un même endroit. Nous allons détecter des ruptures dans la moyenne de cette suite de données, chaque instant de rupture déterminant les débuts et fins d'états de mer stationnaires.

Nous considérons les données comme Gaussiennes et indépendantes, le détecteur est alors très simple:

$$Z(k) = [Z(k-1) + \text{Log } R(X(k))]^+$$

où,

$$R(X(k)) = \exp\left\{ \frac{\mu}{\sigma_2} \left(X(k) - m - \frac{\mu}{2} \right) \right\}$$

m et σ_2 sont les moyenne et variance de X avant l'instant k et sont estimés empiriquement à partir de $X(1), X(2) \dots X(k-1)$. On a supposé le changement de variance négligeable devant celui de moyenne.

μ est un saut minimum de moyenne que l'on cherche à détecter (positif ou négatif), ce paramètre fixe une échelle au problème. Plus μ est petit et meilleure est la résolution, mais on risque alors d'obtenir un découpage trop systématique de nos données.

Nous présentons en Figure 1 plusieurs résultats obtenus avec des données provenant de Mer du Nord, ils correspondent respectivement à 24 et 48 heures d'enregistrement au mois de Janvier 1984, les mesures de $H_{1/3}$ sont effectuées toutes les 20mn. Sont représentés sur les figures:

- Le processus $H_{1/3}$ comme fonction du temps.
- Les instants de ruptures (barres verticales).
- La moyenne estimée du processus $H_{1/3}$ entre deux ruptures.

(Ce travail a été réalisé en collaboration avec E. Rio à IFREMER (France) sous la direction de J. Labeyrie [10].)

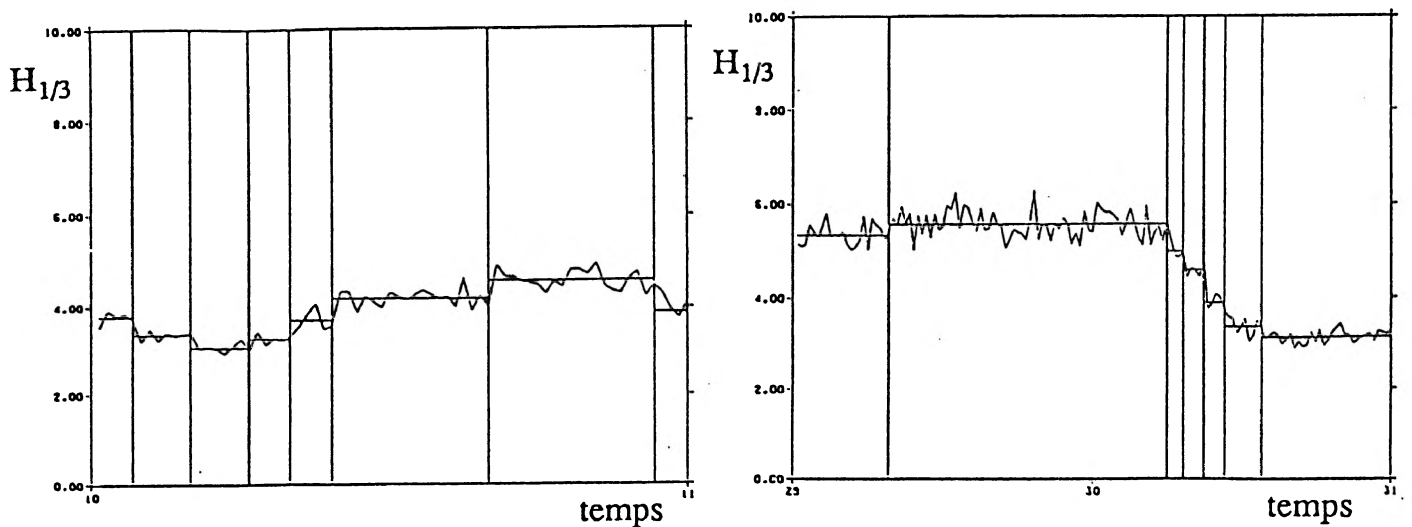


Fig.1 Identification d'états de mer stationnaires

4.3.2 CHANGEMENTS DE LITHOLOGIE DANS LE SOUS-SOL

Nous disposons ici de plusieurs enregistrements effectués dans un puit et mesurant la radio-activité du milieu, sa résistivité, son impédance . . . Sur une portion de ce puit, nous disposons également d'un carottage: on a extrait du puit un échantillon des matériaux, cela permet de savoir le type de lithologie en présence et où se trouvent réellement les changements, on peut ainsi contrôler les résultats obtenus avec notre algorithme.

La présence d'une suite de données multidimensionnelles nous a incité à mettre en place l'algorithme de détection séquentielle de ruptures présenté au paragraphe 4.2.2.b. Toutefois, en comparant les résultats obtenus avec le carottage, il s'est avéré que l'enregistrement mesurant la radio-activité suffisait pour obtenir la presque totalité des changements, les autres enregistrements n'apportant aucune amélioration. Nous sommes donc ramené une nouvelle fois au problème de détection de ruptures dans la moyenne d'une suite de variables unidimensionnelles et l'algorithme utilisé ci-dessus pour identifier des états de mer stationnaires peut être utilisé de manière identique.

Nous présentons en Figure 2 les résultats obtenus, les données ont été fournies par l'INTEVEP (Vénézuéla).

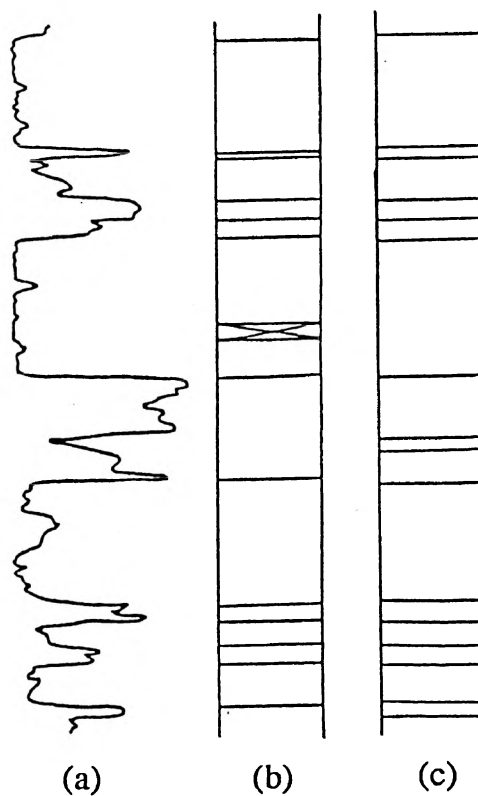


Fig.2 Détection de changements de lithologie dans le sous-sol

- (a) Enregistrement γ -ray (b) carottage (X : données manquantes)
 (c) ruptures détectées

CHAPITRE 5

DETECTION SEQUENTIELLE DE RUPTURES DANS LE SPECTRE D'UN PROCESSUS MULTIDIMENSIONNEL

INTRODUCTION ET NOTATIONS

INTRODUCTION

Nous cherchons ici à détecter des ruptures dans le spectre d'un processus multidimensionnel. Nous supposons que ce processus (dans \mathbb{R}^P) est stationnaire, Gaussien et centré.

