

SPECTRE DU MAXIMUM D'ENTROPIE 1D DANS LE CAS D'UN JEU DE CORRELATIONS LACUNAIRE.

*Christophe LE MARTRET **, *Georges VEZZOSI ***

* CESTA 37 Avenue du Général de Gaulle 35170 BRUZ.

** Laboratoire Traitement du Signal et de l'Image 35042 Rennes Cedex.

RÉSUMÉ

On présente un algorithme primal de calcul du spectre du maximum d'entropie 1D dans le cas d'un jeu de corrélations lacunaire. Puis, on précise une règle de sélection de M corrélations parmi N qui fournit le spectre le plus proche de celui utilisant la totalité de corrélations disponibles. Cette règle s'applique quelle que soit la dimension.

1 - INTRODUCTION

A une dimension, lorsque l'on dispose d'un jeu de $N+1$ corrélations contiguës définies sur l'ensemble des $N+1$ premiers entiers, on sait que le Spectre du Maximum d'Entropie (SME) est identique au spectre AR(N). Le calcul du SME est alors ramené à la résolution d'un problème linéaire pour lequel il existe un algorithme rapide qui exploite la structure de Toeplitz de la matrice de corrélation : l'algorithme de Levinson. Lorsque les corrélations ne sont plus contiguës, le calcul du SME 1D revient à maximiser une fonctionnelle non linéaire. Les différents algorithmes qui ont été proposés [1][2] appartiennent à la classe des algorithmes de descente à base de gradient.

Dans le cadre de l'approche donnée par Rozario et Papoulis [2], nous développons, dans la première partie, une méthode originale permettant de calculer de façon exacte le hessien de la fonctionnelle à maximiser en utilisant la formule de Gohberg. La fonctionnelle étant strictement concave, le problème est alors résolu en peu d'itérations par un algorithme de Newton à pas contrôlé.

Dans la deuxième partie, nous abordons le problème du choix d'un jeu de $M < N$ corrélations parmi N : étant donné un ensemble de $N+1$ corrélations défini sur les $N+1$ premiers entiers, comment choisir, à $M < N$ fixé, les M corrélations de retard non nul donnant le meilleur SME associé ? Nous décrivons une méthode originale donnant des spectres plus résolants que ceux obtenus par la méthode décrite dans [3].

2 - UN ALGORITHME DE MAXIMUM D'ENTROPIE 1D DANS LE CAS D'UN JEU DE CORRELATIONS LACUNAIRE

2.1 - Principe

Soit \mathcal{D}_M une partie finie des entiers > 0 contenant M éléments, et considérons un jeu de corrélations complexes défini sur \mathcal{D}_M complété par l'origine :

$$\{r(0), (r(n), n \in \mathcal{D}_M)\}. \quad (1)$$

ABSTRACT

A primal method is presented for determining the maximum entropy spectrum for a set of 1D non consecutive correlations. Then, a rule is given for selecting a set of M correlations among N , which yields the spectrum which best fits the one calculated with the whole set of correlations. This rule can be applied for all dimensions.

Supposant ce jeu extensible [1], on se propose de calculer le SME associé. On sait que la solution s'exprime comme l'inverse d'un polynôme trigonométrique strictement positif :

$$S(f) = \frac{1}{P(f, \underline{\lambda})},$$

$$P(f, \underline{\lambda}) = \lambda(0) + 2 \operatorname{Re} \sum_{n \in \mathcal{D}_M} \lambda(n) e^{-2i\pi f n}. \quad (2)$$

Elle peut être obtenue pratiquement en appliquant les méthodes générales de dualité. On minimise alors la fonction convexe [1] :

$$J(\underline{\lambda}) = \frac{1}{2} \left\langle \int_{-1/2}^{1/2} \log \frac{1}{P(f, \underline{\lambda})} df + \langle \underline{\lambda}, \underline{r} \rangle - 1 \right\rangle, \quad (3)$$

$$\text{où } \langle \underline{\lambda}, \underline{r} \rangle = \lambda(0) r(0) + 2 \operatorname{Re} \sum_{n \in \mathcal{D}_M} \bar{\lambda}(n) r(n),$$

par rapport aux $2M+1$ inconnues réelles :

$$\{\lambda(0), (\operatorname{Re} \lambda(n), \operatorname{Im} \lambda(n), n \in \mathcal{D}_M)\}.$$

La minimisation s'opère classiquement par des méthodes de descente à base de gradient. Le gradient et le hessien sont alors calculés par des FFT 1D. Cette méthode peut être appliquée quelle que soit la dimension.

