

Approche bayésienne

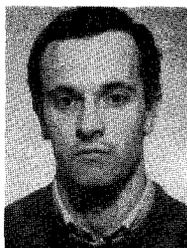
de l'analyse spectrale adaptative :

modèles AR longs

et filtrage de Kalman rapide

Bayesian Adaptive Spectrum Estimation:

Long AR Models and Fast Kalman Filters



Amrane HOUACINE

Laboratoire des Signaux et Systèmes (CNRS/ESE), École Supérieure d'Électricité, plateau du Moulon, 91190, GIF-SUR-YVETTE et Gréco 69 SARTA

Amrane Houacine a obtenu le Diplôme d'Études Supérieures de Physique, spécialité électronique, en 1982 à l'Université des Sciences et de la Technologie d'Alger, le Diplôme d'Études Approfondies en Automatique et Traitement du Signal en 1983, et le Doctorat en Sciences Informatiques en janvier 1987 à l'Université d'Orsay. Ses principaux domaines de recherche scientifique portent sur les algorithmes rapides, les problèmes de stabilité numérique, le filtrage et l'analyse spectrale adaptatifs.

Guy DEMOMENT

Laboratoire des Signaux et Systèmes (CNRS/ESE), École Supérieure d'Électricité, plateau du Moulon, 91190 GIF-SUR-YVETTE et Gréco 69 SARTA.

Guy Demoment est Ingénieur ESE, Docteur-Ingénieur, Docteur ès Sciences. Chargé de Recherche au CNRS et Professeur de probabilités et théorie du signal à l'École Supérieure d'Électricité, il est également Responsable du groupe « Problèmes Inverses en Traitement du Signal et Imagerie » au Laboratoire des Signaux et Systèmes (CNRS/ESE/UPS).

RÉSUMÉ

L'hypothèse d'une stationnarité locale du processus étudié est à la base de l'analyse spectrale adaptative qui utilise des méthodes issues du cas stationnaire avec des fenêtres d'analyse de longueur réduite. Mais le choix usuel de méthodes d'estimation des moindres-carrés interdit l'utilisation de modèles autorégressifs (AR) longs pourtant nécessaires à une bonne résolution.

Nous avons développé une méthode bayésienne utilisant un modèle AR long dont les paramètres ont une loi *a priori* normale traduisant un *a priori* de douceur sur la solution. Cette approche est devenue classique en statistique bayésienne. La principale originalité de notre méthode réside dans le choix de l'ordre p du modèle et dans le calcul de la solution qui est effectué en associant un modèle d'état au problème et en utilisant un filtre de Kalman rapide obtenu par factorisation de Chandrasekhar. La complexité numérique de la méthode est de $O(p)$ par récursion.

Un autre avantage de la méthode réside dans le fait que le filtre de Kalman permet le calcul en ligne de la log-vraisemblance du coefficient de régularisation qui règle l'équilibre entre l'*a priori* et les données. Cette propriété facilite la détermination de sa valeur optimale à partir des données.

L'article est illustré par une application de la méthode aux signaux-tests élaborés par le groupe de travail « Analyse Spectrale » du Gréco « Systèmes Adaptatifs » du CNRS.

MOTS CLÉS

Analyse spectrale adaptative, modèle AR long, approche bayésienne, filtre de Kalman, équations de Chandrasekhar.

SUMMARY

Adaptive spectrum estimation is based on a local stationarity assumption for the studied process, and uses methods of the stationary case with data windows of reduced length. But conventional least squares methods and parsimony principle (for example Akaike's criterion) preclude use of long AR models necessary for a good spectral resolution.

We developed a Bayesian adaptive spectrum estimation method using long AR models and normal prior distributions expressing a smoothness priors on the solution. This is now a classical approach to spectrum estimation. The main originality of our approach lies in the order choice and in the computation of the solution which is performed by a fast Kalman filter of the Chandrasekhar type with a reduced complexity of $O(p)$ per recursion, p being the model order.

The likelihood of the regularizing factor which weights the smoothness priors is maximized to obtain the best data-dependent priors and is computed recursively as a by-product of our fast Kalman filter, which facilitates the determination of the hyperparameters.

The method performances are illustrated by examples of adaptive spectrum estimation for simulated signals.

KEY WORDS

Adaptive spectrum estimation, long AR model, Bayesian approach, Kalman filter, Chandrasekhar equations.

1. Position du problème

Nous développons dans cet article une nouvelle méthode d'analyse spectrale adaptative de signaux non stationnaires. Ces signaux sont supposés être des réalisations de processus aléatoires autorégressifs (AR) décrits par le modèle suivant :

$$(1.1) \quad y(n) = z(n) + b(n), \quad n = 1, 2, \dots$$

$$(1.2) \quad \begin{cases} z(n) = \sum_{i=1}^p a_i(n) y(n-i) = \mathbf{a}(n)^t \mathbf{y}_p(n) \\ \mathbf{a}(n)^t = [a_p(n), \dots, a_1(n)] \\ \mathbf{y}_p(n)^t = [y(n-p), \dots, y(n-1)] \end{cases}$$

où $\mathbf{a}(n)$ est le vecteur des coefficients du modèle dont l'ordre est p , et où $b(n)$ est un bruit blanc gaussien centré de variance σ_b^2 . L'analyse spectrale adaptative repose sur l'hypothèse d'une stationnarité locale du processus étudié, c'est-à-dire d'une variation lente des paramètres $a_i(n)$ vis-à-vis des fluctuations de l'amplitude du signal. Ceci conduit à rechercher un modèle localement stationnaire, c'est-à-dire un modèle invariant équivalent défini par moyennage sur une fenêtre temporelle réduite, en effectuant un compromis entre une trop grande longueur de fenêtre qui ne permet pas de suivre fidèlement les non-stationnarités du signal et une trop courte longueur qui ne permet pas d'estimer le modèle dans de bonnes conditions.

Les méthodes adaptatives communément utilisées s'inspirent donc directement du cas stationnaire en utilisant simplement une fenêtre d'analyse de longueur réduite obtenue par troncature ou par pondération des données observées. Les méthodes des moindres-carrés, par exemple, minimisent :

$$(1.3) \quad J_1(\mathbf{a}) = \sum_{n=1}^N [y(n) - z(n)]^2$$

où N est la longueur de la fenêtre d'analyse (ou longueur équivalente dans le cas pondéré). Elles

conduisent à des solutions très intéressantes dans la pratique en raison du caractère linéaire du modèle (1.2) qui autorise l'emploi d'algorithmes rapides. Mais l'utilisation de ces méthodes est soumise à des exigences contradictoires.