Les statistiques que nous allons utiliser pour détecter ces ruptures sont construites à partir des estimateurs empiriques des fonctions spectrales.

Nous allons tout d'abord étudier la convergence faible du processus empirique vers un processus Gaussien, à accroissements indépendants dans l'espace du temps et l'espace des fréquences (partie 1). Nous calculons la fonction de covariance de ce processus limite. Ceci est une généralisation de résultats similaires obtenus par Ibragimov [9], Mc Neill [11] et Picard [13].

Dans la partie 2, nous montrons que certaines statistiques ont de bonnes propriétés asymptotiques, en effet, les variables limites sont Gaussiennes et indépendantes.

Leur utilisation transforme alors le problème original en un problème de détection de ruptures dans la moyenne et la matrice de covariance d'une suite de variables aléatoires Gaussiennes indépendantes multidimensionnelles. Nous pouvons donc utiliser l'algorithme décrit dans le Chapitre 4.

Plusieurs simulations sont proposées dans la troisième partie. Des Monte-Carlo sont effectués pour évaluer le nombre de bonnes et mauvaises détections suivant les seuils utilisés. Nous présentons des résultats obtenus avec des processus Gaussiens et non Gaussiens, en effet, les résultats obtenus dans la

partie 1 peuvent être étendus à des processus mélangeants, utilisant des techniques développées par Dahlaus [6], Doukhan et León [7] ou Rosenblatt [14].

NOTATIONS

Soit $X^i = \{X^i(k)\}_{k \in \mathbb{N}}$, $1 \leq i \leq p$, un processus Gaussien stationnaire centré ayant une mesure spectrale absolument continue respectivement à la mesure de Lebesgue. On considère le processus p -dimensionnel $X = (X^1, X^2, \dots, X^p)$, nous adopterons les notations suivantes :

$f^{ii}(\lambda)$ est la densité spectrale de X^i .

$f^{ij}(\lambda)$ est la densité spectrale croisée entre X^i et X^j .

$F^{ij}(\lambda) = \int_0^\lambda f^{ij}(\alpha) d\alpha$ est la distribution spectrale croisée associée.

$R^{ij}(k) = E X^i(n) X^j(n+k)$ est la fonction de covariance entre X^i et X^j .

Nous avons alors les relations suivantes:

$$f^{ij}(\lambda) = \sum_{-\infty}^{+\infty} R^{ij}(k) e^{-i\lambda k} \quad , \quad R^{ij}(k) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f^{ij}(\alpha) e^{i\alpha k} d\alpha$$

$$R^{ij}(-k) = R^{ji}(k) \quad \text{et} \quad f^{ij}(-\lambda) = f^{ji}(\lambda) = \overline{f^{ij}(\lambda)} \quad .$$

Soit $\{X^i(k)\}_{1 \leq k \leq n}$ une observation de longueur n du processus X , nous pouvons estimer la densité spectrale au moyen du périodogramme:

$$I_n^{ij}(\lambda) = \frac{1}{2\pi n} \left(\sum_{k=1}^n X^i(k) e^{i\lambda k} \right) \left(\sum_{l=1}^n X^j(l) e^{-i\lambda l} \right)$$

et déduire des estimateurs des distributions spectrales:

$$F_n^{ij}(\lambda) = \int_0^\lambda I_n^{ij}(\alpha) d\alpha$$

(On peut remarquer que $F_n^{ij}(-\lambda) = F_n^{ji}(\lambda) = \overline{F_n^{ij}(\lambda)}$)

5.1 CONVERGENCE DES FONCTIONS SPECTRALES EMPIRIQUES

Soit $\mathcal{C}([0, \pi] \times [0, 1], \mathbb{C})$ l'espace des fonctions continues de $[0, \pi] \times [0, 1]$ dans \mathbb{C} . On considère le processus $\xi_n^{ij}(\lambda, t)$ de $\mathcal{C}([0, \pi] \times [0, 1], \mathbb{C})$ défini par:

$$\text{- Pour } k \leq n, \quad \xi_n^{ij}(\lambda, \frac{k}{n}) = \frac{k}{\sqrt{n}} (F_k^{ij}(\lambda) - F^{ij}(\lambda))$$

$$\text{- Pour tout } t \text{ de } [0, 1] \text{ tel que } \frac{k}{n} < t < \frac{k+1}{n}, \quad \xi_n^{ij}(\lambda, t) \text{ est l'interpolation polygonale de } \xi_n^{ij}(\lambda, \frac{k}{n}).$$

Pour montrer la convergence de $\xi_n^{ij}(\lambda, t)$ et calculer la fonction de covariance du processus limite, nous devons d'abord définir un ensemble de conditions:

CONDITION C1

$$\text{Pour tout couple } (i, j), \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n |k R^{ij}(k)| = 0$$

CONDITION C2

$$\text{Pour un certain } \delta > 0 \text{ et pour tout } i, \quad f^{ii}(\lambda) \in L^{2+\delta}[-\pi, \pi].$$

CONDITION C3

Pour tout i , $X^i = \{X^i(k)\}_{k \in \mathbb{N}}$ a la représentation suivante:

$$X^i(k) = \sum_{l=-\infty}^{+\infty} a^i(l) \varepsilon^i(k-l), \quad \text{où } \sum_{l=-\infty}^{+\infty} (a^i(l))^2 < +\infty.$$

Les suites élémentaires $\{\varepsilon^i(k)\}$ satisfont:

$$\text{i) Pour tout } k, \quad \varepsilon^i(k) \sim \mathcal{N}(0, 1)$$

$$\text{ii) Il existe un entier } M \text{ tel que pour tout } (i, j) \text{ et } |m| > M, \quad E \varepsilon^i(k) \varepsilon^j(k+m) = 0.$$

5.1.1 THEOREME

Sous les conditions C2 et C3, $\xi_n^{ij}(\lambda, t)$ converge faiblement vers un processus Gaussien centré $\xi^{ij}(\dots)$ de $\mathcal{C}([0, \pi] \times [0, 1], \mathbb{C})$.

