Proof of Local Convergence of a New Algorithm for Covariance Matching of Periodic ARMA Models

Bin Zhu and Giorgio Picci

***Abstract*—In a companion paper, the authors have proposed a new algorithm for the partial stochastic realization of vector discrete-time processes from finite covariance data, based on a nonlinear generalization of the classical Yule-Walker equations. In particular the algorithm provides solutions of the covariance matching problem for periodic ARMA models on a finite interval. In this letter, we provide a proof of convergence of the algorithm for scalar periodic ARMA models based on Lyapunov stability theory.**

***Index Terms*—Stochastic systems, identification, modeling.**

I. Introduction

THE MOTIVATION of this letter originates from the rational covariance extension problem also called *partial stochastic realization* [6], [7], [9], when the data are restricted to a finite interval. This problem, called the *circulant* rational covariance extension problem [10], [11] has wide applications in the fields of signal processing and system identification. In a companion paper [14] which logically precedes this one, we have considered a variant of the circulant extension problem stated in terms of ARMA models. We have found that it can be reformulated in terms of a generalization of the Yule-Walker covariance equations without involving the solution of a variational problem as in [10] and [11]. Although the resulting equations turn out to be nonlinear, a very natural iterative solution is apparent from their structure which has been implemented and shown to converge and work well in several examples. The proof of convergence of this iterative algorithm could be approached from the variational point of view by interpreting it as a quasi-Newton type iteration. This idea requires a reformulation of the variational problem in terms of spectral factors which is not completely transparent and seems to require rather sophisticated technical tools, see [4], [5]. In this letter we show local convergence using an elegant Lyapunov-type analysis of the algorithm interpreted as a nonlinear dynamical system. The analysis of convergence is carried out in the scalar case but a generalization to matrix covariance extension problem seems to be possible and is currently being investigated.

The outline of this letter is as follows: In Section II, we review the representation of finite-interval scalar processes by periodic ARMA models and formulate the covariance matching problem. In Section III we approach the problem by deriving a set of nonlinear Yule-Walker equations. An iterative algorithm to compute the solution is described. The main results are presented in Section IV, where we study in detail the local convergence of the algorithm viewed as a nonlinear dynamical system. The convergence is then proven via Lyapunov stability analysis.

Manuscript received February 27, 2017; revised May 7, 2017; accepted May 22, 2017. Date of publication June 1, 2017; date of current version June 21, 2017. This work was supported by the China Scholarship Council under Grant 201506230140. Recommended by Senior Editor G. Yin. *(Corresponding author: Giorgio Picci.)*

The authors are with the Department of Information Engineering, University of Padova, 35131 Padua, Italy (e-mail: zhubin@dei.unipd.it; picci@dei.unipd.it).

Digital Object Identifier 10.1109/LCSYS.2017.2710939

II. The ARMA Covariance Extension Problem

Since this letter can be seen as a continuation of [14], we shall refer to that paper for most of the background material including notations. Here we just recall some preliminaries for readability. Consider a discrete-time zero-mean second order stationary real process $\{y(t)\}$, defined on a finite interval $[-N+1, N]$ of the integer line $\mathbb{Z}$ and extended to all of $\mathbb{Z}$ as a periodic process with period $2N$ so that $y(t+2kN) = y(t)$ almost surely. We shall write it as a random vector

$$\mathbf{y} := \begin{bmatrix} y(t-N+1) & y(t-N+2) & \dots & y(t+N) \end{bmatrix}^\top \tag{1}$$

As shown in [1] in order for the random vector $\mathbf{y}$ to represent the restriction to $[-N+1, N]$ of a periodic process on $\mathbb{Z}$, the covariances $c_k := \mathbb{E}\, y(t+k)y(t)\,;\ k = 0, 1, \dots, N$, must form a *circulant matrix*, namely the matrix $\mathbf{\Sigma} := \mathbb{E}\,\mathbf{y}\mathbf{y}^\top$ must have the form

$$\mathbf{\Sigma} = \begin{bmatrix} c_0 & c_1 & \cdots & c_N & c_{N-1} & \cdots & c_1 \\ c_1 & c_0 & \cdots & c_{N-1} & c_N & \cdots & c_2 \\ \vdots & \vdots & \ddots & \vdots & & & \vdots \\ c_N & c_{N-1} & c_{N-2} & & & \cdots & c_{N-1} \\ \vdots & \vdots & \vdots & \ddots & & & \vdots \\ c_1 & \cdots & c_N & c_{N-1} & \cdots & c_1 & c_0 \end{bmatrix}, \tag{2}$$

$$= \mathrm{Circ}\{c_0, c_1, c_2, \dots, c_N, c_{N-1}, \dots, c_2, c_1\} \tag{3}$$

Circulant matrices [3] will play a key role in the following.

By stationarity y has a spectral representation

$$y(t) = \int_{-\pi}^{\pi} e^{it\theta} d\hat{y}(\theta), \quad \text{where} \quad \mathbb{E}\{|d\hat{y}|^2\} = dF(e^{i\theta}) \tag{4}$$

is the spectral distribution (see [13, p. 74]), so that

$$c_k = \mathbb{E}\{y(t+k)y(t)\} = \int_{-\pi}^{\pi} e^{ik\theta} dF(e^{i\theta}) \,. \tag{5}$$