Dans la réalité, l'hypothèse d'un modèle AR pour le processus à analyser n'est, au mieux, qu'une approximation et il est donc souhaitable, de ce point de vue, de choisir un ordre le plus élevé possible pour se placer dans les conditions d'une bonne approximation. Avec un modèle AR tel que (1.2), une première limite à la dimension p du modèle est le nombre N de données dont on dispose. Si, comme nous le supposons par la suite, p est comparable ou supérieur à N , la solution du problème des moindres-carrés (1.3) est inacceptable en raison de la variance excessive de la solution qui peut aller jusqu'à la singularité de la matrice normale du problème. Il s'agit en fait d'un problème instable ou mal posé [10, 18]. Plusieurs approches s'offrent alors pour le stabiliser :

— On peut conserver le critère des moindres carrés et restreindre fortement la dimension p de la solution, c'est l'approche usuelle. Le problème important est alors de choisir cet ordre. Ce choix est en général guidé par un principe de parcimonie, et l'utilisation du critère d'Akaike en constitue un bon exemple [3]. Mais son application conduit toujours à une « situation de contraste » commune en statistique, c'est-à-dire à une situation dans laquelle le nombre de paramètres inconnus à estimer (ici l'ordre p du modèle) est très inférieur au nombre des données disponibles pour les estimer (ici la longueur N de la fenêtre). La longueur de la fenêtre étant nécessairement réduite pour pouvoir suivre les non-stationnarités du processus, la densité spectrale de puissance (DSP) instantanée qui s'en déduit manque de résolution [8], et la pratique montre [17] que cela conduit à des résultats au mieux équivalents à ceux d'une approche bayésienne.

— On peut reconnaître le caractère mal posé du problème qui se traduit par un mauvais conditionnement de la matrice normale, et décider de projeter la solution sur un sous-espace significatif dont la

dimension est déterminée par ce conditionnement, et dont les vecteurs de base sont les vecteurs singuliers de la matrice normale [19]. Ces méthodes de décomposition tronquée en valeurs singulières ont l'avantage de conduire à des problèmes très bien conditionnés, mais elles sont lourdes à mettre en œuvre du fait de la nécessité de rechercher ces valeurs singulières.

— On peut enfin choisir de régulariser le problème en ajoutant une fonctionnelle stabilisatrice au critère initial. On peut montrer [5] que cette approche présente une certaine équivalence avec la précédente pour ce qui est de la stabilité de la solution, mais elle présente surtout le grand avantage de permettre l'utilisation de techniques récursives rapides comme nous le verrons au chapitre suivant. Il faut aussi remarquer que, contrairement à la première approche, les deux dernières ne nécessitent absolument pas de réduire la dimension du problème initial définie par $p = N$.

C'est pourquoi, plutôt que d'appliquer un principe de parcimonie dans une situation où les données ne contiennent pas assez d'information pour estimer des modèles AR de grande dimension, nous avons choisi une approche bayésienne qui fait l'hypothèse d'une loi de probabilité *a priori* des paramètres et qui fournit comme résultat une loi *a posteriori*.

L'utilisation de méthodes bayésiennes n'est pas nouvelle en analyse spectrale et notre étude peut être replacée dans la lignée des travaux d'Akaike [1], de Kitagawa [15], et de Kitagawa et Gersch [16]. Mais son originalité réside dans le choix de l'ordre du modèle et dans le calcul de la solution qui est effectué en associant un modèle d'état au problème et en utilisant des lois *a priori* permettant de développer un filtre de Kalman rapide à l'aide de techniques de factorisation de Chandrasekhar. La complexité numérique de la méthode n'est alors que de l'ordre de p opérations par récursion.

2. Analyse spectrale bayésienne rapide, dans le cas stationnaire

Plaçons-nous tout d'abord dans le cas stationnaire, l'extension à l'analyse spectrale adaptative étant immédiate. Les méthodes classiques d'analyse spectrale fournissent des résultats satisfaisants tant que l'horizon d'observation du processus est long. Les méthodes paramétriques ont été introduites pour réduire cette longueur, mais, quand l'horizon devient relativement court, la qualité de l'estimation spectrale effectuée se dégrade à cause de la variabilité des statistiques utilisées [lorsque N devient comparable ou inférieur à p dans (1.3) par exemple]. La méthode proposée consiste à utiliser un modèle AR d'ordre p relativement élevé, et à stabiliser par régularisation le problème d'estimation correspondant.

2.1. RÉGULARISATION D'UN PROBLÈME MAL POSÉ

La régularisation d'un problème mal posé est la démarche qui consiste à stabiliser le problème initial en incorporant une information *a priori* sur la solution. Ceci peut être obtenu en minimisant le critère

modifié :

$$(2.1) \quad J_2(\mathbf{a}) = J_1(\mathbf{a}) + \mu \Omega(\mathbf{a}), \quad \mu > 0$$

où $\Omega(\mathbf{a})$ est une fonctionnelle régularisante choisie de manière à traduire cet *a priori*. Il existe une grande variété de fonctionnelles régularisantes [4]. Pour des raisons pratiques évidentes, nous choisissons une fonctionnelle quadratique qui présente l'avantage de conduire à une solution linéaire par rapport aux données :

$$(2.2) \quad J_2(\mathbf{a}) = \sum_{n=1}^N [y(n) - z(n)]^2 + \mu \|\mathbf{a} - \mathbf{a}_0\|_R^2$$

R est une matrice définie positive. Ce critère J_2 comprend donc deux parties; la première J_1 traduit la fidélité du modèle vis-à-vis des données, la seconde traduit la fidélité du modèle vis-à-vis d'un *a priori*. Il s'agit d'une norme euclidienne généralisée sur la solution, ce qui est un moyen classique et simple dans la théorie des problèmes mal posés pour contrôler le comportement de la solution au travers du choix de R . R est souvent la matrice de la norme des différences k -ièmes ($k=0, 1, 2$ le plus souvent), ce qui, en reprenant le vocabulaire consacré de la littérature des problèmes mal posés, traduit un *a priori* physique de « douceur » sur la solution [18]. La constante μ , ou coefficient de régularisation, règle le poids relatif des deux composantes du critère J_2 .