Sous la condition C1, la covariance de $\xi_n^{ij}(\lambda, t)$ est donnée par:

$$E \xi_n^{ij}(\lambda, t) \overline{\xi_n^{gh}(\mu, s)} = 2(t \wedge s) \int_0^{\lambda \wedge \mu} f^{ig}(\alpha) \overline{f^{jh}(\alpha)} d\alpha,$$

($x \wedge y$ est le minimum de x et y).

Remarque: Ce théorème regroupe des résultats de [9], [13] et [11].

I.A. Ibragimov montre dans le cas unidimensionnel la convergence faible du processus $\sqrt{n} (F_n^{ii}(\lambda) - F^{ii}(\lambda))$ vers un processus $\xi^{ii}(\lambda, 1)$ à accroissements indépendants en fréquence et de fonction de covariance:

$$E [\xi^{ii}(\lambda, 1) \xi^{ii}(\mu, 1)] = 2 \int_0^{\lambda \wedge \mu} f^2(\alpha) d\alpha$$

D. Picard montre, également dans le cas unidimensionnel, que le processus limite est aussi à accroissements indépendants dans l'espace du temps:

$$E [\xi^{ii}(\lambda, t) \xi^{ii}(\lambda, s)] = 2(t \wedge s) \int_0^{\lambda} f^2(\alpha) d\alpha$$

I.B. Mc Neill étudie le cas multidimensionnel et montre la convergence du processus $\sqrt{n} (F_n^{ij}(\lambda) - F^{ij}(\lambda))$ vers un processus Gaussien centré $\xi^{ij}(\lambda, 1)$ de covariance:

$$E \xi^{ij}(\lambda, 1) \overline{\xi^{ij}(\mu, 1)} = 2 \int_0^{\lambda \wedge \mu} f^{ii}(\alpha) \overline{f^{jj}(\alpha)} d\alpha$$

5.1.2 DÉMONSTRATION DU THÉOREME

5.1.2.a Convergence

La convergence faible du processus $\xi^{ij}(\lambda, t)$ est une conséquence directe des travaux de [11] et [13]. En effet, sous les conditions C2 et C3, McNeill montre que $\sqrt{k}(F_k^{ij}(\lambda) - F^{ij}(\lambda))$ converge vers un processus Gaussien centré $\zeta(\lambda)$. Maintenant, nous avons la relation:

$$\xi_n^{ij}(\lambda, \frac{k}{n}) = \sqrt{\frac{k}{n}} [\sqrt{k} (F_k^{ij}(\lambda) - F^{ij}(\lambda))],$$

donc $\xi_n^{ij}(\lambda, t)$ converge vers $\sqrt{t} \zeta(\lambda)$.

La tension du processus est une extension au cas multidimensionnel du résultat montré par Picard ([13], p 857-858).

5.1.2.b Identification du processus limite

Pour compléter la matrice de covariance, nous devons calculer les covariances entre ξ^{ij} et ξ^{gh} pour tout (i, j, g, h) de $\{1, 2, \dots, p\}$.

Remarque préliminaire

Dans [9], [13] et [11], il est montré que le biais $E F_n^{ij}(\lambda) - F^{ij}(\lambda)$ converge vers 0. Donc asymptotiquement, $F_n^{ij}(\lambda) - F^{ij}(\lambda)$ peut être remplacé par $F_n^{ij}(\lambda) - E F_n^{ij}(\lambda)$.

Nous noterons donc désormais $\xi_n^{ij}(\lambda, \frac{k}{n}) = \frac{k}{\sqrt{n}} (F_k^{ij}(\lambda) - E F_n^{ij}(\lambda))$

Accroissements indépendants dans l'espace du temps

Soit t dans $[0, 1]$, nous choisissons $k = k(n)$ tel que $\frac{k}{n} \leq t < \frac{k+1}{n}$, donc $t = \lim_{n \rightarrow \infty} \frac{k}{n}$.

Nous allons montrer que pour tout $m > 0$,

$$\lim_{n \rightarrow \infty} E \xi_n^{ij}(\lambda, \frac{k}{n}) \overline{\xi_n^{gh}(\mu, \frac{k+m}{n})} = 2t \int_0^{\lambda \wedge \mu} f^{ig}(\alpha) \overline{f^{jh}(\alpha)} d\alpha.$$

Soit $T = E \xi_n^{ij}(\lambda, \frac{k}{n}) \overline{\xi_n^{gh}(\mu, \frac{k+m}{n})}$, on a alors:

$$T = \frac{k(k+m)}{n} \left[E(F_k^{ij}(\lambda) \overline{F_{k+m}^{gh}(\mu)}) - E F_k^{ij}(\lambda) E \overline{F_{k+m}^{gh}(\mu)} \right]$$

$$T = \frac{k(k+m)}{n} \int_0^\lambda \int_0^\mu \left[E(I_k^{ij}(\alpha_1) \overline{I_{k+m}^{gh}(\alpha_2)}) - E I_k^{ij}(\alpha_1) E \overline{I_{k+m}^{gh}(\alpha_2)} \right] d\alpha_1 d\alpha_2.$$

Soit $A_k^i(\alpha_1) = \sum_{a=1}^k X^i(a) e^{ia\alpha_1}$,

$$T = \frac{1}{4\pi^2 n} \int_0^\lambda \int_0^\mu \left[E[A_k^i(\alpha_1) A_k^j(-\alpha_1) A_{k+m}^g(-\alpha_2) A_{k+m}^h(\alpha_2)] - E[A_k^i(\alpha_1) A_k^j(-\alpha_1)] E[A_{k+m}^g(-\alpha_2) A_{k+m}^h(\alpha_2)] \right] d\alpha_1 d\alpha_2$$

Comme les A_k^i sont Gaussiens,

$$T = \frac{1}{4\pi^2 n} \int_0^\lambda \int_0^\mu \left[E[A_k^i(\alpha_1) A_{k+m}^g(-\alpha_2)] E[A_k^j(-\alpha_1) A_{k+m}^h(\alpha_2)] + E[A_k^i(\alpha_1) A_{k+m}^h(\alpha_2)] E[A_k^j(-\alpha_1) A_{k+m}^g(-\alpha_2)] \right] d\alpha_1 d\alpha_2$$

