

OPTIMISATION DU COMPROMIS BIAIS-VARIANCE  
DANS L'ESTIMATION D'UNE MATRICE INTERSPECTRALE

B. LUMEAU\* et H. CLERGEOT\*\*

\*L2S C.N.R.S.-E.S.E., Plateau du Moulon, 91190 Gif sur Yvette

\*\*LESIR E.N.S.T., 61 Av. du Pr. Wilson, 94230 Cachan

**RESUME** - Nous présentons une méthode non paramétrique pour le calcul de la matrice interspectrale de signaux stationnaires permettant de minimiser l'erreur totale d'estimation, au moins dans le cas de signaux à spectres lentement variables. La méthode repose sur un algorithme récursif de blanchiment et de décorrélation de signaux stationnaires, qui se ramène à chaque étape au calcul des autospectres et de l'interspectre de deux signaux par un lissage fréquentiel avec adaptation automatique de la fenêtre de pondération. Cette méthode tout en assurant le caractère défini positif de la matrice interspectrale permet d'optimiser le compromis biais-variance sur chacun de ses éléments.

**SUMMARY** - A non-parametric method is presented for the computation of the interspectral matrix of stationary signals, intended to minimize the total estimation error, at least in the case of slowly varying spectra. The method is based on a recursive algorithm which achieves, by successive linear transforms, whitening and decorrelation of the signals. Each step involves the estimation of two spectra and one interspectrum, using frequency smoothing with an automatic, optimal choice of the smoothing window. This method preserves the definite non-negative property of the matrix and achieves an optimization of the trade-off between bias and variance.

1. INTRODUCTION

L'une des méthodes les plus classiques en analyse spectrale non paramétrique, le lissage du périodogramme, impose quelques remarques quant au choix de la fenêtre de lissage et à sa largeur. En effet, lorsque nous voulons estimer une matrice interspectrale  $\hat{\Gamma}(m)$ , en discret, à partir des transformées discrètes de Fourier (T.F.D)  $\{\hat{x}_i(m)\}$  d'un ensemble de signaux, nous formons la "matrice périodogramme", dont l'espérance mathématique est voisine de la densité spectrale  $\Gamma(m)$  :

$$\left\{ \begin{aligned} \hat{\Pi}(m) &= \hat{x}(m) \hat{x}^+(m) / N, & \left\{ \begin{aligned} \hat{x}(m) &= [\hat{x}_1(m) \dots \hat{x}_K(m)]^T \\ m &\in \{1, N\}, \end{aligned} \right. \\ E[\hat{\Pi}(m)] &\cong \hat{\Gamma}(m). \end{aligned} \right. \quad (1)$$

Pour réduire la variance on calcule une estimée de  $\hat{\Gamma}(m)$  par la relation de lissage fréquentiel [1] :

$$\hat{\Gamma}(m) = \sum_{l=1}^L \hat{w}(l) \hat{\Pi}(m-l), \quad \sum_{l=1}^L \hat{w}(l) = 1. \quad (2)$$

Les matrices  $\hat{\Pi}(m)$  étant définies non négatives, la même propriété pour  $\hat{\Gamma}(m)$  est garantie si les  $\hat{w}(l)$  sont positifs.

Plus la largeur de la fenêtre de pondération est grande plus l'effet de "lissage" est important : la variance est réduite mais les variations rapides du spectre sont perdues. Le choix de la largeur résulte donc forcément d'un compromis entre variance et biais d'estimation qui est, le plus souvent, obtenu de manière empirique.

Dans le cas scalaire et plus particulièrement dans le cas des autospectres plusieurs approches permettant d'optimiser le compromis biais-variance ont déjà été proposées [1]. Nous donnons, dans la première partie, une présentation générale de l'approche de ce problème d'optimisation. Nous montrons, dans la deuxième partie, par quelle transformation cette approche peut être appliquée au cas des autospectres. Nous étendons, dans la troisième partie, ces résultats au cas de l'interspectre de deux signaux en utilisant la notion de cohérence. L'extension directe au cas de la matrice interspectrale ne peut être envisagée sans risque de perdre le caractère défini positif d'une telle matrice. Afin de répondre à ce problème nous proposons, dans la quatrième partie, un algorithme de blanchiment et de décorrélation récursif.

