



## EQUATIONS DE SCHUR et de CHANDRASEKHAR GENERALISEES POUR DES PROBLEMES GENERAUX D'ESTIMATION LINEAIRE.

**Guy DEMOMENT\* et Roger REYNAUD\*\***

(\*) Laboratoire des Signaux et Systèmes (CNRS/ESE) Plateau du Moulon, 91190 Gif-sur-Yvette, France.

(\*\*) Institut d'Electronique Fondamentale, Université de Paris-Sud, 91400 Orsay, France.  
et Gréco SARTA "Systèmes Adaptatifs".

### RESUME

Le problème traité est celui du lien entre les différentes approches (internes, externes, algébriques, géométriques,...) pour l'estimation de processus non-stationnaires par des algorithmes rapides. Ce lien entre les algorithmes du type Levinson ou Schur et ceux du type Chandrasekhar n'a été clairement établi que dans le cas où l'on dispose d'un modèle d'état invariant du processus. L'utilisation d'un modèle d'état dégénéré, c'est-à-dire d'un modèle ne nécessitant que la covariance du processus, permet d'établir simplement ce lien, sans se limiter, comme jusqu'à présent, au cas des processus stationnaires.

### INTRODUCTION

Le problème de l'estimation linéaire au sens des moindres-carrés a été très largement étudié et de nombreuses méthodes ont été proposées. Elles peuvent être séparées en méthodes utilisant des descriptions "internes" ou d'état, et méthodes utilisant des représentations "externes" ou d'entrée-sortie.

Dans le cas d'un processus décrit par un modèle d'état invariant, les algorithmes du type Chandrasekhar permettent d'obtenir une solution rapide au problème du filtrage et du calcul de l'innovation. Dans le cas d'un processus décrit par sa covariance, les algorithmes de la famille Levinson fournissent aussi une solution rapide dans le cas stationnaire, avec un volume d'opérations de l'ordre de  $N^2$ ,  $N$  étant la dimension de la matrice de covariance à inverser.

On a cru pendant longtemps que la prédiction de processus non-stationnaires nécessitait  $O(N^3)$  opérations, à moins de disposer d'un modèle d'état pour le processus étudié. Mais l'introduction du concept de "rang de déplacement" d'une matrice a permis de développer des algorithmes de Levinson généralisés [5] avec une complexité croissante entre les matrices fortement structurées des covariances stationnaires et les matrices complètement arbitraires.

Le lien entre ces algorithmes de Levinson généralisés et les équations de Chandrasekhar a déjà été établi [5], mais dans le cas où l'on dispose d'un modèle d'état invariant du processus, ce qui n'est pas la règle générale, et par des techniques matricielles assez lourdes.

Ce même lien a été éclairci par d'autres techniques, soit géométriques du type "relations en treillis" [1,4], soit algébriques du type "algorithmes de Schur" [3], mais dans le seul cas d'un processus stationnaire.

Nous proposons une autre approche de ces problèmes, fondée sur une description du type "modèle d'état dégénéré" du processus étudié [2]. Ce type de représentation présente un premier intérêt dans des problèmes de traitement d'image où il permet d'obtenir une solution lissée du problème d'estimation en un seul balayage des données. Mais il présente surtout l'intérêt ici de ne pas nécessiter de représentation réellement "interne" du processus tout en conservant la possibilité de développer des équations du type Chandrasekhar.

### SUMMARY

This work deals with the link between the different approaches (internal, external, algebraic, geometrical,...) commonly used in estimation of non-stationary processes with fast least-squares algorithms. The link between different types of algorithms (i.e. Levinson, Schur or Chandrasekhar) has been investigated only when a constant state-space model is available. Use of a degenerate state-space model (i.e. a model requiring only the covariance information) allows us to explain easily the relationship between these algorithms without being limited to stationary processes.

