Journal of Multivariate Analysis 176 (2020) 104578

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Closed-form expression for finite predictor coefficients of multivariate ARMA processes

Akihiko Inoue

Department of Mathematics, Hiroshima University, Higashi-Hiroshima 739-8526, Japan

ARTICLE INFO

Article history:
Received 24 March 2019
Received in revised form 27 November 2019
Accepted 29 November 2019
Available online 2 December 2019

AMS 2010 subject classifications:
primary 60G25
secondary 60G10
65F05

Keywords:
Finite predictor coefficients
Multivariate ARMA processes
Closed-form expression
Linear-time algorithm

ABSTRACT

We derive a closed-form expression for the finite predictor coefficients of multivariate ARMA (autoregressive moving-average) processes. The expression is given in terms of several explicit matrices that are of fixed sizes independent of the number of observations. The significance of the expression is that it provides us with a linear-time algorithm to compute the finite predictor coefficients. In the proof of the expression, a correspondence result between two relevant matrix-valued outer functions plays a key role. We apply the expression to determine the asymptotic behavior of a sum that appears in the autoregressive model fitting and the autoregressive sieve bootstrap. The results are new even for univariate ARMA processes.

1. Introduction

Let $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$ and $\overline{\mathbb{D}} := \{z \in \mathbb{C} : |z| \le 1\}$ be the unit circle and the closed unit disk, in $\mathbb{C}$, respectively. For $d \in \mathbb{N}$, a d-variate ARMA (autoregressive moving-average) process $\{X_k : k \in \mathbb{Z}\}$ is a $\mathbb{C}^d$-valued, centered, weakly stationary process with spectral density w of the form

$$w(e^{i\theta}) = h(e^{i\theta})h(e^{i\theta})^*, \qquad \theta \in [-\pi, \pi), \tag{1}$$

where $h : \mathbb{T} \to \mathbb{C}^{d\times d}$ satisfies the following condition:

$$\text{the entries of } h(z) \text{ are rational functions in } z \text{ that have no poles in } \overline{\mathbb{D}}, \text{ and } \det h(z) \text{ has no zeros in } \overline{\mathbb{D}}. \tag{2}$$

The finite predictor coefficients $\phi_{n,j} \in \mathbb{C}^{d\times d}$, $j \in \{1, \dots, n\}$, of $\{X_k\}$ are defined by

$$P_{[-n,-1]}X_0 = \phi_{n,1}X_{-1} + \cdots + \phi_{n,n}X_{-n}, \tag{3}$$

where, for $n \in \mathbb{N}$, $P_{[-n,-1]}X_0$ stands for the best linear predictor of the future value X_0 based on the finite past $\{X_{-n}, \dots, X_{-1}\}$ (see Section 2 for the precise definition). The finite predictor coefficients $\phi_{n,j}$ are among the most basic quantities in the prediction theory for $\{X_k\}$.

The main aim of this paper is to derive a closed-form expression for the finite predictor coefficients $\phi_{n,j}$ of a multivariate ARMA process. More precisely, in the main result of this paper, i.e., Theorem 6, we show that the finite predictor coefficients $\phi_{n,j}$ can be expressed in terms of several explicit matrices to be introduced in Section 4, which are of fixed sizes

E-mail address: inoue100@hiroshima-u.ac.jp.

https://doi.org/10.1016/j.jmva.2019.104578

independent of n, unlike, e.g., the matrices that appear in the Yule–Walker equations for $\phi_{n,j}$. See Example 5 that illustrates this point. The significance of the closed-form expression for $\phi_{n,j}$ is that it provides us with a linear-time algorithm to compute $\phi_{n,1},\ldots,\phi_{n,n}$ (see Remark 6).

The closed-form expression for $\phi_{n,j}$ also provides us with a powerful tool to study problems concerning the asymptotic behavior of $\phi_{n,j}$. Among such problems, we show a result on the asymptotic behavior of the sum $\sum_{j=1}^{n}\|\phi_{n,j}-\phi_j\|$ as $n\to\infty$, where ϕ_j are the infinite predictor coefficients; see (18). This sum appears, for example, in proving the consistency of the autoregressive model fitting process and the corresponding autoregressive spectral density estimator (see Berk [3]), and in proving the validity of autoregressive sieve bootstrap (see, e.g., Bühlmann [6] and Kreiss et al. [13]). Because of difficulties in finding the asymptotic behavior of $\sum_{j=1}^{n}\|\phi_{n,j}-\phi_j\|$ itself, Baxter's inequality

$$\sum_{j=1}^{n}\|\phi_{n,j}-\phi_j\|\le K\sum_{j=n+1}^{\infty}\|\phi_j\|,\qquad K\in(0,\infty),$$

in [2] has been used instead. Under a mild condition on the multivariate ARMA process, the closed-form expression for $\phi_{n,j}$ now enables us to determine the precise asymptotic behavior of $\sum_{j=1}^{n}\|\phi_{n,j}-\phi_j\|$ as $n\to\infty$ (see Theorem 8). It turns out that Baxter's inequality gives an asymptotically optimal bound of $\sum_{j=1}^{n}\|\phi_{n,j}-\phi_j\|$ in the sense that

$$\lim_{n\to\infty}\frac{\sum_{j=1}^{n}\|\phi_{n,j}-\phi_j\|}{\sum_{j=n+1}^{\infty}\|\phi_j\|}\in(0,\infty)$$

holds (see Corollary 9).

The proof of the closed-form expression for $\phi_{n,j}$ is long. One important ingredient of the proof is the explicit representation of $\phi_{n,j}$ (see the proof of Theorem 6 in Appendix D), which was obtained recently in Inoue et al. [11], extending the earlier univariate result in Inoue and Kasahara [8]; see also Inoue et al. [10] and Inoue and Kasahara [9] for related work. To explain another important ingredient of the proof of the closed-form expression for $\phi_{n,j}$, we recall that, for $h:\mathbb{T}\to\mathbb{C}^{d\times d}$ satisfying (1) and (2), there exists $h_\sharp:\mathbb{T}\to\mathbb{C}^{d\times d}$ that satisfies (2) and

$$w(e^{i\theta})=h(e^{i\theta})h(e^{i\theta})^*=h_\sharp(e^{i\theta})^*h_\sharp(e^{i\theta}),\qquad\theta\in[-\pi,\pi),\tag{4}$$

and that $h_\sharp$ is unique up to a constant unitary factor (see, e.g., [11]). We may take $h_\sharp=h$ for the univariate case $d=1$ but not so for $d\ge 2$. We show, in Theorem 2, that $h_\sharp^{-1}$ has the same poles with the same multiplicities as h^{-1}. This is a key finding in deriving the closed-form expression for $\phi_{n,j}$ when $d\ge 2$. We remark, however, that the closed-form expression for $\phi_{n,j}$ itself, i.e., Theorem 6, is new even for univariate ($d=1$) ARMA processes.

We explain the difference between the explicit representation of $\phi_{n,j}$ in [11], i.e., Theorem 5.4 in [11], and the closed-form expression of $\phi_{n,j}$ in this paper. The representation in [11] holds both for long and short memory processes, and has several applications such as the proof of Baxter's inequality for multivariate long-memory processes in [11]. The representation of $\phi_{n,j}$ in [11] is, however, not a closed-form expression since it involves infinite series. In this paper, for multivariate ARMA processes, we transform the representation in [11] to a closed-form expression for $\phi_{n,j}$. The advantage of the latter is clear from the fact that it can be viewed as a linear-time algorithm to compute $\phi_{n,1},\ldots,\phi_{n,n}$, as stated above.

This paper is organized as follows. In Section 2, we give preliminary definitions and basic facts. In Section 3, we prove the correspondence between the poles of h^{-1} and $h_\sharp^{-1}$. In Section 4, we introduce several matrices which are to become building blocks for the closed-form expression of $\phi_{n,j}$. In Section 5, we present the main result, i.e., the closed-form expression for $\phi_{n,j}$. In Section 6, we apply the closed-form expression for $\phi_{n,j}$ to derive the asymptotic behavior of $\sum_{j=1}^{n}\|\phi_{n,j}-\phi_j\|$ as $n\to\infty$. Finally, the Appendix contains the omitted proofs.

2. Preliminaries

Let $\mathbb{D}:=\{z\in\mathbb{C}:|z|<1\}$ denote the open unit disk in $\mathbb{C}$. Let $\mathbb{C}^{m\times n}$ be the set of all complex $m\times n$ matrices; we write $\mathbb{C}^d$ for $\mathbb{C}^{d\times 1}$. We write I_n for the $n\times n$ unit matrix. For $a\in\mathbb{C}^{m\times n}$, $a^\top$ denotes the transpose of a, and $\bar{a}$ and a^* the complex and Hermitian conjugates of a, respectively; thus, in particular, $a^*:=\bar{a}^\top$. For $a\in\mathbb{C}^{d\times d}$, we write $\|a\|$ for the norm $\|a\|:=\sup_{u\in\mathbb{C}^d,|u|\le 1}|au|$, where $|u|:=(\sum_{i=1}^{d}|u^i|^2)^{1/2}$ denotes the Euclidean norm of $u=(u^1,\ldots,u^d)^\top\in\mathbb{C}^d$. We denote by $\ell_{2+}^{d\times d}$ the space of $\mathbb{C}^{d\times d}$-valued sequences $\{a_k\}_{k=0}^{\infty}$ such that $\sum_{k=0}^{\infty}\|a_k\|^2<\infty$. For $r\in[1,\infty)$, we write $L_r(\mathbb{T})$ for the Lebesgue space of measurable functions $f:\mathbb{T}\to\mathbb{C}$ such that $\|f\|_r<\infty$, where $\|f\|_r:=\{\int_{-\pi}^{\pi}|f(e^{i\theta})|^r d\theta/(2\pi)\}^{1/r}$. Let $L_r^{m\times n}(\mathbb{T})$ be the space of $\mathbb{C}^{m\times n}$-valued functions on $\mathbb{T}$ whose entries belong to $L_r(\mathbb{T})$.

For $d\in\mathbb{N}$, let $\{X_k\}=\{X_k:k\in\mathbb{Z}\}$ be a $\mathbb{C}^d$-valued, centered, weakly stationary process, defined on a probability space $(\Omega,\mathcal{F},P)$, which we shall simply call a d-variate stationary process. If there exists a positive $d\times d$ Hermitian matrix-valued function w on $\mathbb{T}$, satisfying $w\in L_1^{d\times d}(\mathbb{T})$ and $E[X_mX_n^*]=\int_{-\pi}^{\pi}e^{-i(m-n)\theta}w(e^{i\theta})d\theta/(2\pi)$, $n,m\in\mathbb{Z}$, then we call w the spectral density of $\{X_k\}$. Here and throughout this paper, we assume that $\{X_k\}$ is a d-variate ARMA process in the sense that $\{X_k\}$ satisfies the following condition:

$$\{X_k\}\text{ is a }d\text{-variate stationary process that has spectral density }w\text{ satisfying (1) with (2).}\tag{5}$$

Remark 1. Suppose that $\{X_k\}$ is a d-variate, causal and invertible ARMA process in the sense of [5], that is, a $\mathbb{C}^d$-valued, centered, weakly stationary process described by the ARMA equation

$$\Phi(B)X_n = \Psi(B)Z_n, \qquad n \in \mathbb{Z},$$

where, for $r, s \in \mathbb{N} \cup \{0\}$ and $\Phi_i, \Psi_j \in \mathbb{C}^{d\times d}$, $i \in \{1,\ldots,r\}$, $j \in \{1,\ldots,s\}$,

$$\Phi(z) = I_d - z\Phi_1 - \cdots - z^r\Phi_r, \qquad \Psi(z) = I_d - z\Psi_1 - \cdots - z^s\Psi_s$$

are $\mathbb{C}^{d\times d}$-valued polynomials satisfying $\det\Phi(z) \neq 0$ and $\det\Psi(z) \neq 0$ on $\overline{\mathbb{D}}$, B is the backward shift operator defined by $BX_m = X_{m-1}$, and $\{Z_k : k \in \mathbb{Z}\}$ is a d-variate white noise, that is, a d-variate, centered process such that $E[Z_n Z_m^*] = \delta_{nm}\Sigma$ for some positive-definite $\Sigma \in \mathbb{C}^{d\times d}$. Then, $\{X_k\}$ is a d-variate ARMA process satisfying (1) with (2) for $h(z) = \Phi(z)^{-1}\Psi(z)\Sigma^{1/2}$. Conversely, we can show that any d-variate ARMA process $\{X_k\}$ satisfying (1) with (2) is described by the above type of ARMA equation.

