

$\varepsilon_{1s}, \varepsilon_{2s}$  being  $N(0, 2)$ , the  $l_1$ -periodogram does not perform much worse than the standard  $l_2$ -periodogram.

V. CONCLUSIONS

A new type of the periodogram is developed for observations contaminated by impulse random errors having an unknown heavy-tailed error distribution. A nonquadratic residual loss function used for a fitting of observations is a key point that separates the new periodogram from the standard one. The Huber’s minimax robust statistics are applied for a choice of this residual function. The formulas for the asymptotic bias and variance of the robust  $M$ -periodogram are obtained. The simulation given for the  $l_1$ -periodogram demonstrates a radical improvement in the quality of the periodogram.

ACKNOWLEDGMENT

The author would like to thank the two anonymous referees for their helpful comments.

REFERENCES

[1] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.  
 [2] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outliers Detection*. New York: Wiley, 1987.  
 [3] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*. New York: Wiley, 1995.  
 [4] P. J. Huber, “Robust regression: Asymptotics, conjectures and Monte Carlo,” *Ann. Math. Statist.*, vol. 1, no. 5, pp. 799–821, 1973.  
 [5] B. T. Poljak and J. Z. Tsytkin, “Robust identification,” *Automatika*, vol. 16, pp. 53–63, 1980.

Structural Stability of Least Squares Prediction Methods

Jérôme Idier and Jean-François Giovannelli

**Abstract**—A structural stability condition is sought for least squares linear prediction methods in the given data case. Save the Toeplitz case, the structure of the normal equation matrix yields no acknowledged guarantee of stability. Here, a new sufficient condition is provided, and several least squares prediction methods are shown to be structurally stable.

I. INTRODUCTION

This correspondence addresses stability conditions of linear prediction filters in the *given data* case. A simple condition of strict stability of the prediction filter is proposed, which applies to least squares estimates. Whereas general stability tests [1], as well as simpler sufficient conditions [2], are known to apply to the estimated predictor itself, the proposed condition applies to the normal equation matrix (NEM). As a consequence, it shows that some least squares methods are *structurally stable*, i.e., that they ensure the predictor stability for any data sequence.

Manuscript received February 19, 1998; revised April 16, 1998. The associate editor coordinating the review of this paper and approving it for publication was Dr. Eric Moulines.

The authors are with Laboratoire des Signaux et Systèmes, Supélec, Plateau de Moulon, Gif-sur-Yvette, France (e-mail: idier@lss.supelec.fr).

Publisher Item Identifier S 1053-587X(98)07819-2.

© 1998 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Structural stability of the *autocorrelation method* is a well-known result. Because the NEM is positive definite and Toeplitz, the proof can be identified to that of the stability of the prediction error filter in the *given covariance* case [3]. The *post-windowed* approach is also known to be structurally stable [4], although the associated NEM is not Toeplitz. With regard to other methods, such as the *covariance method*, the *modified covariance method*, and the *prewindowed method* [5], the lack of structural stability is also acknowledged. On the other hand, the question of structural stability remains open for some other methods, such as the *smoothness priors long autoregressive method* of Kitagawa and Gersch [6]. In addition, in the case of *weighted least squares methods*, the effect of a forgetting factor on stability is unknown. In nearly all cases but the autocorrelation approach, the NEM is still positive (semi)definite, but it is not Toeplitz. The main contribution of the paper is to show that positive definite normal equation matrices still provide stable prediction filters, provided that the associated *displacement matrix* is positive semidefinite. Then, in the light of this property, structural stability of classical *least squares methods* is examined (or reexamined).

II. CONDITIONS OF STABILITY

A. Problem Formulation

Let  $M$  be a positive definite matrix of given size  $(P+1) \times (P+1)$  defined as a function of the complex-valued data sequence  $\mathbf{x} = [x_1, \dots, x_n, \dots, x_N]^t$ . Let  $\alpha = [1 \mid -\mathbf{a}^t]^t$  and

$$J(\mathbf{a}) = \alpha^\dagger M \alpha \tag{1}$$

be a quadratic criterion to be minimized with respect to the vector of prediction parameters  $\mathbf{a} = [a_1, \dots, a_P]^t$ . Let us introduce the following partition for  $M$ :

$$M = \begin{bmatrix} \rho & \mathbf{r}^\dagger \\ \mathbf{r} & R \end{bmatrix} \tag{2}$$

so that the minimum of  $J(\mathbf{a})$  is reached by the prediction vector  $\hat{\mathbf{a}} = R^{-1} \mathbf{r}$ .