### POSITION DU PROBLEME

Nous considérons le problème de l'estimation d'un processus aléatoire  $x_k$  de dimension  $n$  à partir d'observations d'un processus lié  $y_k$ , de dimension  $p$ . Si l'on suppose connues les matrices de covariances

$$R_{yy}(N) = [r_{yy}(i,j)], \quad r_{yy}(i,j) = E\{y_i y_j^t\}, \quad i,j=1,2,\dots,N \quad (1)$$

$$R_{xy}(N) = [r_{xy}(i,j)], \quad r_{xy}(i,j) = E\{x_i y_j^t\}, \quad i,j=1,2,\dots,N \quad (2)$$

le meilleur estimateur linéaire (au sens des moindres carrés) de  $x_N$  étant donné le passé (c.a.d.  $y_k$   $k=1,2,\dots,N-1$ ) a la forme suivante

$$\hat{x}_{N/N-1} = \sum_{k=1}^{N-1} h_{xy}(k;N) y_k \quad (3)$$

Le prédicteur optimal à un pas (dans le sens de la récursion, c'est-à-dire dans le sens direct) s'obtient en appliquant la condition bien connue d'orthogonalité sur l'erreur de prédiction

$$x_N - \hat{x}_{N/N-1} \perp y_k \quad k=1,2,\dots,N-1 \quad (4)$$

ce qui conduit à l'équation normale

$$h_{xy}(N) R_{yy}(N-1) = [r_{xy}(N,1) \dots r_{xy}(N,N-1)] \quad (5)$$

$$\text{où } h_{xy}(N) = [h_{xy}(N,1) \dots h_{xy}(N,N-1)] \quad (6)$$

est une matrice de dimensions  $n \times (N-1)p$ . On retrouve évidemment les équations habituelles de la prédiction linéaire dans le cas où  $x=y$  :

$$\hat{x}_{N/N-1} = \sum_{k=1}^{N-1} h_{xy}(N,k) y_k \quad (7)$$

Plus généralement, l'estimation de  $x_i$  pour une valeur de l'indice  $i$  dans l'intervalle  $[1,N]$  nécessite de trouver le lisseur optimal  $H_{xy}(i,k;N)$  tel que

$$\hat{x}_{i/N} = \sum_{k=1}^N H_{xy}(i,k;N) y_k \quad (8)$$



ce qui, avec les mêmes conditions d'orthogonalité, conduit à une autre équation normale :

$$H_{xy}(N) R_{yy}(N) = R_{xy}(N) \tag{9}$$

$$\text{où } H_{xy}(N) = [H_{xy}(i,j;N)] \quad i,j=1,2,\dots,N \tag{10}$$

Ces différents problèmes de prédiction ou de lissage ont donc en commun de nécessiter, pour leur résolution, l'inversion d'une matrice de covariance. Quand le processus observé  $y_k$  est stationnaire, cette matrice est Toeplitz et il est bien connu que l'algorithme de Levinson permet de l'inverser rapidement. Quand le processus  $y_k$  n'est pas stationnaire, l'utilisation du concept de rang de déplacement [7] permet de passer progressivement d'une complexité  $O(N^2)$  dans le cas Toeplitz à une complexité  $O(N^3)$  dans le cas le plus général [5]. L'algorithme rapide de calcul récursif des prédicteurs optimaux  $h_{xy}$  fait appel à des quantités auxiliaires qui sont les matrices de prédiction directe et rétrograde du processus observé  $y_k$ . Mais le calcul récursif de ces prédicteurs se fait par un algorithme de Levinson généralisé dont l'établissement utilise des techniques algébriques relativement lourdes [5].

Nous voulons montrer dans cet article que le choix de prédicteurs direct et rétrograde, aussi naturel soit-il, n'est pas nécessairement le mieux adapté à la résolution du problème initial, et qu'un choix différent, conduisant à un algorithme de Schur généralisé, présente les avantages d'obtenir plus simplement la solution et d'établir un lien avec les algorithmes de Chandrasekhar sans nécessiter de faire l'hypothèse d'un modèle d'état invariant [5] ou d'un processus observé stationnaire [3,4]. On obtient aussi la solution du problème du lissage optimal rapide, qui avait été négligée malgré son importance dans de nombreuses applications, sans faire appel à des techniques à base de matrices polynomiales [8].

**EQUATION DE CHANDRASEKHAR GENERALISEES**

Suivant le cadre général fourni par [5], nous nous limitons par la suite aux problèmes linéaires de la forme :

$$y_k = H x_k + b_k \tag{11}$$

où  $b_k$  est un bruit blanc centré stationnaire indépendant de  $x_k$ . La technique utilisée pour développer notre algorithme peut s'interpréter comme une généralisation de celle qui est utilisée pour établir les algorithmes des moindres carrés récursifs rapides [10]. Elle consiste à introduire un vecteur auxiliaire étendu  $X$ , construit par concaténation de tous les  $x_k$ ,  $k=1,2,\dots,N$  afin de pouvoir exploiter simplement les propriétés d'invariance du modèle (11) lors du calcul du lisseur ou du prédicteur à un pas optimaux.

