Available at WWW.MATHEMATICSWEB.ORG POWERED BY SCIENCE DIRECT®

Journal of Multivariate Analysis 87 (2003) 46–59

Journal of Multivariate Analysis

http://www.elsevier.com/locate/jmva

Characterization of the partial autocorrelation function of nonstationary time series

Serge Dégerine* and Sophie Lambert-Lacroix

Laboratoire LMC-IMAG, Joseph Fourier University, BP 53, 38041 Grenoble cedex 9, France

Received 6 March 2001

Abstract

The second order properties of a process are usually characterized by the autocovariance function. In the stationary case, the parameterization by the partial autocorrelation function is relatively recent. We extend this parameterization to the nonstationary case. The advantage of this function is that it is subject to very simple constraints in comparison with the autocovariance function which must be nonnegative definite. As in the stationary case, this parameterization is well adapted to autoregressive models or to the identification of deterministic processes.

AMS 2000 subject classifications: 60G12; 62M10

Keywords: Nonstationary processes; Discrete time; Second order properties; Partial autocorrelation

1. Introduction

Although the partial correlation notion was introduced many years ago by Yule [19], the parameterization of a stationary time series by the partial autocorrelation function (PACF) is relatively recent. This result is established by Barndorff-Nielsen and Schou [1] for autoregressive processes and by Ramsey [18] for the general stationary case. It is also observed by Burg [2], in the signal processing field, where the partial autocorrelation coefficients are called reflection coefficients. In fact, the

*Corresponding author. Fax: +33-04-76-63-12-63.

E-mail addresses: Serge.Degerine@imag.fr (S. Dégerine), Sophie.Lambert@imag.fr (S. Lambert-Lacroix).

0047-259X/03/$ - see front matter
doi:10.1016/S0047-259X(03)00025-3

one-to-one correspondence between autocorrelation function (ACF) and PACF is a classical result in orthogonal polynomial theory when the spectral measure has an infinite set of growth points (see [12]). The simplicity of the constraints on the PACF gave birth to many autoregressive estimation methods [2,5,9].

The extension of the PACF to the multivariate stationary case is a delicate point. The difficulty to define a partial autocorrelation matrix lies in the square root choice in normalizing the partial autocovariance. Morf et al. [16] propose to use the triangular square root. Dégerine [4] gives a general approach for the multivariate stationary case and introduces the canonical PACF. Estimation techniques are proposed in [6,10,17].

In the nonstationary case, the partial correlation coefficients appear in the generalization of Schur and Levinson–Durbin algorithms for nonlocally deterministic processes by Lev-Ari and Kailath [15]. In this note, we show that the PACF, like the ACF, can be used in order to parametrize any nonstationary time series. We precise its variation domain $\mathscr{D}_\beta$ and present a new algorithm which constructs a process sequence with a prescribed PACF. This construction proves that any element of $\mathscr{D}_\beta$ is a PACF. We extend the Levinson–Durbin algorithm to the general nonstationary situation and use it to prove that the PACF associated with any ACF is unique. We show that the PACF is well adapted to the identification of nonnegative definite functions and give a nice characterization of both deterministic and autoregressive processes. As in the stationary case, we observe that the PACF is subject to only very simple constraints in comparison with the ACF. This suggests to use the PACF in order to estimate, in a nonparametric way, the second order characteristics of a process. Furthermore, this approach has lead to a new estimation method for periodic autoregressive processes [14], which extends the one based on the empirical PACF in the stationary case [5]. Notice that a new time-dependent power spectrum is clearly related to the PACF (see [7]).

The next section is devoted to the parametrization by the PACF and the last one relates some interesting results coming from the use of this function.

2. Partial autocorrelation function

2.1. Notations and definitions

Let $X(\cdot) = \{X(t), t \in \mathbb{Z}\}$ be a scalar complex valued nonstationary process with zero mean and second order moments. In this paper we are only concerned with the second order properties of the process $X(\cdot)$. Consequently it is convenient to use a geometrical approach by considering the following Hilbert space $\mathscr{M}$, with the hermitian product $\langle U, V \rangle = E\{U\bar{V}\} = Cov\{U, V\}$. The elements of $\mathscr{M}$ are the linear combinations, with complex coefficients, of elements of $\{X(t), t \in \mathbb{Z}\}$ and their limits for mean square convergence. So the ACF $R(\cdot,\cdot)$ is defined by

$$R(t,s) = \langle X(t), X(s) \rangle = Cov\{X(t), X(s)\} \quad (t,s) \in \mathbb{Z}^2.$$

This function is nonnegative definite (n.n.d.), that is for all $s \leqslant t$, $R_{s,t} = \{R(s+i, s+j)\}_{i,j=0,\ldots,t-s}$ is n.n.d. (as the covariance matrix of the random vector $[X(s), \ldots, X(t)]^T$). As in [18], it is convenient to consider the decomposition

$$\mathscr{D}_R = \mathscr{I}(\mathscr{D}_R) + \mathscr{B}(\mathscr{D}_R)$$

of the set $\mathscr{D}_R$ of n.n.d. functions, where the interior $\mathscr{I}(\mathscr{D}_R)$ consists of all positive definite (p.d.) functions (all the matrices $R_{s,t}$ are p.d.), while the boundary $\mathscr{B}(\mathscr{D}_R)$ consists of all n.n.d. functions for which some matrix $R_{s,t}$ is singular. In this case, the corresponding process $X(\cdot)$ will be called *locally deterministic* since for some $s \leqslant t$, the components $X(s), \ldots, X(t)$ are almost surely linearly dependent. In the opposite case, the process $X(\cdot)$ will be called *nonlocally deterministic*.