Our first contribution is to propose a simple condition on the structure of matrix  $M$  to ensure the stability of the allpole filter defined by  $\hat{\mathbf{a}}$ . Equivalently, the issue is to guarantee that the roots of the monic polynomial

$$A(z) \triangleq z^P - \sum_{k=1}^P a_k z^{P-k} \tag{3}$$

lie within the unit circle.

A. Sufficient Condition

For any square matrix  $Q$  of size  $n \times n$ , let us denote, respectively,  $\underline{Q}$ ,  $\overline{Q}$ ,  $\overline{Q}$ , and  $\overline{Q}$  as the northwest, southeast, northeast, and southwest matrices of size  $(n-1) \times (n-1)$  extracted from  $Q$ . According to such a notation, the matrix  $R$  introduced in (2) is nothing but  $\overline{M}$ , and

$$\Delta \triangleq \overline{M} - M \tag{4}$$

is the displacement matrix of  $M$ , whose rank defines the distance from Toeplitz matrices [7]. The following result shows that the positivity of the displacement matrix plays a specific role with regard to the stability of the estimated prediction filter.

*Theorem 1:* Let  $M$  be a positive definite matrix. Then, with the notations of (2) and (4),  $\hat{\mathbf{a}} = R^{-1} \mathbf{r}$  defines a stable prediction filter if  $\Delta \geq 0$ .

*Proof:* Let  $A$  be a monic polynomial of degree  $P$ , and let  $z_0$  stand for one of its roots

$$A(z) = (z - z_0)B(z) \quad (5)$$

where  $B$  is a monic polynomial of degree  $P - 1$ . In addition, let

$$\overset{\circ}{A}(z) = (z - z_0/|z_0|)B(z) \quad (6)$$

be the polynomial obtained by shifting  $z_0$  onto the unit circle. Finally, let us denote  $\alpha$ ,  $\beta$ , and  $\overset{\circ}{\alpha}$ , and  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\overset{\circ}{\mathbf{a}}$  as the innovation and prediction vectors corresponding to  $A$ ,  $B$ , and  $\overset{\circ}{A}$ , respectively, in conformity with the notation introduced in (3).

In terms of innovation vectors, (5) reads

$$\alpha = \begin{bmatrix} \beta \\ 0 \end{bmatrix} - z_0 \begin{bmatrix} 0 \\ \beta \end{bmatrix}$$

which provides the following expression for (1):

$$J(\mathbf{a}) = \beta^\dagger (\underline{M} + |z_0|^2 \overline{M} - z_0 \underline{M} - z_0^* \overline{M}) \beta.$$

In the same way, (6) yields

$$J(\overset{\circ}{\mathbf{a}}) = \beta^\dagger \left( \underline{M} + \overline{M} - \frac{z_0}{|z_0|} \underline{M} - \frac{z_0^*}{|z_0|} \overline{M} \right) \beta$$

and a combination of the latter two equations provides the following result:

$$J(\mathbf{a}) = \beta^\dagger \overline{M} \beta |z_0|^2 + (J(\overset{\circ}{\mathbf{a}}) - \beta^\dagger (\overline{M} + \underline{M}) \beta) |z_0| + \beta^\dagger \underline{M} \beta.$$

Since neither  $J(\overset{\circ}{\mathbf{a}})$  nor  $\beta$  depend on  $|z_0|$ ,  $J(\mathbf{a})$  is a quadratic function of  $|z_0|$ . Moreover, since  $M$  is positive,  $\overline{M}$  is also positive, and  $J(\mathbf{a})$  passes through a unique minimum on  $\mathbb{R}_+$ . It is easy to check that

$$\left. \frac{\partial J(\mathbf{a})}{\partial |z_0|} \right|_{|z_0|=1} = J(\overset{\circ}{\mathbf{a}}) + \beta^\dagger \Delta \beta \quad (7)$$

which is strictly positive. As a function of  $|z_0|$ , we can conclude that  $J(\mathbf{a})$  is strictly increasing for any  $|z_0| \geq 1$ . Hence, its unique minimum is necessarily reached *strictly inside* the unit circle. Then, as a function of  $\mathbf{a}$ , since  $M$  is positive,  $J$  passes through a unique minimum that is necessarily achieved for a polynomial  $\overset{\circ}{A}$  with all its roots within the unit circle.  $\square$

In the following, the matrix  $M$  will be said to be *canonical* when the conditions of Theorem 1 are fulfilled.