Les difficultés n'ont pas disparu pour autant et le problème réside maintenant dans le choix de μ et de R . Mais ce problème de moindres-carrés régularisés a une interprétation bayésienne simple [12] puisque la minimisation de (2.2) est équivalente à la maximisation de :

$$(2.3) \quad V(\mathbf{a}) = \exp \left\{ -(1/2 \sigma_b^2) \times \left(\sum [y(n) - z(n)]^2 + \mu \|\mathbf{a} - \mathbf{a}_0\|_R^2 \right) \right\}$$

La solution $\hat{\mathbf{a}}$ est donc en fait la moyenne de la loi *a posteriori* définie par la distribution conditionnelle des données :

$$(2.4) \quad f(y/\sigma_b^2, \mathbf{a}) \sim \exp \left\{ -(1/2 \sigma_b^2) \times \left(\sum [y(n) - z(n)]^2 \right) \right\}$$

et la loi *a priori* :

$$(2.5) \quad \begin{cases} f(\mathbf{a}/\mathbf{a}_0, \Pi_0) \sim \exp \left\{ -1/2 \|\mathbf{a} - \mathbf{a}_0\|_{\Pi_0}^2 \right\} \\ \Pi_0 = (\sigma_b^2/\mu) R^{-1} \end{cases}$$

On peut donc considérer que cette loi *a priori* spécifie une classe d'estimateurs par l'intermédiaire des paramètres \mathbf{a}_0 et Π_0 . Ce sont les hyperparamètres du problème. Comme les lois sont normales, la vraisemblance (marginale) de ces hyperparamètres est aussi normale et est obtenue par intégration :

$$(2.6) \quad V(\sigma_b^2, \mathbf{a}_0, \Pi_0/y) = \int f(y/\mathbf{a}, \sigma_b^2) f(\mathbf{a}/\mathbf{a}_0, \Pi_0) d\mathbf{a}$$

Le modèle finalement retenu est celui qui maximise cette vraisemblance par rapport aux hyperparamètres.

C'est en ce sens que la méthode est adaptative puisque le choix de l'*a priori* dépend des données.

2.2. APPLICATION A L'ANALYSE SPECTRALE

Le problème d'estimation adaptative tel qu'il vient d'être défini est très difficile à résoudre en pratique à cause du nombre élevé d'hyperparamètres. Pour réduire la complexité du problème, on suppose souvent ces hyperparamètres partiellement connus en s'appuyant sur des considérations physiques par exemple. Mais cette démarche est compliquée ici par le fait que les coefficients AR, choisis pour la simplicité du modèle associé, sont sans dimension et n'ont guère de signification physique, contrairement à des pôles ou à des coefficients cepstraux par exemple. Nous avons donc introduit un minimum d'information *a priori* en effectuant le choix suivant :

– $a_0 = 0$ dans le cas d'un signal stationnaire. Par contre, dans le cas du traitement par bloc d'un signal non stationnaire, nous choisissons pour un bloc le résultat du traitement du bloc précédent (cf. chap. 3).

– $\Pi_0 = \sigma_a^2 I$, c'est-à-dire une variance *a priori* uniforme et une absence de corrélation *a priori* entre les coefficients AR.

Ce choix constituera certainement un des points à creuser dans le développement ultérieur de la méthode. Il faut aussi remarquer que cet *a priori* de douceur porte sur les paramètres AR, ce qui n'interdit absolument pas de voir apparaître des pics sur la densité spectrale de puissance qui s'en déduit par :

$$(2.7) \left\{ \begin{array}{l} G_{yy}(f) = \sigma_b^2 \left| 1 - \sum_{k=1}^p a_k \exp(-2i\pi fk) \right|^{-2} \\ -1/2 \leq f \leq 1/2 \end{array} \right.$$

Il ne subsiste plus alors que deux hyperparamètres, σ_b^2 et $\mu = \sigma_b^2 / \sigma_a^2$, qui, nous le verrons par la suite, peuvent être découplés dans le problème. Le coefficient de régularisation μ devient ainsi l'inverse de la variance relative *a priori* des paramètres, il traduit donc bien le poids de l'*a priori* dans le critère (2.2).

La maximisation de la vraisemblance (2.6) reste un problème complexe, même lorsqu'il n'y a plus que deux hyperparamètres, car le problème complet d'estimation des paramètres et des hyperparamètres est non linéaire. C'est pourquoi il est résolu de manière sous-optimale en trois étapes :

– Le domaine de variation de μ est réduit à un ensemble fini de valeurs discrètes choisies au départ.

– On calcule, pour chacune de ces valeurs, la solution optimale \hat{a} et la vraisemblance correspondante $V(\mu/y)$ à partir des données observées. Ce calcul est effectué à l'aide d'un algorithme rapide (cf. infra).

– On retient finalement la solution correspondant à la vraisemblance la plus forte.

Bien que cette idée de régularisation d'un problème mal posé soit assez étrangère au traitement du signal qui est très marqué par la statistique orthodoxe et l'application du principe de parcimonie, cette démarche bayésienne n'est pas en soi nouvelle en analyse spectrale. On verra en effet par la suite que plusieurs ingrédients de notre méthode sont déjà

contenus dans Kitagawa [15] ou dans Clergeot [2] par exemple. Mais ce qui est original, c'est l'esprit dans lequel nous l'avons développée, et les points suivants nous distinguent radicalement :

– L'*a priori* joue un rôle essentiel dans notre méthode, il n'est pas considéré comme un mal nécessaire pour utiliser un filtre de Kalman, et dont on cherche à se débarrasser (en choisissant une matrice de covariance *a priori* très grande) afin d'obtenir une solution qui puisse être remplacée, au moins asymptotiquement, dans un cadre statistique orthodoxe [2].

– L'approche bayésienne d'Akaike [1], de Kitagawa [15], et de Kitagawa et Gersch [16] consiste à augmenter l'ordre du modèle AR par rapport aux méthodes usuelles, mais en maintenant toujours un contraste ($p \ll N$) que rien ne justifie plus. Pourquoi en effet choisir $p = 20$ si $N = 114$, plutôt que $p = 25$ ou $p = 30$ [17]? Nous choisissons à l'opposé de stabiliser le problème mal posé initial, tel qu'il est défini plus haut, c'est-à-dire avec $p = N$.

– Contrairement à Kitagawa et Gersch [16], au lieu d'imposer à chaque paramètre d'avoir une trajectoire douce, nous imposons à chaque instant une contrainte sur l'ensemble des paramètres.

– La dernière originalité du travail est dans la manière dont sont calculés les paramètres AR et la vraisemblance des hyperparamètres. Nous utilisons pour cela une représentation d'état associée au problème et une technique de filtrage de Kalman rapide.