Write $X_k = (X_k^1,\ldots,X_k^d)^\top$, and let V be the complex Hilbert space spanned by all the entries $\{X_k^j : k \in \mathbb{Z},\ j \in \{1,\ldots,d\}\}$ in $L^2(\Omega,\mathcal{F},P)$, which has inner product $(x,y)_V := E[x\overline{y}]$ and norm $\|x\|_V := (x,x)_V^{1/2}$. For $J \subset \mathbb{Z}$ such as $\{n\}$, $(-\infty,n] := \{n,n-1,\ldots\}$, $[n,\infty) := \{n,n+1,\ldots\}$, and $[m,n] := \{m,\ldots,n\}$ with $m \le n$, we write V_J^X for the closed linear span of $\{X_k^j : j \in \{1,\ldots,d\},\ k \in J\}$ in V. Let $(V_J^X)^\perp$ be the orthogonal complement of V_J^X in V, and let P_J and $P_J^\perp$ be the orthogonal projection operators of V onto V_J^X and $(V_J^X)^\perp$, respectively.

Let V^d be the space of $\mathbb{C}^d$-valued random variables on $(\Omega,\mathcal{F},P)$ whose entries belong to V. The norm $\|x\|_{V^d}$ of $x = (x^1,\ldots,x^d)^\top \in V^d$ is given by $\|x\|_{V^d} := (\sum_{i=1}^d \|x^i\|_V^2)^{1/2}$. For $J \subset \mathbb{Z}$ and $x = (x^1,\ldots,x^d)^\top \in V^d$, we write $P_J x$ for $(P_J x^1,\ldots,P_J x^d)^\top$. We define $P_J^\perp x$ in a similar way. For $n \in \mathbb{N}$ and $j \in \{1,\ldots,n\}$, the finite predictor coefficients $\phi_{n,j} \in \mathbb{C}^{d\times d}$ of $\{X_k\}$ are defined by (3). For $x = (x^1,\ldots,x^d)^\top$ and $y = (y^1,\ldots,y^d)^\top$ in V^d, $\langle x,y\rangle := E[xy^*] = ((x^i,y^j)_V)_{1\le i,j\le d} \in \mathbb{C}^{d\times d}$ stands for the Gram matrix of x and y.

For $K \in \mathbb{N}$, let $p_1,\ldots,p_K$ be distinct points in $\mathbb{D}\setminus\{0\}$. For $\mu \in \{1,\ldots,K\}$ and $i \in \mathbb{N}$, we define $p_{\mu,i} : \mathbb{N}\cup\{0\} \to \mathbb{C}$ by

$$p_{\mu,i}(k) := \binom{k}{i-1} p_\mu^{k-i+1}, \qquad k \in \mathbb{N}\cup\{0\}. \tag{6}$$

Notice that $p_{\mu,i}(0) = \binom{0}{i-1}p_\mu^{-i+1} = \delta_{i,1}$. Take $m_\mu \in \mathbb{N}$ for $\mu \in \{1,\ldots,K\}$ and let

$$M := \sum_{\mu=1}^K m_\mu. \tag{7}$$

The next proposition will be used in Section 3 and Appendix B.

Proposition 1. *For $N \in \mathbb{N}\cup\{0\}$, the M vectors $\mathbf{p}_{\mu,i} \in \mathbb{C}^{1\times M}$, $\mu \in \{1,\ldots,K\}$, $i \in \{1,\ldots,m_\mu\}$, defined by*

$$\mathbf{p}_{\mu,i} = (p_{\mu,i}(N), p_{\mu,i}(N+1), \ldots, p_{\mu,i}(N+M-1))$$

are linearly independent.

3. Correspondence between the poles of h^{-1} and $h_\sharp^{-1}$

In this section, we assume that $\{X_k\}$ satisfies (5). Let h and $h_\sharp$ be as in (1) and (4), respectively, both satisfying (2). Since h^{-1} also satisfies (2), we can write $h^{-1}(z)$ in the form

$$h(z)^{-1} = -\rho_0 - \sum_{\mu=1}^K \sum_{j=1}^{m_\mu} \frac{1}{(1-\overline{p}_\mu z)^j}\rho_{\mu,j} - \sum_{j=1}^{m_0} z^j \rho_{0,j}, \tag{8}$$

where

$$\begin{cases} K \in \mathbb{N}\cup\{0\}, \\ p_\mu \in \mathbb{D}\setminus\{0\}, \quad \mu \in \{1,\ldots,K\}, \qquad p_\mu \neq p_\nu, \quad \mu \neq \nu, \\ m_\mu \in \mathbb{N}, \quad \mu \in \{1,\ldots,K\}, \qquad m_0 \in \mathbb{N}\cup\{0\}, \\ \rho_{\mu,j} \in \mathbb{C}^{d\times d}, \quad \mu \in \{0,\ldots,K\},\ j \in \{1,\ldots,m_\mu\}, \qquad \rho_0 \in \mathbb{C}^{d\times d}, \\ \rho_{\mu,m_\mu} \neq 0, \quad \mu \in \{0,\ldots,K\}. \end{cases} \tag{9}$$

In fact, we can obtain the expression (8) from the partial fraction decompositions of the entries of $h(z)^{-1}$; see Example 2. We remark that the convention $\sum_{k=1}^0 = 0$ is adopted in the sums on the right-hand side of (8).

The next theorem shows that $h_\sharp^{-1}$ of a multivariate ARMA process has the same m_0 and the same poles with the same multiplicities as h^{-1}.

Theorem 2. *For m_0, K and $(p_1, m_1), \dots, (p_K, m_K)$ in (8) with (9), $h_\sharp^{-1}$ has the form*

$$h_\sharp(z)^{-1} = -\rho_0^\sharp - \sum_{\mu=1}^{K}\sum_{j=1}^{m_\mu} \frac{1}{(1-\bar{p}_\mu z)^j}\rho_{\mu,j}^\sharp - \sum_{j=1}^{m_0} z^j \rho_{0,j}^\sharp, \tag{10}$$

where

$$\begin{cases} \rho_{\mu,j}^\sharp \in \mathbb{C}^{d\times d}, \quad \mu \in \{0,\dots,K\},\ j \in \{1,\dots,m_\mu\}, \qquad \rho_0^\sharp \in \mathbb{C}^{d\times d}, \\ \rho_{\mu,m_\mu}^\sharp \neq 0, \quad \mu \in \{0,\dots,K\}. \end{cases} \tag{11}$$

Moreover, we have

$$\rho_{\mu,m_\mu} h_\sharp(p_\mu)^* = h(p_\mu)^* \rho_{\mu,m_\mu}^\sharp, \qquad \mu \in \{0,\dots,K\}. \tag{12}$$

The first half of Theorem 2 is a key ingredient of the proof of Theorem 6, while the relations (12) play an important role in the proof of Theorem 8.

Example 2. For $p \in \mathbb{D}$, let

$$h(z) = \begin{pmatrix} 1 & 0 \\ 1/(1-\bar{p}z) & 1 \end{pmatrix}.$$

Then h satisfies (2). For this h, we can take

$$h_\sharp(z) = r\begin{pmatrix} 1-|p|^2 & 1 \\ -1+\dfrac{1-|p|^2}{1-\bar{p}z} & -|p|^2+\dfrac{1}{1-\bar{p}z} \end{pmatrix},$$

where $r := 1/\sqrt{1-|p|^2+|p|^4}$ (see Example 3 in [11]). We have

$$h(z)^{-1} = \begin{pmatrix} 1 & 0 \\ -1/(1-\bar{p}z) & 1 \end{pmatrix}, \quad h_\sharp(z)^{-1} = r\begin{pmatrix} -|p|^2+\dfrac{1}{1-\bar{p}z} & -1 \\ 1-\dfrac{1-|p|^2}{1-\bar{p}z} & 1-|p|^2 \end{pmatrix},$$

so that $K=1$, $m_0=0$, $m_1=1$, $p_1=p$, and

$$\rho_0 = -\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \rho_{1,1} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \rho_0^\sharp = r\begin{pmatrix} |p|^2 & 1 \\ -1 & -1+|p|^2 \end{pmatrix}, \quad \rho_{1,1}^\sharp = r\begin{pmatrix} -1 & 0 \\ 1-|p|^2 & 0 \end{pmatrix}.$$

4. Building block matrices

In this section, we introduce and study some matrices that serve as building blocks for the closed-form expression of $\phi_{n,j}$. We assume that $\{X_k\}$ satisfies (5). Let h and $h_\sharp$ be as in (1) and (4), respectively, both satisfying (2). We also assume that $K \geq 1$ for K in (8). This assumption implies that $\{X_k\}$ is a d-variate ARMA process that is not an AR process; see Remark 3. For $m_1, \dots, m_K$ in (8), we define M by (7).

For $\mu \in \{1,\dots,K\}$, $i \in \{1,\dots,m_\mu\}$, and $n \in \mathbb{N}\cup\{0\}$, we define

$$\mathbf{p}_{\mu,i}(n) := p_{\mu,i}(n) I_d \in \mathbb{C}^{d\times d} \tag{13}$$

using $p_{\mu,i}(n)$ in (6). For $n \in \mathbb{N}\cup\{0\}$, we also define $\mathbf{p}_n \in \mathbb{C}^{dM\times d}$ by the following block representation:

$$\mathbf{p}_n := (\mathbf{p}_{1,1}(n),\dots,\mathbf{p}_{1,m_1}(n) \mid \mathbf{p}_{2,1}(n),\dots,\mathbf{p}_{2,m_2}(n) \mid \cdots \mid \mathbf{p}_{K,1}(n),\dots,\mathbf{p}_{K,m_K}(n))^\top. \tag{14}$$

Notice that

$$\mathbf{p}_0 = (I_d, 0, \dots, 0 \mid I_d, 0, \dots, 0 \mid \cdots \mid I_d, 0, \dots, 0)^\top \in \mathbb{C}^{dM\times d}. \tag{15}$$

We define $\Lambda \in \mathbb{C}^{dM\times dM}$ by

$$\Lambda := \sum_{\ell=0}^{\infty} \mathbf{p}_\ell \mathbf{p}_\ell^*. \tag{16}$$

For $\mu, \nu \in \{1,2,\dots,K\}$, we define $\Lambda^{\mu,\nu} \in \mathbb{C}^{dm_\mu\times dm_\nu}$ by the block representation

$$\Lambda^{\mu,\nu} := \begin{pmatrix} \lambda^{\mu,\nu}(1,1) & \lambda^{\mu,\nu}(1,2) & \cdots & \lambda^{\mu,\nu}(1,m_\nu) \\ \lambda^{\mu,\nu}(2,1) & \lambda^{\mu,\nu}(2,2) & \cdots & \lambda^{\mu,\nu}(2,m_\nu) \\ \vdots & \vdots & \ddots & \vdots \\ \lambda^{\mu,\nu}(m_\mu,1) & \lambda^{\mu,\nu}(m_\mu,2) & \cdots & \lambda^{\mu,\nu}(m_\mu,m_\nu) \end{pmatrix},$$

where, for $i \in \{1, \dots, m_\mu\}$, $j \in \{1, \dots, m_\nu\}$,

$$\lambda^{\mu,\nu}(i,j) := \sum_{r=0}^{j-1} \binom{i-1}{r} \binom{i+j-r-2}{i-1} \frac{p_\mu^{j-r-1} \overline{p}_\nu^{i-r-1}}{(1-p_\mu \overline{p}_\nu)^{i+j-r-1}} I_d \in \mathbb{C}^{d\times d}.$$

Here is a closed-form expression of Λ.