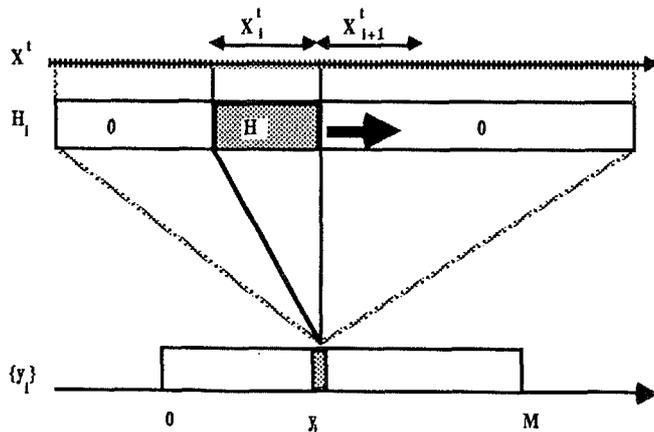


Figure 1 - Définition du vecteur étendu X.

Le vecteur  $X$  est de dimension  $nN$ . Soit  $H_1$  la matrice-bloc contenant  $N$  blocs de dimension  $n \times p$  et construite par concaténation de  $i-1$  matrices nulles, de la matrice  $H$ , et de  $N-1$  matrices nulles. Le modèle (11) s'écrit alors indifféremment

$$y_k = Hx_k + b_k \tag{11}$$

$$= H_k X + b_k \tag{12}$$

La figure 1 illustre les définitions adoptées. L'intérêt du modèle (12) réside dans le fait que la propriété d'invariance du modèle (11) se traduit par la relation suivante

$$H_{k+1}^t = D H_k^t, \quad H_k^t = D^t H_{k+1}^t \tag{13}$$

où  $D$  est la matrice de décalage dans le cas d'une observation scalaire ( $p=1$ ), et de décalage par bloc dans le cas général. Cette relation explicite sera utilisée dans la résolution du problème qui se fait par filtrage de Kalman appliqué au modèle d'état dégénéré, car sans dynamique dans l'état, qui suit :

$$X_{k+1} = X_k = X = \text{Cste} \tag{14}$$

$$y_k = H_k X + b_k \tag{12}$$

$$H_{k+1}^t = D H_k^t \tag{13}$$

Avec les conditions initiales

$$\hat{X}_{1/0} = 0 = E\{X_0\} \tag{15}$$

$$P_{1/0} = \Pi_0 = E\{X_0 X_0^t\} \tag{16}$$

$$E\{b_0\} = 0 \quad E\{b_k b_j^t\} = R \delta_{kj} \tag{17}$$

$$E\{b_k X_0^t\} = 0 \tag{18}$$

Le filtre de Kalman correspondant s'écrit

$$\hat{X}_{k/k} = \hat{X}_{k/k-1} + K_k (R_k^{-1})^{-1} \varepsilon_k \tag{19}$$

$$\varepsilon_k = y_k - H_k \hat{X}_{k/k-1} \tag{20}$$

$$R_k^{-1} = E\{\varepsilon_k \varepsilon_k^t\} \tag{21}$$

$$\hat{X}_{k+1/k} = \hat{X}_{k/k} \tag{22}$$

$$R_k^{-1} = H_k P_{k/k-1} H_k^t + R \tag{23}$$

$$K_k = P_{k/k-1} H_k^t \tag{24}$$

$$P_{k+1/k} = P_{k/k-1} - K_k (R_k^{-1})^{-1} K_k^t \tag{25}$$

Les conditions initiales sont bien équivalentes aux données du problèmes de base puisque

$$r_{yy}(i,j) = H r_{xy}(i,j) + R \delta_{ij} \tag{26}$$

$$r_{xy}(i,j) = Q_i \Pi_0 H_j^t \tag{27}$$

$$Q_k = [0 \dots | I | \dots 0] \tag{28}$$

où  $Q_k$  est la matrice de projection par bloc, dont tous les blocs, de dimension  $n \times n$ , sont nuls à l'exception du  $k$ ème bloc égal à l'identité ("pinning matrix" pour les anglo-saxons)