Because of the periodicity condition, the support of the spectral distribution dF must be contained in the *discrete unit circle* $\mathbb{T}_{2N} \coloneqq \{\zeta_{-N+1}, \zeta_{-N+2}, \dots, \zeta_N\}$, where

$$\zeta_k = e^{ik\pi/N}. \tag{6}$$

As explained in [11], one can represent dF as $dF = \Phi \, d\nu$ where $d\nu$ is a uniform discrete measure supported on $\mathbb{T}_{2N}$, i.e.,

$$d\nu(\theta) = \sum_{j=-N+1}^{N} \delta(e^{i\theta} - \zeta_j) \frac{d\theta}{2N}, \tag{7}$$

and Φ is the discrete Fourier transform (DFT) of the sequence $(c_{-N+1}, \dots, c_N)$, called the *spectral density* of $\mathbf{y}$,

$$\Phi(\zeta) = \sum_{k=-N+1}^{N} c_k \, \zeta^{-k}, \tag{8}$$

which is in fact *the symbol* of the circulant matrix $\mathbf{\Sigma}$. This is a nonnegative function of the discrete variable $\zeta \in \mathbb{T}_{2N}$ which is strictly positive if and only if the $2N \times 2N$ covariance matrix $\mathbf{\Sigma}$ is positive definite, see [2, Proposition 2], that is to say, the process is *full rank* which we shall assume all through this letter. As described above, periodic processes can be conveniently seen as being defined on the finite group $\mathbb{Z}_{2N}$ made of the discrete interval $[-N+1, N]$ with arithmetics modulo $2N$. We are interested in periodic processes which can be represented by unilateral ARMA models of the form

$$\sum_{k=0}^{n} a_k y(t-k) = \sum_{k=0}^{n} b_k w(t-k), \qquad t \in \mathbb{Z}_{2N} \tag{9}$$

where $\{w(t)\}$ is a periodic white noise on $\mathbb{Z}_{2N}$ of unit variance and $\{a_k, b_k\}$ are real parameters. In order to fully specify the model on the finite interval $[-N+1, N]$, we need to impose periodic boundary conditions at the endpoints, i.e.,

$$y(-N) = y(N), \dots, \;\; y(-N-n+1) = y(N-n+1) \tag{10}$$

which, after introducing the vector notation

$$\mathbf{w} \coloneqq \begin{bmatrix} w(-N+1) & w(-N+2) & \cdots & w(N) \end{bmatrix}^\top$$

with $\mathbb{E}\{\mathbf{w}\mathbf{w}^\top\} = \mathbf{I}_{2N}$ (identity), leads to a compact circulant matrix representation of the model (9)

$$\mathbf{A}\mathbf{y} = \mathbf{B}\mathbf{w}, \tag{11}$$

where $\mathbf{A}$ and $\mathbf{B}$ are $2N \times 2N$ nonsingular lower-triangular circulant matrices of bandwidth n denoted

$$\begin{aligned} \mathbf{A} &= \mathrm{Circ}\{a_0, a_1, \dots, a_n, 0, \dots, 0\} \\ \mathbf{B} &= \mathrm{Circ}\{b_0, \, b_1, \dots, b_n, 0, \dots, 0\}. \end{aligned} \tag{12}$$

The symbols of $\mathbf{A}$ and $\mathbf{B}$ are the polynomials $a(\zeta), \; b(\zeta)$ in the indeterminate ζ defined in terms of the model coefficients as

$$a(\zeta) \coloneqq \sum_{k=0}^{n} a_k \, \zeta^{-k}, \quad b(\zeta) \coloneqq \sum_{k=0}^{n} b_k \, \zeta^{-k} \tag{13}$$

where the negative exponent in ζ^{-k} agrees with the interpretation of DFT as a k-steps delay operator in the frequency domain. In terms of the DFT, the model (9) can then be rewritten as

$$a(\zeta)\hat{y}(\zeta) = b(\zeta)\hat{w}(\zeta), \qquad \zeta \in \mathbb{T}_{2N} \tag{14}$$

where $\hat{y}(\zeta) = \sum_{t=-N+1}^{N} y(t)\zeta^{-t}$, $\hat{w}(\zeta) = \sum_{t=-N+1}^{N} w(t)\zeta^{-t}$ are the DFT of the random vectors $\mathbf{y}$ and $\mathbf{w}$. The solution of (14) can formally be written as

$$\hat{y}(\zeta) = \frac{b(\zeta)}{a(\zeta)} \hat{w}(\zeta). \tag{15}$$

Since the DFT $\hat{w}(\zeta_k)$ satisfies $\frac{1}{2N}\mathbb{E}[\hat{w}(\zeta_k)\overline{\hat{w}(\zeta_l)}] = \delta_{kl}$, from (15) it readily follows that the spectral density of $\{y(t)\}$ is

$$\Phi(\zeta) = \frac{1}{2N}\mathbb{E}\Big[\hat{y}(\zeta)\overline{\hat{y}(\zeta)}\Big] = \frac{b(\zeta)b(\zeta^{-1})}{a(\zeta)a(\zeta^{-1})} \coloneqq \frac{P(\zeta)}{Q(\zeta)} \tag{16}$$

which is a rational function, i.e., quotient of two symmetric positive polynomials

$$P(\zeta) \coloneqq b(\zeta)b(\zeta^{-1}) \,, \qquad Q(\zeta) \coloneqq a(\zeta)a(\zeta^{-1}) \tag{17}$$

We consider now the covariance matching problem for periodic ARMA processes.

Problem 1: Suppose that we are given the MA coefficients $\{b_k \,;\, k = 0, 1, 2, \dots, n\}$ of (9) and a partial covariance sequence $c_0, c_1, \dots, c_n$ with $n < N$, such that the Toeplitz matrix

$$\mathbf{T}_n = \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_n \\ c_1 & c_0 & c_1 & \cdots & c_{n-1} \\ c_2 & c_1 & c_0 & \cdots & c_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_n & c_{n-1} & c_{n-2} & \cdots & c_0 \end{bmatrix}, \tag{18}$$

is positive definite. We want to determine the AR coefficients $\{a_k\}$ such that the first $n+1$ covariance lags of the periodic process $\{y(t)\}$ defined by (9) match the sequence $\{c_k\}$.