Let $X^{\mathrm{f}}(t;s), s \leqslant t$, denote the orthogonal projection of $X(t)$ on the closed linear subspace $\mathscr{M}(s;t-1) = \overline{sp}\{X(s), \ldots, X(t-1)\}$, i.e. the linear predictor of $X(t)$ given $X(t-1), \ldots, X(s)$, with the convention $X^{\mathrm{f}}(t;t) = \vec{0}$. The $(t-s)$th-order forward partial innovation is $\varepsilon^{\mathrm{f}}(t;s) = X(t) - X^{\mathrm{f}}(t;s)$ and we put $\sigma^{2\mathrm{f}}(t;s) = ||\varepsilon^{\mathrm{f}}(t;s)||^2 = Var\{\varepsilon^{\mathrm{f}}(t;s)\}$. The associated normalized innovation is defined, for $s \leqslant t$, by

$$\eta^{\mathrm{f}}(t;s) = \frac{\varepsilon^{\mathrm{f}}(t;s)}{\sigma^{\mathrm{f}}(t;s)}, \quad \sigma^{\mathrm{f}}(t;s) = \sqrt{\sigma^{2\mathrm{f}}(t;s)},$$

with the convention $0^{-1} = 0$. Notice that this convention is necessary in the locally deterministic case when $X(t) \in \mathscr{M}(s;t-1)$. All the notions obtained by reversing the time index are denoted by b for "backward"; for instance, for $s \leqslant t$, the backward innovation $\varepsilon^{\mathrm{b}}(s;t) = X(s) - X^{\mathrm{b}}(s;t)$ leads to $\eta^{\mathrm{b}}(s;t) = \varepsilon^{\mathrm{b}}(s;t)/\sigma^{\mathrm{b}}(s;t)$. The PACF $\beta(\cdot,\cdot)$ describes, for all (t,s) of $\mathbb{Z}^2$, the partial correlation coefficient between $X(t)$ and $X(s)$ in the set $\{X(s), \ldots, X(t)\}$.

Definition 1. The partial autocorrelation function $\beta(\cdot,\cdot)$ of $X(\cdot)$ is defined on $\mathbb{Z}^2$ by

$$\beta(t,s) = \begin{cases} \langle \eta^{\mathrm{f}}(t;s+1), \eta^{\mathrm{b}}(s;t-1) \rangle & \text{if } s < t, \\ ||X(t)||^2 & \text{if } s = t, \\ \langle \eta^{\mathrm{b}}(t;s-1), \eta^{\mathrm{f}}(s;t+1) \rangle & \text{if } s > t. \end{cases}$$

Note that, setting $\beta(t,t) = Var\{X(t)\}$ instead of 1 in the above definition, the function $\beta(\cdot,\cdot)$, like the ACF, characterizes the second order properties of $X(\cdot)$.

2.2. PACF variation domain

The advantage of the PACF is that its variation domain can be easily described. For $t \neq s$, we have $|\beta(t,s)| \leqslant 1$ and the equality to 1 implies linear relationships. Indeed, for $s < t$, $|\beta(t,s)| = 1$ if and only if s is the largest integer such that $X(t)$ belongs to the set $\mathscr{M}(s;t-1)$. By convention, the partial correlation is then set equal to 0 everytime it is undefined, i.e. for the points $(t, s-k)$ and $(t+k, s), k \geqslant 1$. In the same way, we have $\beta(t, t-k) = \beta(t+k, t) = 0$ for $k \geqslant 0$, when the variable $X(t)$ is

equal to zero almost surely. Note that our convention differs from that of [18] but is well adapted to the one-to-one correspondence given in Theorem 3 below. Precisely, the PACF $\beta(\cdot,\cdot)$ is in the set $\mathscr{D}_\beta$ defined by the following conditions:

(i) $\beta(s,t)=\overline{\beta(t,s)}$ with $\beta(t,t)\geqslant 0$ and $|\beta(t,s)|\leqslant 1$ if $t\neq s, (t,s)\in\mathbb{Z}^2$,
(ii) $\beta(t,t)=0\Rightarrow\beta(t,s)=0$, $s\in\mathbb{Z}$,
(iii) $|\beta(t,s)|=1$, $s<t\Rightarrow\beta(t,s-k)=\beta(t+k,s)=0$, $k\geqslant 1$.

We also consider the decomposition

$$\mathscr{D}_\beta=\mathscr{I}(\mathscr{D}_\beta)+\mathscr{B}(\mathscr{D}_\beta),$$

where the interior $\mathscr{I}(\mathscr{D}_\beta)$ consists of all $\beta(\cdot,\cdot)\in\mathscr{D}_\beta$ satisfying, for all $t\in\mathbb{Z}$, $\beta(t,t)>0$ and $|\beta(t,s)|<1$ for $s\neq t$. So the boundary $\mathscr{B}(\mathscr{D}_\beta)$ consists of all $\beta(\cdot,\cdot)\in\mathscr{D}_\beta$ satisfying $\beta(t,t)=0$ for some $t\in\mathbb{Z}$, or $|\beta(t,s)|=1$ for some $(t,s)\in\mathbb{Z}^2$ with $s\neq t$. Clearly the PACF of a locally deterministic process is in $\mathscr{B}(\mathscr{D}_\beta)$ and we will see in Section 3 that $\mathscr{I}(\mathscr{D}_\beta)$ corresponds to the nonlocally deterministic case.

We now show that $\mathscr{D}_\beta$ is the PACF variation domain. It means that any function in $\mathscr{D}_\beta$ is the PACF of a nonstationary or stationary process. To do so, we give in Theorem 2 an algorithm which allows to construct a process with any prescribed PACF in $\mathscr{D}_\beta$. The following recursion is the basis of this constructive process.