Dans le cas d'une seule dimension, il existe également une solution primale qui n'opère que sur un nombre fini de corrélations. Soit N l'entier le plus grand contenu dans \mathcal{D}_M , \mathcal{D} l'ensemble des N premiers entiers strictement positifs, \mathcal{D}_g la différence $\mathcal{D} \setminus \mathcal{D}_M$. Comme le polynôme $P(f, \underline{\lambda})$ est une fonction d'une variable unique, il s'exprime d'après le théorème de Fejer-Riesz comme le module au carré d'un polynôme de degré N . La fonction $S(f)$ s'écrit alors :



$$S(f) = \frac{\sigma^2}{\left| 1 + \sum_{n=1}^N a(n) e^{-2i\pi fn} \right|^2},$$

ce qui montre que le SME 1D appartient à la classe des spectres AR(N). On sait que l'entropie d'un tel spectre vaut :

$$H = \frac{1}{2} \int_{-1/2}^{1/2} \log S(f) df = \frac{1}{2} \log \sigma^2.$$

A partir des équations normales :

$$\underline{a} \underline{R} = \sigma^2 \underline{\epsilon}_1^T,$$

où

$$\underline{a} = [1, a(1), \dots, a(N)],$$

$$\underline{\epsilon}_1 = [1, 0, \dots, 0]^T,$$

$$\underline{R} = \begin{pmatrix} r(0) & r(1) & \dots & r(N) \\ \bar{r}(1) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & r(1) \\ \bar{r}(N) & \dots & \bar{r}(1) & r(0) \end{pmatrix},$$

on montre facilement que l'entropie du spectre AR(N) s'écrit :

$$H = \frac{1}{2} \log \sigma^2 = -\frac{1}{2} \log \underline{\epsilon}_1^T \underline{R}^{-1} \underline{\epsilon}_1. \quad (4)$$

La relation (4) exprime l'entropie du SME 1D comme une fonction des corrélations connues $\{r(n), n \in \mathcal{D}_M\}$ et des corrélations inconnues $\{r(n), n \in \mathcal{D}_g\}$. Le problème est alors résolu en maximisant (4) par rapport aux corrélations inconnues.

2.2 - Calcul du gradient et du hessien

Si $\{r(n) = x(n) + i y(n), n \in \mathcal{D}_g\}$ représente les corrélations manquantes, on se propose ici de calculer les dérivées premières et secondes de H par rapport aux variables réelles $\{x(n), y(n), n \in \mathcal{D}_g\}$. On montre que ces dérivées s'expriment sous la forme [5]:

$$\begin{aligned} \frac{\partial H}{\partial x(n)} &= 2 \operatorname{Re} \left(\frac{\partial H}{\partial r(n)} \right), & \frac{\partial H}{\partial y(n)} &= -2 \operatorname{Im} \left(\frac{\partial H}{\partial r(n)} \right), \\ \frac{\partial^2 H}{\partial x(m) \partial x(n)} &= 2 \operatorname{Re} \left(\frac{\partial^2}{\partial r(m) \partial r(n)} + \frac{\partial^2}{\partial r(m) \partial \bar{r}(n)} \right) H, \\ \frac{\partial^2 H}{\partial y(m) \partial y(n)} &= -2 \operatorname{Re} \left(\frac{\partial^2}{\partial r(m) \partial r(n)} + \frac{\partial^2}{\partial r(m) \partial \bar{r}(n)} \right) H, \\ \frac{\partial^2 H}{\partial x(m) \partial y(n)} &= -2 \operatorname{Im} \left(\frac{\partial^2}{\partial r(m) \partial r(n)} - \frac{\partial^2}{\partial r(m) \partial \bar{r}(n)} \right) H. \end{aligned} \quad (5)$$

Il suffit alors d'évaluer les trois dérivées suivantes :

$$\frac{\partial H}{\partial r(n)}, \quad \frac{\partial^2 H}{\partial r(m) \partial r(n)}, \quad \frac{\partial^2 H}{\partial r(m) \partial \bar{r}(n)}.$$

Soit \underline{Z}_n désigne la matrice de décalage :

$$0 \leq i, j \leq N, \quad -N \leq n \leq N, \quad \underline{Z}_n(i, j) = \begin{cases} 1 & \text{si } j-i = n \\ 0 & \text{sinon} \end{cases}$$

$$\text{comme } \frac{\partial \underline{R}^{-1}}{\partial r(n)} = -\underline{R}^{-1} \underline{Z}_n \underline{R}^{-1},$$

on trouve en dérivant (4) et en utilisant les équations normales :

$$\frac{\partial H}{\partial r(n)} = -\frac{1}{2\sigma^2} \underline{a} \underline{Z}_n \underline{a}^*, \quad (6)$$

ce qui montre que le gradient de H s'exprime avec l'autocorrélation du vecteur des coefficients de prédiction.