This problem, once stated in terms of the polynomials (17), is essentially the same moment problem discussed in [11] and [12], where it is proven that for any fixed positive polynomial $P(\zeta)$, there is a unique solution $Q(\zeta)$. Actually the equivalence holds modulo a factorability condition which is stated in the next proposition.

Proposition 1: Assume $N > n$, then the pseudo-polynomial $P(\zeta) = \sum_{-n}^{n} p_k \, \zeta^{-k}$ admits a factorization $P(\zeta) = a(\zeta)a(\zeta^{-1})$ with $a(\zeta)$ as in (13), if and only if $P(z) \coloneqq a(z)a(z^{-1})$ for $z \in \mathbb{T}$, i.e., the usual polynomial factorization holds with the same coefficients.

Proof: Sufficiency is obvious. For the necessity, suppose $P(\zeta) = a(\zeta)a(\zeta^{-1})$ holds and define the pseudo-polynomial in z

$$\check{P}(z) = \sum_{k=-n}^{n} \check{p}_k z^{-k} := a(z^{-1})a(z)\,.$$

Then the two polynomials $P(z)$ and $\check{P}(z)$ of order $2n$ coincides at $2N$ points, i.e., $\zeta_{-N+1}, \dots, \zeta_N$. Since $N > n$, this implies that $P(z) = \check{P}(z)$ and hence admits a usual polynomial factorization. ∎

Positivity of a symmetric polynomial $P(\zeta)$ on $\mathbb{T}_{2N}$ does not necessarily imply the existence of *banded* spectral factors. For this to hold, nonnegativity of the extension $P(z)$ on the whole unit circle is necessary. Although such a requirement may seem restrictive, it is in fact satisfied if N is large enough, which we shall assume for the remaining part of this letter. This point is demonstrated as follows.

Proposition 2: Let $P(\zeta) = \sum_{-n}^{n} p_k \zeta^{-k}$ be positive on $\mathbb{T}_{2N}$. If N is large enough, the extension of $P(\zeta)$ to the unit circle $P(z)\,;\ z \in \mathbb{T}$, must be nonnegative for all $z \in \mathbb{T}$.

Proof: Suppose that for some $z_0 \in \mathbb{T}$, $P(z_0) < 0$; then there must exist an interval neighborhood $\mathscr{I}$ of z_0 in $\mathbb{T}$ having positive measure where $P(e^{i\theta}) < 0$ for any $e^{i\theta} \in \mathscr{I}$. But if N is large enough some $\zeta_k \in \mathbb{T}_{2N}$ must belong to this neighborhood and then $P(\zeta_k)$ must be negative which is impossible. ∎

III. The ARMA Covariance Matching Equation

Let $\gamma := \{\gamma_k\,;\ k = -N+1, \dots, N\}$ denote the inverse DFT of $\frac{b(\zeta)}{a(\zeta)}$. The time-domain version of (15) is a (circulant) convolution representation of $y(t)$ in terms of the input noise $w(t)$

$$y(t) = \sum_{s=-N+1}^{N} \gamma_{t-s} w(s), \quad t \in \mathbb{Z}_{2N}, \tag{19}$$

which can also be written in matrix notation as

$$\mathbf{y} = \Gamma \mathbf{w}, \tag{20}$$

where $\Gamma = \mathrm{Circ}\{\gamma_0, \gamma_1, \dots, \gamma_N, \gamma_{-N+1}, \dots, \gamma_{-1}\}$, has symbol

$$\Gamma(\zeta) := \sum_{t=-N+1}^{N} \gamma_t \zeta^{-t} = \frac{b(\zeta)}{a(\zeta)}. \tag{21}$$

The impulse response samples γ_k and the circulant $\mathbf{\Gamma}$ are denoted W_k and $\mathbf{W}$ in the companion paper [14]. Here we have changed notation to avoid confusion with the white noise process. In circulant matrix notation, from (11) we have

$$\Gamma = \mathbf{A}^{-1}\mathbf{B}. \tag{22}$$

Now, multiplying the model equation (11) on both sides from the right by the transpose of (20) and taking expectations, we obtain an equation for the circulant covariance:

$$\mathbf{A}\mathbf{\Sigma} = \mathbf{B}\Gamma^{\top}. \tag{23}$$

Introduce the vector notation

$$\mathbf{a} = \begin{bmatrix} a_0 & \dots & a_n \end{bmatrix}^{\top}, \quad \mathbf{b} = \begin{bmatrix} b_0 & \dots & b_n \end{bmatrix}^{\top}$$

and denote the upper-left $(n+1) \times (n+1)$ submatrix of $\mathbf{\Sigma}$ by $\mathbf{\Sigma}_n$. Since $\mathbf{b}$ is fixed, the covariance matrix $\mathbf{\Sigma}$ is a function of $\mathbf{a}$ so it is appropriate to denote $\mathbf{\Sigma}_n$ by $\mathbf{\Sigma}_n(\mathbf{a})$. With these notation, our covariance matching equation can be written as

$$\mathbf{T}_n = \mathbf{\Sigma}_n(\mathbf{a}). \tag{24}$$

which is the analog of equation (24) in [14]. Our algorithm is based on a consequence of (24) which is obtained by the same Yule-Walker type calculation of [14] combining the model equation (9) with the one-sided representation (19). It is a nonlinear equation for the coefficient vector $\mathbf{a}$ of the polynomial $a(\zeta)$ having the form:

$$\mathbf{T}_n\mathbf{a} = \mathbf{\Gamma}_n\mathbf{b} \tag{25}$$

where $\mathbf{T}_n$ is the data matrix (18), and

$$\mathbf{\Gamma}_n = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_n \\ \gamma_{-1} & \gamma_0 & & \vdots \\ \vdots & & \ddots & \gamma_1 \\ \gamma_{-n} & \cdots & \gamma_{-1} & \gamma_0 \end{bmatrix}.$$

is the upper-left $(n+1)\times(n+1)$ block of the circulant impulse response matrix $\mathbf{\Gamma}$. For the same reason as done for $\mathbf{\Sigma}_n(\mathbf{a})$, we shall denote $\mathbf{\Gamma}_n$ by $\mathbf{\Gamma}_n(\mathbf{a})$. With this definition of $\Gamma_n(\mathbf{a})$, we have

$$\mathbf{\Gamma}_n(\mathbf{a})\mathbf{b} = \mathbf{\Sigma}_n(\mathbf{a})\mathbf{a}, \tag{26}$$

which obviously agrees with (24) and (25).

It is evident that any $\mathbf{a}$ solving (24) will be a solution to (25). On the other hand, the nonlinear equation (25) for $\mathbf{a}$ has in general several solutions, corresponding to different spectral factors $a(\zeta)$ of $Q(\zeta)$ obtained by flipping zeros about the unit circle. Among them we shall privilege the unique polynomial whose extension obtained by substituting ζ with $z \in \mathbb{C}$ is a Schur (minimum phase) polynomial. Here we slightly modify the definition of a Schur polynomial to accommodate the convention of Fourier transform. Specifically, the set $\mathcal{S}_n$ of Schur polynomials of degree n contains those

$$p(z) = \sum_{k=0}^{n} p_k z^{-k}, \quad p_0 > 0$$

such that $z^n p(z)$ has all its roots strictly inside the unit circle.

We define also the set

$$S_n := \{\mathbf{p} = \begin{bmatrix} p_0 & \dots & p_n \end{bmatrix} \mid p(z) \in \mathcal{S}_n\}$$

to distinguish the polynomials and their coefficients. Before attempting a solution to (25) we recall from [14, eq. (28)] that for any $\mathbf{a} \in \mathscr{A} := \{\mathbf{a} \mid a(e^{i\theta}) \neq 0,\ \forall\, \theta \in [-\pi, \pi]\}$

$$\mathbf{a}^{\top}\mathbf{\Sigma}_n(\mathbf{a})\mathbf{a} = \frac{1}{2N} \sum_{k=-N+1}^{N} P(\zeta_k) := m_P\,,$$

where m_P is a constant once the numerator $P(\zeta)$ of the spectral density $\Phi_{\mathbf{a}}(\zeta)$ is fixed. Thus any solution to (25) must satisfy the constraint

$$\mathbf{a}^{\top}\mathbf{T}_n\mathbf{a} = m_P. \tag{27}$$

We call the model (9) *normalized* if the above constraint for the AR coefficients is satisfied. For any nonzero vector $\mathbf{a}$, the following map achieves the normalization

$$\mathbf{s} : \mathbf{a} \mapsto \sqrt{\frac{m_P}{\mathbf{a}^\top \mathbf{T}_n \mathbf{a}}}\,\mathbf{a}. \tag{28}$$

Now, consider the following iterative algorithm to solve numerically the nonlinear equation (25) which is just the scalar process analogue of that presented in [14].

Algorithm (Fixed Point Iteration With Renormalization):

1) Initialize $\mathbf{a}^{(0)}$, e.g., as the output of the Levinson algorithm for the ordinary covariance extension. Set a threshold τ to decide convergence
2) Iterate $\mathbf{a}^{(k+1)} = \mathbf{T}_n^{-1}\mathbf{\Gamma}_n(\mathbf{a}^{(k)})\mathbf{b}$
3) Rescale the result by setting $\mathbf{a}^{(k+1)} = \mathbf{s}(\mathbf{a}^{(k+1)})$
4) If $\|\mathbf{a}^{(k+1)} - \mathbf{a}^{(k)}\| > \tau$, go to step 2.

The algorithm above has a connection with the variational approach formulated in the next proposition, in which we shall introduce the objective function $\mathbb{J}_P$ from [11, Th. 2].

Proposition 3: Step 2 of the algorithm can be interpreted as a quasi-Netwon step for the minimization of the function

$$\mathbb{J}_P(\mathbf{a}) = \mathbf{a}^\top \mathbf{T}_n \mathbf{a} - \int_{-\pi}^{\pi} b(e^{i\theta})b(e^{-i\theta}) \log\left[a(e^{i\theta})a(e^{-i\theta})\right] d\nu. \tag{29}$$

Proof: We first compute the gradient of $\mathbb{J}_P(\mathbf{a})$:

$$2\mathbf{T}_n\mathbf{a} - \int_{-\pi}^{\pi} b(e^{i\theta})b(e^{-i\theta}) \left\{ \frac{1}{a(e^{i\theta})}\bar{\mathbf{u}}(e^{i\theta}) + \frac{1}{a(e^{-i\theta})}\mathbf{u}(e^{i\theta}) \right\} d\nu,$$

where for convenience we have introduced the column vectors

$$\begin{aligned} \mathbf{u}(z) &= \begin{bmatrix} 1 & z & \dots & z^n \end{bmatrix}^\top, \\ \bar{\mathbf{u}}(z) &= \begin{bmatrix} 1 & z^{-1} & \dots & z^{-n} \end{bmatrix}^\top. \end{aligned} \tag{30}$$