*Remark 1:* The conditions of Theorem 1 are  $M > 0$  and  $\Delta \geq 0$ , but the slightly modified conditions  $\overline{M} \geq 0$  and  $\Delta > 0$  are also sufficient, as is apparent from (7) (note that  $\Delta > 0 \Rightarrow \overline{M} > 0$ ).

*Remark 2:* Let  $\overset{\circ}{A}(z) = (z - 1/z_0^*)B(z)$  the polynomial obtained by “reflecting”  $z_0$  with respect to the unit circle, and let  $\overset{\circ}{\mathbf{a}}$  be the corresponding prediction vector. Then, it is easy to show that

$$J(\mathbf{a}) - |z_0|^2 J(\overset{\circ}{\mathbf{a}}) = (|z_0|^2 - 1) \beta^\dagger \Delta \beta. \quad (8)$$

This provides a simple alternative to (7) to conclude that  $\hat{A}$  has no roots outside the unit circle, but it does not prove that the roots are *strictly* interior.

*Remark 3:* The condition  $M > 0$  is clearly too restrictive: Positivity of  $\alpha^\dagger M \alpha$  could be required for “innovation-type” vectors  $\alpha = [1 \mid -\mathbf{a}^\dagger]^\dagger$  only. On the other hand,  $\Delta \geq 0$  depends on the value of the upperleft entry  $\rho$ , whereas the estimate  $\hat{\mathbf{a}} = (\overline{M})^{-1} \mathbf{r}$  does not depend on it. Actually, it can be shown that the conditions of Theorem 1 can be relaxed under the following form:  $\overline{\Delta} > 0$  and  $\tilde{\Delta} \geq 0$ , where  $\tilde{\Delta} = \Delta$ , save that  $\mathbf{r}^\dagger \hat{\mathbf{a}}$  is the upper-left entry of  $\tilde{\Delta}$ . Yet, such broader conditions are not necessary, whereas they do not enjoy the same simplicity as the original conditions of Theorem 1.

*Example 1—Toeplitz Case:* If matrix  $M$  is Toeplitz, then  $\Delta = 0$ , and (8) boils down to the simpler form  $J(\mathbf{a}) = |z_0|^2 J(\overset{\circ}{\mathbf{a}})$ . It is interesting to notice that in the given covariance case, the latter relation has a direct counterpart in terms of mean-squared prediction error, which classically ensures the stability of the prediction error filter [3].

*Example 2—Diagonal Case:* If matrix  $M$  is diagonal, the conditions of Theorem 1 are fulfilled for any increasing series of positive diagonal coefficients. This is a trivial example of a non-Toeplitz canonical matrix.

*Example 3—Mixed Case:* It is easy to check that the set of canonical matrices forms a convex cone. As a consequence, a positive definite Toeplitz matrix whose diagonal entries are augmented by any increasing positive sequence remains canonical.

Viewed as new possibilities of testing stability, the conditions of Theorem 1 or the broader conditions of Remark 3 are only of moderate interest since testing the positivity of a matrix is not simpler than directly testing the stability of the estimated predictor with a standard stability test. Moreover, such conditions are only sufficient, and they are mainly restricted to normal equation approaches. Nonetheless, they provide a new tool for the study of *structural stability* for some prediction methods, as shown in the following section.

### III. APPLICATION TO LEAST SQUARES PREDICTION ESTIMATION METHODS

#### A. Basic Cases

The most classical least squares prediction estimation methods correspond to quadratic forms  $J(\mathbf{a}) = \|X\alpha\|^2$ . By construction, the normal matrix  $M = X^\dagger X$  is positive semidefinite, and the data matrix  $X$  differs according to the windowing assumption. The four classical cases correspond to the *autocorrelation method* (AC), the *post-windowed method* (POST), the *covariance method* (COV), and the *prewindowed method* (PRE) [5]. Simple calculations yield, respectively

$$\begin{cases} \Delta^{\text{AC}} = \mathbf{0} \\ \Delta^{\text{POST}} = \mathbf{x}_P^* \mathbf{x}_P^\dagger \\ \Delta^{\text{COV}} = \mathbf{x}_P^* \mathbf{x}_P^\dagger - \mathbf{x}_N^* \mathbf{x}_N^\dagger \\ \Delta^{\text{PRE}} = -\mathbf{x}_N^* \mathbf{x}_N^\dagger \end{cases}$$

where  $\mathbf{x}_n \triangleq [x_n, \dots, x_{n-P+1}]^\dagger$ . Obviously, matrix  $M^{\text{AC}}$  is canonical; given Remark 1,  $M^{\text{POST}}$  is also canonical if  $\mathbf{x}_P \neq \mathbf{0}$ . On the other hand, neither  $M^{\text{COV}}$  nor  $M^{\text{PRE}}$  are canonical (unless  $\mathbf{x}_N = \alpha \mathbf{x}_P$ , with  $|\alpha| \leq 1$ , or  $\mathbf{x}_P = \mathbf{0}$ , respectively). In fact, the existence of counterexamples shows that the covariance and the prewindowed methods are not structurally stable [5].