Lemma 3. *The matrix Λ has the following block representation:*

$$\Lambda = \begin{pmatrix} \Lambda^{1,1} & \Lambda^{1,2} & \cdots & \Lambda^{1,K} \\ \Lambda^{2,1} & \Lambda^{2,2} & \cdots & \Lambda^{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda^{K,1} & \Lambda^{K,2} & \cdots & \Lambda^{K,K} \end{pmatrix}.$$

We define

$$\tilde{h}(z) := \{h_\sharp(\overline{z})\}^*. \tag{17}$$

Then $\tilde{h}$ satisfies (2). We define, respectively, the forward MA and AR coefficients c_k and a_k of $\{X_k\}$ by

$$h(z) = \sum_{k=0}^{\infty} z^k c_k, \qquad -h(z)^{-1} = \sum_{k=0}^{\infty} z^k a_k, \qquad z \in \mathbb{D},$$

and the backward MA and AR coefficients $\tilde{c}_k$ and $\tilde{a}_k$ of $\{X_k\}$ by

$$\tilde{h}(z) = \sum_{k=0}^{\infty} z^k \tilde{c}_k, \qquad -\tilde{h}(z)^{-1} = \sum_{k=0}^{\infty} z^k \tilde{a}_k, \qquad z \in \mathbb{D}.$$

All of $\{c_k\}$, $\{a_k\}$, $\{\tilde{c}_k\}$ and $\{\tilde{a}_k\}$ are $\mathbb{C}^{d\times d}$-valued sequences that decay exponentially fast to zero, and we have $c_0 a_0 = \tilde{c}_0 \tilde{a}_0 = -I_d$. We have the AR representation $\sum_{k=-\infty}^{n} a_{n-k} X_k + \varepsilon_n = 0$ and the infinite prediction formula $P_{(-\infty,-1]} X_0 = \sum_{k=1}^{\infty} \phi_k X_{-k}$, where

$$\phi_k := c_0 a_k \in \mathbb{C}^{d\times d}, \qquad k \in \mathbb{N}. \tag{18}$$

We call ϕ_k the infinite predictor coefficients of $\{X_k\}$.

Remark 3. If K in (9) satisfies $K = 0$, then $a_0 = \rho_0$, $a_k = \rho_{0,k}$ $(1 \le k \le m_0)$ and $a_k = 0$ $(k \ge m_0 + 1)$. In particular, we have $\sum_{k=0}^{m_0} a_k X_{n-k} + \varepsilon_n = 0$ for $n \in \mathbb{Z}$. This implies that $P_{[-n,-1]} X_0 = \phi_1 X_{-1} + \cdots + \phi_{m_0} X_{-m_0}$ for $n \ge \max(m_0, 1)$ and ϕ_k in (18). Therefore, the finite predictor coefficients $\phi_{n,j}$ in (3) are trivially obtained. By this reason, we assume $K \ge 1$ in Sections 4–6.

For $\tilde{h}$ in (17), we see from Theorem 2 that

$$\tilde{h}(z)^{-1} = -\tilde{\rho}_0 - \sum_{\mu=1}^{K} \sum_{j=1}^{m_\mu} \frac{1}{(1-p_\mu z)^j} \tilde{\rho}_{\mu,j} - \sum_{j=1}^{m_0} z^j \tilde{\rho}_{0,j}, \tag{19}$$

where

$$\tilde{\rho}_0 := (\rho_0^\sharp)^*, \qquad \tilde{\rho}_{\mu,j} := (\rho_{\mu,j}^\sharp)^*, \quad \mu \in \{0,\dots,K\},\ j \in \{1,\dots,m_\mu\}.$$

Proposition 4. *We have*

$$a_n = \sum_{\mu=1}^{K} \sum_{j=1}^{m_\mu} \binom{n+j-1}{j-1} \overline{p}_\mu^n \rho_{\mu,j}, \qquad n \ge m_0 + 1, \tag{20}$$

$$\tilde{a}_n = \sum_{\mu=1}^{K} \sum_{j=1}^{m_\mu} \binom{n+j-1}{j-1} p_\mu^n \tilde{\rho}_{\mu,j}, \qquad n \ge m_0 + 1. \tag{21}$$

Moreover, if $m_0 \ge 1$, then we have

$$a_n = \rho_{0,n} + \sum_{\mu=1}^{K} \sum_{j=1}^{m_\mu} \binom{n+j-1}{j-1} \overline{p}_\mu^n \rho_{\mu,j}, \qquad n \in \{1,\dots,m_0\}, \tag{22}$$

$$\tilde{a}_n = \tilde{\rho}_{0,n} + \sum_{\mu=1}^{K} \sum_{j=1}^{m_\mu} \binom{n+j-1}{j-1} p_\mu^n \tilde{\rho}_{\mu,j}, \qquad n \in \{1,\dots,m_0\}. \tag{23}$$

Proof. Since

$$\frac{1}{(1-qz)^j}=\sum_{n=0}^{\infty}\binom{n+j-1}{j-1}q^n z^n,\qquad q,z\in\mathbb{D},\ j\in\mathbb{N},\tag{24}$$

(19) gives

$$\tilde{h}(z)^{-1}=-\tilde{\rho}_0-\sum_{n=0}^{\infty}z^n\sum_{\mu=1}^{K}\sum_{j=1}^{m_\mu}\binom{n+j-1}{j-1}p_\mu^n\tilde{\rho}_{\mu,j}-\sum_{j=1}^{m_0}z^j\tilde{\rho}_{0,j}.$$

Thus, (21) and (23) follow. Similarly, we obtain (20) and (22) from (8) and (24). □

For $n\in\mathbb{N}$, we define $v_n,\tilde{v}_n\in\mathbb{C}^{dM\times d}$ by

$$v_n:=\sum_{\ell=0}^{\infty}\mathbf{p}_\ell a_{n+\ell},$$
$$\tilde{v}_n:=\sum_{\ell=0}^{\infty}\bar{\mathbf{p}}_\ell\tilde{a}_{n+\ell}.$$

To give closed expressions for v_n and $\tilde{v}_n$, we introduce some matrices. For $n\in\mathbb{N}$ and $\mu,\nu\in\{1,2,\ldots,K\}$, we define $\Xi_n^{\mu,\nu}\in\mathbb{C}^{dm_\mu\times dm_\nu}$ by the block representation

$$\Xi_n^{\mu,\nu}:=\begin{pmatrix}\xi_n^{\mu,\nu}(1,1) & \xi_n^{\mu,\nu}(1,2) & \cdots & \xi_n^{\mu,\nu}(1,m_\nu)\\ \xi_n^{\mu,\nu}(2,1) & \xi_n^{\mu,\nu}(2,2) & \cdots & \xi_n^{\mu,\nu}(2,m_\nu)\\ \vdots & \vdots & \ddots & \vdots\\ \xi_n^{\mu,\nu}(m_\mu,1) & \xi_n^{\mu,\nu}(m_\mu,2) & \cdots & \xi_n^{\mu,\nu}(m_\mu,m_\nu)\end{pmatrix},$$

where, for $n\in\mathbb{N}$, $i\in\{1,\ldots,m_\mu\}$, $j\in\{1,\ldots,m_\nu\}$, $\xi_n^{\mu,\nu}(i,j)\in\mathbb{C}^{d\times d}$ is defined by

$$\xi_n^{\mu,\nu}(i,j):=\sum_{r=0}^{j-1}\binom{n+i+j-2}{r}\binom{i+j-r-2}{i-1}\frac{p_\mu^{j-r-1}\overline{p}_\nu^{n+i+j-r-2}}{(1-p_\mu\overline{p}_\nu)^{i+j-r-1}}I_d.$$

For $n\in\mathbb{N}$, we define $\Xi_n\in\mathbb{C}^{dM\times dM}$ by

$$\Xi_n:=\begin{pmatrix}\Xi_n^{1,1} & \Xi_n^{1,2} & \cdots & \Xi_n^{1,K}\\ \Xi_n^{2,1} & \Xi_n^{2,2} & \cdots & \Xi_n^{2,K}\\ \vdots & \vdots & \ddots & \vdots\\ \Xi_n^{K,1} & \Xi_n^{K,2} & \cdots & \Xi_n^{K,K}\end{pmatrix}.$$

We also define $\rho\in\mathbb{C}^{dM\times d}$ and $\tilde{\rho}\in\mathbb{C}^{dM\times d}$ by the block representations

$$\rho:=(\rho_{1,1}^\top,\ldots,\rho_{1,m_1}^\top\mid\rho_{2,1}^\top,\ldots,\rho_{2,m_2}^\top\mid\cdots\mid\rho_{K,1}^\top,\ldots,\rho_{K,m_K}^\top)^\top$$

and

$$\begin{aligned}\tilde{\rho}&:=(\tilde{\rho}_{1,1}^\top,\ldots,\tilde{\rho}_{1,m_1}^\top\mid\tilde{\rho}_{2,1}^\top,\ldots,\tilde{\rho}_{2,m_2}^\top\mid\cdots\mid\tilde{\rho}_{K,1}^\top,\ldots,\tilde{\rho}_{K,m_K}^\top)^\top\\ &=\left(\overline{\rho_{1,1}^\sharp},\ldots,\overline{\rho_{1,m_1}^\sharp}\mid\overline{\rho_{2,1}^\sharp},\ldots,\overline{\rho_{2,m_2}^\sharp}\mid\cdots\mid\overline{\rho_{K,1}^\sharp},\ldots,\overline{\rho_{K,m_K}^\sharp}\right)^\top,\end{aligned}$$

respectively.

Here are closed-form expressions for v_n and $\tilde{v}_n$.

Lemma 5. *We have*

$$v_n=\Xi_n\rho,\qquad n\ge m_0+1,\tag{25}$$
$$\tilde{v}_n=\overline{\Xi}_n\tilde{\rho},\qquad n\ge m_0+1.\tag{26}$$

Moreover, if $m_0\ge 1$, then we have

$$v_n=\Xi_n\rho+\sum_{\ell=0}^{m_0-n}\mathbf{p}_\ell\rho_{0,n+\ell},\qquad n\in\{1,\ldots,m_0\},\tag{27}$$

$$\tilde{v}_n = \overline{\Xi}_n \tilde{\rho} + \sum_{\ell=0}^{m_0-n} \overline{\mathbf{p}}_\ell \tilde{\rho}_{0,n+\ell}, \qquad n \in \{1, \ldots, m_0\}. \tag{28}$$

We define

$$h^\dagger(z) := h(1/\overline{z})^* \tag{29}$$

For $\mu \in \{0, 1, \ldots, K\}$, $j \in \{1, \ldots, m_\mu\}$, we put

$$\theta_{\mu,j} := -\lim_{z \to p_\mu} \frac{1}{(m_\mu - j)!} \frac{d^{m_\mu - j}}{dz^{m_\mu - j}} \left\{ (z - p_\mu)^{m_\mu} h_\sharp(z) h^\dagger(z)^{-1} \right\} \in \mathbb{C}^{d \times d}, \tag{30}$$

where $p_0 := 0$. We define the block-diagonal matrix $\Theta \in \mathbb{C}^{dM \times dM}$ by

$$\Theta := \begin{pmatrix} \Theta_1 & 0 & \cdots & 0 \\ 0 & \Theta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Theta_K \end{pmatrix}, \tag{31}$$

where, for $\mu \in \{1, \ldots, K\}$, $\Theta_\mu \in \mathbb{C}^{dm_\mu \times dm_\mu}$ is defined by

$$\Theta_\mu := \begin{pmatrix} \theta_{\mu,1} & \theta_{\mu,2} & \cdots & \theta_{\mu,m_\mu - 1} & \theta_{\mu,m_\mu} \\ \theta_{\mu,2} & \theta_{\mu,3} & \cdots & \theta_{\mu,m_\mu} & \\ \vdots & \vdots & & & \\ \theta_{\mu,m_\mu - 1} & \theta_{\mu,m_\mu} & & & \\ \theta_{\mu,m_\mu} & & & & \mathbf{0} \end{pmatrix} \tag{32}$$

using $\theta_{\mu,j}$ in (30) with (29).