II - METHODE DE LISSAGE OPTIMAL

La méthode de lissage optimal est introduite pour un modèle statistique consistant en un signal  $\hat{s}(m)$  à estimer perturbé par un bruit blanc additif  $\hat{b}(m)$  de variance connue  $\sigma^2$ . Nous montrerons dans les paragraphes suivants comment on se ramène à un tel modèle à partir du périodogramme d'un signal scalaire stationnaire ou de "l'interpériodogramme" de deux signaux. La partie signal est estimée de manière paramétrique : nous représentons  $\hat{s}(m)$  par un développement tronqué sur une base orthonormée dont l'ordre est estimé à l'aide d'un critère d'erreur quadratique minimale. L'expression de cette erreur totale dans la base fait apparaître deux termes : un terme de variance qui croît en fonction de l'ordre et un terme de biais qui décroît en fonction de l'ordre. La minimisation de cette erreur convenablement estimée nous donne la valeur optimale du rang et par suite la largeur optimale de la fenêtre de lissage. Ainsi, nous définissons un test automatique assurant le meilleur compromis biais-variance.

A partir de M observations (l'indice fréquence sera omis), le signal observé s'écrit sous forme vectorielle :

$$\hat{u} = \hat{s} + \hat{b}, \quad \hat{u} = [u_1, \dots, u_M]^T. \quad (3)$$

Décomposons chacun des signaux composant (3) dans la base orthonormée  $\{\underline{v}(0), \underline{v}(1), \dots, \underline{v}(Q-1)\}$  :

$$\left\{ \begin{aligned} \hat{u} &= \sum_{l=0}^{M-1} \alpha(l) \underline{v}(l), \\ \hat{s} &= \sum_{l=0}^{M-1} a(l) \underline{v}(l), \\ \hat{b} &= \sum_{l=0}^{M-1} \beta(l) \underline{v}(l). \end{aligned} \right. \quad (4)$$

Nous en déduisons entre les projections la relation :

$$\alpha(l) = a(l) + \beta(l). \quad (5)$$

La transformation entre  $b(l)$  et  $\beta(l)$  étant orthogonale il s'en suit que  $\beta(l)$  est lui même un bruit blanc avec :

$$E\{|\beta(l)|^2\} = E\{|b(l)|^2\} = \sigma^2. \quad (6)$$

2.1. Développement paramétrique tronqué à l'ordre  $p < M$

Par une connaissance a priori du signal, la base peut être choisie pour que la puissance du signal soit concentrée



sur les premières composantes  $\alpha(p)$ . C'est le cas en présence d'un signal à variations lentes si l'on utilise comme transformation la T.F.D., où les vecteurs de base sont des exponentielles de fréquence croissante. Le signal peut alors être estimé sous forme d'un développement tronqué :

$$\hat{\underline{s}}(p) = \sum_{l=0}^{p-1} \hat{\alpha}(l) \underline{v}(l), \quad (7)$$

en minimisant la distance  $\|\hat{\underline{s}}(p) - \hat{\underline{u}}\|$  à l'observation  $\hat{\underline{u}}$ , ce qui conduit à la solution :

$$\hat{\alpha}(l) = a(l) + \beta(l), \quad l \in \{0, \dots, p-1\}. \quad (8)$$

L'écart entre la meilleure estimée  $\hat{\underline{s}}(p)$  et le signal  $\underline{s}$

est alors :

$$\hat{\underline{s}}(p) - \underline{s} = \sum_{l=0}^{p-1} \beta(l) \underline{v}(l) - \sum_{l=p}^{M-1} a(l) \underline{v}(l). \quad (9)$$

Dans ces conditions l'erreur totale d'estimation est :

$$Q(p) = E\{|\hat{\underline{s}}(p) - \underline{s}|^2\} = p \sigma^2 + \sum_{l=p}^{M-1} |a(l)|^2. \quad (10)$$

Cette expression fait apparaître deux termes :

- un terme  $p\sigma^2$  de variance, croissant linéairement en fonction de  $p$ ,
- un terme  $\sum |a(l)|^2$  de biais, décroissant en fonction de  $p$ .