Nous avons la décomposition $A_{k+m}^i(\alpha_1) = A_k^i(\alpha_1) + \tilde{A}_{k,m}^i(\alpha_1)$ où :

$$\tilde{A}_{k,m}^i(\alpha_1) = \sum_{a=k+1}^{k+m} X^i(a) e^{ia\alpha_1}.$$

Soit $\tilde{T} = T - E \xi^{ij}(\lambda, \frac{k}{n}) \overline{\xi^{gh}(\mu, \frac{k}{n})}$, pour montrer que \tilde{T} converge vers 0, il est suffisant de montrer que :

$$\lim_{n \rightarrow \infty} \frac{1}{4\pi^2 n} \int_0^\lambda \int_0^\mu E[A_k^i(\alpha_1) \tilde{A}_{k,m}^j(\alpha_2)] d\alpha_1 d\alpha_2 = 0.$$

Les inégalités suivantes permettent de conduire au résultat voulu :

$$\begin{aligned} \frac{1}{n} |E[A_k^i(\alpha_1) \tilde{A}_{k,m}^j(\alpha_2)]| &\leq \frac{1}{n} \sum_{a_1=1}^k \sum_{a_2=k+1}^{k+m} |E X^i(a_1) X^j(a_2)| \\ &\leq \frac{m}{n} \sum_{a=m}^k |R^{ij}(a)| + \frac{1}{n} \sum_{a=1}^m |a R^{ij}(a)| + \frac{1}{n} \sum_{a=1}^m |a R^{ij}(k+m-a)|. \end{aligned}$$

La condition C1 est suffisante pour que les 3 termes de cette somme convergent vers 0, donc:

$$\lim_{n \rightarrow \infty} E \xi_n^{ij}(\lambda, \frac{k}{n}) \overline{\xi_n^{gh}(\mu, \frac{k+m}{n})} = \lim_{n \rightarrow \infty} E \xi_n^{ij}(\lambda, \frac{k}{n}) \overline{\xi_n^{gh}(\mu, \frac{k}{n})}.$$

Accroissements indépendants dans l'espace des fréquences

Nous avons à présent:

$$\lim_{n \rightarrow \infty} T = \lim_{n \rightarrow \infty} \frac{1}{4\pi^2 n} \int_0^\lambda \int_0^\mu \left[E[A_k^i(\alpha_1)A_k^g(-\alpha_2)] E[A_k^j(-\alpha_1)A_k^h(\alpha_2)] + E[A_k^i(\alpha_1)A_k^h(\alpha_2)] E[A_k^j(-\alpha_1)A_k^g(-\alpha_2)] \right] d\alpha_1 d\alpha_2$$

Calculons:

$$A_k^i(\alpha_1)A_k^j(\alpha_2) = \sum_{a_1=1}^k \sum_{a_2=1}^k E X_{(a_1)}^i X_{(a_2)}^j e^{ia_1\alpha_1} e^{ia_2\alpha_2}$$

$$A_k^i(\alpha_1)A_k^j(\alpha_2) = \int_{-\pi}^{\pi} \left(\sum_{a_1=1}^k e^{ia_1(\alpha_1-\gamma)} \right) \left(\sum_{a_2=1}^k e^{ia_2(\alpha_2-\gamma)} \right) f^{ij}(\gamma) d\gamma$$

En posant:

$$e_1(\gamma, \alpha_1, \alpha_2) = \frac{\sin \frac{k}{2}(\alpha_1-\gamma) \sin \frac{k}{2}(\alpha_2-\gamma)}{\sin \frac{\alpha_1-\gamma}{2} \sin \frac{\alpha_2-\gamma}{2}}, \quad e_2(\gamma, \alpha_1, \alpha_2) = \frac{\sin \frac{k}{2}(\alpha_1-\gamma) \sin \frac{k}{2}(\alpha_2+\gamma)}{\sin \frac{\alpha_1-\gamma}{2} \sin \frac{\alpha_2+\gamma}{2}},$$

nous pouvons écrire $A_k^i(\alpha_1)A_k^j(\alpha_2) = \exp\{i \frac{k+1}{2}(\alpha_1+\alpha_2)\} \int_{-\pi}^{\pi} e_2(\gamma, \alpha_1, \alpha_2) f^{ij}(\gamma) d\gamma$, et,

$$E \xi_n^{ij}(\lambda, \frac{k}{n}) \overline{\xi_n^{gh}(\mu, \frac{k}{n})} = \frac{1}{4\pi^2 n} \int_0^\lambda \int_0^\mu \left[\int_{-\pi}^{\pi} f^{ig}(\beta) e_1(\beta, \alpha_1, \alpha_2) d\beta \int_{-\pi}^{\pi} \overline{f^{jh}(\gamma)} e_1(\gamma, \alpha_1, \alpha_2) d\gamma + \int_{-\pi}^{\pi} f^{ih}(\beta) e_2(\beta, \alpha_1, \alpha_2) d\beta \int_{-\pi}^{\pi} \overline{f^{jg}(\gamma)} e_2(\gamma, \alpha_1, \alpha_2) d\gamma \right] d\alpha_1 d\alpha_2$$

L'expression ci-dessus peut être réécrite:

$$E \xi_n^{ij}(\lambda, \frac{k}{n}) \overline{\xi_n^{gh}(\mu, \frac{k}{n})} = \frac{k}{n} \left(\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f^{ig}(\beta) \overline{f^{jh}(\gamma)} \left[\Psi_k^{(1)}(\beta, \gamma) + \Psi_k^{(2)}(\beta, \gamma) \right] d\beta d\gamma + \right. \\ \left. + \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f^{ih}(\beta) \overline{f^{jg}(\gamma)} \left[\Psi_k^{(3)}(\beta, \gamma) + \Psi_k^{(4)}(\beta, \gamma) \right] d\beta d\gamma \right)$$