For $n \in \mathbb{N} \cup \{0\}$, we define the block-diagonal matrix $\Pi_n \in \mathbb{C}^{dM \times dM}$ by

$$\Pi_n := \begin{pmatrix} \Pi_{1,n} & 0 & \cdots & 0 \\ 0 & \Pi_{2,n} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Pi_{K,n} \end{pmatrix}, \tag{33}$$

where, for $\mu \in \{1, \ldots, K\}$ and $n \in \mathbb{N} \cup \{0\}$, $\Pi_{\mu,n} \in \mathbb{C}^{dm_\mu \times dm_\mu}$ is defined by

$$\Pi_{\mu,n} := \begin{pmatrix} \mathbf{p}_{\mu,1}(n) & \mathbf{p}_{\mu,2}(n) & \mathbf{p}_{\mu,3}(n) & \cdots & \mathbf{p}_{\mu,m_\mu}(n) \\ & \mathbf{p}_{\mu,1}(n) & \mathbf{p}_{\mu,2}(n) & \cdots & \mathbf{p}_{\mu,m_\mu - 1}(n) \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & \mathbf{p}_{\mu,2}(n) \\ \mathbf{0} & & & & \mathbf{p}_{\mu,1}(n) \end{pmatrix} \tag{34}$$

using $\mathbf{p}_{\mu,i}(n)$ in (13).

5. Closed-form expression for finite predictor coefficients

In this section, we assume that $\{X_k\}$, h and $h_\sharp$ are as in Section 4. Thus $\{X_k\}$ is a d-variate ARMA process satisfying (5) and $K \geq 1$ for K in (8). Recall the finite predictor coefficients $\phi_{n,k} \in \mathbb{C}^{d \times d}$ of the d-variate ARMA process $\{X_k\}$ from (3). For $n \in \mathbb{N} \cup \{0\}$, we define $G_n, \tilde{G}_n \in \mathbb{C}^{dM \times dM}$ by

$$G_n := \Pi_n \Theta \Lambda, \tag{35}$$

$$\tilde{G}_n := (\Pi_n \Theta)^* \Lambda^\top. \tag{36}$$

Here is the main theorem of this paper, which gives a closed-form expression for $\phi_{n,j}$.

Theorem 6. *For $n \geq \max(m_0, 1)$ and $j \in \{1, \ldots, n\}$, we have*

$$\phi_{n,j} = c_0 a_j + c_0 \mathbf{p}_0^\top (I_{dM} - \tilde{G}_n G_n)^{-1} (\Pi_n \Theta)^* \{\Lambda^\top \Pi_n \Theta v_j + \tilde{v}_{n-j+1}\}. \tag{37}$$

Recall the assumption for Theorem 6 from the beginning of this section; $\{X_k\}$ in Theorem 6 is a general d-variate ARMA process that is not an AR process (see Remark 3). We remark that, from Lemma 19, $I_{dM} - \tilde{G}_n G_n$ is invertible for $n \geq m_0$.

Corollary 7. *If $m_0 = 0$, then, for $n \geq 1$ and $j \in \{1, \dots, n\}$, we have*

$$\phi_{n,j} = c_0 a_j + c_0 \mathbf{p}_0^\top (I_{dM} - \tilde{G}_n G_n)^{-1} (\Pi_n \Theta)^* \{\Lambda^\top \Pi_n \Theta \Xi_j \rho + \overline{\Xi}_{n-j+1} \tilde{\rho}\}. \tag{38}$$

Proof. The corollary follows immediately from Theorem 6 and Lemma 5. □

The matrices a_j, $\mathbf{p}_0$, Π_n, and Θ in (37) are given by the closed-form expressions (20) and (22), (15), (33) with (34), and (31) with (32), respectively. The closed-form expressions of Λ, v_n and $\tilde{v}_n$ are given by Lemmas 3 and 5, and those of G_n and $\tilde{G}_n$ by (35) and (36), respectively. Moreover, the matrix c_0 is given by $c_0 = h(0) = -\{\rho_0 + \sum_{\mu=1}^{K} \sum_{j=1}^{m_\mu} \rho_{\mu,j}\}^{-1}$. Therefore, (37) gives a complete closed-form expression for $\phi_{n,j}$. Notice that the sizes of all the matrices are fixed and independent of n.

Remark 4. Notice that $c_0 a_j = \phi_j$ in (37) is the infinite predictor coefficient.

Example 5. Suppose that $m_\mu = 1$, $\mu \in \{1, \dots, K\}$ and $m_0 = 0$, that is,

$$h(z)^{-1} = -\rho_0 - \sum_{\mu=1}^{K} \frac{1}{1 - \overline{p}_\mu z} \rho_{\mu,1}, \qquad h_\sharp(z)^{-1} = -\rho_0^\sharp - \sum_{\mu=1}^{K} \frac{1}{1 - \overline{p}_\mu z} \rho_{\mu,1}^\sharp.$$

Then, Corollary 7 holds with $a_j = \sum_{\mu=1}^{K} \overline{p}_\mu^j \rho_{\mu,1}$ for $j \geq 1$, $\mathbf{p}_0^\top = (I_d, \dots, I_d) \in \mathbb{C}^{d \times dK}$,

$$\Theta = \begin{pmatrix} p_1 h_\sharp(p_1) \rho_{1,1}^* & 0 & \cdots & 0 \\ 0 & p_2 h_\sharp(p_2) \rho_{2,1}^* & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_K h_\sharp(p_K) \rho_{K,1}^* \end{pmatrix} \in \mathbb{C}^{dK \times dK},$$

$$\Lambda = \begin{pmatrix} \frac{1}{1-p_1\overline{p}_1} I_d & \frac{1}{1-p_1\overline{p}_2} I_d & \cdots & \frac{1}{1-p_1\overline{p}_K} I_d \\ \frac{1}{1-p_2\overline{p}_1} I_d & \frac{1}{1-p_2\overline{p}_2} I_d & \cdots & \frac{1}{1-p_2\overline{p}_K} I_d \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{1-p_K\overline{p}_1} I_d & \frac{1}{1-p_K\overline{p}_2} I_d & \cdots & \frac{1}{1-p_K\overline{p}_K} I_d \end{pmatrix} \in \mathbb{C}^{dK \times dK},$$

$$\Pi_n = \begin{pmatrix} p_1^n I_d & 0 & \cdots & 0 \\ 0 & p_2^n I_d & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_K^n I_d \end{pmatrix} \in \mathbb{C}^{dK \times dK}, \qquad n \geq 0,$$

$$\Xi_n = \begin{pmatrix} \frac{\overline{p}_1^n}{1-p_1\overline{p}_1} I_d & \frac{\overline{p}_2^n}{1-p_1\overline{p}_2} I_d & \cdots & \frac{\overline{p}_K^n}{1-p_1\overline{p}_K} I_d \\ \frac{\overline{p}_1^n}{1-p_2\overline{p}_1} I_d & \frac{\overline{p}_2^n}{1-p_2\overline{p}_2} I_d & \cdots & \frac{\overline{p}_K^n}{1-p_2\overline{p}_K} I_d \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\overline{p}_1^n}{1-p_K\overline{p}_1} I_d & \frac{\overline{p}_2^n}{1-p_K\overline{p}_2} I_d & \cdots & \frac{\overline{p}_K^n}{1-p_K\overline{p}_K} I_d \end{pmatrix} \in \mathbb{C}^{dK \times dK}, \qquad n \geq 1,$$

$$\rho = (\rho_{1,1}^\top, \rho_{2,1}^\top, \dots, \rho_{K,1}^\top)^\top \in \mathbb{C}^{dK \times d}, \qquad \tilde{\rho} = \left(\overline{\rho_{1,1}^\sharp}, \overline{\rho_{2,1}^\sharp}, \dots, \overline{\rho_{K,1}^\sharp}\right)^\top \in \mathbb{C}^{dK \times d}$$

and $G_n = \Pi_n \Theta \Lambda$, $\tilde{G}_n = (\Pi_n \Theta)^* \Lambda^\top \in \mathbb{C}^{dK \times dK}$.

Remark 6. We define the block-diagonal matrix $J \in \mathbb{C}^{dM \times dM}$ by

$$J := \begin{pmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_K \end{pmatrix},$$

where, for $\nu \in \{1, \ldots, K\}$, $J_\nu \in \mathbb{C}^{dm_\nu \times dm_\nu}$ is defined by

$$J_\nu := \begin{pmatrix} \overline{p}_\nu I_d & I_d & & & \mathbf{0} \\ & \overline{p}_\nu I_d & I_d & & \\ & & \ddots & \ddots & \\ & & & \ddots & I_d \\ \mathbf{0} & & & & \overline{p}_\nu I_d \end{pmatrix}, \quad m_\nu \geq 2, \qquad := \overline{p}_\nu I_d, \quad m_\nu = 1.$$

Then it is easy to see that $\Xi_{n+1} = \Xi_n J$ for $n \in \mathbb{N}$. By this recursion, we can compute $\Xi_1, \ldots, \Xi_n$ in $O(n)$ arithmetic operations. The other matrices in (37) and (38) can also be computed in $O(n)$ operations. Therefore, we see that the complexity of the algorithm to compute $\phi_{n,1}, \ldots, \phi_{n,n}$ that is provided by Theorem 6 or Corollary 7 is only $O(n)$, which is the best possible. Notice that $(\phi_{n,n}, \phi_{n,n-1} \ldots, \phi_{n,1})$ is the solution to the Yule–Walker equation

$$(\phi_{n,n}, \phi_{n,n-1}, \ldots, \phi_{n,1}) T_n(w) = (\gamma(-n), \gamma(-n+1), \ldots, \gamma(-1))$$

or

$$T_n(w)(\phi_{n,n}, \phi_{n,n-1}, \ldots, \phi_{n,1})^* = (\gamma(-n), \gamma(-n+1), \ldots, \gamma(-1))^*,$$

where $T_n(w)$ is the truncated block Toeplitz matrix defined by

$$T_n(w) := \begin{pmatrix} \gamma(0) & \gamma(-1) & \cdots & \gamma(-n+1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(-n+2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{pmatrix} \in \mathbb{C}^{dn \times dn}.$$

Also notice that the multivariate Durbin–Levinson recursion solves the Yule–Walker equation in $O(n^2)$ time (see, e.g., Brockwell and Davis [5]). Algorithms for Toeplitz linear systems that run faster than $O(n^2)$ are called superfast; see Xi et al. [19] and the references therein.