2.3. ÉQUATIONS D'ÉTAT ASSOCIÉES AU PROBLÈME

On sait que le filtrage de Kalman permet de calculer récursivement la solution du problème défini par (2.2). Mais son utilisation sous forme standard fait appel à une équation de Riccati qui n'exploite pas la propriété de stationnarité locale du modèle qui est à la base des méthodes adaptatives, ni les propriétés de décalage des vecteurs $y_p(n)$ successifs. Pour y remédier [13], on commence par définir un vecteur paramètre étendu à m composantes :

$$(2.8) \quad \mathbf{A}_m(i) = [\mathbf{O}_{i-1}^t, a_p, \dots, a_1, 0, \dots, 0]^t$$

et on considère le vecteur global des données :

$$(2.9) \quad \mathbf{Y}_m = [y(-p+1), y(-p+1), y(-p+2), \dots, y(0), \dots, y(m-p)]^t$$

Le modèle linéaire défini par les équations (1.1) et (1.2) peut alors se mettre sous la forme du modèle d'état suivant :

$$(2.10) \quad \mathbf{A}_m(i+1) = \mathbf{D}^t \mathbf{A}_m(i), \quad i = 1, 2, \dots$$

$$(2.11) \quad y(i) = \mathbf{Y}_m^t \mathbf{A}_m(i) + b(i)$$

où \mathbf{D} est l'opérateur de décalage :

$$(2.12) \quad \mathbf{D} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & & \\ \cdot & & & & \\ 0 & & & & 1 \\ \cdot & \cdot & \dots & & 0 \end{pmatrix}$$

Le calcul récursif de la solution peut être effectué par le filtre de Kalman :

$$(2.13) \quad \hat{\mathbf{A}}_m(i+1/i) = \mathbf{D}^t \hat{\mathbf{A}}_m(i/i-1) + \mathbf{K}_m(i) r(i)^{-1} [y(i) - \mathbf{Y}_m^t \hat{\mathbf{A}}_m(i/i-1)]$$

$$(2.14) \quad r(i) = \mathbf{Y}_m^t \mathbf{P}_m(i/i-1) \mathbf{Y}_m + \sigma_b^2$$

$$(2.15) \quad \mathbf{K}_m(i) = \mathbf{D}^t \mathbf{P}_m(i/i-1) \mathbf{Y}_m$$

$$(2.16) \quad \mathbf{P}_m(i+1/i) = \mathbf{D}^t \mathbf{P}_m(i/i-1) \mathbf{D} - \mathbf{K}_m(i) r(i)^{-1} \mathbf{K}_m(i)^t$$

avec les conditions initiales :

$$(2.17) \quad \hat{\mathbf{A}}_m(1/0) = [\mathbf{a}_0^t, 0, \dots, 0]$$

$$(2.18) \quad r(0) = \mathbf{Y}_m^t \mathbf{P}_m(1/0) \mathbf{Y}_m + \sigma_b^2$$

$$(2.19) \quad \mathbf{K}_m(0) = \mathbf{D}^t \mathbf{P}_m(1/0) \mathbf{Y}_m$$

$$(2.20) \quad \mathbf{P}(1/0) = \begin{pmatrix} \Pi_0 & 0 \\ 0 & 0 \end{pmatrix}$$

Il peut paraître surprenant que, pour réduire le volume des calculs, on commence par augmenter de façon importante les dimensions du vecteur paramètre ou vecteur d'état. Il ne s'agit en fait que d'un artifice destiné à faire apparaître explicitement les propriétés d'invariance que l'on cherche à exploiter. On peut d'ailleurs vérifier qu'à chaque récursion, le nombre de coordonnées non nulles du vecteur gain $\mathbf{K}_m(i)$ reste égal à p . Ce modèle d'état étant invariant (\mathbf{D} et \mathbf{Y}_m constants) et à bruit stationnaire, on obtient directement un algorithme rapide en remplaçant les équations standards (2.14)-(2.16) par des équations de Chandrasekhar. En plus de la rapidité, ces équations de Chandrasekhar permettent d'aboutir naturellement à des algorithmes en racine carrée offrant une meilleure stabilité numérique. C'est la seconde différence avec l'approche de Kitagawa et Gersch dans laquelle la condition de douceur porte sur l'évolution temporelle de chacun des paramètres AR, ce qui impose d'introduire un terme de forçage stochastique dans l'équation d'état, et ce qui interdit par la suite d'utiliser des algorithmes rapides. C'est ce qui explique aussi sans doute pourquoi ces auteurs conservent dans leurs exemples d'application un contraste $p \ll N$ par ailleurs injustifié !

2.4. ALGORITHME B-CAR (Block-Chandrasekhar Adaptive Regularizer)

Les équations de Chandrasekhar exploitent les propriétés d'invariance locale du modèle (2.10)-(2.11). Elles ne propagent que les incréments des quantités (2.14)-(2.16) qui sont eux-mêmes factorisés. L'application directe de ces techniques de factorisation au filtre (2.13)-(2.16) conduit aux équations suivantes [13] :

$$(2.21) \quad \mathbf{v}(i) = \mathbf{B}(i)^t \mathbf{Y}_m$$

$$(2.22) \quad \mathbf{w}(i) = \mathbf{M}(i) \mathbf{v}(i)$$

$$(2.23) \quad \mathbf{M}(i+1) = \mathbf{M}(i) + \mathbf{w}(i) r(i)^{-1} \mathbf{w}(i)^t$$

$$(2.24) \quad \mathbf{K}_m(i+1) = \mathbf{K}_m(i) + \mathbf{D}^t \mathbf{B}(i) \mathbf{w}(i)$$

$$(2.25) \quad r(i+1) = r(i) + \mathbf{v}(i)^t \mathbf{w}(i)$$

$$(2.26) \quad \mathbf{B}(i+1) = \mathbf{D}^t \mathbf{B}(i) - \mathbf{K}_m(i+1) r(i+1)^{-1} \mathbf{v}(i)^t$$

Une écriture condensée de l'algorithme est obtenue par projection, en ne faisant apparaître que les composantes intervenant effectivement à chaque récursion dans le calcul. On définit pour cela un vecteur gain effectif $\mathbf{k}(i)$ de dimension p contenant les coordonnées non nulles de $\mathbf{K}_m(i)$:

$$(2.27) \quad \mathbf{K}_m(i) = [\mathbf{O}_{\alpha, i}^t, \mathbf{k}(i)^t, \mathbf{O}_{\alpha, m-p-i-1}^t]^t$$

De la même façon, si α est la dimension de $\mathbf{M}(i)$, on peut aussi écrire :

$$(2.28) \quad \mathbf{B}(i) = [\mathbf{O}_{\alpha, i}, \mathbf{S}(i)^t, \mathbf{O}_{\alpha, m-p-i-1}]^t$$