Où, pour $\lambda \leq \mu$:

$$\Psi_k^{(1)}(\beta, \gamma) = \frac{1}{4\pi^2 k} \int_0^{\lambda} \int_0^{\lambda} e_1(\beta, \alpha_1, \alpha_2) e_1(\gamma, \alpha_1, \alpha_2) d\alpha_1 d\alpha_2.$$

$$\Psi_k^{(2)}(\beta, \gamma) = \frac{1}{4\pi^2 k} \int_0^{\lambda} \int_{\lambda}^{\mu} e_1(\beta, \alpha_1, \alpha_2) e_1(\gamma, \alpha_1, \alpha_2) d\alpha_1 d\alpha_2.$$

$$\Psi_k^{(3)}(\beta, \gamma) = \frac{1}{4\pi^2 k} \int_0^{\lambda} \int_0^{\lambda} e_2(\beta, \alpha_1, \alpha_2) e_2(\gamma, \alpha_1, \alpha_2) d\alpha_1 d\alpha_2.$$

$$\Psi_k^{(4)}(\beta, \gamma) = \frac{1}{4\pi^2 k} \int_0^{\lambda} \int_{\lambda}^{\mu} e_2(\beta, \alpha_1, \alpha_2) e_2(\gamma, \alpha_1, \alpha_2) d\alpha_1 d\alpha_2.$$

Dans [9], Ibragimov étudie les noyaux $\Psi_k^{(a)}$, $a=1,2,3,4$ et montre que pour toutes fonctions u et v de $L^2[-\pi, \pi]$:

$$\lim_{k \rightarrow \infty} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} u(\beta) v(\gamma) \Psi_k^{(a)}(\beta, \gamma) d\beta d\gamma = \begin{cases} 2 \int_0^{\lambda} u(\alpha) v(\alpha) d\alpha & \text{si } a = 1 \\ 0 & \text{sinon} \end{cases}$$

$$\text{Finalement, } \lim_{n \rightarrow \infty} E \xi_n^{ij}(\lambda, \frac{k}{n}) \overline{\xi_n^{gh}(\mu, \frac{k}{n})} = 2t \int_0^{\lambda} f^{ig}(\alpha) \overline{f^{jh}(\alpha)} d\alpha.$$

5.2 STATISTIQUES DE TEST

Dans [13], D. Picard considère une rupture unique dans un processus unidimensionnel, notre problème diffère en deux points importants:

- * Le processus est multidimensionnel.
- * Le nombre de ruptures est inconnu.

Le second point nous a conduit à construire un algorithme séquentiel de détection de ruptures, où à chaque instant, nous devons décider si il y a rupture ou non dans le spectre du processus.

Le premier problème consiste à choisir une bonne statistique de test.

5.2.1 CAS UNIDIMENSIONNEL

Pour des raisons de simplicité, nous allons d'abord considérer un processus unidimensionnel $X = \{X(t)\}_{1 \leq t \leq n}$. L'estimateur empirique de la distribution spectrale est alors:

$$F_n(\lambda) = \int_0^\lambda I_n(\alpha) d\alpha.$$

$\sqrt{n} (F_n(\lambda) - EF(\lambda))$ converge vers un processus Gaussien $\zeta(\lambda)$.

Nous avons $F_n(\lambda) = \frac{1}{n} \sum_{t=1}^n Y(\lambda, t)$ où,

$$Y(\lambda, t) = \lambda X^2(t) + 2X(t) \sum_{s=1}^{t-1} \frac{\sin(\lambda s)}{s} X(t-s)$$

$F_n(\lambda)$ est la moyenne empirique de n variables aléatoires, on pourrait alors penser utiliser Y comme statistique de test, mais ces variables ne sont pas équidistribuées et ne sont pas indépendantes. Même sous l'hypothèse nulle (en l'absence de ruptures), leur distribution est particulièrement difficile à calculer.

Nous allons maintenant voir comment utiliser les propriétés asymptotiques de $F_n(\lambda)$, montrées dans la 1^{ère} partie, pour construire une suite de variables aléatoires asymptotiquement équidistribuées, normales et indépendantes. Ces variables nous serviront comme statistiques de test.

5.2.1.a Equidistribution

Soit $m_1(n)$ une suite de \mathbb{N} telle que $m_1(n) \rightarrow +\infty$ et $\frac{m_1(n)}{n} \rightarrow 0$. Nous allons alors définir $\tilde{Y}(\lambda, t)$, une version tronquée de $Y(\lambda, t)$:

$$\tilde{Y}(\lambda, t) = \lambda X^2(t) + 2X(t) \sum_{s=1}^{m_1(n)} \frac{\sin(\lambda s)}{s} X(t-s) \quad , \quad t > m_1(n)$$

En remarquant que pour un λ donné, les $\tilde{Y}(\lambda, i)$ ont tous la même distribution, on définit \tilde{F}_n comme la troncature de F_n :

$$\tilde{F}_n(\lambda) = \frac{1}{n-m_1(n)} \sum_{m_1(n)+1}^n \tilde{Y}(\lambda, t).$$

Nous avons $\lim_{n \rightarrow \infty} \tilde{F}_n(\lambda) = \lim_{n \rightarrow \infty} F_n(\lambda)$, donc $\sqrt{n}(\tilde{F}_n(\lambda) - EF(\lambda))$ converge vers une copie de $\zeta(\lambda)$.

5.2.1.b Indépendance dans l'espace du temps

La convergence de $\xi_n^{ii}(\lambda, \frac{k}{n})$ vers un processus $\xi^{ii}(\cdot, \cdot)$ à accroissements indépendants dans le temps nous permet maintenant de construire une suite de variables asymptotiquement indépendantes :

Soit $m_2(n)$ une suite \mathbb{N} telle que $m_2(n) \rightarrow +\infty$ et $\frac{m_2(n)}{n} \rightarrow 0$, on définit :

$$\tilde{S}_n(\lambda, k) = \frac{1}{m_2(n)} \sum_{t=km_2(n)+1}^{(k+1)m_2(n)} \tilde{Y}(\lambda, t)$$

En l'absence de ruptures dans le spectre de X , pour tout k et tout λ , $\tilde{S}_n(\lambda, k)$ et $\tilde{F}_n(\lambda)$ ont la même loi limite: $\sqrt{m_2(n)}(\tilde{S}_n(\lambda, k) - F(\lambda))$ converge vers un produit de copies indépendantes de $\zeta(\lambda)$.

5.2.1.c Indépendance dans l'espace des fréquences

Soit $\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_F)$ une partition de $[0, \pi]$, $\int_{\lambda_{m-1}}^{\lambda_m} f(\alpha) d\alpha$ est la distribution spectrale de X dans la bande de fréquences $[\lambda_{m-1}, \lambda_m]$. Si Λ est bien choisi, un changement dans le spectre de X peut être considéré comme un changement dans une ou plusieurs de ces bandes de fréquences.