Remark 7. From the discussions in Remark 6, we are naturally led to the problem of finding linear-time algorithms to compute the solution $x \in \mathbb{C}^{dn \times d}$ of the general block Toeplitz system $T_n(w)x = b$ for $b \in \mathbb{C}^{dn \times d}$ and w satisfying (1) with (2). This problem will be solved in [7].

Remark 8. One possible application of Theorem 6 is model fitting. More precisely, suppose that we are given a dataset $x_1, \ldots, x_N$ as a realization of the underlying process $\{X_k\}$. Then, for suitable n, we search for the parameters of the ARMA model that minimize the least squares error $\sum_{m=n+1}^{N} |x_m - \sum_{k=1}^{n} \phi_{n,k} x_{m-k}|^2$, using Theorem 6. In this way, we simultaneously fit the ARMA model to the data and estimate the predictor coefficients $\phi_{n,1}, \ldots, \phi_{n,n}$, without estimating the autocovariance function γ. The validity of this method will be discussed in future work.

6. Application

We continue to assume that $\{X_k\}$ is a d-variate ARMA process satisfying (5) and $K \geq 1$ for K in (8). In this section, we further assume

$$|p_1| > \max\{|p_\mu| : \mu \in \{2, \ldots, K\}\}, \tag{39}$$

and apply Theorem 6 to determine the asymptotic behavior of $\sum_{j=1}^{n} \|\phi_{n,j} - \phi_j\|$ as $n \to \infty$. We write $s_n \sim t_n$ as $n \to \infty$ to mean that $\lim_{n\to\infty} s_n/t_n = 1$.

Theorem 8. *We assume (39). Then*

$$\sum_{j=1}^{n} \|\phi_{n,j} - \phi_j\| \sim \frac{C_1}{(m_1-1)!} n^{m_1-1} |p_1|^n \quad \text{as } n \to \infty, \tag{40}$$

where C_1 is a positive constant given by $C_1 := \sum_{k=1}^{\infty} \|c_0 h(p_1)^* \rho_{1,m_1}^{\sharp} H \tilde{v}_k\|$ with $H := (I_d, 0, \ldots, 0) \in \mathbb{C}^{d \times dM}$.

Proof. First we show that the constant C_1 is in $(0, \infty)$. We define $C_{1,k} := \|c_0 h(p_1)^* \rho_{1,m_1}^{\sharp} H \tilde{v}_k\|$ for $k \in \mathbb{N}$, so that $C_1 := \sum_{k=1}^{\infty} C_{1,k}$ holds. Then, the sum converges since $C_{1,k}$ decays exponentially fast as $k \to \infty$. Therefore, it is enough to show that $C_{1,k} > 0$ for k large enough. By Lemma 5, we have, for $k \geq m_0 + 1$,

$$C_{1,k} = \|c_0 h(p_1)^* \rho^\sharp_{1,m_1} H \overline{\Xi}_k \tilde{\rho}\| = \left\| c_0 h(p_1)^* \rho^\sharp_{1,m_1} \left(\sum_{\nu=1}^{K} \sum_{j=1}^{m_\nu} \overline{\xi}_k^{1,\nu}(1,j)(\rho^\sharp_{\nu,j})^* \right) \right\|$$
$$= \left\| \sum_{\nu=1}^{K} \sum_{j=1}^{m_\nu} \overline{\xi}_k^{1,\nu}(1,j) c_0 h(p_1)^* \rho^\sharp_{1,m_1} (\rho^\sharp_{\nu,j})^* \right\| \geq k^{m_1-1}|p_1|^k (A_k - B_k),$$

where

$$A_k := \left\| (k^{m_1-1}|p_1|^k)^{-1} \overline{\xi}_k^{1,1}(1,m_1) c_0 h(p_1)^* \rho^\sharp_{1,m_1} (\rho^\sharp_{1,m_1})^* \right\|,$$
$$B_k := \left\| \sum_{(\nu,j)\neq(1,m_1)} (k^{m_1-1}|p_1|^k)^{-1} \overline{\xi}_k^{1,\nu}(1,j) c_0 h(p_1)^* \rho^\sharp_{1,m_1} (\rho^\sharp_{\nu,j})^* \right\|.$$

The main term in

$$\overline{\xi}_k^{1,1}(1,m_1) = \sum_{r=0}^{m_1-1} \binom{k+m_1-1}{r} \frac{\overline{p}_1^{m_1-r} p_1^{k+m_1-1-r}}{(1-|p_1|^2)^{m_1-r}} I_d$$

is $\binom{k+m_1-1}{m_1-1} p_1^k (1-|p_1|^2)^{-1} I_d$ for $r = m_1 - 1$ and we have $\lim_{k\to\infty} (k^{m_1-1}|p_1|^k)^{-1} \overline{\xi}_k^{1,1}(1,m_1) = \{(m_1-1)!(1-|p_1|^2)\}^{-1} I_d$, so that $\lim_{k\to\infty} A_k = A_\infty$, where $A_\infty := \{(m_1-1)!(1-|p_1|^2)\}^{-1} \|c_0 h(p_1)^* \rho^\sharp_{1,m_1} (\rho^\sharp_{1,m_1})^*\|$. Since $c_0 h(p_1)^*$ is invertible and $\rho^\sharp_{1,m_1}(\rho^\sharp_{1,m_1})^* \neq 0$, we have $A_\infty > 0$. On the other hand, (39) implies $\lim_{k\to\infty}(k^{m_1-1}|p_1|^k)^{-1}\overline{\xi}_k^{1,\nu}(1,j) = 0$ for $(\nu,j) \neq (1,m_1)$. Hence $\lim_{k\to\infty} B_k = 0$. Combining, we see that $C_{1,k} > 0$ for k large enough, as desired.

Next we prove (40). Recall $p_{\mu,i}(n)$ and $\mathbf{p}_{\mu,i}(n)$ from (6) and (13), respectively. Since (39) implies

$$\lim_{n\to\infty} \frac{1}{p_{1,m_1}(n)} \mathbf{p}_{\mu,i}(n) = \begin{cases} I_d, & \mu = 1,\ i = m_1, \\ 0, & \text{otherwise}, \end{cases}$$

we have $\lim_{n\to\infty}(1/p_{1,m_1}(n))\Pi_n = \Delta$, where $\Delta \in \mathbb{C}^{dM\times dM}$ is defined by

$$\Delta := \begin{pmatrix} \Delta_1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \qquad \Delta_1 := \begin{pmatrix} 0 & \cdots & 0 & I_d \\ 0 & \cdots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{pmatrix} \in \mathbb{C}^{dm_1\times dm_1}.$$

Hence, by Theorem 6 and the dominated convergence theorem, we get

$$\frac{1}{|p_{1,m_1}(n)|} \sum_{j=1}^{n} \|\phi_{n,j} - \phi_j\| = \frac{1}{|p_{1,m_1}(n)|} \sum_{j=1}^{n} \|c_0 \mathbf{p}_0^\top (I_{dM} - \tilde{G}_n G_n)^{-1} (\Pi_n \Theta)^* \{\Lambda^\top \Pi_n \Theta v_j + \tilde{v}_{n-j+1}\}\|$$
$$= \sum_{k=1}^{n} \left\| c_0 \mathbf{p}_0^\top (I_{dM} - \tilde{G}_n G_n)^{-1} \left(\frac{1}{p_{1,m_1}(n)} \Pi_n \Theta \right)^* \{\Lambda^\top \Pi_n \Theta v_{n+1-k} + \tilde{v}_k\} \right\| \quad \to \quad \sum_{k=1}^{\infty} \|c_0 \mathbf{p}_0^\top (\Delta\Theta)^* \tilde{v}_k\|, \quad n \to \infty.$$

By simple calculations, we have

$$\Delta\Theta = \begin{pmatrix} (p_1)^{m_1} h_\sharp(p_1) \rho^*_{1,m_1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \in \mathbb{C}^{dM\times dM},$$

so that $\mathbf{p}_0^\top (\Delta\Theta)^* = (p_1)^{m_1} \rho_{1,m_1} h_\sharp(p_1)^* H$. However, (12) implies that $\rho_{1,m_1} h_\sharp(p_1)^* = h(p_1)^* \rho^\sharp_{1,m_1}$. Hence, we see that $\sum_{k=1}^{\infty} \|c_0 \mathbf{p}_0^\top (\Delta\Theta)^* \tilde{v}_k\| = C_1$. Thus (40) follows. □

Corollary 9. *We assume* (39). *Then*

$$\lim_{n\to\infty} \frac{\sum_{j=1}^{n} \|\phi_{n,j} - \phi_j\|}{\sum_{k=n+1}^{\infty} \|\phi_k\|} = \frac{(1-|p_1|)C_1}{|p_1| \cdot \|c_0 \rho_{1,m_1}\|}. \tag{41}$$

Proof. By (18), Proposition 4 and (39), we have

$$\|\phi_k\| = \|c_0 a_k\| = \left\| \sum_{\mu=1}^{K} \sum_{j=1}^{m_\mu} \binom{k+j-1}{j-1} \overline{p}_\mu^k c_0 \rho_{\mu,j} \right\| \sim \|c_0 \rho_{1,m_1}\| \binom{k+m_1-1}{m_1-1} |p_1|^k, \qquad k \to \infty.$$

Hence, $\sum_{k=n+1}^{\infty} \|\phi_k\| \sim \|c_0\rho_{1,m_1}\| \sum_{k=n+1}^{\infty} \binom{k+m_1-1}{m_1-1} |p_1|^k$ as $n \to \infty$. From

$$\sum_{k=n+1}^{\infty} \binom{k+m_1-1}{m_1-1} x^k = \frac{1}{(m_1-1)!} \left(\frac{d}{dx}\right)^{m_1-1} \left(\frac{x^{n+m_1}}{1-x}\right), \qquad |x| < 1,$$

and Leibniz's rule, we have, as $k \to \infty$,

$$\sum_{k=n+1}^{\infty} \binom{k+m_1-1}{m_1-1} |p_1|^k \sim \binom{n+m_1}{m_1-1} \frac{|p_1|^{n+1}}{1-|p_1|} \sim \frac{n^{m_1-1}|p_1|^{n+1}}{(m_1-1)!(1-|p_1|)}.$$

Thus

$$\sum_{k=n+1}^{\infty} \|\phi_k\| \sim \frac{\|c_0\rho_{1,m_1}\|}{(m_1-1)!(1-|p_1|)} n^{m_1-1} |p_1|^{n+1}, \qquad n \to \infty. \tag{42}$$

The assertion (41) follows from (42) and Theorem 8. □

Remark 9. To explain the assumption (39), we consider two parameters $p_1 = x_1 + iy_1$ and $p_2 = x_2 + iy_2$ belonging to the space $A := \{(p_1, p_2) \in (\mathbb{D} \setminus \{0\})^2 : |p_1| \geq |p_2|\}$. Then, the arrangement $|p_1| > |p_2|$ is generic in the sense that the complement

$$\{(p_1, p_2) \in A : |p_1| = |p_2|\} = \{(p_1, p_2) \in A : x_1^2 + y_1^2 = x_2^2 + y_2^2\}$$

forms a hypersurface, hence its 4-dimensional Lebesgue measure is zero. In the same sense, the arrangement of $(p_1, \dots, p_K)$ given by (39) is generic. For, without loss of generality, we may assume $|p_1| \geq \max\{|p_\mu| : \mu = 2, \dots, K\}$. Then, if $(p_1, \dots, p_K)$ does not satisfy (39), then we have $|p_1| = |p_\mu|$ for some $\mu \in \{2, \dots, K\}$. Here, it should be noticed that the special choice of p_1 in (39) is just for the sake of simplicity; an analogue of Theorem 8, hence Corollary 9, still holds even if we replace (39) by, e.g.,

$$|p_K| > \max\{|p_\mu| : \mu \in \{1, \dots, K-1\}\}.$$

Still, it will be interesting to pursue analogues of Theorem 8 and Corollary 9 when (39) fails to hold, hence oscillations of the type $k_1 p_1^n + k_2 (e^{i\theta} p_1)^n$ occur as $n \to \infty$.

