Stabilité structurelle de méthodes de prédiction linéaire

Jérôme Idier et Jean-François Giovannelli

Laboratoire des Signaux et Systèmes,
SUPÉLEC, Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France
idier@lss.supelec.fr, giovannelli@lss.supelec.fr

RÉSUMÉ

Cette communication concerne la recherche d'une condition systématique de stabilité pour les méthodes de prédiction linéaire de type moindres carrés, dans le cas d'un extrait de signal observé. Sauf le cas d'une matrice normale de Toeplitz, il existe très peu de résultats liés à la structure de la matrice normale, et non pas au prédicteur lui-même. Nous obtenons une condition suffisante générale puis des conséquences originales, qui concernent en particulier la méthode régularisée de Kitagawa et Gersch.

ABSTRACT

In this communication, a structural stability condition is sought for least squares linear prediction methods, in the given data case. Save the Toeplitz case, very few results are known in connection with the structure of the sample covariance matrix, rather than the predictor itself. A general sufficient condition is provided, and several original consequences are drawn. Special attention is paid to Kitagawa and Gersch regularized prediction method.

1 Introduction

Cette étude concerne le problème classique de la prédiction linéaire dans le cas pratique où seul un extrait $\boldsymbol{x} = [x_1, \dots, x_n, \dots, x_N]^{\mathrm{t}}$ du signal est observé. Nous proposons une condition systématique de stabilité stricte du filtre prédicteur qui s'applique aux méthodes de prédiction linéaire interprétables selon le principe des moindres carrés. Contrairement aux tests de stabilité classiques [1] ou à d'autres conditions suffisantes [2], cette condition s'applique directement à la matrice normale, et non pas au prédicteur lui-même. Elle permet finalement de distinguer certaines méthodes *structurellement stables*, c'est-à-dire qui ne produisent que des filtres prédicteurs stables.

La stabilité de la méthode *fenêtré* (c'est-à-dire *pré- et post-fenêtré*, ou encore méthode *autocorrélation*, ou approche *Yule-Walker*) est un résultat classique, lié au caractère Toeplitz positif de la matrice normale [3, 4]. Le cas *post-fenêtré* est également connu pour être stable, bien que la matrice normale associée ne soit pas de Toeplitz [5]. D'autres méthodes sont connues pour ne pas présenter de garantie de stabilité, en particulier la méthode *covariance*, la méthode *covariance modifiée* (ou encore *aller-retour*, ou méthode *des moindres carrés*) et la méthode pré-fenêtrée [6]. Pour d'autres encore, telles que la méthode d'estimation régularisée introduite par Kitagawa et Gersch [7], la question de la stabilité n'est pas tranchée. Dans ces différents cas, la matrice normale est encore positive mais n'est plus de Toeplitz.

2 Conditions de canonicité

2.1 Formulation du problème

Soit M une matrice hermitienne positive fonction de $\boldsymbol{x}$, et

$$J(\boldsymbol{a}) = \boldsymbol{\alpha}^{\dagger} M \boldsymbol{\alpha} \qquad (1)$$

un critère quadratique à minimiser par rapport à l'innovateur $\boldsymbol{\alpha} = [1 \mid -\boldsymbol{a}^{\mathrm{t}}]^{\mathrm{t}}$, ou au prédicteur $\boldsymbol{a}$de taille P. Notre objectif est de mettre en évidence des conditions portant sur la structure de la matrice M pour garantir la stabilité du filtre récursif défini par les coefficients du prédicteur solution $\hat{\boldsymbol{a}} \triangleq R^{-1}\boldsymbol{r}$, en notant

$$M = \left[\begin{array}{c|c} \rho & \boldsymbol{r}^{\dagger} \\ \hline \boldsymbol{r} & R \end{array}\right]. \qquad (2)$$