L'algorithme peut alors se mettre sous la forme réduite :

$$(2.29) \quad \mathbf{v}(i) = \mathbf{S}(i)^t \begin{bmatrix} y(i-p) \\ \mathbf{y}_p(i+1) \end{bmatrix}$$

$$(2.30) \quad \mathbf{w}(i) = \mathbf{M}(i) \mathbf{v}(i)$$

$$(2.31) \quad \mathbf{M}(i+1) = \mathbf{M}(i) + \mathbf{w}(i) r(i)^{-1} \mathbf{w}(i)^t$$

$$(2.32) \quad r(i+1) = r(i) + \mathbf{v}(i)^t \mathbf{w}(i)$$

$$(2.33) \quad \begin{bmatrix} 0 \\ \mathbf{k}(i+1) \end{bmatrix} = \begin{bmatrix} \mathbf{k}(i) \\ 0 \end{bmatrix} + \mathbf{S}(i) \mathbf{w}(i)$$

$$(2.34) \quad \mathbf{S}(i+1) = \mathbf{S}(i) - \begin{bmatrix} 0 \\ \mathbf{k}(i+1) \end{bmatrix} r(i+1)^{-1} \mathbf{v}(i)^t$$

la remise à jour des paramètres se faisant selon :

$$(2.35) \quad \hat{\mathbf{a}}(i+1/i) = \hat{\mathbf{a}}(i/i-1) + \mathbf{k}(i) r(i)^{-1} [y(i) - \mathbf{y}_p(i)^t \hat{\mathbf{a}}(i/i-1)]$$

On reconnaît sous cette forme un algorithme du type « Schur généralisé » qui présente l'avantage appréciable de s'appliquer aussi bien à des problèmes pré-fenêtrés (traitement du cas stationnaire ou du premier bloc dans le cas non stationnaire) qu'à des problèmes du type covariance (traitement d'un bloc courant dans le cas non stationnaire).

L'initialisation de l'algorithme est conditionnée par la matrice de covariance *a priori* Π_0 dont se déduisent $\mathbf{B}(1)$ et $\mathbf{M}(1)$ lors de la factorisation de l'incrément initial

$$\delta \mathbf{P}(1) \triangleq \mathbf{P}(2/1) - \mathbf{P}(1/0) = \mathbf{B}(1) \mathbf{M}(1) \mathbf{B}(1)^t.$$

Dans notre méthode, Π_0 est choisie diagonale, et le rang de $\delta \mathbf{P}(1)$ est $\alpha=3$. La factorisation initiale n'étant pas unique, nous avons choisi pour $\mathbf{M}(1)$ la matrice de signature de $\delta \mathbf{P}(1)$ qui fournit un algorithme sous forme racine carrée dont les qualités numériques sont meilleures [20]. Nous avons donc :

$$(2.36) \quad \mathbf{M}(1) = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$(2.37) \quad \mathbf{S}(1) = [\bar{\mathbf{k}}(1), \mathbf{b}, \mathbf{c}]$$

$$(2.38) \quad \bar{\mathbf{k}}(1) = [0, \mathbf{k}(1)^t r(1)^{-1/2}]^t$$

$$(2.39) \quad \mathbf{b} = [\sigma_a, 0, \dots, 0]^t$$

$$(2.40) \quad \mathbf{c} = [0, \dots, 0, \sigma_a]^t$$

Le décompte des opérations arithmétiques élémentaires, limité aux seules multiplications, montre que la complexité de cet algorithme est de $O(11p)$ par récursion. Comme on choisit $p=N$, on voit que la complexité totale pour traiter N données est de $O(N^2)$, ce qui est normal pour un algorithme de cette famille.

2.5. CALCUL DE LA VRAISEMBLANCE DES HYPERPARAMÈTRES

L'utilisation d'un filtre de Kalman permet de calculer récursivement la vraisemblance V (ou son logarithme L) des hyperparamètres à l'aide des quantités intervenant dans l'algorithme. On a en effet, par application séquentielle de la règle de Bayes à la relation (2.6) :

$$(2.41) \quad V(\mu, \sigma_b^2/y) = f[y(1)] \prod_{n=2}^N f[y(n)/n-1]$$

On utilise la densité marginale conditionnelle de $y(n)$ étant données les observations $y(1), y(2), \dots, y(n-1)$:

$$(2.42) \quad f[y(n)/n-1] = \int f[y(n)/\mathbf{a}] f[\mathbf{a}/n-1] d\mathbf{a} = (2\pi\sigma_b^2)^{-1/2} \times \exp\left\{-\left[y(n) - \sum_{i=1}^p a(i)y(n-i)\right]^2 / 2\sigma_b^2\right\}$$

Mais, pour calculer cette densité conditionnelle de cette manière, il faudrait connaître σ_b^2 et \mathbf{a} qui sont inconnus. Or on sait aussi que :

$$(2.43) \quad y(n) = \mathbf{y}_p(n)^t \hat{\mathbf{a}}(n/n-1) + e(n)$$

où $\hat{\mathbf{a}}(n/n-1)$ est l'estimée optimale [c'est-à-dire conditionnelle au passé $y(1), y(2), \dots, y(n-1)$] et où $e(n)$ est l'innovation du processus observé. On a donc aussi :

$$(2.44) \quad f[y(n)/n-1] = [2\pi\sigma_e(n)^2]^{-1/2} \times \exp[-e(n)^2 / 2\sigma_e(n)^2]$$

Le problème est alors d'évaluer la variance de l'innovation σ_e^2 . Mais on sait qu'un filtre de Kalman n'est sensible qu'à la seule variance relative du bruit d'observation et du bruit d'état. Le calcul récursif de $\hat{\mathbf{a}}(n)$ ne dépendant donc que de μ et pas de la valeur de σ_b^2 , cette dernière est fixée arbitrairement à l'unité dans cet algorithme, et l'on a ainsi :

$$(2.45) \quad \sigma_e(n)^2 = \sigma_b^2 r(n) = \sigma_b^2 [\mathbf{Y}_m^t \mathbf{P}_m(n) \mathbf{Y}_m + 1]$$

D'où l'on tire :

$$(2.46) \quad f[y(n)/n-1] = [2\pi\sigma_b^2 r(n)]^{-1/2} \times \exp[-(1/2)e(n)^2 \sigma_b^{-2} r(n)^{-1}]$$

On vérifie bien qu'il existe un découplage entre les deux hyperparamètres. L'estimée de σ_b^2 selon le maximum de vraisemblance (marginale) est donc :

$$(2.47) \quad \hat{\sigma}_b^2 = (1/N) \sum_{i=1}^N e(i)^2 r(i)^{-1}$$

et la vraisemblance (marginale) de μ est :

$$(2.48) \quad L(\mu/y) = -1/2 \sum_{i=1}^N [e(i)^2 r(i)^{-1} + \text{Log}|r(i)|] + \text{Cte}$$

2.6. RÉSUMÉ DE LA MÉTHODE

La mise en œuvre de la méthode se fait donc de la manière suivante :

- Pour chacune des valeurs discrètes de μ choisies (cf. § 2.1), on estime un vecteur $\hat{\mathbf{a}}(N)$ avec l'algorithme (2.29)-(2.35) dans lequel on fait $\sigma_b^2=1$.
- La suite des innovations $e(n)$ et de leurs variances relatives $r(n)$ permet d'estimer la variance absolue $\hat{\sigma}_b^2$ et de calculer la log-vraisemblance de μ .
- On retient les paramètres $\hat{\mathbf{a}}$ et les hyperparamètres $\hat{\sigma}_b^2$ et $\hat{\mu}$ correspondant à la log-vraisemblance la plus forte.
- On calcule enfin la densité spectrale de puissance correspondante à l'aide de la relation (2.7).