La valeur optimale  $L$  de l'ordre  $p$  du développement serait celle minimisant  $Q(p)$ .

## 2.2. Optimisation du compromis biais-variance

D'après (5) et compte tenu du fait que les  $\{a(l)\}$  sont inconnus, une estimation non biaisée de  $Q(p)$  nous est donnée en remplaçant dans (10)  $|a(l)|^2$  par  $|\alpha(l)|^2 - \sigma^2$  :

$$\hat{Q}(p) = (2p-M) \sigma^2 + \sum_{l=p}^{M-1} |\alpha(l)|^2. \quad (11)$$

La valeur  $\hat{L}$  de  $p$  minimisant cette expression sera utilisée comme estimée de la largeur optimale de la fenêtre de troncature.

## 2.4. Détermination d'une fenêtre optimum

Un estimateur de  $\underline{s}$  plus général peut-être obtenu par fenêtrage de  $\hat{\underline{u}}$ .

$$\hat{\underline{s}} = \sum_{l=0}^{M-1} w(l) \alpha(l) \underline{v}(l). \quad (12)$$

L'erreur totale d'estimation sur  $\hat{\underline{s}}$  est donnée d'après

$$(12) \text{ par :} \quad (13)$$

$$\hat{Q}(w(\cdot)) = E\{|\hat{\underline{s}} - \underline{s}|^2\} = \sigma^2 \sum_{l=0}^{M-1} w^2(l) + \sum_{l=0}^{M-1} |a(l)|^2 (w(l)-1)^2.$$

Il est facile de vérifier d'après (13) que le minimum de  $\hat{Q}(w(\cdot))$  est obtenu pour :

$$w(l) = \frac{|a(l)|^2}{|a(l)|^2 + \sigma^2}. \quad (14)$$

D'après (14)  $w(l)$  tend vers 1 dès que  $|a(l)|^2$  est significativement plus grand que  $\sigma^2$  ce qui se produit à peu près à partir de la valeur  $\hat{L}$  détectée par le critère  $\hat{Q}(p)$ . En pratique la largeur de  $w(l)$  est choisie égale à  $\hat{L}$  et la valeur de la coupure est choisie en fonction des variations de  $\alpha^2(l)/\sigma^2$  au voisinage de  $\hat{L}$ . Dans le cas où la T.F.D. est utilisée comme transformation orthogonale, le produit par  $w(l)$  de  $\alpha(l)$  correspond à la convolution ("le lissage") de  $\alpha(l)$  par la transformée  $\hat{w}(l)$  qui est bien ainsi déterminée de façon optimale.

## III. LISSAGE OPTIMAL POUR LES AUTOSPECTRES

Il est bien connu que la moyenne et l'écart type du périodogramme sont proportionnels à la densité spectrale : le périodogramme s'interprète comme le spectre à estimer noyé dans un "bruit multiplicatif". On retrouve un modèle avec "bruit additif" pour le Log-périodogramme [2]. On constate cependant une augmentation notable de variance à fenêtre donnée si le spectre est estimé par lissage du Log-périodogramme. Pour éviter cet inconvénient nous avons proposé [3] de faire un "prémoyennage" sur le périodogramme avant de prendre le logarithme.