CRediT authorship contribution statement

Akihiko Inoue: Conceptualization.

Acknowledgments

The author would like to thank the Editor and referees for their helpful comments. The author was supported by the JSPS Grant-in-Aid 17K05302 (Inoue, A., Hiroshima Univ.).

Appendix A. Proof of Proposition 1

For $f : \mathbb{N} \cup \{0\} \to \mathbb{C}$ and $\mu \in \{1, 2, \dots, K\}$, we define $D_\mu f : \mathbb{N} \cup \{0\} \to \mathbb{C}$ by

$$D_\mu f(k) := f(k+1) - p_\mu f(k), \qquad k \in \mathbb{N} \cup \{0\}.$$

Proposition 10. *For $\mu, \nu \in \{1, 2, \dots, K\}$, $i \in \mathbb{N}$ and $k \in \mathbb{N} \cup \{0\}$,*

$$D_\nu p_{\mu,i}(k) = (p_\mu - p_\nu) p_{\mu,i}(k) + p_{\mu,i-1}(k), \tag{A.1}$$

where $p_{\mu,0} \equiv 0$.

Proof. Since

$$D_\nu p_{\mu,1}(k) = p_\mu^{k+1} - p_\nu p_\mu^k = (p_\mu - p_\nu) p_{\mu,1}(k),$$

(A.1) holds for $i=1$. If $i \geq 2$, then, Pascal's rule $\binom{k+1}{i-1} = \binom{k}{i-1} + \binom{k}{i-2}$ implies that

$$D_\nu p_{\mu,i}(k) = \binom{k+1}{i-1} p_\mu^{k-i+2} - \binom{k}{i-1} p_\nu p_\mu^{k-i+1} = (p_\mu - p_\nu) p_{\mu,i}(k) + p_{\mu,i-1}(k).$$

Thus (A.1) follows. □

Proof of Proposition 1. Let $\gamma_{\mu,i} \in \mathbb{C}$, $\mu \in \{1,\dots,K\}$, $i \in \{1,\dots,m_\mu\}$ and suppose that $\sum_{\mu=1}^K \sum_{i=1}^{m_\mu} \gamma_{\mu,i} p_{\mu,i}(k) = 0$, $k \in \{N,\dots,N+M-1\}$. By Proposition 10, we have

$$0 = \left(D_1^{m_1-1} D_2^{m_2} \cdots D_K^{m_K} \sum_{\mu=1}^K \sum_{i=1}^{m_\mu} \gamma_{\mu,i} p_{\mu,i} \right)(N) = \gamma_{1,m_1} p_1^N \prod_{\mu=2}^K (p_1 - p_\mu)^{m_\mu}.$$

Hence $\gamma_{1,m_1} = 0$. Repeating this procedure, we find that $\gamma_{\mu,i} = 0$, $\mu \in \{1,\dots,K\}$, $i \in \{1,\dots,m_\mu\}$. Thus $p_{\mu,i}$'s are linearly independent. □

Appendix B. Proof of Theorem 2

As in Section 3, we assume that $\{X_k\}$ satisfies (5). Let h and $h_\sharp$ be as in (1) and (4), respectively, both satisfying (2).

We consider the unitary matrix valued function $h^* h_\sharp^{-1} = h^{-1} h_\sharp^*$ on $\mathbb{T}$, called the phase function of $\{X_k\}$ (see p. 428 in Peller [15]). We define a sequence $\{\beta_k\}_{k=-\infty}^\infty$ as the (minus of the) Fourier coefficients of $h^* h_\sharp^{-1} = h^{-1} h_\sharp^*$:

$$\beta_k = -\int_{-\pi}^{\pi} e^{-ik\theta} h(e^{i\theta})^* h_\sharp(e^{i\theta})^{-1} \frac{d\theta}{2\pi} = -\int_{-\pi}^{\pi} e^{-ik\theta} h(e^{i\theta})^{-1} h_\sharp(e^{i\theta})^* \frac{d\theta}{2\pi}, \qquad k \in \mathbb{Z}. \tag{B.1}$$

From (B.1), we have

$$\beta_k^* = -\int_{-\pi}^{\pi} e^{ik\theta} \{h_\sharp(e^{i\theta})^*\}^{-1} h(e^{i\theta}) \frac{d\theta}{2\pi} = -\int_{-\pi}^{\pi} e^{ik\theta} h_\sharp(e^{i\theta}) \{h(e^{i\theta})^*\}^{-1} \frac{d\theta}{2\pi}, \qquad k \in \mathbb{Z}. \tag{B.2}$$

The proof of Theorem 2 is based on the calculations of β_k in two different ways.

Recall $h^\dagger$ from (29). From (8), we have

$$h^\dagger(z)^{-1} = -\rho_0^* - \sum_{\mu=1}^K \sum_{j=1}^{m_\mu} \frac{z^j}{(z-p_\mu)^j} \rho_{\mu,j}^* - \sum_{j=1}^{m_0} z^{-j} \rho_{0,j}^*. \tag{B.3}$$

Since $h(e^{i\theta})^* = h^\dagger(e^{i\theta})$, we see from (B.2) that

$$\beta_k^* = -\int_{-\pi}^{\pi} e^{ik\theta} h_\sharp(e^{i\theta}) h^\dagger(e^{i\theta})^{-1} \frac{d\theta}{2\pi}, \qquad k \in \mathbb{Z}. \tag{B.4}$$

Notice that the entries of $h_\sharp(z) h^\dagger(z)^{-1}$ are rational functions of $z \in \mathbb{C}$.

Recall $\theta_{\mu,j}$ from (30).

Proposition 11. *The matrix function $h_\sharp(z) h^\dagger(z)^{-1}$ has the form*

$$h_\sharp(z) h^\dagger(z)^{-1} = -\sum_{\mu=1}^K \sum_{j=1}^{m_\mu} \frac{1}{(z-p_\mu)^j} \theta_{\mu,j} - \sum_{j=1}^{m_0} z^{-j} \theta_{0,j} - R(z),$$

where $R(z)$ is a $d \times d$ matrix function whose entries are rational functions of z with no poles in $\overline{\mathbb{D}}$. Moreover, we have

$$\theta_{\mu,m_\mu} = \begin{cases} (p_\mu)^{m_\mu} h_\sharp(p_\mu) \rho_{\mu,m_\mu}^* \neq 0, & \mu \in \{1,\dots,K\}, \\ h_\sharp(0) \rho_{0,m_0}^* \neq 0, & \mu = 0. \end{cases} \tag{B.5}$$

Proof. From (B.3), we have

$$-h_\sharp(z) h^\dagger(z)^{-1} = h_\sharp(z)\rho_0^* + \sum_{\mu=1}^K \sum_{j=1}^{m_\mu} \frac{1}{(z-p_\mu)^j} z^j h_\sharp(z) \rho_{\mu,j}^* + \sum_{j=1}^{m_0} z^{-j} h_\sharp(z) \rho_{0,j}^* = \sum_{\mu=1}^K \sum_{j=1}^{m_\mu} \frac{1}{(z-p_\mu)^j} \theta_{\mu,j} + \sum_{j=1}^{m_0} z^{-j} \theta_{0,j} + R(z),$$

where $R(z)$ is a $d \times d$ matrix valued function whose entries are rational functions of z with no poles in $\overline{\mathbb{D}}$. In particular, we have $\theta_{0,m_0} = h_\sharp(0)\rho_{0,m_0}^*$ and $\theta_{\mu,m_\mu} = (p_\mu)^{m_\mu} h_\sharp(p_\mu) \rho_{\mu,m_\mu}^*$, $\mu \in \{1,\dots,K\}$. Since $\rho_{0,m_0} \neq 0$ and $h_\sharp(0)$ is invertible, we see that $\theta_{0,m_0} \neq 0$. Similarly, $\theta_{\mu,m_\mu} \neq 0$, $\mu \in \{1,\dots,K\}$. □

Proposition 12. *We have* $\beta_{n+1}^{*} = \sum_{\mu=1}^{K}\sum_{j=1}^{m_\mu}\binom{n}{j-1}p_\mu^{n-j+1}\theta_{\mu,j} + \sum_{j=1}^{m_0}\delta_{n+1,j}\theta_{0,j}$ *for* $n \in \mathbb{N}\cup\{0\}$. *In particular,* $\beta_{n+1}^{*} = \sum_{\mu=1}^{K}\sum_{j=1}^{m_\mu}\binom{n}{j-1}p_\mu^{n-j+1}\theta_{\mu,j}$ *for* $n \geq m_0$.

Proof. By (B.4), Proposition 11 and Cauchy's formula, we have, for $n \in \mathbb{N}\cup\{0\}$,

$$\beta_{n+1}^{*} = -\int_{\mathbb{T}}\zeta^n h_\sharp(\zeta)h^\dagger(\zeta)^{-1}\frac{d\zeta}{2\pi i} = \sum_{\mu=1}^{K}\sum_{j=1}^{m_\mu}\int_{\mathbb{T}}\frac{\zeta^n}{(\zeta-p_\mu)^j}\frac{d\zeta}{2\pi i}\theta_{\mu,j} + \sum_{j=1}^{m_0}\int_{\mathbb{T}}\zeta^{n-j}\frac{d\zeta}{2\pi i}\theta_{0,j} + \int_{\mathbb{T}}\zeta^n R(\zeta)\frac{d\zeta}{2\pi i}$$
$$= \sum_{\mu=1}^{K}\sum_{j=1}^{m_\mu}\binom{n}{j-1}p_\mu^{n-j+1}\theta_{\mu,j} + \sum_{j=1}^{m_0}\delta_{n+1,j}\theta_{0,j}.$$

Thus, the proposition follows. $\square$

Proof of Theorem 2. As in (8) with (9), we can write $h_\sharp(z)^{-1}$ in the form

$$h_\sharp(z)^{-1} = -\sigma_0 - \sum_{\mu=1}^{L}\sum_{j=1}^{n_\mu}\frac{1}{(1-\overline{r}_\mu z)^j}\sigma_{\mu,j} - \sum_{j=1}^{n_0}z^j\sigma_{0,j},$$

where

$$\begin{cases} L \in \mathbb{N}\cup\{0\}, \\ r_\mu \in \mathbb{D}\setminus\{0\}, \quad \mu \in \{1,\dots,L\}, \qquad r_\mu \neq r_\nu, \quad \mu \neq \nu, \\ n_\mu \in \mathbb{N}, \quad \mu \in \{1,\dots,L\}, \qquad n_0 \in \mathbb{N}\cup\{0\}, \\ \sigma_{\mu,j} \in \mathbb{C}^{d\times d}, \quad \mu \in \{0,\dots,L\},\ j \in \{1,\dots,n_\mu\}, \qquad \sigma_0 \in \mathbb{C}^{d\times d}, \\ \sigma_{\mu,n_\mu} \neq 0, \quad \mu \in \{0,\dots,L\}. \end{cases}$$