2.7 EXEMPLE D'APPLICATION

L'exemple choisi à la figure 1 est celui du signal-test stationnaire élaboré par le groupe de travail « Analyse

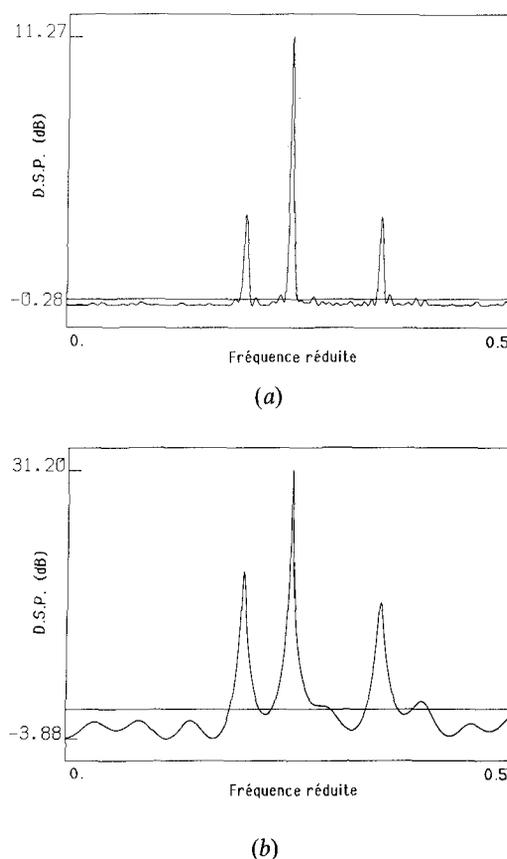


Fig. 1. - Analyse spectrale d'un signal synthétique stationnaire obtenu par superposition de trois sinusoides et d'un bruit blanc. Les fréquences réduites respectives des trois sinusoides sont $f_1=0,20$, $f_2=0,25$ et $f_3=0,35$. Leurs amplitudes respectives sont $A_1=5$, $A_2=10$, $A_3=5$; et leurs phases initiales respectives sont $\Phi_1=\pi/4$, $\Phi_2=\pi/2$, $\Phi_3=0$. Le rapport signal-à-bruit défini comme le rapport de la puissance moyenne des sinusoides à la puissance moyenne du bruit, est fixé à 10 dB. Le signal analysé comporte $N=128$ points. La figure a montre le résultat obtenu par notre méthode avec $p=128$. La figure b illustre la dégradation apportée par le choix d'un ordre $p < N$. Dans cet exemple, $p=20$.

spectrale » du Gréco « Systèmes adaptatifs » du CNRS. Ce n'est pas un signal purement aléatoire puisqu'il est constitué de trois sinusoides et d'un bruit blanc. Il est observé sur 128 échantillons. Nous avons donc choisi un modèle AR d'ordre 128. Les paramètres sont estimés par la méthode précédemment décrite. L'initialisation du traitement se fait en choisissant pour \mathbf{a}_0 le vecteur nul. Le résultat obtenu est présenté à la figure 1 a. La figure 1 b illustre dans ce cas la dégradation entraînée par l'emploi de modèles AR d'ordre réduit et confirme les remarques faites au paragraphe 2. 2.

3. Cas non stationnaire

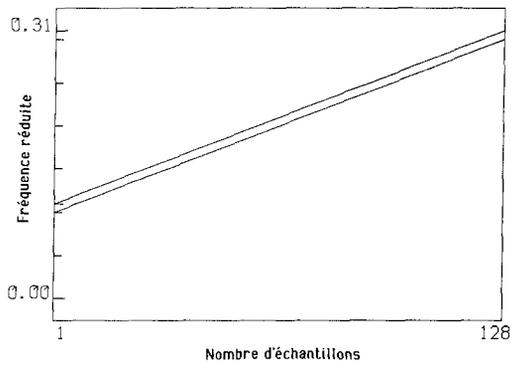
L'extension de la méthode précédente au cas non stationnaire est immédiate. Elle repose simplement sur une information *a priori* supplémentaire constituée par le temps de quasi-stationnarité du processus étudié. Les données à traiter sont alors organisées en blocs consécutifs adjacents de longueur N_s égale à ce temps de quasi-stationnarité, le résultat du traitement d'un bloc étant utilisé comme moyenne *a priori* \mathbf{a}_0 pour le bloc suivant. On voit donc qu'en toute généralité N_s doit être considéré lui aussi comme un hyperparamètre du problème. Le principe de la méthode reste inchangé, mais la mise en œuvre se complique puisque l'on doit maximiser la vraisemblance de trois hyperparamètres, $V(N_s, \mu, \sigma_b^2)$.

On obtient ainsi un modèle « local » à la fin de chaque bloc traité. Mais l'estimation des paramètres étant faite en ligne, il est possible de calculer à chaque instant la DSP instantanée du processus avec la relation (2.6) pour obtenir une représentation temps-fréquence sous la forme d'une nappe. Il faudrait alors, en toute rigueur, calculer une solution lissée $\hat{\mathbf{a}}(n/N)$ sur chaque bloc. Une autre façon, plus simple, de procéder consisterait à utiliser une fenêtre glissante sur les données, à l'aide par exemple d'un algorithme S-CAR (Sliding window-Chandrasekhar Adaptive Regularizer [14], mais au prix du calcul de deux gains à chaque récurrence. Dans ce travail préliminaire, nous nous sommes contentés d'effectuer un traitement par bloc avec un simple filtrage.

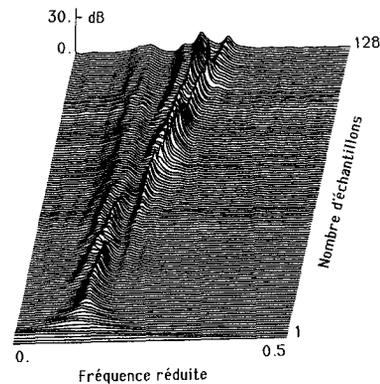
Les signaux-tests non stationnaires du Gréco « Systèmes adaptatifs » sont utilisés pour illustrer le comportement de la méthode qui a été appliquée de manière simplifiée, N_s étant fixé *a priori*, et non pas de manière adaptative pour la longueur des blocs. Les résultats apparaissent sur les figures 2 à 6 où l'on présente à la fois le motif idéal et la nappe de la représentation temps-fréquence estimée.