Theorem 1. *The partial innovations satisfy, for $s<t$, the recurrence formulae*

$$\begin{aligned}\varepsilon^{\mathrm{f}}(t;s)&=\varepsilon^{\mathrm{f}}(t;s+1)-\beta(t,s)\tfrac{\sigma^{\mathrm{f}}(t;s+1)}{\sigma^{\mathrm{b}}(s;t-1)}\varepsilon^{\mathrm{b}}(s;t-1),\\ \varepsilon^{\mathrm{b}}(s;t)&=\varepsilon^{\mathrm{b}}(s;t-1)-\beta(s,t)\tfrac{\sigma^{\mathrm{b}}(s;t-1)}{\sigma^{\mathrm{f}}(t;s+1)}\varepsilon^{\mathrm{f}}(t;s+1),\end{aligned}\tag{1}$$

and the residual variances are given by

$$\begin{aligned}\sigma^{2\mathrm{f}}(t;s)&=\beta(t,t)\prod_{j=1}^{t-s}[1-|\beta(t,t-j)|^2],\\ \sigma^{2\mathrm{b}}(s;t)&=\beta(s,s)\prod_{j=1}^{t-s}[1-|\beta(s,s+j)|^2].\end{aligned}\tag{2}$$

Proof. From the projection of $X(t)$ on the orthogonal decomposition

$$\mathscr{M}(s;t-1)=\mathscr{M}(s+1;t-1)\oplus\overline{sp}\{\eta^{\mathrm{b}}(s;t-1)\},$$

we have

$$X^{\mathrm{f}}(t;s)=X^{\mathrm{f}}(t;s+1)+\beta(t,s)\sigma^{\mathrm{f}}(t;s+1)\eta^{\mathrm{b}}(s;t-1),$$

and consequently the first recurrence formula relationship holds. Since the variables $\varepsilon^{\mathrm{f}}(t;s)$ and $\varepsilon^{\mathrm{b}}(s;t-1)$ are uncorrelated, we obtain

$$\sigma^{2\mathrm{f}}(t;s)=\sigma^{2\mathrm{f}}(t;s+1)[1-|\beta(t,s)|^2]$$

and then the expression of the residual variance. The second relationships in (1) and (2) are proved in a similar way. □

Our constructive process is the following.

Theorem 2. *Let* $\{Z(t), t\in\mathbb{Z}\}$ *be a white noise sequence with unit variance and* $\beta(\cdot,\cdot)$ *an element of* $\mathscr{D}_\beta$. *Then, starting from any* s *in* $\mathbb{Z}$, *the sequence* $\{X(t), t\geqslant s\}$ *defined by, for* $t = s, s+1, \ldots$:

$$\eta^{\mathrm{f}}(t;s) = Z(t) \quad \textit{if } \sigma^{2\mathrm{f}}(t;s) = \beta(t,t)\prod_{j=1}^{t-s}[1-|\beta(t,t-j)|^2] > 0, \textit{ else } \vec{0},$$

for $k = s, \ldots, t-1$:

$$\eta^{\mathrm{f}}(t;k+1) = [1-|\beta(t,k)|^2]^{\frac{1}{2}}\eta^{\mathrm{f}}(t;k) + \beta(t,k)\eta^{\mathrm{b}}(k;t-1),$$

$$\eta^{\mathrm{b}}(k;t) = [1-|\beta(k,t)|^2]^{-\frac{1}{2}}\{\eta^{\mathrm{b}}(k;t-1) - \beta(k,t)\eta^{\mathrm{f}}(t;k+1)\},$$

$$X(t) = [\beta(t,t)]^{\frac{1}{2}}\eta^{\mathrm{f}}(t;t), \quad \eta^{\mathrm{b}}(t;t) = \eta^{\mathrm{f}}(t;t),$$

admits a PACF which coincides with $\beta(\cdot,\cdot)$ *on* $\{(u,t)\in\mathbb{Z}^2, u, t = s, s+1, \ldots\}$.

Proof. We suppose that the sequence $\{X(u), u = s, \ldots, t-1\}$ has been constructed in terms of the variables $Z(u), u = s, \ldots, t-1$, in such a way that its PACF coincides with $\beta(\cdot,\cdot)$ on $\{(u,v)\in\mathbb{Z}^2, u, v = s, \ldots, t-1\}$.

Furthermore, the basis $\{\eta^{\mathrm{b}}(k;t-1), k = s, \ldots, t-1\}$ of the corresponding space $\mathscr{M}(s;t-1)$ is available. Note that these hypotheses are satisfied for $t = s+1$ after the first recurrence step:

$$\eta^{\mathrm{f}}(s;s) = \eta^{\mathrm{b}}(s;s) = Z(s), \quad X(s) = \beta(s,s)^{\frac{1}{2}}\eta^{\mathrm{f}}(s;s).$$

Through the first recurrence formula, the algorithm defines the variable $X(t)$ as

$$X(t) = \sum_{k=s}^{t-1}\sigma^{\mathrm{f}}(t;k+1)\beta(t,k)\eta^{\mathrm{b}}(k;t-1) + \sigma^{\mathrm{f}}(t;s)\eta^{\mathrm{f}}(t;s),$$

where $\eta^{\mathrm{f}}(t;s) = Z(t)$ is orthogonal to $\mathscr{M}(s;t-1)$. This definition is equivalent to $X(t) = X^{\mathrm{f}}(t;s) + \sigma^{\mathrm{f}}(t;s)\eta^{\mathrm{f}}(t;s)$. According to the $\mathscr{M}(s;t-1)$ basis choice, we have

$$\begin{aligned}\varepsilon^{\mathrm{f}}(t;s+1) &= X(t) - X^{\mathrm{f}}(t;s+1)\\ &= \sigma^{\mathrm{f}}(t;s+1)\beta(t,s)\eta^{\mathrm{b}}(s;t-1) + \sigma^{\mathrm{f}}(t;s)\eta^{\mathrm{f}}(t;s),\end{aligned}$$

$$||\varepsilon^{\mathrm{f}}(t;s+1)||^2 = \sigma^{2\mathrm{f}}(t;s+1)|\beta(t,s)|^2 + \sigma^{2\mathrm{f}}(t;s) = \sigma^{2\mathrm{f}}(t;s+1),$$

$$\langle \varepsilon^{\mathrm{f}}(t;s+1), \eta^{\mathrm{b}}(s;t-1)\rangle = \sigma^{\mathrm{f}}(t;s+1)\beta(t,s).$$