En dérivant une seconde fois par rapport à $r(m)$, on obtient :

$$\begin{aligned} \frac{\partial^2 H}{\partial r(m) \partial r(n)} &= \frac{1}{2} [p(m, n) + p(n, m)] + 2 \frac{\partial H}{\partial r(m)} \cdot \frac{\partial H}{\partial r(n)}, \\ \frac{\partial^2 H}{\partial r(m) \partial \bar{r}(n)} &= \frac{1}{2} [p(m, -n) + p(-n, m)] + 2 \frac{\partial H}{\partial r(m)} \cdot \frac{\partial H}{\partial \bar{r}(n)}, \end{aligned} \quad (7)$$

où

$$p(m, n) = -\frac{1}{\sigma^2} \underline{a} \underline{Z}_m \underline{R}^{-1} \underline{Z}_n \underline{a}^*. \quad (8)$$

Calcul de $p(m, n)$

Disposant des coefficients de prédiction à l'ordre N, on peut exprimer dans (8) \underline{R}^{-1} par la formule de Gohberg [4]:

$$\underline{R}^{-1} = \frac{1}{\sigma^2} (\underline{A}_1 \underline{A}_1^* - \underline{A}_3 \underline{A}_3^*),$$

où

$$\underline{A}_1 = \begin{pmatrix} 1 & & & & & \\ \bar{a}_1 & \ddots & & \mathbf{0} & & \\ \vdots & \ddots & \ddots & & & \\ \vdots & & & \ddots & & \\ \bar{a}_N & \dots & \dots & \bar{a}_1 & 1 & \end{pmatrix},$$

$$\underline{A}_3 = \begin{pmatrix} 0 & & & & & \\ \bar{a}_N & \ddots & & \mathbf{0} & & \\ \vdots & \ddots & \ddots & & & \\ \vdots & & & \ddots & & \\ \bar{a}_1 & \dots & \dots & \bar{a}_N & 0 & \end{pmatrix}.$$

En posant

$$\underline{\epsilon}_m = \underline{A}_1^* \underline{Z}_m \underline{a}^*, \quad \underline{\epsilon}_m = \underline{A}_3^* \underline{Z}_m \underline{a}^*, \quad (9)$$

(8) s'écrit alors :

$$p(m,n) = -\frac{1}{\sigma^4} (\underline{c}_{-m}^* \underline{c}_n - \underline{d}_{-m}^* \underline{d}_n). \quad (10)$$

Les relations (5-7) et (9-10) résolvent le problème posé.

2.3 - L'algorithme

Nous décrivons ici un algorithme de Newton à pas contrôlé qui converge beaucoup plus rapidement que l'algorithme du gradient. En contrôlant correctement le pas, la convergence globale de l'algorithme est assurée du fait de la concavité stricte de la fonction H [5] par rapport aux corrélations inconnues.

Si $\underline{u}_k = [\underline{x}_k(n), \underline{y}_k(n)]^T$, $n \in \mathcal{D}_g$ désigne le vecteur des parties réelles et imaginaires des corrélations inconnues, \underline{g}_k le gradient et \underline{A}_k le hessien à l'étape k , l'algorithme se décrit alors de la manière suivante :

- 1) Initialisation
- 2) Calcul des dérivées premières et secondes
- 3) Test d'arrêt, si $\|\underline{g}_k\| < \epsilon$, aller en 9)
- 4) Calcul du vecteur de montée $\underline{d}_k = -\underline{A}_k^{-1} \underline{g}_k$
- 5) Calcul du pas α_k qui maximise $H(\underline{u}_k + \alpha_k \underline{d}_k)$
- 6) Accroissement du vecteur des corrélations inconnues
 $\underline{u}_{k+1} = \underline{u}_k + \alpha_k \underline{d}_k$
- 7) Calcul des coefficients de prédiction et de σ^2 par Levinson
- 8) Aller en 2)
- 9) Fin