La figure 2 montre que nous ne sommes pas parvenus à estimer correctement la DSP idéale. La non-stationnarité impose une faible longueur aux blocs de données, ce qui a deux conséquences :

- Un manque de résolution pour pouvoir séparer correctement deux sinusoides aussi proches.
- Un « évanouissement » de la méthode au premier tiers du motif, qui coïncide avec un battement des deux fréquences entraînant une réduction sensible du rapport signal-à-bruit sur le bloc contemporain, et qui est dû à l'adaptativité de la méthode.



(a)



(b)

Fig. 2. — Signal non-stationnaire constitué de deux sinusoides modulées linéairement en fréquence et d'un bruit blanc. L'écart de fréquence réduite des deux sinusoides est constant et égal à 0,01, leurs phases initiales respectives sont $\pi/4$ et $\pi/2$. Leurs amplitudes sont égales et le rapport signal-à-bruit est fixé à 15 dB. La figure a représente le motif idéal. La figure b montre le résultat obtenu avec $p=36$, $N_s=36$, pour $N=128$ points observés. La résolution de la méthode est insuffisante. On observe un « évanouissement » au premier tiers du motif. C'est la conséquence de l'adaptativité de la méthode et de l'existence d'un battement entre les deux fréquences qui entraîne une diminution sensible du rapport signal-à-bruit pour le bloc contemporain.

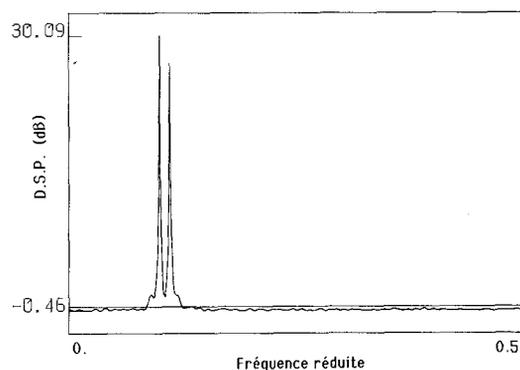
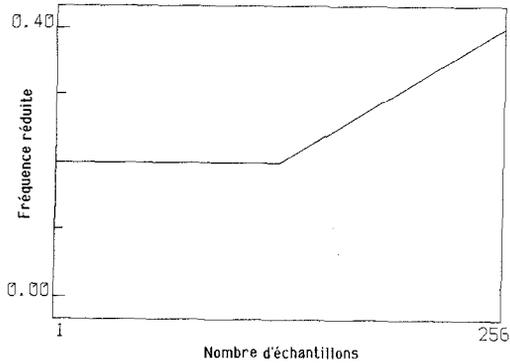


Fig. 3. — La résolution en fréquence des méthodes adaptatives est directement liée au temps de quasi-stationnarité du processus étudié. Dans cet exemple, les deux sinusoides précédentes ont des fréquences constantes et sont observées sur $N=128$ points. Le choix d'un modèle d'ordre $p=128$ permet alors de les séparer parfaitement.

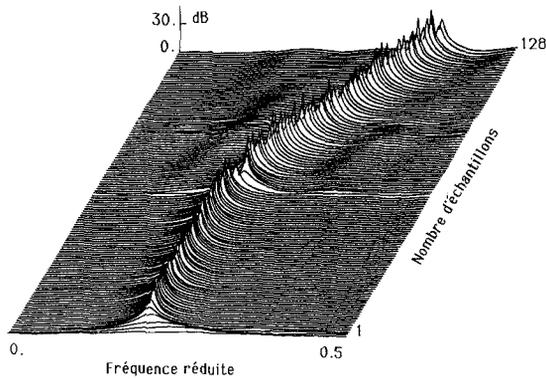
La figure 3 montre, comme nous venons de l'indiquer, que cette résolution est directement liée au temps de stationnarité N_s , et qu'elle est sensiblement améliorée lorsque l'accroissement de N_s permet d'augmenter l'ordre du modèle.

Les figures 4 et 5 n'appellent pas de remarques particulières, sinon que l'on gagnerait probablement en résolution dans les zones stationnaires ou faiblement

non stationnaires (première moitié des deux motifs) si le choix de la longueur des blocs était fait lui-aussi de manière adaptative.

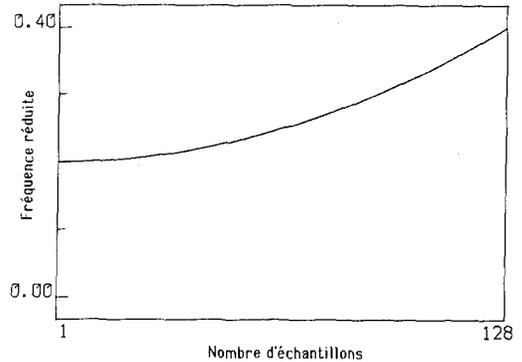


(a)

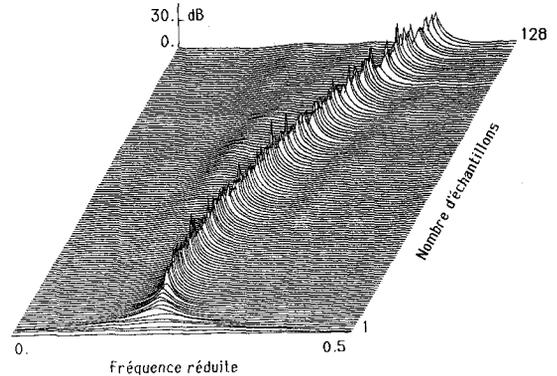


(b)

Fig. 4. — Signal non-stationnaire constitué d'une sinusoïde de fréquence constante puis déferlante, et d'un bruit blanc. L'amplitude de la sinusoïde est constante, la phase à l'origine est $\pi/4$, la différence de phase au début du déferlement est proche de $\pi/2$. Le rapport signal-à-bruit est fixé à 20 dB. La figure a représente le motif idéal. La figure b montre le résultat obtenu avec $p=10$, $N_s=10$, pour $N=256$ points observés.



(a)



(b)

Fig. 5. — Signal non-stationnaire constitué d'une seule sinusoïde avec une modulation parabolique de fréquence. L'amplitude est constante, la phase initiale est nulle. La figure a représente le motif idéal. La figure b montre le résultat obtenu avec $p=12$, $N_s=12$, pour $N=128$ points observés.