Pour une raison de place nous ne ferons qu'exposer de façon schématique les étapes du passage permettant d'aboutir à un modèle analogue à celui du paragraphe II pour le calcul de l'autospectre. Les approximations successives sont les suivantes :

- a) les composantes  $\hat{x}(m)$  de la T.F.D. sont des variables complexes gaussiennes centrées décorréelées avec partie réelle et imaginaire décorréelées et de même variance  $N\gamma(m)/2$ ;
- b) le périodogramme prémoyenné est défini par

$$\hat{\pi}_p(m) = \frac{1}{NP} \sum_{p=m-P/2}^{m+P/2} |\hat{x}(m-p)|^2 \quad (15)$$

et se met sous la forme :

$$\hat{\pi}_p(m) = \tilde{\gamma}_p(m) \chi_{2p}^2(m) \quad (16)$$

où  $\chi_{2p}^2(m)$  est une variable en chi-deux d'ordre  $2P$ ;  $\tilde{\gamma}_p(m)$  reste voisin de  $\gamma(m)$  si  $P$  est petit;

- c) le Log-périodogramme s'écrit :

$$\text{Log } \hat{\pi}_p(m) = \text{Log } \tilde{\gamma}_p(m) + \text{Log } \chi_{2p}^2(m), \quad (17)$$

où

$$\begin{cases} E[\text{Log } \chi_{2p}^2(m)] = \eta_p, \\ E[\text{Log } \chi_{2p}^2(m) - \eta_p]^2 = \frac{1}{P} (1 + \delta_p); \end{cases} \quad (18)$$

le coefficient  $\delta_p$  représente l'augmentation de variance liée à l'utilisation du Log-périodogramme, il tend rapidement vers zéro (ainsi que  $\eta_p$ ) lorsque  $P$  augmente ( $P = 1, \delta_p = 0,65; P = 3, \delta_p = 0,18$ ).

Pour obtenir le modèle du paragraphe II nous posons:

$$\begin{cases} \hat{u}(m) = \text{Log } \hat{\pi}_p(m) - \eta_p = \hat{s}(m) + \hat{b}(m), \\ \hat{s}(m) = \text{Log } \tilde{\gamma}_p(m), \\ \hat{b}(m) = \text{Log } \chi_{2p}^2(m) - \eta_p, \quad \sigma_b^2 \approx \frac{1}{P} \quad (P \geq 3). \end{cases} \quad (19)$$

Pour former  $\hat{u}$  on prend les valeurs de  $\hat{u}(m)$  de  $P$  en  $P$  pour garder l'indépendance des composantes.

Pour a) et b) on fait l'hypothèse que le spectre étudié est "lentement variable", dans le sens où les variations relatives de  $\gamma(m)$  à  $\gamma(m+1)$  sont faibles. L'approximation a) est universellement connue et utilisée par tous les statisticiens; elle est discutée en détail dans [3]. L'approximation b) utilise le caractère gaussien de  $\hat{x}(m)$ , qui n'impose pas que le signal  $\hat{x}(m)$  soit lui-même gaussien [4]. Le point c) est discuté dans [3].

## IV. LISSAGE OPTIMAL POUR DES INTERSPECTRES

Dans ce paragraphe nous montrons comment, à partir de "l'interpériodogramme" et pour estimer un interspectre, il est possible, en utilisant la notion de cohérence, de se ramener au modèle défini par la relation (3).

Si nous considérons deux signaux  $\hat{x}_i(m)$  et  $\hat{x}_j(m)$ , leurs autospectres respectifs  $\hat{\gamma}_{ii}(m)$  et  $\hat{\gamma}_{jj}(m)$  et leur interspectre



$\hat{\gamma}_{ij}(m)$ , nous savons que la cohérence de ces signaux est définie par :

$$\hat{c}_{ij}(m) = \hat{\gamma}_{ij}(m) / (\hat{\gamma}_{ii}(m) \hat{\gamma}_{jj}(m))^{1/2} \quad (20)$$

Nous pouvons obtenir une estimation de la cohérence :

$$\hat{c}_{ij}(m) = \hat{x}'_i(m) \hat{x}_j^*(m), \quad (21)$$

avec

$$\hat{x}'_i(m) = \hat{x}_i(m) / \hat{\gamma}_{ii}^{\Delta 1/2}(m), \quad \hat{x}_j(m) = \hat{x}_j(m) / \hat{\gamma}_{jj}^{\Delta 1/2}(m) \quad (22)$$