Remarque sur le calcul du pas : L'étape 5) nécessite quelques précautions. En effet, la matrice de corrélations \underline{R} construite avec le vecteur $\underline{u}_k + \alpha_k \underline{d}_k$ n'est pas nécessairement positive $\forall \alpha_k$. Comme pour le problème dual [5], on montre que \underline{R} reste positive ssi $\alpha_k \in I = [0, \alpha_a]$ où α_a est une asymptote verticale de la fonction $H(\underline{u}_k + \alpha_k \underline{d}_k)$. α_a n'est pas calculable de manière simple, mais l'algorithme de Levinson (implicitement utilisé lors de cette étape pour le calcul du critère) permet de tester pratiquement l'appartenance de α_k à l'intervalle I en contrôlant le module des coefficients de réflexion. Ceci permet de conserver la positivité de \underline{R} tout au long de l'algorithme.

3 - SELECTION D'UN JEU DE CORRELATIONS

3.1 - Position du problème

Le problème suivant se pose aussi bien en une ou en plusieurs dimensions. Pour la cohérence de l'exposé, nous le présentons ici en 1D, étant entendu que la solution proposée peut être appliquée quelle que soit la dimension.

Soit \mathcal{D} une partie finie des entiers >0 contenant N éléments. Considérons un jeu de corrélations extensible défini sur \mathcal{D} complété par l'origine : $\{r(0), (r(n), n \in \mathcal{D})\}$, et soit $S(f)$ le SME associé. $S(f)$ est l'inverse d'un polynôme, et peut être obtenu en minimisant la fonction convexe (2).

Parmi les corrélations qui servent à construire le SME $S(f)$, toutes ne contribuent pas de façon égalitaire au résultat final. Par exemple, si le signal est composé d'un bruit blanc et d'une somme de fréquences pures, les corrélations associées aux grands retards contribuent le plus à la résolution. A $M < N$ fixé, peut-on identifier de façon générale les M corrélations les plus significatives ? Soit

$\mathcal{D}_M = \mathcal{D} \setminus \mathcal{D}_g$ l'une des C_N^M parties de \mathcal{D} contenant M éléments, et $S_M(f)$ le SME associé. Par construction, l'entropie de $S_M(f)$ est plus grande ou égale à celle de $S(f)$. Nous considérons que le meilleur choix de \mathcal{D}_M est celui qui donne le SME dont l'entropie est la plus proche de celle de $S(f)$. Cette procédure s'apparente

donc à un critère de minimax : parmi les C_N^M dsp d'entropie maximum construites avec tous les sous ensembles \mathcal{D}_M , on retient celle dont l'entropie est la plus petite.

La mise en œuvre de cette procédure suppose toutefois le calcul préalable de C_N^M SME et n'est donc pas envisageable pratiquement. Nous donnons ci-dessous deux solutions approximatives.

3.2 - Approximation quadratique du critère

Considérons le SME $S(f) = [P(f, \underline{\lambda}_0)]^{-1}$ calculé avec la totalité des corrélations disponibles. Au point $\underline{\lambda}_0$, la fonction convexe (2) qui définit le problème dual est alors minimum, et égale à l'entropie de $S(f)$. Elle admet un développement quadratique autour de $\underline{\lambda}_0$:

$$J(\underline{\lambda}) = J(\underline{\lambda}_0) + \frac{1}{2} (\underline{\lambda} - \underline{\lambda}_0)^T \underline{Q} (\underline{\lambda} - \underline{\lambda}_0), \quad (11)$$

où $\underline{\lambda}$ désigne le vecteur réel d'ordre $2N+1$:

$$\underline{\lambda} = [\lambda(0), \{\text{Re } \lambda(n), \text{Im } \lambda(n), n \in \mathcal{D}\}]^T$$

Calculer le SME $S_M(f)$ associé à $\mathcal{D}_M = \mathcal{D} \setminus \mathcal{D}_g$ équivaut à minimiser la fonction (2) dans laquelle les inconnues $\lambda(n)$ associées à \mathcal{D}_g ont été mises à zéro. Au lieu de minimiser cette fonction, on peut aussi minimiser son développement quadratique (11), ce qui conduit à des formules explicites. Par une permutation convenable des inconnues, on peut toujours supposer que les variables mises à zéro sont placées à la fin du vecteur $\underline{\lambda}$. Si nous posons :