$\int_{\lambda_{m-1}}^{\lambda_m} f(\alpha) d\alpha$ est estimé par $F_n(\lambda_m) - F_n(\lambda_{m-1})$, les variables $F_n(\lambda_m) - F_n(\lambda_{m-1})$ étant asymptotiquement

indépendantes, on définit une nouvelle variable $\tilde{B}_n(\lambda_m, k)$:

$$\tilde{B}_n(\lambda_m, k) = \tilde{S}_n(\lambda_m, k) - \tilde{S}_n(\lambda_{m-1}, k)$$

Posons $N(n) = \frac{n}{m_2(n)}$,

comme $\tilde{F}_n(\lambda_m) = \frac{1}{N(n)} \sum_{k=1}^{N(n)} \tilde{S}_n(\lambda_m, k)$, $\int_{\lambda_{m-1}}^{\lambda_m} f(\alpha) d\alpha$ peut être estimé par $\frac{1}{N(n)} \sum_{k=1}^{N(n)} \tilde{B}_n(\lambda_m, k)$.

$\sqrt{m_2(n)} \left(\tilde{B}_n(\lambda_m, k) - \int_{\lambda_{m-1}}^{\lambda_m} f(\alpha) d\alpha \right)$ converge vers une variable Gaussienne centrée de covariance:

$$\lim_{n \rightarrow \infty} \frac{1}{N(n)} \text{Cov}(\tilde{B}_n(\lambda_m, k) \tilde{B}_n(\lambda_{m'}, k')) = \begin{cases} \int_{\lambda_{m-1}}^{\lambda_m} f^2(\alpha) d\alpha & \text{si } (m, k) = (m', k') \\ 0 & \text{sinon} \end{cases}$$

5.2.2 CAS MULTIDIMENSIONNEL

Dans le cas d'un processus p -dimensionnel, on définit un sous-ensemble I de $[1, p] \times [1, p]$. En effet, nous allons tâcher de détecter des ruptures dans les distributions spectrales $\{F^{ij}(\lambda)\}_{(i,j) \in I}$

Comme nous l'avons fait dans le cas unidimensionnel, nous allons construire une variable $\tilde{B}_n^{ij}(\lambda_m, k)$ de la façon suivante:

$$\tilde{B}_n^{ij}(\lambda_m, k) = \frac{1}{m_2(n)} \sum_{t=km_2(n)+1}^{(k+1)m_2(n)} [Y^{ij}(\lambda_m, t) - Y^{ij}(\lambda_{m-1}, t)]$$

Où,

$$\tilde{Y}^{ij}(\lambda, t) = \lambda X^i(t)X^j(t) + \sum_{s=1}^{m_1(n)} [X^i(t)X^j(t-s)e^{i\lambda s} + X^i(t-s)X^j(t)e^{-i\lambda s}] , \quad t > m_1(n)$$

Nous avons:

$$\lim_{n \rightarrow \infty} \tilde{B}_n^{ij}(\lambda_m, k) = \int_{\lambda_{m-1}}^{\lambda_m} f^{ij}(\alpha) d\alpha \quad \text{et,}$$

$$\lim_{n \rightarrow \infty} \frac{1}{N(n)} \text{Cov}(\tilde{B}_n^{ij}(\lambda_m, k) \tilde{B}_n^{gh}(\lambda_{m'}, k')) = \begin{cases} \int_{\lambda_{m-1}}^{\lambda_m} f^{ig}(\alpha) \overline{f^{jh}(\alpha)} d\alpha & \text{si } (m, k) = (m', k') \\ 0 & \text{sinon} \end{cases}$$

On définit à présent une variable multidimensionnelle $\mathfrak{B}_n(\lambda_m, k) = \{\tilde{B}_n^{ij}(\lambda_m, k)\}_{(i,j) \in I}$.

- * $\mathfrak{B}_n(\lambda_m, k)$ converge vers $\{F^{ij}(\lambda_m) - F^{ij}(\lambda_{m-1})\}_{(i,j) \in I}$
- * Normalisés et à l'asymptotique, les $\mathfrak{B}_n(\lambda_m, k)$ sont des variables Gaussiennes, centrées et indépendantes.

Nous allons donc utiliser la suite $\{B_n(k)\} = \{\mathfrak{B}_n(\lambda_1, k), \mathfrak{B}_n(\lambda_2, k), \dots, \mathfrak{B}_n(\lambda_F, k)\}$ comme technique pour détecter d'éventuelles ruptures dans le spectre du processus X.

Soit $L(B_n)$ la vraisemblance de $(B_n(1), B_n(2), \dots, B_n(N(n)))$, asymptotiquement:

$$* L(B_n) = \prod_{k=1}^{N(n)} L(B_n(k)) = \prod_{k=1}^{N(n)} \prod_{m=1}^F L(\mathfrak{B}_n(\lambda_m, k))$$

- * $L(\mathfrak{B}_n(\lambda_m, k))$ est la vraisemblance d'un vecteur Gaussien centré.

5.2.3 TEST DE DETECTION DE RUPTURES

Nous voulons détecter des ruptures dans la série $\{X(t)\}$, une rupture à l'instant t_0 signifie que les distributions spectrales du processus avant et après t_0 sont différentes. Nous allons utiliser la loi limite de B_n pour construire un test séquentiel et détecter ces changements.

Au temps t_0 , nous allons tester les hypothèses:

$$H_0 : \mathbf{B}_n(t_0) \sim \mathcal{N}(m_0, \Gamma_0)$$

$$H_1 : \mathbf{B}_n(t_0) \sim \mathcal{N}(m_1, \Gamma_1)$$

Où $(m_0, \Gamma_0) = (E\mathbf{B}_n(t), \text{Var}\mathbf{B}_n(t))$ pour $t < t_0$, et $(m_1, \Gamma_1) \neq (m_0, \Gamma_0)$.

Enfin, le problème original de détection de ruptures dans le spectre d'un processus est ramené à celui de détection de ruptures dans les paramètres d'un vecteur Gaussien, pour lequel la structure de covariance est connue. Les ruptures affectent la moyenne et la matrice de covariance qui doivent être estimés. Les variations de ces paramètres sont inconnus et doivent aussi être estimées. Nous sommes alors confronté au type de problème proposé dans le chapitre précédent, l'algorithme présenté va donc pouvoir être utilisé ici.