Si comme précédemment nous supposons que les variables  $\hat{x}_i(m)$  et  $\hat{x}_j(m)$  sont gaussiennes, centrées, circulaires et si l'on pose

$$\tilde{c}_{ij}(m) = E[\hat{c}_{ij}(m)], \quad (23)$$

il est facile de montrer que :

$$E\{|\hat{c}_{ij}(m)|^2\} = |\tilde{c}_{ij}(m)|^2 + 1 \quad (24)$$

Ainsi, l'estimateur de la cohérence peut se mettre sous la forme :

$$\hat{c}_{ij}(m) = \tilde{c}_{ij}(m) + \hat{\epsilon}(m) \quad (25)$$

où  $\hat{\epsilon}(m)$  est un bruit blanc centré de variance unité.

Par suite, si nous faisons un lissage de "l'interpériodogramme" sur P composantes et prenons les valeurs de P en P, nous sommes ramenés aux résultats du paragraphe précédent, à savoir :

$$\hat{u}(l) = \hat{s}(l) + \hat{b}(l),$$

expression dans laquelle  $\hat{u}(l)$  représente une estimation de la cohérence à la fréquence l,  $\hat{s}(l)$  la cohérence moyenne que nous voulons estimer et  $\hat{b}(l)$  un bruit blanc centré de variance connue 1/P.

Pour tenir compte d'un éventuel retard entre les signaux la méthode de détermination de  $\hat{w}(l)$  peut-être modifiée: en accord avec (14) un algorithme recherche les zones où les valeurs  $\alpha(l)$  sont significativement supérieures au bruit et en déduit les valeurs  $\hat{w}(l)$ .

### V. LISSAGE OPTIMAL POUR UNE MATRICE INTERSPECTRALE

Dans ce paragraphe nous présentons un algorithme itératif d'estimation spectrale multivariable. Cet algorithme a pour tâche d'optimiser l'estimation de chacun des éléments d'une matrice interspectrale  $\hat{\Gamma}(m)$  en utilisant les résultats précédents (obtenus dans le cas scalaire) tout en maintenant le caractère défini non négatif de la matrice.

Dans le cas scalaire la recherche de la densité spectrale d'un signal  $\hat{x}(m)$  peut se ramener à la recherche du filtre blanchisseur  $\hat{g}(m)$  transformant  $\hat{x}(m)$  en  $\hat{\epsilon}(m)$  tel que :

$$\hat{\gamma}_{xx}(m) = \hat{\gamma}_{\epsilon\epsilon}(m) [|\hat{g}(m)|^2]^{-1}, \quad (26)$$

où  $\hat{\epsilon}(m)$  est un bruit blanc de variance unité.

Dans le cas multivariable la recherche de la matrice interspectrale d'un signal vectoriel  $\hat{x}(m)$  peut se ramener à la recherche de la matrice "blanchisseuse"  $\hat{G}(m)$  telle que

$$\hat{\Gamma}_{xx}(m) = E[\hat{x}(m) \hat{x}^+(m)] = \hat{G}^{-1}(m) \hat{\Gamma}_{\epsilon\epsilon}(m) [\hat{G}^+(m)]^{-1} \quad (27)$$