$$\underline{\lambda}_0 = [\underline{\lambda}_{0,1}, \underline{\lambda}_{0,2}]^T, \\ \underline{Q} = \begin{pmatrix} \underline{Q}_{1,1} & \underline{Q}_{1,2} \\ \underline{Q}_{2,1} & \underline{Q}_{2,2} \end{pmatrix}, \quad \underline{Q}^{-1} = \begin{pmatrix} \underline{Q}^{1,1} & \underline{Q}^{1,2} \\ \underline{Q}^{2,1} & \underline{Q}^{2,2} \end{pmatrix}$$

un calcul élémentaire montre que le minimum de la forme quadratique (11) quand les $2(N-M)$ dernières composantes du vecteur $\underline{\lambda}$ sont mises à zéro a lieu pour :

$$\underline{\lambda}_{M,1} = \underline{\lambda}_{0,1} + \underline{Q}_{1,1}^{-1} \underline{Q}_{1,2} \underline{\lambda}_{0,2},$$

et que l'accroissement d'entropie correspondant vaut :

$$\Delta J_M = J(\underline{\lambda}_{M,1}, 0) - J(\underline{\lambda}_0) = \frac{1}{2} \underline{\lambda}_{0,2}^T \underline{Q} \underline{\lambda}_{0,2}, \quad (12)$$

$$\text{où } \underline{Q} = [\underline{Q}^{2,2}]^{-1} = \underline{Q}_{2,2} - \underline{Q}_{2,1} \underline{Q}_{1,1}^{-1} \underline{Q}_{1,2}.$$

Ces relations constituent la base d'une méthode possible de sélection des corrélations :

(i) on calcule le SME correspondant à la totalité des corrélations disponibles, puis le hessien du problème dual et son inverse.



(ii) on calcule les C_N^M formes quadratiques (12), et on retient l'ensemble \mathcal{D}_M qui donne le plus petit accroissement ΔJ_M .

Si le calcul des C_N^M formes quadratiques est jugé trop coûteux, on pourra se contenter d'une *procédure diagonale* : étant donné la variable complexe $\lambda(n)$, il lui correspond dans la matrice Q^{-1} un bloc diagonal 2×2 de la forme :

$$Q_n = \begin{pmatrix} q_{1,1} & q_{1,2} \\ q_{2,1} & q_{2,2} \end{pmatrix},$$

auquel correspond l'accroissement d'entropie

$$\alpha_n = \frac{1}{2} \underline{u}_n^T Q_n^{-1} \underline{u}_n,$$

avec $\underline{u}_n = [\text{Re } \lambda(n), \text{Im } \lambda(n)]^T$.

Les corrélations retenues sont alors celles qui fournissent les M accroissements α_n les plus grands. Lorsque le hessien est proche d'une matrice colinéaire à l'identité, ceci revient à sélectionner les M corrélations associées aux multiplicateurs $\lambda(n)$ de plus grand module, et non, comme il est proposé dans [3], les corrélations de plus grand module.

3.3 - Exemple

Soit $r(n)$ la suite des corrélations d'un signal composé de deux sinusoides complexes d'amplitude 1, de fréquences normalisées $f_1=0.25$ et $f_2=0.3$, noyé dans du bruit blanc de variance $\sigma^2=0.2$:

$$r(n) = e^{2i\pi f_1 n} + e^{2i\pi f_2 n} + \sigma^2 \delta(n).$$

Considérons le cas $N=9$, $M=5$, et comparons les résultats obtenus par la méthode du § précédent et celle décrite dans [3]. La méthode de Lagunas sélectionne les corrélations de plus grand module. Dans le cas présent, la suite des modules $|r(n)|$ étant strictement décroissante pour $1 \leq n \leq 9$, le spectre correspondant est le spectre AR(5) calculé avec les 6 premières corrélations. Pour notre méthode, les corrélations retenues sont $r(0)$, $r(2)$, $r(3)$, $r(4)$, $r(5)$, $r(9)$, quelle que soit l'approximation considérée au § 3.2.

La Fig.1 représente les spectres obtenus par les deux méthodes. On constate que la nôtre permet de séparer les deux sources et l'autre pas.

La Fig.2 montre que les spectres $S_5(f)$ et $S(f)$ sont très proches.