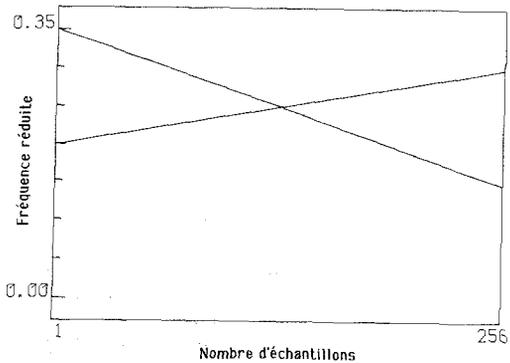
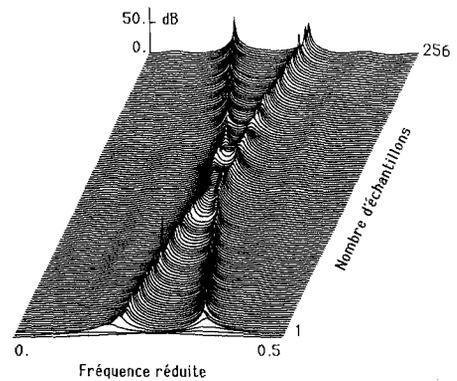


Fig. 6. — Signal non-stationnaire constitué de deux sinusoïdes modulées linéairement en fréquence, mais dont les motifs se croisent sur la durée de l'observation. Leurs amplitudes sont égales et leurs phases à l'origine sont nulles. La figure a représente le motif idéal. La figure b montre le résultat obtenu avec $p=12$, $N_s=12$,



(b)

pour $N=256$ points observés. La résolution de la méthode se dégrade lorsque les deux motifs se croisent. Elle redevient correcte lorsque les fréquences des deux sinusoïdes sont à nouveau suffisamment séparées.

La figure 6 confirme la remarque faite à la figure 2 quant à la difficulté à séparer correctement deux sinusoides de fréquences très voisines. La partie centrale du motif manque de résolution.

4. Conclusion

Nous avons présenté une nouvelle méthode d'analyse spectrale adaptative de signaux non stationnaires qui démontre l'intérêt d'une utilisation combinée de modèles AR longs et d'une approche statistique bayésienne. Les résultats obtenus sur des signaux-tests, la faible complexité algorithmique, et le caractère relativement grossier de l'information *a priori* exigée par la méthode nous font penser qu'elle peut au moins constituer une bonne méthode générale, utilisable pendant l'étude préliminaire d'un problème, avant l'introduction d'une information plus fine mais plus compromettante, telle que l'existence de sinusoides dans un bruit aléatoire.

Manuscrit reçu le 7 octobre 1986.

BIBLIOGRAPHIE

- [1] H. AKAIKE, *Likelihood and the Bayes procedure*, in *Bayesian statistics*, BERNARDO et al. éd., University Press, Valencia, Spain, 1980, p. 141-166.
- [2] H. CLERGEOT, Estimation du spectre d'un signal aléatoire gaussien par le critère du maximum de vraisemblance ou du maximum de probabilité *a posteriori*, *Thèse de Doctorat d'État*, Université de Paris-Sud, n° 2637, 1982.
- [3] H. CLERGEOT, Filter Order Selection in Adaptive Maximum Likelihood Estimation, *IEEE Trans.*, IT-30, 1984, p. 199-210.
- [4] J. CULLUM, The effective choice of the smoothing norm in regularization, *Math. Comp.*, 33, 1979, p. 149-170.
- [5] J. CULLUM, Ill-posed deconvolutions: regularizations and singular value decompositions, *Proc. 19th IEEE Conf. Decision and Control*, Albuquerque, Dec. 1980, p. 29-35.
- [6] G. DEMOMENT, *Algorithmes rapides pour le filtrage, la prédiction et l'identification des processus stochastiques à temps discret*, Rapport interne L2S n° 86/034 et Cours de l'ESE n° M009/86, 1986, 70 p.
- [7] G. DEMOMENT et R. REYNAUD, Equations de Schur et de Chandrasekhar généralisées pour des problèmes généraux d'estimation linéaire, *Actes du 11^e Colloque GRETSI*, Nice, mai 1987.
- [8] M. BOUVET et H. CLERGEOT, Spectral lines detection using parametric methods: on the advantages of overestimating the model order, *Proc. Conf. Inform. Sci. Systems*, 1986, Princeton.
- [9] G. FAVIER, *Filtrage, modélisation et identification des systèmes linéaires stochastiques à temps discret*, CNRS, 1982.
- [10] J. N. FRANKLIN, Well-posed stochastic extensions of ill-posed linear problems, *J. Math. Anal. Appl.*, 31, 1970, p. 682-716.
- [11] C. GUEGUEN, An introduction to displacement ranks and related fast algorithms, *Lectures Notes*, NATO Summer School on Digital Signal Processing, Les Houches, 1985.
- [12] Y. C. HO et R. C. K. LEE, A Bayesian approach to problems in stochastic estimation and control, *IEEE Trans.*, AC-9, 1964, p. 333-339.
- [13] A. HOUACINE et G. DEMOMENT, Chandrasekhar adaptive regularizer for adaptive filtering, *Proc. IEEE ICASSP 86*, Tokyo, 1986, p. 2967-2970.
- [14] A. HOUACINE, Algorithmes des moindres carrés récursifs rapides régularisés pour le filtrage et l'analyse spectrale adaptatifs, *Thèse de Doctorat*, Université de Paris-Sud, Centre d'Orsay, n° 183, 9 janvier 1987.
- [15] G. KITAGAWA, Changing Spectrum Estimation, *J. Sound Vibration*, 89, 1983, p. 433-445.
- [16] G. KITAGAWA et W. GERSCH, A Smoothness Priors Time-Varying AR Coefficient Modeling of Nonstationary Covariance Time Series, *IEEE Trans.*, AC-30, 1985, p. 48-56.
- [17] G. KITAGAWA et W. GERSCH, A Smoothness Priors Long AR Model Method for Spectral Estimation, *IEEE Trans.*, AC-30, 1985, p. 57-65.
- [18] M. Z. NASHED, Operator-Theoretic and Computational Approaches to Ill-Posed Problems with Applications to Antenna Theory, *IEEE Trans.*, AP-29, 1981, p. 220-231.
- [19] D. TUFTS et R. KUMARESAN, Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform Like Maximum Likelihood, *Proceedings IEEE*, 70, 1982, p. 975-989.
- [20] M. VERHAEGHEN et P. VAN DOOREN, Numerical aspects of different Kalman filter implementations, *IEEE Trans.*, AC-31, 1986, p. 907-917.