So $\sigma^{2f}(t;s+1)$ is the norm of $\varepsilon(t;s+1)$ and $\beta(t,s)$ is the partial correlation between $X(t)$ and $X(s)$. For $j=s+2,\ldots,t$, this recurrence carries on with

$$\varepsilon^{f}(t;j)=\sum_{k=s}^{j-1}\sigma^{f}(t;k+1)\beta(t,k)\eta^{b}(k;t-1)+\sigma^{f}(t;s)\eta^{f}(t;s),$$

$$\begin{aligned}||\varepsilon^{f}(t;j)||^2&=\sum_{k=s}^{j-1}\sigma^{2f}(t;k+1)|\beta(t,k)|^2+\sigma^{2f}(t;s)\\&=\sum_{k=s+1}^{j-1}\sigma^{2f}(t;k+1)|\beta(t,k)|^2+\sigma^{2f}(t;s+1)=\cdots=\sigma^{2f}(t;j),\end{aligned}$$

$$\langle\varepsilon^{f}(t;j),\eta^{b}(j-1;t-1)\rangle=\sigma^{f}(t;j)\beta(t,j-1).$$

Thus $\sigma^{2f}(t;j)$ is the norm of $\varepsilon(t;j)$ and $\beta(t,j-1)$ is the partial correlation between $X(t)$ and $X(j-1)$. Finally

$$||X(t)||^2=\sum_{k=s}^{t-1}\sigma^{2f}(t;k+1)|\beta(t,k)|^2+\sigma^{2f}(t;s)=\beta(t,t).$$

The first step of the recurrence hypotheses at time t is true. Consequently the second one constructs effectively (see Theorem 1) the new basis $\{\eta^{b}(k;t),k=s,\ldots,t\}$ of the space $\mathscr{M}(s;t)$. □

This algorithm transforms a white noise sequence $\{Z(t),t\geqslant 0\}$ into a process $\{X(t),t\geqslant 0\}$ with specified values of $\beta(\cdot,\cdot)$ on $\mathbb{N}^2$. At each time t, the new sample $X(t)$ is obtained from $Z(t)$, from the past, advisedly stored in the form $\eta^{b}(s;t-1),s=0,\ldots,t-1$, and from the necessary new coefficients $\beta(t,s),s=0,\ldots,t$. Starting from $t=0$, a two-sided sequence $\{X(t),t\in\mathbb{Z}\}$, in which $X(t)$ and $X(-t)$ are generated alternatively, can be associated in a similar way to any $\beta(\cdot,\cdot)$ of $\mathscr{D}_\beta$. In such a construction, we have at time $t, X(-t+1),\ldots,X(t-1)$, that is $\eta^{b}(k;t-1)$ and $\eta^{f}(-k;-t+1)$ for $k=-t+1,\ldots,t-1$. The new variables $X(t)$ and $X(-t)$ are obtained in the following way:

$$\begin{aligned}X(t)&=X^{f}(t;-t+1)+\varepsilon^{f}(t;-t+1)\\&=\sum_{k=-t+1}^{t-1}\sigma^{f}(t;k+1)\beta(t,k)\eta^{b}(k;t-1)+\sigma^{f}(t;-t+1)Z(t),\end{aligned}$$

$$\begin{aligned}X(-t)&=X^{b}(-t;t)+\varepsilon^{b}(-t;t)\\&=\sum_{k=-t+1}^{t}\sigma^{b}(-t;k-1)\beta(-t,k)\eta^{f}(k;-t+1)+\sigma^{b}(-t;t)Z(-t),\end{aligned}$$

where $\sigma^{2f}(t;k)$ and $\sigma^{2b}(-t;k)$ are determined by means of the relations in (2). Therefore from any element $\beta(\cdot,\cdot)$ of $\mathscr{D}_\beta$, it is possible to construct a nonstationary

sequence $\{X(t), t\in\mathbb{Z}\}$ which admits $\beta(\cdot,\cdot)$ as PACF. This shows that the application $R(\cdot,\cdot)\to\beta(\cdot,\cdot)$ maps $\mathscr{D}_R$ onto $\mathscr{D}_\beta$.

2.3. The generalized Levinson–Durbin (GLD) Algorithm

In order to prove the one-to-one correspondence between $\mathscr{D}_R$ and $\mathscr{D}_\beta$, we extend the Levinson–Durbin Algorithm of Lev-Ary and Kailath [15] for nonstationary processes to the locally deterministic case. Let us introduce the following notations for $s\leqslant t$:

$$\varepsilon^{\mathrm{f}}(t;s)=\sum_{j=0}^{t-s} a_t^{\mathrm{f}}(t-s,j)X(t-j),\quad a_t^{\mathrm{f}}(t-s,0)=1,$$

$$\varepsilon^{\mathrm{b}}(s;t)=\sum_{j=0}^{t-s} a_t^{\mathrm{b}}(t-s,j)X(s+j),\quad a_t^{\mathrm{b}}(t-s,0)=1. \tag{3}$$

The coefficients in these decompositions are not uniquely defined in the locally deterministic case. In the algorithm below, they are selected recursively by the process itself, but the correspondence between $R(\cdot,\cdot)$ and $\beta(\cdot,\cdot)$ given by (4) is satisfied with any set of coefficients.

Theorem 3 (GLD Algorithm). *The correspondence between $R(\cdot,\cdot)$ and $\beta(\cdot,\cdot)$ on $[s,\ldots,t]^2$ is realized as follows.*

For $k=s,\ldots,t$,

$$\sigma^{2\mathrm{f}}(k;k)=\sigma^{2\mathrm{b}}(k;k)=\beta(k,k)=R(k,k).$$

For $n=1,\ldots,t-s$,

for $k=s+n,\ldots,t$, with the conventions $\sum_{j=1}^{0}\ldots=0$ and $0^{-1}=0$:

$$\beta(k,k-n)=\frac{R(k,k-n)+\sum_{j=1}^{n-1} a_k^{\mathrm{f}}(n-1,j)R(k-j,k-n)}{\sigma^{\mathrm{f}}(k;k-n+1)\sigma^{\mathrm{b}}(k-n;k-1)}, \tag{4}$$

if $n\neq t-s$ and $k\geqslant s+n+1$:

$$\sigma^{2\mathrm{f}}(k;k-n)=[1-|\beta(k,k-n)|^2]\sigma^{2\mathrm{f}}(k;k-n+1),$$

$$\sigma^{2\mathrm{b}}(k-n-1;k-1)=[1-|\beta(k-n-1,k-1)|^2]\sigma^{2\mathrm{b}}(k-n-1;k-2), \tag{5}$$