#### B. Regularized Methods

Kitagawa and Gersch [6] have proposed a *smoothness priors long autoregressive method*, which is based on a penalized least squares

criterion based on the prewindowed approach

$$J_{\text{KG}}^{\text{PRE}}(\mathbf{a}) = \boldsymbol{\alpha}^\dagger M^{\text{PRE}} \boldsymbol{\alpha} + \lambda \sum_{p=1}^P p^{2k} a_p^2 \quad (9)$$

where  $\lambda$  is a regularization parameter, and  $k$  is the so-called *smoothness order*. The justification stems from the Parseval's relation [6]

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

Criterion (9) can be put into the form of (1), which yields

$$M_{\text{KG}}^{\text{PRE}} = M^{\text{PRE}} + \lambda \text{diag}\{p^{2k}\}_{p=0,\dots,P}.$$

The same regularization technique applies to the other windowing alternatives. In particular, the regularized form of the autocorrelation method has been studied in [8] in the context of Doppler spectral analysis. Since the associated NEM  $M_{\text{KG}}^{\text{COV}}$  has the mixed structure of Example 3, we can conclude that the *regularized autocorrelation method* is structurally stable for any smoothness order  $k \geq 0$  and any  $\lambda \geq 0$ . Furthermore, it remains stable if the penalizing term incorporates several terms corresponding to different smoothness orders. Finally, the smoothness order need not be restricted to entire values. For instance, the canonical matrix obtained for  $k = 1/2$  has a null second-order displacement rank [7], which is a potentially interesting property with a view to fast inversion.

The following corollary shows that the original regularized prewindowed method of Kitagawa and Gersch becomes structurally stable beyond a certain level of regularization. Similar results can be derived for the regularized versions of the covariance and modified covariance methods.

*Corollary 1:* For any  $k > 0$ ,  $M_{\text{KG}}^{\text{PRE}}$  is canonical if  $\lambda > \sum_{p=1}^P |x_{N+1-p}|^2 / (p^{2k} - (p-1)^{2k})$ .

*Proof:* Matrix  $M_{\text{KG}}^{\text{PRE}}$  is positive definite. Its displacement matrix reads  $\Delta_{\text{KG}}^{\text{PRE}} = \lambda D - \mathbf{x}_N^* \mathbf{x}_N^t$ , with  $D \triangleq \text{diag}\{p^{2k} - (p-1)^{2k}\}_{p=1,\dots,P}$ . From [9, Th. 32, p. 45]

$$\det \Delta_{\text{KG}}^{\text{PRE}} = \lambda^P (1 - \lambda^{-1} \mathbf{x}_N^* D^{-1} \mathbf{x}_N^t) \prod_{p=1}^P (p^{2k} - (p-1)^{2k})$$

and it is apparent that the condition  $\lambda \geq \mathbf{x}_N^* D^{-1} \mathbf{x}_N^t$  is necessary to the positive semidefiniteness of  $\Delta_{\text{KG}}^{\text{PRE}}$ . Actually, it is also sufficient since the  $P-1$  other conditions that express the positivity of the minors are similar but less restrictive than  $\det \Delta_{\text{KG}}^{\text{PRE}} \geq 0$ .  $\square$

The particular case  $k = 0$  provides a method that has been proposed *per se* in the context of *linear minimum free energy estimation* by Silverstein [10]. It basically reduces to adding a positive constant  $\lambda$  to the main diagonal of the NEM. Obviously, the autocorrelation version is still canonical since the NEM remains Toeplitz, positive definite. On the other hand, the case  $k = 0$  is excluded from the canonicity condition of Corollary 1. Yet, it is intuitive that such a method becomes structurally stable for large values of  $\lambda$ . This is actually so, since, from the sufficient condition of stability  $\|\hat{\mathbf{a}}\| < 1/P$  [2], it is possible to deduce that  $\lambda > \|\mathbf{r}\| \sqrt{P}$  ensures that  $\hat{\mathbf{a}}$  defines a stable prediction filter.