We put $r_0 := 0$ and $h_\sharp^\dagger(z) := \{h_\sharp(1/\overline{z})\}^*$. We follow the argument in the proof of Proposition 12 by using $\beta_k^* = -\int_{-\pi}^{\pi}e^{ik\theta}\{h_\sharp(e^{i\theta})^*\}^{-1}h(e^{i\theta})d\theta/(2\pi)$ instead of $\beta_k^* = -\int_{-\pi}^{\pi}e^{ik\theta}h_\sharp(e^{i\theta})\{h(e^{i\theta})^*\}^{-1}d\theta/(2\pi)$ to calculate β_{n+1}^*. Then,

$$\beta_{n+1}^{*} = \sum_{\mu=1}^{L}\sum_{j=1}^{n_\mu}\binom{n}{j-1}r_\mu^{n-j+1}\lambda_{\mu,j} + \sum_{j=1}^{n_0}\delta_{n+1,j}\lambda_{0,j}, \qquad n \in \mathbb{N}\cup\{0\}, \tag{B.6}$$

where

$$\lambda_{\mu,j} = -\lim_{z\to r_\mu}\frac{1}{(n_\mu-j)!}\frac{d^{n_\mu-j}}{dz^{n_\mu-j}}\left\{(z-r_\mu)^{n_\mu}h_\sharp^\dagger(z)^{-1}h(z)\right\} \in \mathbb{C}^{d\times d}, \qquad \mu \in \{0,\dots,L\},\ j \in \{1,\dots,n_\mu\}.$$

We also obtain

$$\lambda_{\mu,n_\mu} = \begin{cases} (r_\mu)^{n_\mu}\sigma_{\mu,n_\mu}^{*}h(r_\mu) \neq 0, & \mu \in \{1,\dots,L\}, \\ \sigma_{0,n_0}^{*}h(0) \neq 0, & \mu = 0. \end{cases} \tag{B.7}$$

From Proposition 12 and (B.6), we have

$$\sum_{\mu=1}^{K}\sum_{j=1}^{m_\mu}\binom{n}{j-1}p_\mu^{n-j+1}\theta_{\mu,j} + \sum_{j=1}^{m_0}\delta_{n+1,j}\theta_{0,j} = \sum_{\mu=1}^{L}\sum_{j=1}^{n_\mu}\binom{n}{j-1}r_\mu^{n-j+1}\lambda_{\mu,j} + \sum_{j=1}^{n_0}\delta_{n+1,j}\lambda_{0,j}, \qquad n \in \mathbb{N}\cup\{0\}.$$

In particular, $\sum_{\mu=1}^{K}\sum_{j=1}^{m_\mu}\binom{n}{j-1}p_\mu^{n-j+1}\theta_{\mu,j} = \sum_{\mu=1}^{L}\sum_{j=1}^{n_\mu}\binom{n}{j-1}r_\mu^{n-j+1}\lambda_{\mu,j}$ for $n \geq \max(m_0, n_0)$. This and Proposition 1 yield $K = L$, $p_\mu = r_{f(\mu)}$, $m_\mu = n_{f(\mu)}$ and $\theta_{\mu,j} = \lambda_{f(\mu),j}$ for $\mu \in \{1,\dots,K\}$, $j \in \{1,\dots,m_\mu\}$ and some bijection $f : \{1,\dots,K\} \to \{1,\dots,K\}$. We now have $\sum_{j=1}^{m_0}\delta_{n+1,j}\theta_{0,j} = \sum_{j=1}^{n_0}\delta_{n+1,j}\lambda_{0,j}$ for $n \in \mathbb{N}\cup\{0\}$, and this gives $m_0 = n_0$ (as well as $\theta_{0,j} = \lambda_{0,j}$, $j \in \{1,\dots,m_0\}$). Thus, (10) and (11) hold with $\rho_0^\sharp = \sigma_0$ and $\rho_{\mu,j}^\sharp = \sigma_{f(\mu),j}$, $\mu \in \{0,\dots,K\}$, $j \in \{1,\dots,m_\mu\}$. Finally, we obtain (12) from $\theta_{\mu,m_\mu} = \lambda_{f(\mu),m_\mu}$, (B.5) and (B.7). $\square$

Appendix C. Proofs of Lemmas 3 and 5

To prove Lemma 3, we use the next proposition.

Proposition 13. *For* $i, j, n \in \mathbb{N}\cup\{0\}$ *and* $x, y \in \mathbb{D}$, *we have*

$$\sum_{\ell=0}^{\infty}\binom{\ell}{i}\binom{\ell+n}{j}x^{\ell-i}y^{\ell+n-j} = \sum_{r=0}^{j}\binom{n+i}{r}\binom{i+j-r}{i}\frac{x^{j-r}y^{n+i-r}}{(1-xy)^{i+j+1-r}}.$$

Proof. Let $i, j, n \in \mathbb{N} \cup \{0\}$ and $x, y \in \mathbb{D}$. Since $y^n/(1-xy) = \sum_{\ell=0}^{\infty} x^\ell y^{n+\ell}$, we have

$$\frac{1}{i!j!}\left(\frac{\partial}{\partial y}\right)^j \left(\frac{\partial}{\partial x}\right)^i \frac{y^n}{1-xy} = \sum_{\ell=0}^{\infty} \binom{\ell}{i}\binom{n+\ell}{j} x^{\ell-i} y^{n+\ell-j}.$$

On the other hand, since $(1/r!)(d/dy)^r y^{n+i} = \binom{n+i}{r} y^{n+i-r}$ and

$$\frac{1}{(j-r)!}\left(\frac{\partial}{\partial y}\right)^{j-r} \frac{1}{(1-xy)^{i+1}} = \binom{i+j-r}{j-r}\frac{x^{j-r}}{(1-xy)^{i+j+1-r}} = \binom{i+j-r}{i}\frac{x^{j-r}}{(1-xy)^{i+j+1-r}}, \qquad j \ge r,$$

we have

$$\begin{aligned}
\frac{1}{i!j!}\left(\frac{\partial}{\partial y}\right)^j \left(\frac{\partial}{\partial x}\right)^i \frac{y^n}{1-xy} &= \frac{1}{j!}\left(\frac{\partial}{\partial y}\right)^j \frac{y^{n+i}}{(1-xy)^{i+1}} \\
&= \sum_{r=0}^{j} \binom{j}{r}\binom{j}{r}^{-1} \left\{\frac{1}{r!}\left(\frac{\partial}{\partial y}\right)^r y^{n+i}\right\} \left\{\frac{1}{(j-r)!}\left(\frac{\partial}{\partial y}\right)^{j-r} \frac{1}{(1-xy)^{i+1}}\right\} \\
&= \sum_{r=0}^{j} \binom{n+i}{r}\binom{i+j-r}{i} \frac{x^{j-r}y^{n+i-r}}{(1-xy)^{i+j+1-r}}.
\end{aligned}$$

Comparing, we obtain the proposition. □

Remark 10. Notice that Proposition 13 with $n = 0$ implies

$$\sum_{r=0}^{j} \binom{i}{r}\binom{i+j-r}{i} \frac{x^{j-r}y^{i-r}}{(1-xy)^{i+j+1-r}} = \sum_{r=0}^{i} \binom{j}{r}\binom{i+j-r}{j} \frac{x^{j-r}y^{i-r}}{(1-xy)^{i+j+1-r}}.$$

Also, notice that $\binom{i}{r}\binom{i+j-r}{i} = \binom{j}{r}\binom{i+j-r}{j}$.

Proof of Lemma 3. The proof is immediate from (16) and Proposition 13 with $n = 0$, and i and j replaced by $i-1$ and $j-1$, respectively. □

Proof of Lemma 5. If $n \ge m_0 + 1$, then Proposition 13 yields, for $\mu \in \{1, \ldots, K\}$, $i \in \{1, \ldots, m_\mu\}$,

$$\sum_{\ell=0}^{\infty} \mathbf{p}_{\mu,i}(\ell) a_{\ell+n} = \sum_{\nu=1}^{K}\sum_{j=1}^{m_\nu} \left\{ \sum_{\ell=0}^{\infty} \binom{\ell}{i-1}\binom{n+\ell+j-1}{j-1} p_\mu^{\ell-i+1} \overline{p}_\nu^{n+\ell} \right\} \rho_{\nu,j} = \sum_{\nu=1}^{K}\sum_{j=1}^{m_\nu} \xi_n^{\mu,\nu}(i,j) \rho_{\nu,j}$$

and

$$\sum_{\ell=0}^{\infty} \overline{\mathbf{p}}_{\mu,i}(\ell) \tilde{a}_{\ell+n} = \sum_{\nu=1}^{K}\sum_{j=1}^{m_\nu} \left\{ \sum_{\ell=0}^{\infty} \binom{\ell}{i-1}\binom{n+\ell+j-1}{j-1} \overline{p}_\mu^{\ell-i+1} p_\nu^{n+\ell} \right\} \tilde{\rho}_{\nu,j} = \sum_{\nu=1}^{K}\sum_{j=1}^{m_\nu} \overline{\xi}_n^{\mu,\nu}(i,j) \tilde{\rho}_{\nu,j}.$$

Thus, (25) and (26) follow. If $m_0 \ge 1$ and $1 \le n \le m_0$, then, similarly, we have (27) and (28). □

Appendix D. Proof of Theorem 6

To prove Theorem 6, we first prepare some propositions and lemmas. Recall $\mathbf{p}_n$ from (14).

Proposition 14. *For $N \in \mathbb{N} \cup \{0\}$, the matrix $(\mathbf{p}_N, \mathbf{p}_{N+1}, \ldots, \mathbf{p}_{N+M-1}) \in \mathbb{C}^{dM \times dM}$ is invertible.*

Proof. For $k \in \mathbb{N} \cup \{0\}$, we define $p(k) \in \mathbb{C}^M$ by

$$p(k) = (p_{1,1}(k), \ldots, p_{1,m_1}(k) | p_{2,1}(k), \ldots, p_{2,m_2}(k) | \cdots | p_{K,1}(k), \ldots, p_{K,m_K}(k))^\top.$$

Then, by the definition of determinant, we have

$$\det(\mathbf{p}_N, \mathbf{p}_{N+1}, \ldots, \mathbf{p}_{N+M-1}) = \{\det(p(N), p(N+1), \ldots, p(N+M-1))\}^d.$$

Since Proposition 1 implies that $\det(p(N), p(N+1), \ldots, p(N+M-1)) \neq 0$, the assertion follows. □

The next proposition will be used in the proof of Lemma 19.

Proposition 15. *The matrix Λ is positive definite. In particular, Λ is invertible.*

Proof. Clearly, Λ is a Hermitian matrix. Suppose that $v\Lambda v^* = 0$ for $v \in \mathbb{C}^{1\times dM}$. Since $v\mathbf{p}_\ell \mathbf{p}_\ell^* v^* = v\mathbf{p}_\ell (v\mathbf{p}_\ell)^* \geq 0$, we see that $v\mathbf{p}_\ell = 0$ for any $\ell \in \mathbb{N}\cup\{0\}$. This implies $v(\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_{M-1}) = 0$. Since $(\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_{M-1}) \in \mathbb{C}^{dM\times dM}$ is invertible by Proposition 14, we have $v = 0$. Thus, Λ is positive definite. □

Let $X_k = \int_{-\pi}^{\pi} e^{-ik\theta}\eta(d\theta)$, $k \in \mathbb{Z}$, be the spectral representation of $\{X_k\}$, where η is a $\mathbb{C}^d$-valued random spectral measure. We define a d-variate stationary process $\{\varepsilon_k : k \in \mathbb{Z}\}$, called the forward innovation process of $\{X_k\}$, by

$$\varepsilon_k := \int_{-\pi}^{\pi} e^{-ik\theta} h(e^{i\theta})^{-1}\eta(d\theta), \qquad k \in \mathbb{Z}.$$

Then, $\{\varepsilon_k\}$ satisfies $\langle \varepsilon_n, \varepsilon_m\rangle = \delta_{nm} I_d$ and $V^X_{(-\infty,n]} = V^\varepsilon_{(-\infty,n]}$ for $n \in \mathbb{Z}$, hence

$$(V^X_{(-\infty,n]})^\perp = V^\varepsilon_{[n+1,\infty)}, \qquad n \in \mathbb{Z}. \tag{D.1}$$