The left term of the integrand can be written as

$$\frac{b(e^{i\theta})}{a(e^{i\theta})}\bar{\mathbf{u}}(e^{i\theta})\mathbf{u}(e^{i\theta})^\top \mathbf{b},$$

so that this part of the integral becomes the sum

$$\frac{1}{2N}\sum_{j=-N+1}^{N} \frac{b(\zeta_j)}{a(\zeta_j)} \begin{bmatrix} 1 & \zeta_j & \dots & \zeta_j^n \\ \zeta_j^{-1} & 1 & & \vdots \\ \vdots & & \ddots & \zeta_j \\ \zeta_j^{-n} & \dots & \zeta_j^{-1} & 1 \end{bmatrix} \mathbf{b} = \mathbf{\Gamma}_n(\mathbf{a})\mathbf{b}$$

by the definition of γ. The computation involving the other term in the integral is similar, yielding in fact the same result, so that

$$\nabla \mathbb{J}_P(\mathbf{a}) = 2[\mathbf{T}_n\mathbf{a} - \mathbf{\Gamma}_n(\mathbf{a})\mathbf{b}]. \tag{31}$$

which, in force of (26) is equivalent to

$$\frac{1}{2}\nabla \mathbb{J}_P(\mathbf{a}) = [\mathbf{T}_n - \mathbf{\Sigma}_n(\mathbf{a})]\,\mathbf{a}. \tag{32}$$

The iteration in Step 2 of the algorithm can therefore be written as

$$\mathbf{T}_n[\mathbf{a}^{(k+1)} - \mathbf{a}^{(k)}] = \mathbf{\Gamma}_n(\mathbf{a}^{(k)})\mathbf{b} - \mathbf{T}_n\mathbf{a}^{(k)} = -\frac{1}{2}\nabla \mathbb{J}_P(\mathbf{a}^{(k)})$$

which is a quasi-Newton step

$$\mathbf{a}^{(k+1)} = \mathbf{a}^{(k)} - \frac{1}{2}\mathbf{T}_n^{-1}\nabla \mathbb{J}_P(\mathbf{a}^{(k)}). \tag{33}$$

■

Now the functional $\mathbb{J}_P$ parametrized as a function of $Q(\zeta)$ is strictly convex [11, Th. 2] and one may guess that (29) should be locally strictly convex about each normalized solution $\mathbf{a}$ of (24) since each such solution corresponds 1:1 to a spectral factor $a(\zeta)$ of $Q(\zeta)$. It can be shown (but we cannot report the full proof here) that this is actually true under a non-singularity condition of a certain matrix $\mathbf{M}(\mathbf{a})$ which will be stated in the next proposition.

Recall that a polynomial $a(z)$ is called *unmixing* if it has no reciprocal zeros. In particular, the Schur property implies unmixing.

Proposition 4: If the coefficient vector $\mathbf{a}$ that solves the nonlinear equation (25) defines an unmixing spectral factor $a(z)$ then it is a critical point of $\mathbb{J}_P(\mathbf{a})$ yielding a minimum; in fact the functional has in general a finite number of equivalent such minimizers all giving the same minimum value. In particular the solution $\hat{\mathbf{a}} \in S_n$ is one of such minimizers which defines a periodic ARMA process (9) whose covariance matrix is a circulant extension of the data $\mathbf{T}_n$.

Proof: Note that the term $\mathbf{T}_n\hat{\mathbf{a}}$ can be written as $\mathbf{M}(\hat{\mathbf{a}})\mathbf{c}$, where $\mathbf{M}(\mathbf{a}) =$

$$\begin{bmatrix} a_0 & a_1 & a_2 & \dots & a_n \\ a_1 & a_2 & \dots & a_n & 0 \\ a_2 & \dots & a_n & 0 & 0 \\ \vdots & \iddots & \iddots & & \vdots \\ a_n & 0 & \dots & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & a_0 & 0 & \dots & 0 \\ 0 & a_1 & a_0 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & a_{n-1} & \dots & a_1 & a_0 \end{bmatrix}$$

is the so-called *Jury matrix* mentioned in [8] whose determinant is

$$\prod_{j=1}^{n}\prod_{k=1}^{n}(1 - r_j r_k), \tag{34}$$

where r_j is the j-th root of the polynomial $a(z)$. Hence $\mathbf{M}(\hat{\mathbf{a}})$ is nonsingular iff $a(z)$ is unmixing; in particular if it is a Schur polynomial. Consider then the equation in the unknown $\mathbf{c}$

$$\mathbf{M}(\hat{\mathbf{a}})\mathbf{c} = \mathbf{\Gamma}_n(\hat{\mathbf{a}})\mathbf{b}. \tag{35}$$

with $\hat{\mathbf{a}}$ and the corresponding $\gamma = \gamma(\hat{\mathbf{a}})$ fixed. This is a linear equation which has a unique solution vector $\mathbf{c} = \begin{bmatrix} c_0 \dots c_n \end{bmatrix}^\top$, whose components are exactly the first $n+1$ covariance lags of the periodic ARMA process (9). ■

IV. Proof of Local Convergence

Taking into account the normalization step (28), each iteration in our algorithm can be written as a composition of two maps

$$\mathbf{a}^{(k+1)} = \mathbf{g}(\mathbf{a}^{(k)}) := \mathbf{s}(\mathbf{f}(\mathbf{a}^{(k)}))\,, \tag{36}$$

where

$$\mathbf{f}(\mathbf{a}^{(k)}) := \mathbf{T}_n^{-1}\mathbf{\Sigma}_n(\mathbf{a}^{(k)})\mathbf{a}^{(k)}. \tag{37}$$

The convergence of the iterative algorithm can be studied via the stability analysis of the system (36) around its equilibria and we shall do so through linearization. The following well-known theorem is mentioned in [16, p. 194].