On dira alors que la matrice M et le vecteur $\hat{\boldsymbol{a}}$ sont *canoniques*. L'idée développée ici exploite directement la condition nécessaire et suffisante de réalisabilité pour les filtres linéaires : un filtre est réalisable (*i.e.*,, causal stable) si et seulement si les pôles de sa transformée en z, $\mathcal{H}(z)$, sont de module inférieur à un. Dans le cas présent,

$$\mathcal{H}(z) = \frac{1}{1 - \sum_{k=1}^{P} a_k z^{-k}} = \frac{z^P}{A(z)},$$

en notant

$$A(z) \triangleq z^P - \sum_{k=0}^{P-1} a_{P-k} z^k. \qquad (3)$$

L'étude porte sur certaines garanties d'appartenance au disque unité des zéros du polynôme normalisé $A(z)$ de degré P, non pas directement à partir de ses coefficients (ce problème est très classique [1, 2]), mais des coefficients de la matrice M.

2.2 Condition suffisante

Pour une matrice carrée M de taille $n \times n$, notons respectivement $\underline{M}\rfloor$, $\lceil\overline{M}$, $\lfloor\underline{M}$ et $\overline{M}\rceil$ les matrices $(n-1) \times (n-1)$ nord-ouest, sud-est, nord-est et sud-ouest extraites de P. Selon cette notation, la matrice R introduite en (2) n'est autre que $\lceil\overline{M}$, et

$$\Delta \triangleq \lceil\overline{M} - \underline{M}\rfloor$$

est la matrice de déplacement dont le rang définit la *distance à Toeplitz* [8].

Théorème 2.1 — *Si $M > 0$ et $\underline{M}\rfloor \leqslant \lceil\overline{M}$, alors M est canonique.*

Démonstration : Le principe de la démonstration rappelle à son début celui de la prédiction dans L^2 [3, 4], dont il constitue en fait une généralisation. Soit $A(z)$ un polynôme normalisé dont au moins un zéro z_0 est de module supérieur à un :

$$A(z) = (z - z_0)B(z), \tag{4}$$

où $B(z)$ est un polynôme normalisé de degré $P-1$, et soit

$$\tilde{A}(z) = (z - 1/z_0^\star)B(z) \tag{5}$$

le polynôme normalisé de degré P obtenu en « repliant » z_0 en $1/z_0^\star$, c'est-à-dire à l'intérieur du cercle unité, et tel que A et $\tilde{A}$ ont même restriction sur le cercle unité, d'après le résultat classique suivant :

$$x\forall z, z_0 \in \mathbb{C},\ z_0 \neq 0 \text{ et } |z| = 1 \ \Rightarrow\ |z - z_0| = |1 - z/z_0^\star|. \tag{6}$$

x Notons enfin $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ et $\tilde{\boldsymbol{\alpha}}$, et $\boldsymbol{a}$, $\boldsymbol{b}$ et $\tilde{\boldsymbol{a}}$ les innovateurs et les prédicteurs associés respectivement à A, B et $\tilde{A}$, conformément à la notation (3). L'objectif est de montrer que dans les conditions du Théorème 2.1, le critère J défini par (1) diminue systématiquement après « repliement » d'un pôle, ce qui garantit que le minimum est atteint à l'intérieur du cercle unité.

En termes d'innovateurs, la relation (4) s'écrit

$$\boldsymbol{\alpha} = \left[\begin{array}{c} \boldsymbol{\beta} \\ \hline 0 \end{array}\right] - z_0 \left[\begin{array}{c} 0 \\ \hline \boldsymbol{\beta} \end{array}\right],$$

qui, reportée dans (1), donne l'expression :

$$J(\boldsymbol{a}) = \boldsymbol{\beta}^\dagger(\underline{M}\rfloor + |z_0|^2 \lceil\overline{M} - z_0 \overline{M}\rceil - z_0^\star \lfloor\underline{M})\boldsymbol{\beta}.$$