5.3 SIMULATIONS

5.3.1 COMPORTEMENT DU DETECTEUR

Le premier exemple montre le comportement du détecteur. Nous avons simulé un processus Gaussien (X, Y) :

$$\begin{cases} X(k) = aX(k-1) + \varepsilon(k) \\ Y(k) = bX(k) + \eta(k) \end{cases}$$

ε et η sont 2 bruits blancs Gaussiens de variance égale à 1,

$$\text{pour } 1 \leq i \leq 2000, \quad (a, b) = (2, 0)$$

$$\text{pour } 2001 \leq i \leq 4000, \quad (a, b) = (-2, 0)$$

$$\text{pour } 4001 \leq i \leq 6000, \quad (a, b) = (-2, 0.5).$$

- * Les distributions spectrales ont été estimées sur des fenêtres telles que $m_1 = 20$ et $m_2 = 40$.
- * Nous avons fixé $(\delta, T) = (2, 5)$.

Nous présentons en Figure 1 la distance $K(i)$ et le détecteur $Z(i)$.

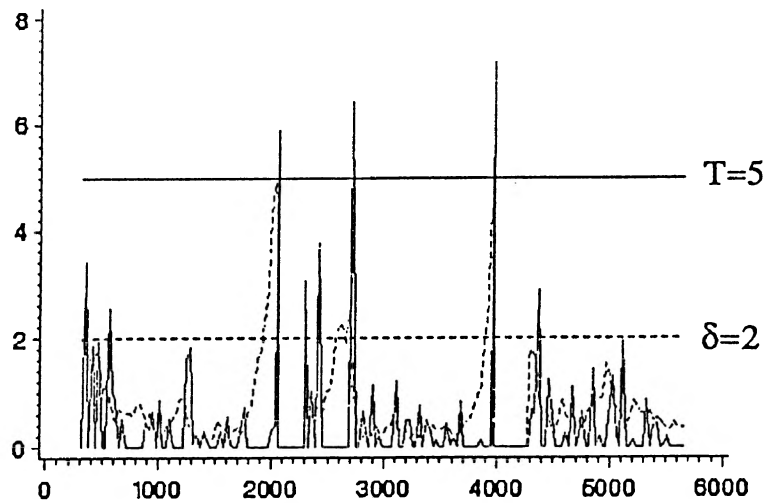


Fig.1 Comportement des détecteurs $Z(l)$ et $K(l)$.

Les 2 ruptures en 2000 et 4000 sont bien détectées (Z et K passent tous deux leur seuil respectif). L'avantage d'un double détecteur est évident vers 2700, en effet, nous avons à la fois $Z(i) > T$ et $K(i) < \delta$: aucune rupture n'est détectée, le changement est trop petit pour être retenu.

5.3.2 EVALUATION DES NOMBRES DE BONNES ET MAUVAISES DETECTIONS

Dans chacun des exemples que nous présentons maintenant, nous avons simulé un processus bidimensionnel (X, Y) dont la distribution dépend d'un paramètre θ .

On considère une réalisation du processus $\{X(k), Y(k)\}_{1 \leq k \leq n}$ avec $n = 4000$. Une rupture est présente à l'instant 2000. Cette rupture, due à un changement de θ , affecte la structure de covariance du processus, mais pas sa moyenne. Nous appellerons θ_1 le paramètre avant le changement et θ_2 le paramètre après le changement.

Dans chaque exemple, et pour différentes valeurs de (δ, T) , nous avons simulé 20 réalisations du processus, et ensuite, nous avons utilisé l'algorithme de détection de ruptures.

- * La partition de $[0, \pi]$ utilisée est $\Lambda = (0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi)$. Cela signifie que nous cherchons à détecter des ruptures dans 4 bandes de fréquences de longueur $\frac{\pi}{4}$.
- * Le sous-ensemble de $[1, p] \times [1, p]$ choisi est $I = \{(1,1), (2,2)\}$. Cela signifie que nous ne cherchons à détecter les ruptures que dans les distributions F^x et F^y , et non dans les distributions croisées F^{xy} .

Pour chaque exemple, nous présentons les nombres de mauvaises et bonnes détections réalisées, nous avons appliqué cette procédure à des processus Gaussiens et non Gaussiens:

- * Dans les exemples 1.a, 2.a et 3.a, ε et η sont deux bruits blancs indépendants Gaussiens, de variance égale à 1.
- * Dans les exemples 1.b, 2.b et 3.b, ε et η sont deux bruits blancs indépendants Bernoulli :

$$P(\varepsilon = 1) = P(\varepsilon = -1) = P(\eta = 1) = P(\eta = -1) = 0.5$$

Exemple 1

$$\begin{cases} X(k) = \varepsilon(k) + a\varepsilon(k-1) + b\varepsilon(k-2) \\ Y(k) = cX(k) + \eta(k) \end{cases} \quad \theta = (a, b, c)$$

$$\theta_1 = (-0.5, 0.6, 0.4) \quad \theta_2 = (-0.3, 0.8, 0.6)$$

T \ δ	0.5	1	2.5
10	19 15	18 11	15 5
20	17 12	20 5	15 1
50	20 6	20 4	7 1

(a)

Processus Gaussien

T \ δ	0.5	1	2.5
10	18 24	20 18	19 2
20	19 5	20 6	13 3
50	19 2	19 2	13 0

(b)

Processus non Gaussien

Bonnes détections
Mauvaises détections

Exemple 2

$$\begin{cases} X(k) = aX(k-1) + \varepsilon(k) \\ Y(k) = cX(k) + \eta(k) \end{cases} \quad \theta = (a, c)$$

La rupture affecte seulement la distribution de X:

$$\theta_1 = (-0.2, 0) \quad \theta_2 = (+0.2, 0)$$

T\δ	0.5	1	2.5
10	20 18	20 11	19 3
20	18 5	19 6	18 0
50	20 1	20 2	16 2

(a)

Processus Gaussien

Bonnes détections
Mauvaises détections

T\δ	0.5	1	2.5
10	19 20	20 11	19 5
20	19 9	20 7	15 1
50	19 6	18 1	17 0

(b)

Processus non Gaussien

Exemple 3

$$\begin{cases} X(k) = \varepsilon(k) + a\varepsilon(k-1) + b\varepsilon(k-2) \\ Y(k) = cX(k) + \eta(k) \end{cases} \quad \theta = (a, b, c)$$

La rupture affecte seulement la distribution de Y:

$$\theta_1 = (-0.3, 0.5, 0.3) \quad \theta_2 = (-0.3, 0.5, 0.7)$$

T\δ	0.5	1	2.5
10	20 24	20 15	20 2
20	20 9	20 5	17 2
50	20 4	20 2	19 0

(a)

Processus Gaussien

Bonnes détections
Mauvaises détections

T\δ	0.5	1	2.5
10	18 21	20 13	18 6
20	20 7	20 1	20 0
50	20 6	20 4	7 1

(b)

Processus non Gaussien

CONCLUSION

Nous obtenons la même qualité de résultat avec des processus Gaussiens ou non Gaussiens. En effet, sous des hypothèses fortes de mélange et quelques conditions sur les moments d'ordre supérieur, les résultats montrés dans la 1^{ère} partie restent valables pour une grande famille de processus [7], [6], [14]. De petits changements dans le spectre sont bien détectés, tandis que le nombre de fausses alarmes est bien contrôlé avec un choix approprié des seuils.