$$a_k^{\mathrm{f}}(n,n)=-\beta(k,k-n)\frac{\sigma^{\mathrm{f}}(k;k-n+1)}{\sigma^{\mathrm{b}}(k-n;k-1)}, \tag{6}$$

$$a_k^{\mathrm{b}}(n,n)=-\beta(k-n,k)\frac{\sigma^{\mathrm{b}}(k-n;k-1)}{\sigma^{\mathrm{f}}(k;k-n+1)}, \tag{7}$$

for $j=1,\ldots,n-1$:

$$a_k^{\mathrm{f}}(n,j)=a_k^{\mathrm{f}}(n-1,j)+a_k^{\mathrm{f}}(n,n)a_{k-1}^{\mathrm{b}}(n-1,n-j), \tag{8}$$

$$a_k^{\mathrm{b}}(n,j)=a_{k-1}^{\mathrm{b}}(n-1,j)+a_k^{\mathrm{b}}(n,n)a_k^{\mathrm{f}}(n-1,n-j). \tag{9}$$

Proof. Let $X(\cdot)$ be a process with ACF and PACF given by their geometrical definitions. The algebraic relationships between $R(\cdot,\cdot)$ and $\beta(\cdot,\cdot)$ on $[s,\ldots,t]^2$ are obtained by determining, recursively from $n=1$ to $t-s-1$, the decompositions (3) of all the nth-order partial innovations defined on this domain. Relations (5)–(9) follow from Theorem 1 and

$$\beta(k,k-n)=\frac{\langle \varepsilon^{\mathrm{f}}(k;k-n+1),X(k-n)\rangle}{\sigma^{\mathrm{f}}(k;k-n+1)\sigma^{\mathrm{b}}(k-n;k-1)}$$

leads to (4). When the denominator of the above fraction is equal to zero, the numerator is also equal to zero. So the convention in the geometrical definition of $\beta(k,k-n)$ agrees with $0^{-1}=0$ in the algorithm. □

Clearly, the GLD Algorithm shows that two different ACF lead to distinct PACF. Notice that the application $\beta(\cdot,\cdot)\to R(\cdot,\cdot)$ is simply obtained by writing (4) in the form

$$R(k,k-n)=\sigma^{\mathrm{f}}(k;k-n+1)\sigma^{\mathrm{b}}(k-n;k-1)\beta(k,k-n)-\sum_{j=1}^{n-1}a_k^{\mathrm{f}}(n-1,j)R(k-j,k-n),$$

even in the locally deterministic case.

3. Miscellaneous results

Here are presented some straightforward consequences of the parameterization given by the PACF.

3.1. On the nonnegative definiteness property

The extension of the Levinson–Durbin Algorithm to the locally deterministic case allows to determine a Cholesky decomposition of a generalized inverse of any n.n.d. hermitian matrix. Indeed, let us recall that $R_{s,t}=\{R(s+i,s+j)\}_{i,j=0,\ldots,t-s}$ represents the covariance matrix of the random vector $[X(s),\ldots,X(t)]^T$ and let us note that

$$\varepsilon_{s,t}^{\mathrm{f}}=\begin{pmatrix}\varepsilon^{\mathrm{f}}(s;s)\\ \varepsilon^{\mathrm{f}}(s+1;s)\\ \vdots\\ \varepsilon^{\mathrm{f}}(t;s)\end{pmatrix}=A_{s,t}^{\mathrm{f}}\begin{pmatrix}X(s)\\ X(s+1)\\ \vdots\\ X(t)\end{pmatrix},$$

where

$$A^{\mathrm{f}}_{s,t} = \begin{pmatrix} 1 & & & \\ a^{\mathrm{f}}_{s+1}(1,1) & 1 & (0) & \\ \vdots & \ddots & \ddots & \\ a^{\mathrm{f}}_{t}(t-s,t-s) & \cdots & a^{\mathrm{f}}_{t}(t-s,1) & 1 \end{pmatrix}$$

and

$$E\{\varepsilon^{\mathrm{f}}_{s,t}\varepsilon^{\mathrm{f}*}_{s,t}\} = \begin{pmatrix} \sigma^{2\mathrm{f}}(s;s) & & & \\ & \sigma^{2\mathrm{f}}(s+1;s) & (0) & \\ & (0) & \ddots & \\ & & & \sigma^{2\mathrm{f}}(t;s) \end{pmatrix} = \Sigma^{2\mathrm{f}}_{s,t},$$

where $*$ denotes the conjugate transpose. Then a decomposition of a generalized inverse $R^{-}_{s,t}$ of the covariance matrix $R_{s,t}$ is given by $R^{-}_{s,t} = A^{\mathrm{f}*}_{s,t}\Sigma^{2\mathrm{f}^{+}}_{s,t}A^{\mathrm{f}}_{s,t}$, where the matrices $\Sigma^{2\mathrm{f}}_{s,t}$ and $A^{\mathrm{f}}_{s,t}$ are provided by the algorithm. The diagonal matrix $\Sigma^{2\mathrm{f}^{+}}_{s,t}$ is the Moore–Penrose inverse of $\Sigma^{2\mathrm{f}}_{s,t}$ obtained by inverting the terms which are not equal to zero. This decomposition is in fact a by-product of the GLD Algorithm which determines the coefficients of all the innovations $\varepsilon^{\mathrm{f}}(v;u)$ and $\varepsilon^{\mathrm{b}}(u;v)$, $s \leqslant u \leqslant v \leqslant t$. This process, which allows to obtain a Cholesky decomposition of a n.n.d. hermitian matrix, is not common but follows the one proposed by Delsarte et al. [8] in the nonlocally deterministic case. Furthermore, the extension of the Levinson–Durbin Algorithm to the locally deterministic case provides an easy way to check that a given matrix $R_{s,t}$ is n.n.d. and to describe all n.n.d. functions $R(\cdot,\cdot)$ which extend this set of values.