De la même façon, la relation (5) fournit

$$|z_0|^2 J(\tilde{\boldsymbol{a}}) = \boldsymbol{\beta}^\dagger(|z_0|^2 \underline{M}\rfloor + \lceil\overline{M} - z_0 \overline{M}\rceil - z_0^\star \lfloor\underline{M})\boldsymbol{\beta},$$

et la différence des deux expressions précédentes fournit le résultat fondamental :

$$J(\boldsymbol{a}) - |z_0|^2 J(\tilde{\boldsymbol{a}}) = (|z_0|^2 - 1)\boldsymbol{\beta}^\dagger \Delta \boldsymbol{\beta}, \tag{7}$$

qui permet de conclure immédiatement $J(\boldsymbol{a}) > J(\tilde{\boldsymbol{a}})$, donc que le polynôme $\hat{A}$ qui minimise J n'admet aucun zéro hors du cercle unité. Montrons de plus que les zéros de $\hat{A}$ sont strictement à l'intérieur du cercle unité. Soit $\tilde{B}(z) = (z - z_0/|z_0|)B(z)$ le polynôme obtenu en déplaçant z_0 sur le cercle unité. Un calcul analogue au précédent donne

$$x\, J(\boldsymbol{a}) - |z_0| J(\tilde{\boldsymbol{b}}) = J(\tilde{\boldsymbol{b}}) - |z_0| J(\tilde{\boldsymbol{a}}) = (|z_0| - 1)\boldsymbol{\beta}^\dagger \Delta \boldsymbol{\beta}, \tag{8}$$

x par conséquent $J(\boldsymbol{a}) > J(\tilde{\boldsymbol{b}}) > J(\tilde{\boldsymbol{a}})$, et les zéros de $\hat{A}$ sont strictement à l'intérieur du cercle unité. Autrement dit, $\hat{\boldsymbol{a}}$ et M sont canoniques. ■

Exemple 2.1 (Cas Toeplitz). Dans le cas où M est une matrice de Toeplitz, on a $\Delta = 0$, et le résultat (7) se ramène à la forme classique $J(\boldsymbol{a}) = |z_0|^2 J(\tilde{\boldsymbol{a}})$. Le Théorème 2.1 s'applique si $M > 0$.

Exemple 2.2 (Cas diagonal). Dans le cas d'une matrice M diagonale, les conditions du Théorème 2.1 sont satisfaites par toute suite croissante de coefficients diagonaux positifs. C'est un exemple trivial d'une matrice canonique non Toeplitz.

Exemple 2.3 (Cas mixte). La somme de deux matrices vérifiant séparément les conditions du Théorème 2.1 les vérifie également. Ainsi, une matrice de Toeplitz dont la diagonale est augmentée par une suite croissante positive reste canonique.

2.3 Conditions suffisantes raffinées

Pour une matrice M donnée, la validité des deux conditions $M > 0$ et $\underline{M}\rfloor \leqslant \lceil\overline{M}$ dépend de la valeur de ρ, tandis que le prédicteur solution $\hat{\boldsymbol{a}} = \lceil\overline{M}^{-1}\boldsymbol{r}$ n'en dépend pas ! Cette situation n'est pas contradictoire ; elle indique simplement que le théorème démontré admet une forme relaxée, qui ne dépend plus de ρ.

Corollaire 2.2 — *Soit $M(m) \triangleq M + (m - \rho)\boldsymbol{v}\boldsymbol{v}^{\mathrm{t}}$, où $\boldsymbol{v} = [1, 0, \ldots, 0]^{\mathrm{t}}$. Alors M est canonique si les conditions du Théorème 2.1 s'appliquent à $M(m)$, pour au moins un $m \in \mathrm{e}_+^*$.*

La démonstration est évidente. En pratique, si M ne vérifie pas la condition $M > 0$, il faut choisir $m > \rho$ pour espérer obtenir $M(m) > 0$. D'autre part, si M ne vérifie pas $\underline{M}\rfloor \leqslant \lceil\overline{M}$, il faut choisir $m < \rho$ pour espérer obtenir $\underline{M(m)}\rfloor \leqslant \lceil\overline{M}$. Ainsi, la vérification simultanée des deux conditions, si elle est possible, se produit dans un intervalle $]m_1, m_2]$ tel que $M(m_1)$ et $\lceil\overline{M} - \underline{M(m_2)}\rfloor$ sont positives au sens large.