$$r_{yy}(i,j) = H_i \Pi_0 H_j^t + R \delta_{ij} \tag{29}$$

$$\text{d'où } R_{yy}(N) = H \Pi_0 H^t + \text{bloc-diag}\{R\} \tag{30}$$

et où  $H$  est la matrice bloc dont les blocs-lignes sont les  $H_i$ . De même

$$R_{xy}(N) = \Pi_0 H^t \tag{31}$$



La résolution du problème initial de prédiction à un pas ou de filtrage est alors obtenue simplement par projection de (21), c'est-à-dire par

$$\hat{x}_{k/k} = Q_k \hat{x}_{k/k} \quad k=1,2,\dots,N \quad (32)$$

La résolution du problème initial de lissage se fait de manière analogue à partir de  $X_{N/N}$

$$\hat{x}_{k/N} = Q_k \hat{x}_{N/N} \quad k=1,2,\dots,N \quad (33)$$

Une première conséquence intéressante de la méthode est qu'on obtient ainsi un lisseur optimal en une passe, ce qui est important dans des applications en traitement d'images [11]. Mais, étant donné les dimensions élevées du problème auxiliaire (12), tout ce qui précède serait sans intérêt si le calcul du gain de Kalman et de la covariance de l'innovation utilisés dans (19) devait se faire à l'aide d'équations de Riccati (23-25). Or le calcul du gain peut être fait de manière rapide en utilisant les propriétés :

- d'invariance par décalage du modèle (12) explicité dans (13)
- de stationnarité du bruit  $b_k$  (invariance de R).

L'idée générale est la même que celle utilisée dans [9]. On ne manipule que les incréments des quantités (23-25) intervenant dans le filtre de Kalman au lieu des quantités nominales, ce qui conduit à des équations de Chandrasekhar. La difficulté provient du fait que le modèle (12-14) n'étant pas strictement invariant, on ne peut utiliser les équations de Chandrasekhar usuelles. Il faut les modifier pour prendre en compte une invariance par l'opérateur D. Il suffit de généraliser la définition des incréments

$$\partial P_k = P_{k/k-1} - D P_{k-1/k-2} D^t \quad (34)$$

$$\partial K_k = K_k - D K_{k-1} \quad (35)$$

$$\partial R_k^\epsilon = R_k^\epsilon - R_{k-1}^\epsilon \quad (36)$$

Les deux factorisations possibles

$$\begin{aligned} \partial P_{k+1} &= [I - K_k (R_k^\epsilon)^{-1} H_k] [\partial P_k + \partial P_k H_k^t (R_{k-1}^\epsilon)^{-1} H_k \partial P_k^t] \\ &\quad [I - K_k (R_k^\epsilon)^{-1} H_k]^t \\ &= Y_k^1 M_k^1 Y_k^{1t} \end{aligned} \quad (37-1)$$

$$\begin{aligned} \partial P_{k+1} &= [I - D K_{k-1} (R_{k-1}^\epsilon)^{-1} H_k] [\partial P_k - \partial P_k H_k^t (R_k^\epsilon)^{-1} H_k \partial P_k^t] \\ &\quad [I - D K_{k-1} (R_{k-1}^\epsilon)^{-1} H_k]^t \\ &= Y_k^2 M_k^2 Y_k^{2t} \end{aligned} \quad (37-2)$$

conduisent alors à deux algorithmes

$$R_{k+1}^\epsilon = R_k^\epsilon + H_{k+1}^{1,2} Y_k^{1,2} M_k^{1,2} Y_k^{1,2t} H_{k+1}^{1t} \quad (38)$$

$$K_{k+1} = D K_k + Y_k^{1,2} M_k^{1,2} Y_k^{1,2t} H_{k+1}^{1t} \quad (39)$$

$$Y_k^1 = [I - K_k (R_k^\epsilon)^{-1} H_k] Y_{k-1}^1 \quad (40-1)$$

$$M_k^1 = M_{k-1}^1 + M_{k-1}^1 Y_{k-1}^{1t} H_k^t (R_{k-1}^\epsilon)^{-1} H_k Y_{k-1}^1 M_{k-1}^{1t} \quad (41-1)$$

$$Y_k^2 = [I - D K_{k-1} (R_{k-1}^\epsilon)^{-1} H_k] Y_{k-1}^2 \quad (40-2)$$

$$M_k^2 = M_{k-1}^2 - M_{k-1}^2 Y_{k-1}^{2t} H_k^t (R_{k-1}^\epsilon)^{-1} H_k Y_{k-1}^2 M_{k-1}^{2t} \quad (41-2)$$