### C. Adaptive Versions

In order to extend least squares prediction methods to adaptive contexts, the normal approach is to reweight the successive terms of the criterion according to a forgetting factor. The resulting NEM reads  $M = X^\dagger \Gamma X$ , where  $\Gamma$  is a diagonal matrix with geometrically increasing positive entries on its main diagonal. For instance, let us

define  $\Gamma_{\gamma}^{\text{AC}} = \text{diag}\{\gamma^{N-k}\}_{k=1,\dots,N+P}$  in the autocorrelation case and  $\Gamma_{\gamma}^{\text{COV}} = \text{diag}\{\gamma^{N-k}\}_{k=P,\dots,N-1}$  in the covariance case, with  $0 < \gamma \leq 1$ . Then, we can deduce

$$\Delta_{\gamma}^{\text{AC}} = (1 - \gamma) \underline{M}^{\text{AC}} \quad (10a)$$

$$\Delta_{\gamma}^{\text{COV}} = (1 - \gamma) \underline{M}^{\text{COV}} + \gamma^{N-P} \mathbf{x}_P^* \mathbf{x}_P^t - \mathbf{x}_N^* \mathbf{x}_N^t. \quad (10b)$$

As a consequence, structural stability is preserved by the adaptive version of the autocorrelation method. In the same way, this could be shown for the adaptive postwindowed method. On the other hand, the adaptive version of the covariance method is not guaranteed to be structurally stable. However, from (10b), it becomes stable if  $\gamma$  is chosen such as

$$(1 - \gamma) \mathbf{x}_N^* \underline{M}^{\text{COV}} \mathbf{x}_N + \gamma^{N-P} |\mathbf{x}_P^t \mathbf{x}_P|^2 > |\mathbf{x}_N^t \mathbf{x}_N|^2.$$

## IV. CONCLUSION

In the framework of least squares prediction in the given data case, the estimated prediction vector  $\hat{\mathbf{a}}$  is the solution of a normal equation. In order to compute  $\hat{\mathbf{a}}$ , it is a classical result that the complexity of the appropriate generalized Levinson algorithm linearly increases with respect to the distance of the normal equation matrix to Toeplitz, i.e., the rank of the displacement matrix [7]. In this paper, we have shown that the positive definiteness of the displacement matrix ensures that the estimated prediction filter is stable (provided that the normal equation matrix is also positive definite). This result provides a unifying sufficient condition that proves that some classical least squares prediction methods are structurally stable: the autocorrelation method, the postwindowed method, and the autocorrelation version of the regularized method proposed by [6]. It also provides a simple lower bound on the regularization parameter for the original (prewindowed) version to be structurally stable.

## REFERENCES

- [1] Y. Bistritz, "Zero location with respect to the unit circle of discrete-time linear system polynomials," *Proc. IEEE*, vol. 72, pp. 1131–1142, Sept. 1984.
- [2] B. Picinbono and M. Benidir, "Some properties of lattice autoregressive filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 342–349, Apr. 1986.
- [3] S. Lang and J. McClellan, "A simple proof of stability for all-pole linear prediction model," *Proc. IEEE*, vol. 67, pp. 860–861, May 1979.
- [4] B. Friedlander, "Lattice filters for adaptive processing," *Proc. IEEE*, vol. 70, pp. 829–867, Aug. 1982.
- [5] S. M. Kay and S. L. Marple, "Spectrum analysis—A modern perspective," *Proc. IEEE*, vol. 69, pp. 1380–1419, Nov. 1981.
- [6] G. Kitagawa and W. Gersch, "A smoothness priors long AR model method for spectral estimation," *IEEE Trans. Automat. Contr.*, vol. AC-30, pp. 57–65, Jan. 1985.
- [7] B. Friedlander, M. Morf, T. Kailath, and L. Ljung, "New inversion formulas for matrices classified in terms of their distances from Toeplitz matrices," *Linear Algebra Appl.*, vol. 27, pp. 31–60, 1979.
- [8] J.-F. Giovannelli, A. Herment, and G. Demoment, "A Bayesian method for long AR spectral estimation: A comparative study," *IEEE Trans. Ultrason. Ferroelect., Freq. Contr.*, vol. 43, pp. 220–233, Mar. 1996.
- [9] P. Lascaux and R. Theodor, *Analyse Numérique Matricielle Appliquée à l'Art de l'Ingénieur*. Paris, France: Masson, 1986, vol. 1.
- [10] S. D. Silverstein, "Linear minimum free energy estimation: A computationally efficient noise suppression spectral estimation algorithm," *IEEE Trans. Signal Processing*, vol. 39, pp. 1348–1359, June 1991.