On peut aussi remarquer que la condition $M > 0$ est trop restrictive : pour tous les développements effectués, la positivité de $\boldsymbol{\alpha}^\dagger M \boldsymbol{\alpha}$ importe pour des vecteurs du type innovateur $\boldsymbol{\alpha} = [1 \mid -\boldsymbol{a}^{\mathrm{t}}]^{\mathrm{t}}$ seulement. On en déduit une condition suffisante plus lâche, dont la démonstration exploite simultanément le Corrolaire 2.2 :

Théorème 2.3 — *Si $\lceil\overline{M} > 0$ et $\underline{M(\boldsymbol{r}^\dagger \lceil\overline{M}^{-1}\boldsymbol{r})}\rfloor \leqslant \lceil\overline{M}$, alors M est canonique.*

Dans l'idée de caractériser les matrices canoniques, on pourrait se demander si les conditions du Théorème 2.3 sont des conditions nécessaires. Il est facile de vérifier que tel n'est pas le cas, en exhibant un contre-exemple. Ainsi, si M est une matrice de Toeplitz positive, considérons la matrice $Q = M - \varepsilon \boldsymbol{v}\boldsymbol{w}^{\mathrm{t}}$, où $\boldsymbol{v} = [0, \ldots, 0, 1]^{\mathrm{t}}$. Pour des valeurs de ε positives et proches de zéro, la matrice Q reste positive et les zéros du prédicteur associé restent à l'intérieur du cercle unité (par argument de continuité). Pourtant la positivité est mise en défaut pour la matrice de déplacement $\lceil\overline{Q} - \underline{Q}\rfloor = -\varepsilon \boldsymbol{v}\boldsymbol{w}^{\mathrm{t}}$, ainsi que pour toute matrice $\lceil\overline{Q} - \underline{Q(m)}\rfloor = m\boldsymbol{v}\boldsymbol{v}^{\mathrm{t}} - \varepsilon \boldsymbol{v}\boldsymbol{w}^{\mathrm{t}}$.

2.4 Stabilité structurelle

En tant que tel, les conditions de canonicité mises à jour sont d'un intérêt modeste par rapport aux tests déjà existants, car il n'est pas plus simple de tester la positivité d'une matrice que la stabilité d'un prédicteur. De plus, le calcul de $\hat{\boldsymbol{a}} = \lceil\overline{M}^{-1}\boldsymbol{r}$

est un préalable nécessaire à la vérification des conditions du Théorème 2.3. En revanche, ces conditions permettent d'examiner la *stabilité structurelle* de certaines méthodes de prédiction linéaire. L'objectif est de diviser ces méthodes en deux groupes :

– les méthodes structurellement stables ne produisent que des filtres prédicteurs stables ;
– les autres produisent des prédicteurs parfois instables.

3 Application en prédiction linéaire

3.1 Cas générique

Les cas les plus classiques se rapportant à notre étude (Sections 3.2, 3.3 et 3.4) correspondent à des formes quadratiques J du type $J(\boldsymbol{a}) = \|X\boldsymbol{\alpha}\|^2$, où les variantes se distinguent par la forme de la matrice X. Par identification, on a $M = X^\dagger X$, matrice hermitienne positive au sens large par construction.

3.2 Forme *fenêtrée*

Dans ce cas, M est une matrice de Toeplitz, strictement positive si $\boldsymbol{x} \neq \boldsymbol{0}$, dont les coefficients sont

$$M_{mn} = \sum_{k=1+\max(m,n)}^{N+\min(m,n)} x^\star_{k-m} x_{k-n}.$$

Les conditions de canonicité sont donc remplies. Démontré dans le cadre de la prédiction linéaire à passé fini des processus stationnaires d'ordre deux, ce résultat est classique. En effet, M s'interprète comme la covariance d'un tronçon (de taille $P+1$) d'un processus dont $\hat{\boldsymbol{a}}$ est le prédicteur d'ordre P, solution des équations de Yule-Walker.