Theorem 1 (First Method of Lyapunov for Discrete-Time Autonomous Systems): Let $\mathbf{x}^* = 0$ be an equilibrium of the discrete-time autonomous system

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k), \tag{38}$$

where $\mathbf{f} : \mathscr{D} \to \mathbb{R}^n$ is continuously differentiable in a neighborhood of the origin $\mathscr{D} \subset \mathbb{R}^n$, and let $J = [\partial \mathbf{f}/\partial \mathbf{x}_k]_{\mathbf{x}_k=\mathbf{0}}$ be the Jacobian of the system evaluated at the origin. If all the eigenvalues of J are strictly less than one in absolute value, then the system is asymptotically stable about its zero equilibrium.

Before stating the main theorem of this letter, we need more lemmas in order to compute the Jacobian of the maps $\mathbf{f}$ and $\mathbf{g}$ and study its eigenvalues.

Lemma 1: Let

$$\Xi(\mathbf{a}) \coloneqq \frac{d}{d\mathbf{a}}[-\mathbf{\Sigma}_n(\mathbf{a})\mathbf{a}], \tag{39}$$

then the Hessian of $\mathbb{J}_P(\mathbf{a})$ is the $(n+1) \times (n+1)$ matrix

$$\nabla^2 \mathbb{J}_P(\mathbf{a}) = 2(\mathbf{T}_n + \Xi(\mathbf{a})) \tag{40}$$

where, with $\bar{\mathbf{u}}(e^{i\theta})$ defined as in (30),

$$\Xi(\mathbf{a}) = \int_{-\pi}^{\pi} \bar{\mathbf{u}}(e^{i\theta})\bar{\mathbf{u}}(e^{i\theta})^\top \frac{|b(e^{i\theta})|^2}{a(e^{i\theta})^2} d\nu\,. \tag{41}$$

Proof: Clearly (40) follows from (32). From the definition of γ_t, it is straightforward to check that

$$\frac{\partial \gamma_t}{\partial a_j} = -\sum_{k=-N+1}^{N} \zeta_k^{t-j} \frac{b(\zeta_k)}{a(\zeta_k)^2} \frac{1}{2N}.$$

Then consider again the j-th entry of $\mathbf{\Gamma}_n(\mathbf{a})\mathbf{b}$, by interchanging the order of summation,

$$\frac{\partial}{\partial a_k}\left(\sum_{k=0}^{n} b_k \gamma_{-j+k}\right) = -\sum_{\ell=-N+1}^{N} \zeta_\ell^{-j-k} \frac{b(\zeta_\ell^{-1})b(\zeta_\ell)}{a(\zeta_\ell)^2} \frac{1}{2N},$$

which is the (j,k) element of the matrix

$$\frac{d}{d\mathbf{a}}[\mathbf{\Gamma}_n(\mathbf{a})\mathbf{b}].$$

With the help of (26), it is also straightforward to verify the matrix form (41). ∎

Lemma 2: Consider a function defined by the ratio of two polynomials

$$f(z) = \frac{\sum_{k=0}^{n} d_k z^k}{\sum_{k=0}^{n} a_k z^k},$$

where the denominator polynomial has all its zeros strictly outside the unit circle. Then if f takes real values on the unit circle, we must have $\mathbf{d} = \kappa \mathbf{a}$ for some $\kappa \in \mathbb{R}$, which in turn gives $f(z) = \kappa$ for any $z \in \mathbb{C}$.

Proof: Under the condition of the lemma, $f(z)$ must be holomorphic in some open region of the complex plane containing the closed unit disk $\overline{\mathbb{D}}$. For $z = x + iy$, we can write $f(z) = u(x,y) + iv(x,y)$. Then it is a fact that both u and v are *harmonic functions*, which follows directly from the Cauchy-Riemann equations. By the maximum/minimum principle for harmonic functions [15, Ch. 4], the function v can only achieve its maximum and minimum over $\overline{\mathbb{D}}$ at the boundary, i.e., the unit circle, where it is constantly zero. Hence, we have

$$v(z) = 0, \ \ \forall z \in \overline{\mathbb{D}}.$$

Furthermore, by the Cauchy-Riemann equations, the claim of the lemma follows. ∎

Lemma 3: For any $\mathbf{a} \in S_n$, the matrix $\Xi(\mathbf{a})$ satisfies the inequality

$$|\mathbf{d}^\top \Xi(\mathbf{a})\mathbf{d}| \le \mathbf{d}^\top \mathbf{\Sigma}_n(\mathbf{a})\mathbf{d} \tag{42}$$

which holds in a strict sense for all $\mathbf{d} = [d_0, \dots, d_n]^\top \in \mathbb{R}^{n+1}$ except for the one dimensional subspace of vectors which are proportional to $\mathbf{a}$.

Proof: By (41) and the triangle inequality,

$$\begin{aligned} |\mathbf{d}^\top \Xi(\mathbf{a})\mathbf{d}| &= \left| \int_{-\pi}^{\pi} \mathbf{d}^\top \bar{\mathbf{u}}(e^{i\theta}) \frac{P(e^{i\theta})}{a(e^{i\theta})^2} \bar{\mathbf{u}}(e^{i\theta})^\top \mathbf{d}\, d\nu \right| \\ &= \left| \int_{-\pi}^{\pi} \frac{P(e^{i\theta})}{a(e^{i\theta})^2}\, d(e^{i\theta})^2\, d\nu \right| \\ &\le \int_{-\pi}^{\pi} \frac{P(e^{i\theta})}{|a(e^{i\theta})|^2} |d(e^{i\theta})|^2\, d\nu \ = \mathbf{d}^\top \mathbf{\Sigma}_n(\mathbf{a})\mathbf{d}, \end{aligned}$$