4 - CONCLUSION

Etant donné $M+1$ corrélations complexes incluant celle de retard nul, la détermination du spectre d'entropie maximale passant par ces corrélations s'effectue en général en traitant le problème dual. Le spectre s'obtient alors en minimisant une fonction convexe de $2M+1$ variables réelles. Nous avons montré qu'à une dimension, la solution peut être obtenue également en résolvant le problème primal. Si N désigne le retard maximum des corrélations disponibles, on est alors conduit à maximiser une fonction convexe à $2(N-M)$ variables. Les techniques d'optimisation mises en jeu dans les deux cas et les vitesses de convergence étant comparables, on a intérêt à utiliser l'approche primale si $2(N-M) < 2M+1$.

Le problème du choix des corrélations les plus significatives dans les méthodes de maximum d'entropie a été traité en observant qu'ôter des corrélations fait augmenter l'entropie, ce qui suggère d'éliminer les corrélations qui produisent l'augmentation d'entropie

la plus faible. Dans l'approche duale, ôter des corrélations revient à mettre à zéro les multiplicateurs de Lagrange correspondants, et nous avons proposé une procédure approchée qui permet d'identifier les corrélations d'entropie la plus faible. Ceci peut servir à la réduction de modèle, et à la représentation d'un spectre par un nombre minimal de paramètres.

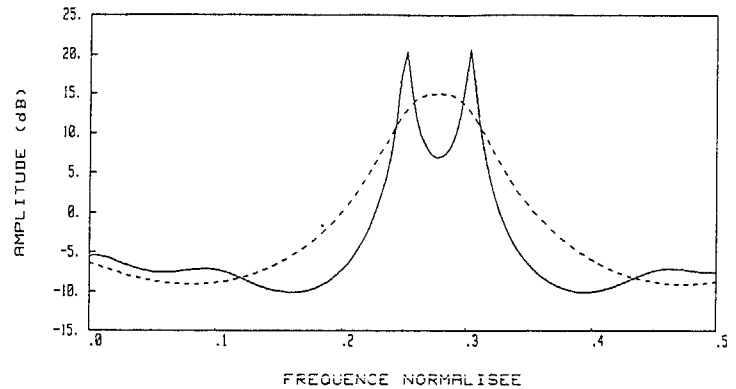


Fig.1. Spectres du maximum d'entropie calculés pour deux jeux de corrélations. En pointillés : spectre calculé avec les corrélations obtenues par la méthode de Lagunas (spectre AR(5)). En trait continu : spectre $S_5(f)$ calculé avec les corrélations obtenues par notre méthode.

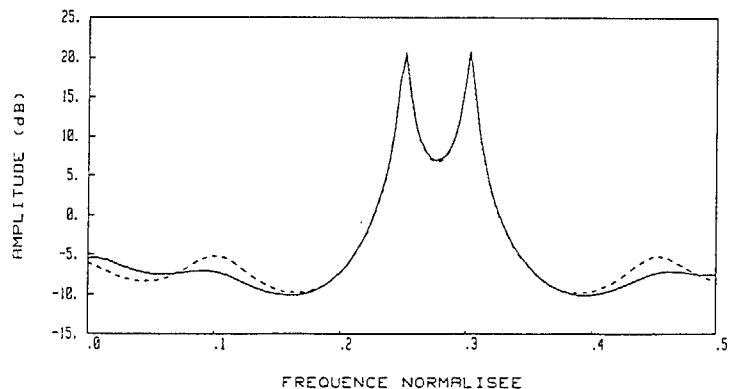


Fig.2. Comparaison du spectre $S_5(f)$ calculé avec les corrélations obtenues par notre méthode (trait continu) et le SME 1D calculé avec toutes les corrélations (pointillés) qui est identique au spectre AR(9).

Références :

- [1] S.W. LANG, J.H. Mc CLELLAN, "Multidimensional Spectral Estimation", IEEE Trans. on ASSP, vol. ASSP-33, no. 6, December 1982.
- [2] N. ROZARIO, A. PAPOULIS, "Spectral Estimation from Non-consecutive Data", IEEE Trans. on IT, vol. IT-33, no. 6, November 1987.
- [3] M.A. LAGUNAS-HERNANDEZ, "Use of Most Significant Autocorrelation Lags in Iterative ME Spectral Estimation", IEEE ASSP, vol. ASSP-32, no. 2, April 1984.
- [4] T. KAILATH, A. VIERA, A. MORF, "Inverse of Toeplitz Operators, Innovations and Orthogonal Polynomials", SIAM Rev. vol. 20, pp. 106-119, January 1978.
- [5] C.J. LE MARTRET, "Contribution à l'estimation spectrale 2D par la méthode du maximum d'entropie", Thèse de l'Université de RENNES I, Novembre 1990.