La complexité de ces algorithmes est directement liée à la dimension de la matrice  $M_k^{1,2}$ , c'est-à-dire au rang de l'incrément initial

$$\partial P_1 = P - D P_{1/0} D^t = \Pi_0 - D \Pi_0 D^t = Y_0^{1,2} M_0^{1,2} Y_0^{1,2t} \quad (42)$$

Il est remarquable que ce rang soit justement le rang de déplacement de la matrice de covariance a priori  $\Pi_0$ , et donc aussi, compte-tenu de (30) et de (13), celui de  $R_{yy}(N)$ . On retrouve donc bien, sous-jacente, la même mesure de complexité que dans [5]. Mais, ce qui est remarquable aussi, c'est que l'établissement de ces algorithmes ne prend pas en compte la nature exacte de l'opérateur D. On obtiendrait donc des algorithmes analogues pour des modèles présentant des périodicités (D circulante unitaire), des affaiblissements exponentiels (D  $\alpha$ -décalante), etc.... Ce point avait déjà été remarqué [6].

### INTERPRETATION GEOMETRIQUE

La forme 2 des équations de Chandrasekhar généralisées peut s'écrire sous la forme suivante

$$\begin{bmatrix} R_{k+1}^\epsilon & 0 \\ K_{k+1} & Y_{k+1} \\ 0 & R_{k+1}^r \end{bmatrix} = \begin{bmatrix} R_k^r & H_{k+1} Y_k \\ D K_k & Y_k \\ Y_k^t H_{k+1}^t & R_k^r \end{bmatrix} \Theta_k \quad (43)$$

où  $R_k^r = -(M_k^2)^{-1} \quad (44)$

et où

$$\Theta_k = \begin{bmatrix} I_1 & -(R_k^\epsilon)^{-1} H_{k+1} Y_k \\ -(R_k^r)^{-1} Y_k^t H_{k+1}^t & I_2 \end{bmatrix} \quad (45)$$

est une matrice J-orthogonale, c'est-à-dire une matrice telle que

$$\Theta_k R_k \Theta_k = R_{k+1} \quad (46)$$

où  $R_k = \begin{bmatrix} R_k^\epsilon & 0 \\ 0 & -R_k^r \end{bmatrix} \quad (47)$

$$\begin{aligned} R_k &= \begin{bmatrix} (R_k^\epsilon)^{1/2} & 0 \\ 0 & -(R_k^r)^{1/2} \end{bmatrix} \begin{bmatrix} I_p & 0 \\ 0 & -S \end{bmatrix} \begin{bmatrix} (R_k^\epsilon)^{1/2} & 0 \\ 0 & -(R_k^r)^{1/2} \end{bmatrix} \\ &= R_k^{1/2} J R_k^{t/2} \end{aligned} \quad (48)$$

S est la matrice de signature de  $M_k$ . La relation (46) peut être réécrite sous forme normalisée en posant :

$$\tilde{\Theta}_k = R_k^{t/2} \Theta_k R_{k+1}^{1/2} \quad (49)$$

Ce qui nous donne :

$$\tilde{\Theta}_k J \tilde{\Theta}_k^t = J \quad (50)$$

L'analogie avec la forme matricielle compacte de l'algorithme de Levinson généralisé [5] est flagrante. La différence essentielle provient du fait que les quantités calculées à chaque récursion ne sont plus, comme dans [5], les matrices des prédicteurs directs et rétrogrades de  $y_k$ , mais  $K_k$  et  $Y_k$  qui sont les matrices de covariance mutuelle entre le vecteur d'état et les innovations directes et rétrogrades de  $y_k$ , respectivement. Il s'agit donc, suivant la classification introduite dans [6] d'un algorithme de Schur généralisé.

On retrouve donc ainsi des relations d'orthogonalisation en treillis, ce qui ne saurait surprendre pour un algorithme rapide de la famille des moindres carrés [1] mais ce qui, à notre connaissance, restait à établir pour un problème de lissage.