where $d(z) \coloneqq \sum_{k=0}^{n} d_k z^{-k}$. This proves the inequality. To prove the other statement, we first note that for any $\mathbf{a} \in \mathscr{A}$ it holds that

$$\Xi(\mathbf{a})\mathbf{a} = \mathbf{\Sigma}_n(\mathbf{a})\mathbf{a} \tag{43}$$

which readily follows from the representation (41) of $\Xi(\mathbf{a})$ since $\bar{\mathbf{u}}(e^{i\theta})^\top \mathbf{a} = a(e^{i\theta})$. To show that (42) can hold equal only for vectors $\mathbf{d} = \kappa \mathbf{a}$, $\kappa \in \mathbb{R}$, we argue as follows. For n complex numbers $z_1, \dots, z_n$, the condition for the equality

$$|z_1 + \cdots + z_n| = |z_1| + \cdots + |z_n|$$

to hold is that for any j, $z_j = r_j z_0$ with real $r_j \ge 0$ (or all $r_i \le 0$) and some common $z_0 \in \mathbb{C}$. Applying this to our case, i.e.,

$$z_j = \frac{P(\zeta_j)}{a(\zeta_j)^2} d(\zeta_j)^2, \quad j = -N+1, \dots, N,$$

given N large enough, it amounts to requiring that the ratio of two polynomials $d(e^{i\theta})/a(e^{i\theta})$ takes real values for any $\theta \in [-\pi, \pi]$ since $d(\zeta_0)/a(\zeta_0)$ is real. By Lemma 2, this cannot happen in general unless $\mathbf{d}$ is proportional to $\mathbf{a}$. ∎

Theorem 2: The algorithm in Section III converges locally to a vector of AR coefficients $\hat{\mathbf{a}} \in S_n$ that is a solution to (25).

Proof: Take $\hat{\mathbf{a}}$ as the coefficient vector of the unique Schur polynomial that solves Problem 1. It is easy to check that $\hat{\mathbf{a}}$ is a fixed point of the function $\mathbf{g}(\cdot) = \mathbf{s}(\mathbf{f}(\cdot))$ since

$$\mathbf{s}(\hat{\mathbf{a}}) = \hat{\mathbf{a}}, \quad \mathbf{f}(\hat{\mathbf{a}}) = \hat{\mathbf{a}}.$$

Next use the representation $\mathbf{f}(\mathbf{a}) = \mathbf{a} - \frac{1}{2}\mathbf{T}_n^{-1}\nabla\mathbb{J}_P(\mathbf{a})$ from Proposition 3 to compute the Jacobian matrix by the chain rule and evaluate it at $\hat{\mathbf{a}}$

$$J := \left.\frac{d\mathbf{g}}{d\mathbf{a}^{(k)}}\right|_{\mathbf{a}^{(k)}=\hat{\mathbf{a}}} = \frac{1}{m_P}\hat{\mathbf{a}}\hat{\mathbf{a}}^\top\Xi(\hat{\mathbf{a}}) - \mathbf{T}_n^{-1}\Xi(\hat{\mathbf{a}}) = \frac{1}{m_P}\hat{\mathbf{a}}\hat{\mathbf{a}}^\top\mathbf{T}_n - \mathbf{T}_n^{-1}\Xi(\hat{\mathbf{a}}) \tag{44}$$

where we have also used Lemma 1 to get

$$\nabla\mathbf{f}(\mathbf{a}) = -\mathbf{T}_n^{-1}\Xi(\mathbf{a}).$$

The last equality in (44) comes from (43).

In order to apply Theorem 1 to assert stability of the equilibrium $\hat{\mathbf{a}}$, we proceed to show that all eigenvalues of (44) have absolute value smaller than 1. To this end, there is a result in linear algebra stating that the spectral radius of a complex matrix A is less than 1 if and only if $A^k \to 0$ as $k \to \infty$, whose proof uses Jordan normal form [18, p. 180]. One can then show by induction that for $k = 1, 2, \ldots$

$$J^{2k} = -\frac{1}{m_P}\hat{\mathbf{a}}\hat{\mathbf{a}}^\top\mathbf{T}_n + \left[\mathbf{T}_n^{-1}\Xi(\hat{\mathbf{a}})\right]^{2k},$$
$$J^{2k+1} = \frac{1}{m_P}\hat{\mathbf{a}}\hat{\mathbf{a}}^\top\mathbf{T}_n - \left[\mathbf{T}_n^{-1}\Xi(\hat{\mathbf{a}})\right]^{2k+1}. \tag{45}$$

Therefore, it is equivalent to show that

$$\lim_{k\to\infty}\left[\mathbf{T}_n^{-1}\Xi(\hat{\mathbf{a}})\right]^k = \frac{1}{m_P}\hat{\mathbf{a}}\hat{\mathbf{a}}^\top\mathbf{T}_n. \tag{46}$$

Naturally, we consider the eigenvalue problem

$$\mathbf{T}_n^{-1}\Xi(\hat{\mathbf{a}})\mathbf{v} = \lambda\mathbf{v}, \quad \mathbf{v} \neq 0.$$

It is equivalent to the generalized eigenvalue problem of the ordered pair $(\Xi(\hat{\mathbf{a}}), \mathbf{T}_n)$

$$\Xi(\hat{\mathbf{a}})\mathbf{v} = \lambda\mathbf{T}_n\mathbf{v}.$$

From [17, Th. 15.3.3, p. 345], in the present case where $\Xi(\hat{\mathbf{a}})$ and $\mathbf{T}_n$ are symmetric with $\mathbf{T}_n$ positive definite, all the eigenvalues λ are real and it is guaranteed that there exists a basis of generalized eigenvectors. Moreover, eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ with distinct eigenvalues are $\mathbf{T}_n$-orthogonal ($\mathbf{v}_1^\top\mathbf{T}_n\mathbf{v}_2 = 0$). The eigenvalues can be expressed in terms of the eigenvectors as Rayleigh quotients

$$\lambda = \frac{\mathbf{v}^\top\Xi(\hat{\mathbf{a}})\mathbf{v}}{\mathbf{v}^\top\mathbf{T}_n\mathbf{v}}. \tag{47}$$

By Lemma 3, we must have $|\lambda| \leq 1$. Furthermore, there is exactly one (generalized) eigenvalue equal to 1 since we have taken the corresponding $\hat{a}(z) \in \mathcal{S}_n$.