Dans les simulations présentées, nous n'avons pas cherché à optimiser la partition Λ de $[0, \pi]$. Face à des données réelles, une connaissance à priori du processus et de son spectre doit permettre une meilleure sélection des fréquences: les changements ne seront détectés que sur les bandes de fréquences les plus intéressantes.

De la même manière, le choix des processus à utiliser dépend de l'application. Si les ruptures n'affectent que les fonctions d'autocorrélation $R^{ii}(k)$, nous utiliserons les distributions spectrales $F^{ii}(\lambda)$. Si on soupçonne l'existence de changements dans les fonctions de corrélation croisée $R^{ij}(k)$, on utilisera les distributions spectrales croisées $F^{ij}(\lambda)$.

BIBLIOGRAPHIE

- [1] M. Basseville, 1986 "*On-line detection of jumps in mean*" [Basseville, Benveniste, 1986] p 11-26.
- [2] M. Basseville, A. Benveniste, 1986 "*Detection of abrupt changes in signals and dynamical systems*" LNCIS 77 Springer Verlag.
- [3] A. Benveniste, M. Metivier, P. Priouret, 1987 "*Algorithmes adaptatifs et approximations stochastiques*" Techniques stochastiques Masson.
- [4] O. de Cambry, 1984 Thèse de troisième cycle Université de Paris-Sud.
- [5] D. Dacunha Castelle, M. Duflo, 1983, "*Probabilités et statistiques : problèmes à temps mobile*" Masson.
- [6] R. Dahlaus, 1985 "*Asymptotic normality of spectral estimates*" Journal of Multivariate Analysis Vol 16 p 412-431.
- [7] P. Doukhan, J.R. León, 1989 "*Cumulants for stationary mixing random sequences and applications to empirical spectral density*" Probability and mathematical statistics Vol 10 N° 1 p 11-26.
- [8] Hinkley, 1971 "*Inference about the Change Point in a Sequence of Random Variables*" Biometrika Vol 57 N° 1 p 1-17.
- [9] I.A. Ibragimov, 1962 "*On estimation of the spectral function of a stationary Gaussian process*" Theory Prob. Appl. Vol 8 p 366-400.
- [10] M. Lavielle, E. Rio, 1987 "*Identification d'états de mer stationnaires ou de transition par détection de ruptures d'un modèle*" Rapport de Recherche LEMPAB N°1 Juin 1987. Rapport de stage de DEA 1987 Université de Paris-Sud.
- I.B. Mac Neill, 1971 "*Limit processes for co-spectral and quadrature spectral distribution functions*" The Annals of Mathematical Statistics Vol 41 N° 1 p 81-96.
- [12] I.V. Nikiforov, 1986 "*Sequential detection of changes in stochastic systems*" [Basseville, Benveniste, 1986] p 216-258.
- [13] D. Picard, 1985 "*Testing and estimating change-points in time series*" Adv. Appl. Prob. Vol 17 p 841-867.
- [14] M. Rosenblatt, 1984 "*Asymptotic normality, strong mixing and spectral density estimates*" The annals of Probability Vol 12 N° 4 p 1167-1180.
- [15] A. N. Shiryaev, 1978 "*Optimal stopping rules*" Springer Verlag.
- [16] A.S. Willsky, 1986 "*Detection of abrupt changes in dynamic systems*" [Basseville, Benveniste, 1986] p 27-49.

TABLE DES MATIERES

INTRODUCTION	2
 PARTIE I DECONVOLUTION 	
Introduction	4
 Chapitre 1 : Algorithme d'estimation 	
INTRODUCTION	8
1.1 ALGORITHME STOCHASTIQUE D'ESTIMATION	9
1.2 ALGORITHME DÉTERMINISTE D'ESTIMATION	12
1.3 REMARQUES	16
1.4 EXEMPLES	18
 ANNEXE	
1.A1 L'échantillonneur de Gibbs	26
1.A2 Calcul de la loi conditionnelle de $Y(i)$	26
 Chapitre 2 : Techniques Bayésiennes de déconvolution Application à des signaux Bernoulli-Gaussiens 	
INTRODUCTION	28
2.1 PRÉDICTION DANS LE CAS D'UN MÉLANGE DE POPULATIONS	29
2.2 DÉCONVOLUTION D'UN MÉLANGE DE DEUX GAUSSIENNES	33
2.3 DÉCONVOLUTION DE SIGNAUX BERNOULLI-GAUSSIENS	37
2.4 SIMULATIONS	40
 Chapitre 3 : Déconvolution 2-D de traces sismiques 	
INTRODUCTION	44
3.1 MODÉLISATION	47
3.2 IDENTIFICATION DU MODÈLE	57
3.3 ESTIMATION DES RÉFLECTEURS	60
4.4 RÉSULTATS	69
CONCLUSION	77
 ANNEXE	
3.A1 Calcul de la loi conditionnelle $P(Y_g/Z, Y_t, t \neq s)$	78
3.A2 L'algorithme de recuit simulé	83
 Bibliographie	 85

PARTIE II
DETECTION DE RUPTURES

Introduction	88
Chapitre 4 : Détection séquentielle de ruptures dans la distribution d'une suite de variables	
INTRODUCTION	89
4.1 DÉTECTION D'UNE RUPTURE UNIQUE	90
4.2 DÉTECTION SÉQUENTIELLE DE RUPTURES	91
4.3 APPLICATIONS	96
Chapitre 5 : Détection séquentielle de ruptures dans le spectre d'un processus multidimensionnel	
INTRODUCTION ET NOTATIONS	99
5.1 CONVERGENCE DES FONCTIONS SPECTRALES EMPIRIQUES	101
5.2 STATISTIQUES DE TEST	106
5.3 SIMULATIONS	111
CONCLUSION	115
Bibliographie	116