The decomposition $\Sigma^{2\mathrm{f}}_{s,t} = A^{\mathrm{f}}_{s,t}R_{s,t}A^{\mathrm{f}*}_{s,t}$ implies that the determinants of the matrices $R_{s,t}$ and $\Sigma^{2\mathrm{f}}_{s,t}$ are equal and then,

$$|R_{s,t}| = \prod_{k=s}^{t} \sigma^{2\mathrm{f}}(k;s),$$

that is, in terms of partial correlations (see Theorem 1),

$$|R_{s,t}| = \left[\prod_{k=s}^{t} \beta(t,t)\right] \prod_{k=s+1}^{t} \prod_{j=1}^{k-s} [1 - |\beta(k,k-j)|^2].$$

When the process is stationary, we obtain the formula

$$|R_{t-s}| = |R_n| = \beta(0)^{n+1} \prod_{j=1}^{n} [1 - |\beta(j)|^2]^{n+1-j},$$

where $R_{t-s} = R_{s,t}, n = t-s$, and $\beta(j) = \beta(t,t-j)$, $j \geqslant 0$. $R_{s,t}$ is p.d. if and only if $|R_{s,s+j}| > 0$, $j = 0, \ldots, t-s$ [11, Theorem 19, p. 337]. So the $|R_{s,t}|$ expression and the GLD Algorithm give immediately the one-to-one correspondence between $R(\cdot,\cdot)$ and $\beta(\cdot,\cdot)$ in the nonlocally deterministic case and then $\mathscr{I}(\mathscr{D}_\beta)$ corresponds to

$\mathscr{I}(\mathscr{D}_R)$. This argument was used by Ramsey [18] in the analogous situation for stationary processes. It is also the proof given in [12] where this situation only is considered. However, the condition $|R_{s,t}| \geqslant 0$, for all (s,t), is not sufficient for the nonnegative definiteness property of $R(\cdot,\cdot)$. On the other hand, if $R(\cdot,\cdot)$ and $\beta(\cdot,\cdot)$ are related by the GLD Algorithm, the Cholesky factorization $R_{s,t} = (A^{\mathrm{f}}_{s,t})^{-1}\Sigma^{2\mathrm{f}}_{s,t}(A^{\mathrm{f}*}_{s,t})^{-1}$ holds and proves that $R_{s,t}$ is n.n.d. when $\beta(\cdot,\cdot)$ belongs to $\mathscr{D}_\beta$. This is the proof used by Burg [2] in the stationary case where the coefficients in $A^{\mathrm{f}}_{s,t}$ are easily uniquely selected for a singular Tœplitz matrix $R_{s,t}$. In the nonstationary case, the Cholesky factorization of $R_{s,t}$, which follows immediately from our approach, is more delicate to prove directly. Notice that Theorem 2 shows that any $\beta(\cdot,\cdot) \in \mathscr{D}_\beta$ is a PACF without the GLD Algorithm. Furthermore, it provides an elegant way to simulate a process with a given PACF. In the stationary case, a PACF $\beta(\cdot)$ belongs to $\mathscr{B}(\mathscr{D}_\beta)$ when $\beta(0) = 0$ or when there exists $p \geqslant 1$ such that $|\beta(p)| = 1$. That is the process $X(\cdot)$ satisfies the stochastic difference equation (cf. (3))

$$\sum_{j=0}^{p} a(p,j)X(t-j) = 0, \quad a(p,0) = 1.$$

The Levinson–Durbin Algorithm gives $R(0), \ldots, R(p-1)$ and also $R(p)$ so that the Toeplitz matrix $R_{p-1} = \{R(i-j)\}_{i,j=0,\ldots,p-1}$ is p.d. but the above stochastic difference equation must be used in order to prove that the Toeplitz matrix R_p, using $R(p)$, is n.n.d. as the covariance matrix of $\{X(1), \ldots, X(p+1)\}$. So in the sufficiency part of the proof of [18] (cf. Case IV), $\{X(1), \ldots, X(p)\}$ must be considered first instead of $\{X(1), \ldots, X(p+1)\}$.

3.2. Deterministic processes

As in the stationary case (cf. [18]), a deterministic process is easily described. Iterating the first relation of (1) in Theorem 1, we have

$$\varepsilon^{\mathrm{f}}(t; t-n) = X(t) - \sum_{k=1}^{n} \beta(t, t-k)\sigma^{\mathrm{f}}(t; t-k+1)\eta^{\mathrm{b}}(t-k; t-1).$$

Then, using mean square convergence, the innovation process $\varepsilon(t)$ satisfies

$$\varepsilon(t) = X(t) - \sum_{k=1}^{+\infty} \beta(t, t-k)\sigma^{\mathrm{f}}(t; t-k+1)\eta^{\mathrm{b}}(t-k; t-1),$$

and its variance is given by

$$\sigma^2_\varepsilon(t) = \lim_{n \to +\infty} \sigma^{2\mathrm{f}}(t; t-n) = \beta(t,t) \prod_{k=1}^{+\infty} [1 - |\beta(t, t-k)|^2].$$

So the process $X(\cdot)$ is deterministic if and only if its PACF satisfies one of the following conditions:

(i) $\beta(t,t) = 0$,

(ii) $|\beta(t,s)|=1$ for some $s<t$,
(iii) $\sum_{k=1}^{+\infty}|\beta(t,t-k)|^2=+\infty$,

for every t in $\mathbb{Z}$.

3.3. Autoregressive processes

In the nonstationary case, we will call $X(\cdot)$ an autoregressive process of order p, denoted by $AR(p)$, if for all $t\in\mathbb{Z}$, there exist some constants $a_t(k)$, $k=1,\ldots,p$ such that

$$\sum_{k=0}^{p} a_t(k)X(t-k)=\varepsilon(t), \quad a_t(0)=1, \tag{10}$$

where $\varepsilon(\cdot)=\{\varepsilon(t),t\in\mathbb{Z}\}$ is the innovation process and p is the smallest integer for which these relationships hold. Let us recall that $\varepsilon(\cdot)$ is a sequence of zero-mean uncorrelated variables such that $\varepsilon(t)$, with variance $\sigma_\varepsilon^2(t)\geqslant 0$, is uncorrelated with $X(s)$, $s<t$. Contrarily to the stationary case, we do not know if a process satisfying (10) with any white noise sequence is $AR(p)$. Furthermore, the process $X(\cdot)$ can be locally deterministic because the variance $\sigma_\varepsilon^2(t)$ can vanish. As in the stationary case (see [18]), the PACF characterizes in a simple way this family of processes.