3.3 Forme *covariance*

Dans ce cas, la matrice M est *proche de Toeplitz.* A partir de ses coefficients : $M_{mn} = \sum_{k=P+2}^{N+1} x^\star_{k-m} x_{k-n}$, ceux de la matrice Δ s'en déduisent aisément. En notant $\boldsymbol{x}_n = [x_n, \ldots, x_{n-P+1}]^{\mathrm{t}}$, on obtient ainsi $\Delta = \boldsymbol{x}_P \boldsymbol{x}_P^\dagger - \boldsymbol{x}_N \boldsymbol{x}_N^\dagger$, matrice de rang 2. Ce résultat est connu car il détermine la distance à Toeplitz de M, donc la complexité des algorithmes de Levinson généralisés qui permettent de l'inverser rapidement [8]. En revanche, il exprime aussi que Δ n'est pas, en général, une matrice positive. En exhibant des contre-exemples, il est par ailleurs connu que la forme covariance ne fournit pas toujours un prédicteur stable [6].

La forme covariance *modifiée* associe la prédiction avant et arrière en rendant persymétrique la matrice M par addition de la forme covariance et de son expression retournée (obtenue par inversement du temps). On obtient ainsi une matrice Δ de rang 4, qui n'est pas nécessairement une matrice positive. Cette méthode n'est pas structurellement stable, contrairement à la méthode de Burg [9] dont elle se rapproche. La pratique montre néanmoins que cette forme modifiée est plus souvent stable que la forme covariance standard.

3.4 Forme *pré-fenêtrée* et forme *post-fenêtrée*

Ces deux formes sont intermédiaires entre la forme covariance et la forme fenêtrée. On vérifie sans difficulté que la matrice de déplacement Δ vaut respectivement $-\boldsymbol{x}_N \boldsymbol{x}_N^\dagger$ et $\boldsymbol{x}_P \boldsymbol{x}_P^\dagger$. Ce dernier cas assure donc la canonicité, résultat mentionné sans démonstration dans [5].

3.5 Formes *régularisées*

Kitagawa et Gersch [7] ont proposé une méthode d'analyse spectrale auto-régressive dont la solution minimise un critère des moindres carrés régularisés. Il s'agit de la forme pré-fenêtrée, augmentée d'un terme régularisant qui favorise la douceur du spectre recherché :

$$J_{\text{KG}}(\boldsymbol{a}) = J(\boldsymbol{a}) + \lambda \sum_{p=1}^{P} p^{2k} a_p^2. \tag{9}$$

λ est un paramètre de régularisation (positif) et k détermine l'*ordre* de la douceur considérée, conformément à l'identité [7]

$$\int_{-1/2}^{1/2} \left| \frac{d^k A(e^{2i\pi f})}{df^k} \right|^2 df = (2\pi)^{2k} \sum_{p=1}^{P} p^{2k} a_p^2.$$

On vérifie facilement qu'il est possible d'exprimer la matrice M_{KG} à partir de la matrice M de la forme standard en ajoutant une suite croissante positive à la diagonale :

$$x M_{\text{KG}} = M + \lambda \ \operatorname{diag}\{p^{2k}\}_{p=0,..,P}. \tag{10}$$

x Évidemment, le même principe de régularisation s'applique également aux autres formes de fenêtrage. La forme fenêtrée possède l'intérêt particulier de fournir la structure matricielle « Toeplitz + diagonale » de l'Exemple 2.3. Nous obtenons ainsi le résultat suivant :

Corollaire 3.1* — *La variante fenêtrée de la méthode de prédiction régularisée proposée par Kitagawa et Gersch [7] est structurellement stable, pour tout ordre de douceur $k \geqslant 0$ et tout $\lambda \geqslant 0$.

En pratique, Kitagawa et Gersch incorporent simultanément plusieurs ordres de douceur en ajoutant les termes adéquats au critère (9). Cette extension reste évidemment stable. De la même façon, on peut étendre l'approche de Kitagawa et Gersch à des valeurs de k non entières, sans remettre en cause la stabilité structurelle de la méthode.