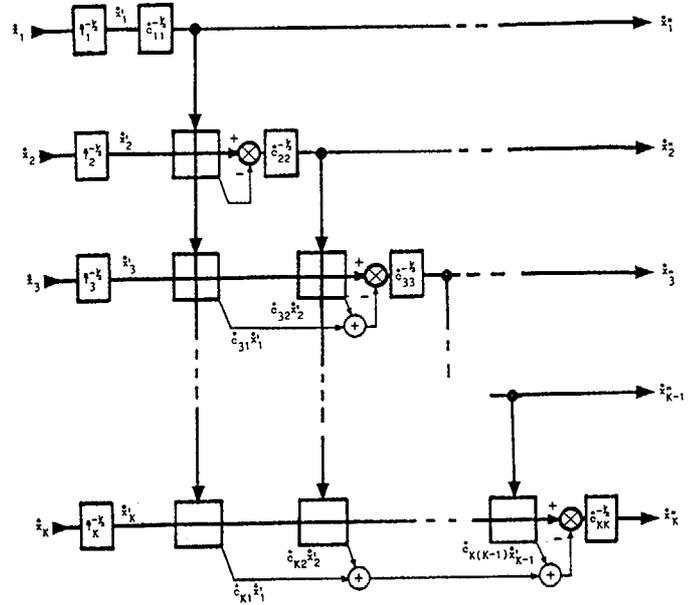
où  $E[\hat{\epsilon}(m) \hat{\epsilon}^+(m)]$  est égal à la matrice identité.

En fait, il existe une infinité de matrice  $\hat{G}(m)$  répondant à la question, cette infinité correspondant à l'ensemble des factorisations possibles de la matrice  $\hat{\Gamma}_{xx}(m)$ . L'algorithme que nous proposons revient à factoriser  $\hat{\Gamma}_{xx}(m)$  sous la forme du produit de deux matrices triangulaires dans

lesquelles les termes diagonaux sont réels et positifs ce qui assure l'unicité de la factorisation.

Formellement l'algorithme est équivalent à un algorithme d'orthogonalisation de Gram-Schmidt pour une série de K vecteurs [5].

La méthode consiste à blanchir et à décorréler les signaux par transformations linéaires successives comme le montre le schéma ci-dessous (l'indice fréquence sera omis par la suite pour alléger les notations).



La première étape de cet algorithme consiste à blanchir les K signaux obtenus par T.F.D. :

$$\hat{x}'_i = \hat{x}_i / \hat{\gamma}_{ii}^{1/2}, \quad i \in \{1, K\}, \quad (28)$$

l'estimation des  $\hat{x}'_i$  se déduisant de celle des  $\hat{\gamma}_{ii}$  selon la méthode proposée au paragraphe II.

La deuxième étape consiste à extraire de chaque signal  $\hat{x}'_i$  la partie décorrelée de l'ensemble des autres signaux à l'aide des cohérences. Par exemple, à partir de la connaissance de la cohérence entre les signaux  $\hat{x}'_2$  et  $\hat{x}'_1$  :

$$\hat{c}_{21} = E\{\hat{x}'_2 \hat{x}'_1^*\}, \quad (29)$$

cette cohérence étant estimée selon la méthode proposée au paragraphe III, nous pouvons écrire :

$$\hat{x}'_2 = \hat{c}_{21} \hat{x}'_1 + \hat{x}_{d2}, \quad (30)$$

expression dans laquelle  $\hat{x}_{d2}$  est la partie du signal  $\hat{x}'_2$  décorrelée avec  $\hat{x}'_1$ .

La troisième étape consiste à blanchir les signaux ainsi décorrelés :

$$\hat{x}''_i = \hat{x}_{di} / \hat{\gamma}_{di}^{1/2}, \quad i \in \{1, K\}. \quad (31)$$

En réitérant ce raisonnement sur l'ensemble des signaux d'entrée nous obtenons :

$$\hat{x}'_j = \hat{c}_{jj} \hat{x}''_j + \sum_{i=1}^{j-1} \hat{c}_{ji} \hat{x}''_i = \sum_{i=1}^j \hat{c}_{ji} \hat{x}''_i \quad (32)$$

Si nous posons :

$$\begin{cases} \hat{\underline{x}} = [\hat{x}'_1, \dots, \hat{x}'_K]^T, \\ \hat{\underline{x}}' = [\hat{x}''_1, \dots, \hat{x}''_K]^T, \\ \hat{\underline{c}} = [\hat{c}_{11}, \dots, \hat{c}_{KK}]^T, \end{cases} \quad (33)$$