Theorem 4. *A process $X(\cdot)$ is autoregressive of order p if and only if its PACF $\beta(\cdot,\cdot)$ satisfies*

$$\beta(t,t-k)=0,\ \forall t\in\mathbb{Z},\ \forall k>p; \qquad \exists t\in\mathbb{Z},\ \beta(t,t-p)\neq 0.$$

Proof. If $X(\cdot)$ is $AR(p)$, then the process $\varepsilon(\cdot)$ in (10) is the innovation process. Moreover $\varepsilon^{\mathrm{f}}(t;t-n)=\varepsilon^{\mathrm{f}}(t;t-p)=\varepsilon(t)$ for $n\geqslant p$ and the variances equality

$$\begin{aligned}\sigma^{2\mathrm{f}}(t;t-n)&=\beta(t,t)\prod_{k=1}^{n}[1-|\beta(t,t-k)|^2]\\&=\beta(t,t)\prod_{k=1}^{p}[1-|\beta(t,t-k)|^2]=\sigma^{2\mathrm{f}}(t;t-p),\end{aligned}$$

shows that $\beta(t,t-k)=0$ for $k>p$. This is obvious if $\sigma^{2\mathrm{f}}(t;t-p)>0$, otherwise this variance is equal to zero because $\beta(t,t)=0$ or because there exists $k\leqslant p$ such that $|\beta(t,t-k)|=1$. For both situations, the conventions used imply $\beta(t,t-j)=0$ for $j>p$. Now let $\tilde{p}$ be the largest integer k less than p for which we have $\beta(t,t-k)\neq 0$ when t belongs to $\mathbb{Z}$. Then $X(\cdot)$ satisfies (10) with $p=\tilde{p}$ and $\varepsilon(t)=\varepsilon^{\mathrm{f}}(t;t-\tilde{p})$, $t\in\mathbb{Z}$. This shows that $p=\tilde{p}$, by definition of the model order and the existence of $t\in\mathbb{Z}$ such that $\beta(t,t-p)\neq 0$ is satisfied. Moreover, these last points establish clearly the sufficient condition of the theorem. □.

A fundamental difference with the stationary case concerns the Wold–Cramér decomposition (see [3])

$$X(t) = U(t) + V(t), \quad t \in \mathbb{Z},$$

where $U(\cdot)$ is purely nondeterministic and $V(\cdot)$ is deterministic. When $X(\cdot)$ is a stationary $AR(p)$ process, only one of these two components is possible according to σ_ε^2 is strictly positive or not. Furthermore, the roots of the polynomial equation

$$\sum_{k=0}^{p} a(k)z^k = 0$$

lie outside the unit circle if $X(\cdot) = U(\cdot)$ and the parameters $a(1), \ldots, a(p)$, together with σ_ε^2, characterize the PACF. If $X(\cdot) = V(\cdot)$, we can suppose that all the roots lie on the unit circle, but $a(1), \ldots, a(p)$ characterize only the support of the spectral measure. In the nonstationary case, the two components $U(\cdot)$ and $V(\cdot)$ can coexist. In this case $X(\cdot)$ and $U(\cdot)$ satisfy (10) with the same parameters $a_t(k)$ and the same variances $\sigma_\varepsilon^2(t)$ but with different PACF. The process $V(\cdot)$ also satisfies (10) with the same parameters $a_t(k)$ as $X(\cdot)$ or $U(\cdot)$ but with $\sigma_\varepsilon^2(t) = 0$ and a third PACF. Let us illustrate this problem by the following example. The model is

$$X(t) + aX(t-1) = \varepsilon(t), \quad t \in \mathbb{Z},$$

with $|a| < 1$ and $\sigma_\varepsilon^2(t) = 1$ for all t in $\mathbb{Z}$. Let $U(\cdot)$ be the stationary $AR(1)$ process associated with these parameters. We have, for all $t \in \mathbb{Z}$,

$$\beta_U(t,t) = \frac{1}{1-a^2}, \quad \beta_U(t,t-1) = -a, \quad \beta_U(t,t-k) = 0, \ \ k > 1.$$

Now, let $V(\cdot)$ be defined by $V(t) + aV(t-1) = 0, \ t \in \mathbb{Z}$, where $V(0)$ is a zero-mean random variable with unit variance and uncorrelated with $U(\cdot)$.

Then $\beta_V(\cdot,\cdot)$ is given by

$$\beta_V(t,t) = a^{2t}, \quad \beta_V(t,t-1) = -1, \quad \beta_V(t,t-k) = 0, \ \ k > 1, \ \ t \in \mathbb{Z}.$$

Using $R_X(\cdot,\cdot) = R_U(\cdot,\cdot) + R_V(\cdot,\cdot)$ and the GLD Algorithm, we obtain:

$$\beta_X(t,t) = \frac{1}{1-a^2} + a^{2t}, \quad \beta_X(t,t-1) = -a\frac{(1+a^{2(t-1)} - a^{2t})^{\frac{1}{2}}}{(1+a^{2t} - a^{2(t+1)})^{\frac{1}{2}}},$$
$$\beta_X(t,t-k) = 0, \ \ k > 1, \ \ t \in \mathbb{Z}.$$