Quant à la forme pré-fenêtrée régularisée originale, on montre qu'une certaine quantité de régularisation suffit à la rendre stable, conformément au résultat suivant (qui se généralise aisément à d'autres variantes) :

Corollaire 3.2 — *Pour $k > 0$, le prédicteur issu de la forme* pré-fenêtrée régularisée *est stable si*

$$\lambda \geqslant \sum_{p=1}^{P} \frac{|x_{N+1-p}|^2}{p^{2k} - (p-1)^{2k}}.$$

Démonstration : La matrice de déplacement de M_{KG} s'écrit $\Delta_{\text{KG}} = \lambda\delta - \boldsymbol{x}_N \boldsymbol{x}_N^\dagger$, avec

$$\delta \triangleq \operatorname{diag}\left\{p^{2k} - (p-1)^{2k}\right\}_{p=1,\ldots,P}.$$

D'après l'expression du déterminant

$$\det \Delta_{\text{KG}} = \lambda^p \left(1 - \lambda^{-1}\mathbf{x}_N^\dagger \delta^{-1}\mathbf{x}_N\right) \prod_1^P \left(p^{2k} - (p-1)^{2k}\right),$$

la condition $\lambda \geqslant \mathbf{x}_N^\dagger \delta^{-1}\mathbf{x}_N$ est nécessaire à la positivité de Δ_{KG}. Elle est aussi suffisante, car les $P-1$ conditions issues des mineures s'avèrent moins restrictives. ■

4 Conclusion

Dans le cadre de la théorie et des algorithmes pour la prédiction linéaire, il est bien connu qu'une matrice normale M s'inverse d'autant plus rapidement que sa *distance à Toeplitz*, c'est-à-dire le rang de sa matrice de déplacement Δ, est faible. Dans cette communication, nous montrons que la positivité de la matrice Δ permet d'assurer la stabilité du filtre prédicteur estimé. Parmi les méthodes classiques de prédiction linéaire basées sur le principe des moindres carrés, ce résultat permet de démontrer dans un cadre unificateur la stabilité structurelle des méthodes fenêtrée et post-fenêtrée, résultats déjà connus, mais également de la variante fenêtrée de la méthode régularisée de Kitagawa et Gersch [7]. Il fournit également une borne inférieure du paramètre de régularisation rendant stable la méthode originale de Kitagawa et Gersch.

Références

[1] Y. Bistritz, « Zero location with respect to the unit circle of discrete-time linear system polynomials », *Proc. IEEE*, vol. 72, 9, pp. 1131–1142, Sep. 1984.

[2] B. Picinbono et M. Benidir, « Some properties of lattice autoregressive filters », *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-34, 2, pp. 342–349, Apr. 1986.

[3] L. Pakula et S. Kay, « Simple proofs of the minimum phase property of the oprimal prediction error filter. », *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-31, 2, pp. 501, Apr. 1983.

[4] S. Orfanidis, « A proof of the minimal phase property of the prediction error filter », *Proc. IEEE*, vol. 71, 7, pp. 905, July 1983.

[5] B. Friedlander, « Lattice filters for adaptative processing », *Proc. IEEE*, vol. 70, pp. 829–867, Aug. 1982.

[6] S. M. Kay et S. L. Marple, « Spectrum analysis–a modern perpective », *Proc. IEEE*, vol. 69, 11, pp. 1380–1419, Nov. 1981.

[7] G. Kitagawa et G. W., « A smoothness priors long AR model method for spectral estimation », *IEEE Trans. Automat. Contr.*, vol. AC-30, 1, pp. 57–65, Jan. 1985.

[8] B. Friedlander, M. Morf, T. Kailath, et L. Ljung, « New inversion formulas for matrices classified in terms of their distances from Toeplitz matrices », *Linear Algebra and its Applications*, vol. 27, pp. 31–60, 1979.

[9] J. P. Burg, « Maximum entropy spectral analysis », in *Proc. of the 37th Meeting of the Society of Exploration Geophysicists*, Oklahoma City, Oct. 1967, pp. 34–41.