Cette interprétation géométrique permet aussi d'éclairer un point particulier des algorithmes de Chandrasekhar. On leur a souvent reproché leur manque d'unicité lors de la factorisation initiale  $dP_1 = Y_1 M_1 Y_1^t$ . En fait  $\alpha = \dim M \geq p$  est une mesure de non-stationnarité de  $y_k$  et, en général, on a  $\alpha > 1$ . Ce manque d'unicité n'est donc pas un défaut propre aux algorithmes de Chandrasekhar, mais il s'agit tout simplement du vieux problème de l'absence d'unicité dans le choix d'une base orthonormée dans un espace de dimension supérieure à 1 [1]. Ce "défaut" est commun à tous les algorithmes rapides de résolution de problèmes multidimensionnels.

### CONCLUSIONS

Nous avons donc pu ainsi, à partir de la seule covariance du processus à estimer, développer des équations de Chandrasekhar généralisées qui présentent les avantages suivants:

- leur dérivation est immédiate et ne nécessite pas de manipulations matricielles complexes. On retrouve d'ailleurs exactement les équations "LWR étendues" de Friedlander et al. [5] pour la résolution du problème initial. En effet, l'équation de mise à jour du gain (39) :

$$K_{k+1} = D K_k + Y_k M_k Y_k^t H_{k+1}^t \quad (39)$$

est l'équation F-13 de [5] :

$$h^{m+1}_{xy} = [o, h^m_{xy}] + E_m M_m^t B_m^t \quad (F-13)$$

- elles font apparaître explicitement le rang de déplacement du processus étudié, ce qui est une conséquence de l'absence de description interne du processus et de la modification des incréments des quantités à factoriser;

- elles peuvent se mettre sous une forme matricielle compacte dans laquelle on reconnaît un algorithme de Schur généralisé;

- cette forme matricielle fait intervenir une matrice de transformation J-orthogonale qui permet d'étudier simplement les liens avec les méthodes géométriques de la théorie des circuits;

- elles permettent de résoudre à la fois le problème de la prédiction et du filtrage, en évitant d'avoir à résoudre une équation de Riccati, mais aussi le problème du lissage, en évitant de résoudre une équation de Fredholm;

- elles permettent d'obtenir une solution récursive rapide, en une seule passe, du problème du lissage optimal, sans être limité strictement au cas stationnaire, et sans utiliser de techniques à base de matrices polynomiales [8].

### BIBLIOGRAPHIE

- [1] Benvéniste A., "Méthodes d'orthogonalisation en treillis pour le problème de la réalisation stochastique", in Algorithmes rapides pour le traitement des systèmes dynamiques linéaires, A. Benvéniste et P. Dewilde eds., INRIA, 1981, pp.2.1-2.42.
- [2] Demoment G. and Reynaud R. "Fast Minimum-Variance Deconvolution", IEEE Trans., vol.ASSP-33, pp.1324-1326, 1985.
- [3] Dugré J.P., Scharf L.L. and Guéguen C., "Exact Likelihood for Stationary Vector Autoregressive Moving Average Processes", Signal Processing, vol.11, pp.105-118, 1986.
- [4] Favier G., "Filtrage, modélisation et identification des systèmes linéaires stochastiques à temps discret", Editions du CNRS, Paris, 1982.
- [5] Friedlander B., Kailath T., Morf M. and Ljung L., "Extended Levinson and Chandrasekhar Equations for General Discrete-Time Linear Estimation Problems", IEEE Trans., vol.AC-23, pp.653-659, 1978.
- [6] Guéguen C., "An introduction to displacement ranks and related fast algorithms", Lectures Notes, TO Summer School on Digital Signal Processing, Les Houches, 1985.
- [7] Kailath T., Kung S.Y. and Morf M., "Displacement ranks of a matrix", Bull. Am. Math. Soc., vol.1, pp769-773, 1979.
- [8] Moir T.J., "Optimal Deconvolution Smoother", IEE Proc. vol.133, Pt.D, pp.13-18, 1986.
- [9] Morf M., Sidhu G.S. and Kailath T., "Some new algorithms for recursive estimation in constant, linear, discrete-time systems", IEEE trans., vol.AC-19, pp315-323, 1974.
- [10] Morf M., Dickinson B., Kailath T. and Viera A., "Efficient solution of covariance equation for linear prediction", IEEE trans., vol.ASSP-25, pp.429-434, 1977.
- [11] Saint-Félix D., Du X.C. et Demoment G., "Restauration rapide d'images : utilisation d'un modèle non-causal et d'équations de Chandrasekhar modifiées", Actes du 2ème colloque Image, CESTA, Nice, 1986, pp.776-783.