We are now ready to show (46). First notice that

$$\Xi(\hat{\mathbf{a}})\mathbf{P} = \mathbf{T}_n\mathbf{P}\mathbf{D}, \tag{48}$$

where the columns of $\mathbf{P}$ are the eigenvectors and $\mathbf{D}$ is the diagonal matrix of eigenvalues. The $\mathbf{T}_n$-orthogonal relation can be written as

$$\mathbf{P}^\top\mathbf{T}_n\mathbf{P} = \mathbf{I}, \tag{49}$$

where the eigenvectors are normalized with the $\mathbf{T}_n$-norm. Specifically, the eigenvector corresponding to the eigenvalue 1 is $\frac{1}{\sqrt{m_P}}\hat{\mathbf{a}}$. Hence,

$$\lim_{k\to\infty}\left[\mathbf{T}_n^{-1}\Xi(\hat{\mathbf{a}})\right]^k = \lim_{k\to\infty}\mathbf{P}\mathbf{D}^k\mathbf{P}^\top\mathbf{T}_n = \frac{1}{m_P}\hat{\mathbf{a}}\hat{\mathbf{a}}^\top\mathbf{T}_n, \tag{50}$$

which concludes the proof of local convergence. ■

V. Conclusion

We have proved local convergence of a new iterative algorithm to solve the covariance matching problem for periodic ARMA models. The algorithm can be seen as a nonlinear dynamical system that is asymptotically stable about its equilibrium. The proof is presented for scalar processes but there are good reasons to believe that it can be extended to vector-valued processes. From our simulations *global* convergence to a solution of (24) appears to be true also for vector processes but a proof is left to future work.

References

[1] F. P. Carli, A. Ferrante, M. Pavon, and G. Picci, "A Maximum entropy solution of the covariance extension problem for reciprocal processes," *IEEE Trans. Autom. Control*, vol. AC-56, no. 9, pp. 1999–2012, Sep. 2012.

[2] F. P. Carli and T. T. Georgiou, "On the covariance completion problem under a circulant structure," *IEEE Trans. Autom. Control*, vol. 56, no. 4, pp. 918–922, Apr. 2011.

[3] P. Davis, *Circulant Matrices*. New York, NY, USA: Wiley, 1979.

[4] P. Enqvist, "A homotopy approach to rational covariance extension with degree constraint," *Int. J. Appl. Math. Comput. Sci.*, vol. 11, no. 5, pp. 1173–1201, 2001.

[5] P. Enqvist, "A convex optimization approach to ARMA(n,m) model design from covariance and cepstral data," *SIAM J. Control Optim.*, vol. 43, no. 3, pp. 1011–1036, 2004.

[6] T. T. Georgiou, "Partial realization of covariance sequences," Ph.D. dissertation, CMST, Univ. Florida, Gainesville, FL, USA, 1983.

[7] T. T. Georgiou, "Realization of power spectra from partial covariance sequences," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 4, pp. 438–449, Apr. 1987.

[8] C. J. Demeure and C. T. Mullis, "The Euclid algorithm and the fast computation of cross-covariance and autocovariance sequences," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 4, pp. 545–552, Apr. 1989.

[9] R. E. Kalman, "Realization of covariance sequences," in *Proc. Toeplitz Memorial Conf.*, Tel Aviv, Israel, 1981, pp. 331–342.

[10] A. Lindquist, C. Masiero, and G. Picci, "On the multivariate circulant rational covariance extension problem," in *Proc. IEEE Conf. Decis. Control*, Florence, Italy, 2013, pp. 7155–7161.

[11] A. G. Lindquist and G. Picci, "The circulant rational covariance extension problem: The complete solution," *IEEE Trans. Autom. Control*, vol. 58, no. 11, pp. 2848–2861, Nov. 2013.

[12] A. Lindquist and G. Picci, "Modeling of stationary periodic time series by ARMA representations," in *Optimization and Its Applications in Control and Data Sciences* (Springer Optimization and Its Applications), vol. 115. Cham, Switzerland: Springer, 2016, pp. 281–314.

[13] A. Lindquist and G. Picci, *Linear Stochastic Systems: A Geometric Approach to Modeling, Estimation and Identification* (Springer series in Contemporary Mathematics). Heidelberg, Germany: Springer Verlag, 2015.

[14] G. Picci and B. Zhu, "Approximation of vector processes by covariance matching with applications to smoothing," *IEEE Control Syst. Lett.*, vol. 1, no. 1, pp. 200–205, Jul. 2017.

[15] C. A. Berenstein and R. Gay, *Complex Variables: An Introduction* (Graduate Texts in Mathematics). New York, NY, USA: Springer, 1997.

[16] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.

[17] B. N. Parlett, *The Symmetric Eigenvalue Problem*. Philadelphia, PA, USA: Soc. Ind. Appl. Math., 1998.

[18] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 2013.