The above example shows that the set of parameters $\{a_t(k), \ k = 1, \ldots, p; \ \sigma_\varepsilon^2(t)\}_{t \in \mathbb{Z}}$, does not specify the second order properties of $X(\cdot)$. This question is still open if we restrict ourself to purely nondeterministic processes. Another problem is to characterize the set of coefficients $\{a_t(k), \ k = 1, \ldots, p\}_{t \in \mathbb{Z}}$ for which a solution of (10) exists. A sufficient condition is given in [13], using the theory of linear difference equations, when $a_t(p) \neq 0$ and $\sigma_\varepsilon^2(t) = 1$ for all $t \in \mathbb{Z}$. Then this solution is purely nondeterministic and corresponds to $U(\cdot)$ in our example. Note that no problem arises when only unilateral sequences $\{X(t), \ t \geqslant 0\}$ are considered. In such a case the second order properties of $X(\cdot)$ are characterized by $\{\beta(t,s), \ s,t \geqslant 0\}$, with the

condition of Theorem 4, or equivalently by any initial condition $\{\beta(t,s),\ 0\leqslant s,\ t<p\}$ and any set of parameters $\{a_t(k),\ k=1,\ldots,p;\ \sigma_\varepsilon^2(t)\}_{t\geqslant p}$, with $a_t(p)\neq 0$ for some $t\geqslant p$. The correspondence between these two parameterizations is clearly one-to-one in the nonlocally deterministic case. Then the coefficients $\{a_t(k),\ k=1,\ldots,p\}_{t\geqslant p}$ are uniquely defined and can take any values. Otherwise the unicity is no longer true.

3.4. Estimation procedure

Notice that the PACF is well adapted for estimating the second order structure of a nonstationary observed sequence in a nonparametric way. Indeed, each function $\beta_t(k)=\beta(t,t,-k),\ k\in\mathbb{N}$, is the PACF of a stationary process. Furthermore, in the nonlocally deterministic case, these functions can be estimated separately because there exist no relationships between them. So we can apply any estimating method of the stationary case in a sliding window or using a forgetting factor. Then the estimated parameters give $\hat{\beta}_t(k)$ for various k and each fixed t. For instance, the maximum entropy method [2] should suggest to fit an evolutive autoregressive process on the nonstationary observed sequence. On each window, we can estimate the order $\hat{p}_t$ of the model using any classical method of the stationary case. Then, the coefficients $\hat{\beta}_t(k),\ k=0,\ldots,\hat{p}_t$, can be estimated by the empirical partial autocorrelation coefficients of [5]. We have been able to observe that this approach leads to an estimated evolutive instantaneous spectrum comparable with the one of [7]. Let us point out that the PACF cannot be replaced, in these approaches, by the ACF since $\hat{R}(\cdot,\cdot)$ given by $\hat{R}(t,t-k)=\hat{R}_t(k)$ is not necessary p.d., even if $\hat{R}_t(k)$ is p.d. as a function of k. For instance, if we estimate $\hat{R}_t(k),\ k=0,1,\ldots$, by the usual Yule–Walker method, we must compute the corresponding $\hat{\beta}_t(k),\ k=0,1,\ldots$, in order to obtain $\hat{\beta}(\cdot,\cdot)$ by $\hat{\beta}(t,t-k)=\hat{\beta}_t(k)$. Now $\hat{R}(\cdot,\cdot)$ associated with this $\hat{\beta}(\cdot,\cdot)$ will be p.d. but generally $\hat{R}(t,t-k)$ will be different of $\hat{R}_t(k)$.

Acknowledgments

The authors thank the referees for useful comments which improve the presentation of the paper.

References

[1] O. Barndorff-Nielsen, G. Schou, On the parametrization of autoregressive models by partial autocorrelations, J. Multivariate Anal. 3 (1973) 408–419.
[2] J.P. Burg, Maximum entropy spectral analysis, Ph.D. Dissertation, Department of Geophysics, Standford University, 1975.
[3] H. Cramér, On some classes of nonstationary stochastic processes, Proceeding of Fourth Berkeley Symposium on Mathematics, Statistic and Probability, Vol. 2, University of California Press, California, 1961, pp. 57–78.

[4] S. Dégerine, Canonical partial autocorrelation function of a multivariate time series, Ann. Statist. 18 (1990) 961–971.
[5] S. Dégerine, Sample partial autocorrelation function, IEEE Trans. Signal Process. 41 (1993) 403–407.
[6] S. Dégerine, Sample partial autocorrelation function of a multivariate time series, J. Multivariate Anal. 50 (1994) 294–313.
[7] S. Dégerine, S. Lambert-Lacroix, Evolutive instantaneous spectrum associated with partial autocorrelation function, J. Time Series Anal. 23 (2002) 377–399.
[8] P.H. Delsarte, Y. Genin, Y. Kamp, A method of matrix inverse triangular decomposition based on contiguous principal submatrices, Linear Algebra Appl. 31 (1980) 199–212.
[9] B.W. Dickinson, Autoregressive estimation using residual energies ratios, IEEE Trans. Inform. Theory IT-24 (1978) 503–506.
[10] B.W. Dickinson, Estimation of partial correlation matrices using Cholesky decomposition, IEEE Trans. Automat. Control AC-24 (1979) 302–305.
[11] F.R. Gantmacher, Matrix Theory, Vol. I, Chelsea Publishing Company, New York, N.Y. 1960.
[12] Y.L. Geronimus, Polynomials Orthogonal on a Circle and Interval, Pergamon, London, 1960.
[13] M. Hallin, G. Mélard, Indéterminabilité pure et inversibilité des processus autorégressifs-moyenne mobile à coefficients dépendant du temps, Cahier du CERO 19 (1977) 385–392.
[14] S. Lambert-Lacroix, On periodic autoregressive processes estimation, IEEE Trans. Signal Processing SP-48 (2000) 1800–1803.
[15] H. Lev-Ari, T. Kailath, Schur and Levinson algorithms for non-stationary processes, Proceedings of the 1981 International Conference on Acoustics, Speech and Signal Processing, Atantla, GA, April 1981, pp. 860–864.
[16] M.A. Morf, A. Vieira, T. Kailath, Covariance characterization by autocorrelation matrices, Ann. Statist. 6 (1978) 643–648.
[17] M.A. Morf, A. Vieira, D. Lee, T. Kailath, Recursive multichannel maximum entropy spectral estimation, IEEE Trans. Geosci. Electron. GE-16 (1978) 85–94.
[18] F.L. Ramsey, Characterization of the partial autocorrelation function, Ann. Statist. 2 (1974) 1296–1301.
[19] G.U. Yule, On the theory of correlation for any number of variables, treated by a new system of notation, Roy. Soc. Proc. 79 (1907) 85–96.