We also define the backward innovation process $\{\tilde{\varepsilon}_k : k \in \mathbb{Z}\}$ of $\{X_k\}$ by

$$\tilde{\varepsilon}_k := \int_{-\pi}^{\pi} e^{ik\theta} \{h_\sharp(e^{i\theta})^*\}^{-1}\eta(d\theta), \qquad k \in \mathbb{Z}.$$

Then, $\{\tilde{\varepsilon}_k\}$ satisfies $\langle \tilde{\varepsilon}_n, \tilde{\varepsilon}_m\rangle = \delta_{nm} I_d$ and $V^X_{[-n,\infty)} = V^{\tilde{\varepsilon}}_{(-\infty,n]}$ for $n \in \mathbb{Z}$, hence

$$(V^X_{[-n,\infty)})^\perp = V^{\tilde{\varepsilon}}_{[n+1,\infty)}, \qquad n \in \mathbb{Z}. \tag{D.2}$$

For $n \in \mathbb{N}\cup\{0\}$, we define $\mathcal{H}_n : (V^X_{[-n,\infty)})^\perp \to (V^X_{(-\infty,-1]})^\perp$ by

$$\mathcal{H}_n x := P^\perp_{(-\infty,-1]} x, \qquad x \in (V^X_{[-n,\infty)})^\perp,$$

and $\tilde{\mathcal{H}}_n : (V^X_{(-\infty,-1]})^\perp \to (V^X_{[-n,\infty)})^\perp$ by

$$\tilde{\mathcal{H}}_n x := P^\perp_{[-n,\infty)} x, \qquad x \in (V^X_{(-\infty,-1]})^\perp.$$

We denote by $\|\mathcal{H}_n\|$ (resp., $\|\tilde{\mathcal{H}}_n\|$) the operator norm of $\mathcal{H}_n$ (resp., $\tilde{\mathcal{H}}_n$).

Proposition 16. *For $n \in \mathbb{N}\cup\{0\}$, we have $\|\mathcal{H}_n\| = \|\tilde{\mathcal{H}}_n\| < 1$.*

Proof. Let $\{X'_k : k \in \mathbb{Z}\}$ be the dual process of $\{X_k\}$, which is a d-variate stationary process characterized by the biorthogonality relation $\langle X_j, X'_k\rangle = \delta_{jk} I_d$; see Masani [14] and Section 5 in [11]. The process $\{X'_k\}$ admits the two MA representations $X'_n = -\sum_{k=0}^\infty a_k^* \varepsilon_{n+k}$ and $X'_{-n} = -\sum_{k=0}^\infty \tilde{a}_k^* \tilde{\varepsilon}_{n+k}$ for $n \in \mathbb{Z}$. Moreover, for the spectral density w of $\{X_k\}$, $\{X'_k\}$ has the spectral density w^{-1}. For $n \geq 0$, let

$$\rho_n := \sup\{|(x,y)_V| : x \in V^{X'}_{(-\infty,-n-1]},\ y \in V^{X'}_{[0,\infty)},\ \|x\|_V \leq 1,\ \|y\|_V \leq 1\}$$

be the cosine of angle between $V^{X'}_{(-\infty,-n-1]}$ and $V^{X'}_{[0,\infty)}$ (see, e.g., Treil and Volberg [17,18], Pourahmadi [16], and Bingham [4]). Since both w and w^{-1} are continuous, hence bounded, on $\mathbb{T}$, w^{-1} satisfies the matrix Muckenhoupt condition

$$\sup_I \left\| \left(\frac{1}{m(I)}\int_I w^{-1} dm\right)^{1/2} \left(\frac{1}{m(I)}\int_I w\, dm\right)^{1/2} \right\| < \infty,$$

where m is the normalized ($m(\mathbb{T}) = 1$) Lebesgue measure on $\mathbb{T}$ and the supremum is taken over all subarcs I of $\mathbb{T}$. Therefore, by Treil and Volberg [17] (see also Peller [15], Arov and Dym [1], and Bingham [4]), we have $\rho_n < 1$ for $n \geq 0$. Since both $-\sum_{k=0}^\infty z^k a_k^* = \{h(\bar{z})^*\}^{-1}$ and $-\sum_{k=0}^\infty z^k \tilde{a}_k^* = h_\sharp(z)^{-1}$ are outer (see, e.g., Katsnelson and Kirstein [12] and Section 2 in [11]), we see from (D.1) and (D.2) that $V^{X'}_{[0,\infty)} = V^\varepsilon_{[0,\infty)} = (V^X_{(-\infty,-1]})^\perp$ and that $V^{X'}_{(-\infty,-n-1]} = V^{\tilde{\varepsilon}}_{[n+1,\infty)} = (V^X_{[-n,\infty)})^\perp$. Therefore,

$$\rho_n = \sup\{|(x,y)_V| : x \in (V^X_{[-n,\infty)})^\perp,\ y \in (V^X_{(-\infty,-1]})^\perp, \|x\|_V \leq 1,\ \|y\|_V \leq 1\} = \|\mathcal{H}_n\| = \|\tilde{\mathcal{H}}_n\|$$

(see Remark 11 for the second and third equalities), so that $\|\mathcal{H}_n\| = \|\tilde{\mathcal{H}}_n\| < 1$ for $n \geq 0$, as desired. □

Remark 11. For two closed subspaces A and B of a Hilbert space L, let $P_A : L \to A$ be the orthogonal projection operator and $P_A|_B$ the restriction of P_A to B. Then we have $\sup\{|(x,y)| : x \in A,\ y \in B,\ \|x\| \leq 1,\ \|y\| \leq 1\} = \|P_A|_B\|$.

The next lemma plays a key role in the arguments below.

Lemma 17. *For $n \geq m_0$ and $k, \ell \in \mathbb{N}\cup\{0\}$, we have $\beta^*_{n+k+\ell+1} = \mathbf{p}_\ell^\top \Pi_n \Theta \mathbf{p}_k$, hence $\beta_{n+k+\ell+1} = \mathbf{p}_k^* (\Pi_n\Theta)^* \overline{\mathbf{p}}_\ell$.*

Proof. We have

$$\sum_{j=1}^{\infty}\binom{n+k+\ell}{j-1}x^{j-1}=(1+x)^{n+k+\ell}=(1+x)^{k}(1+x)^{\ell}(1+x)^{n}=\sum_{j=1}^{\infty}\left\{\sum_{r=0}^{j-1}\binom{k}{j-1-r}\sum_{s=0}^{r}\binom{\ell}{s}\binom{n}{r-s}\right\}x^{j-1}$$
$$=\sum_{j=1}^{\infty}\left\{\sum_{i=1}^{j}\binom{k}{j-i}\sum_{q=1}^{i}\binom{\ell}{q-1}\binom{n}{i-q}\right\}x^{j-1},$$

where we have used the substitutions $i=r+1$ and $q=s+1$. Hence $\binom{n+k+\ell}{j-1}=\sum_{i=1}^{j}\binom{k}{j-i}\sum_{q=1}^{i}\binom{\ell}{q-1}\binom{n}{i-q}$ for $j\in\mathbb{N}$. Since $\mathbf{p}_\ell^\top\Pi_n\Theta\mathbf{p}_k=\mathbf{p}_\ell^\top\Pi_n\times\Theta\mathbf{p}_k$, this and Proposition 12 yield, for $n\geq m_0$,

$$\mathbf{p}_\ell^\top\Pi_n\Theta\mathbf{p}_k=\sum_{\mu=1}^{K}\sum_{i=1}^{m_\mu}\left\{\sum_{q=1}^{i}\binom{\ell}{q-1}p_\mu^{\ell-q+1}\binom{n}{i-q}p_\mu^{n-i+q}I_d\right\}\left\{\sum_{j=i}^{m_\mu}\binom{k}{j-i}p_\mu^{k+i-j}\theta_{\mu,j}\right\}$$
$$=\sum_{\mu=1}^{K}\sum_{j=1}^{m_\mu}\left\{\sum_{i=1}^{j}\binom{k}{j-i}\sum_{q=1}^{i}\binom{\ell}{q-1}\binom{n}{i-q}\right\}p_\mu^{n+\ell+k+1-j}\theta_{\mu,j}$$
$$=\sum_{\mu=1}^{K}\sum_{j=1}^{m_\mu}\binom{n+k+\ell}{j-1}p_\mu^{n+\ell+k+1-j}\theta_{\mu,j}=\beta^*_{n+k+\ell+1},$$

as desired. □

For $n\in\mathbb{N}\cup\{0\}$, we define $H_n:\{(V^X_{[-n,\infty)})^\perp\}^d\to\{(V^X_{(-\infty,-1]})^\perp\}^d$ by

$$H_nx:=(\mathcal{H}_nx^1,\ldots,\mathcal{H}_nx^d)^\top,\qquad x=(x^1,\ldots,x^d)^\top\in(V^X_{[-n,\infty)})^\perp,$$

and $\tilde{H}_n:\{(V^X_{(-\infty,-1]})^\perp\}^d\to\{(V^X_{[-n,\infty)})^\perp\}^d$ by

$$\tilde{H}_nx:=(\tilde{\mathcal{H}}_nx^1,\ldots,\tilde{\mathcal{H}}_nx^d)^\top,\qquad x=(x^1,\ldots,x^d)^\top\in\{(V^X_{(-\infty,-1]})^\perp\}^d.$$

Then, by Lemma 4.2 in [11], we have, for $\{s_\ell\}\in\ell_{2+}^{d\times d}$,

$$H_n\left(\sum_{\ell=0}^{\infty}s_\ell\tilde{\varepsilon}_{n+\ell+1}\right)=-\sum_{j=0}^{\infty}\left(\sum_{\ell=0}^{\infty}s_\ell\beta^*_{n+j+\ell+1}\right)\varepsilon_j\quad\text{and}\quad\tilde{H}_n\left(\sum_{\ell=0}^{\infty}s_\ell\varepsilon_\ell\right)=-\sum_{j=0}^{\infty}\left(\sum_{\ell=0}^{\infty}s_\ell\beta_{n+j+\ell+1}\right)\tilde{\varepsilon}_{n+j+1}.\tag{D.3}$$

Proposition 18. *For $n\geq m_0$ and $v\in\mathbb{C}^{dM\times d}$,*

$$H_n\left(\sum_{\ell=0}^{\infty}(v^\top\overline{\mathbf{p}}_\ell)\tilde{\varepsilon}_{n+\ell+1}\right)=-\sum_{j=0}^{\infty}(v^\top\Lambda^\top\Pi_n\Theta\mathbf{p}_j)\varepsilon_j,\tag{D.4}$$
$$\tilde{H}_n\left(\sum_{\ell=0}^{\infty}(v^\top\mathbf{p}_\ell)\varepsilon_\ell\right)=-\sum_{j=0}^{\infty}(v^\top\Lambda(\Pi_n\Theta)^*\overline{\mathbf{p}}_j)\tilde{\varepsilon}_{n+j+1}.\tag{D.5}$$

Proof. First, we see from Lemma 17 that, for $n\geq m_0$ and $j\in\mathbb{N}\cup\{0\}$,

$$\sum_{\ell=0}^{\infty}v^\top\overline{\mathbf{p}}_\ell\beta^*_{n+j+\ell+1}=v^\top\left(\sum_{\ell=0}^{\infty}\overline{\mathbf{p}}_\ell\mathbf{p}_\ell^\top\right)\Pi_n\Theta\mathbf{p}_j=v^\top\Lambda^\top\Pi_n\Theta\mathbf{p}_j.$$

This and the first equality in (D.3) yield (D.4). Next, we see from Lemma 17 that, for $n\geq m_0$ and $j\in\mathbb{N}\cup\{0\}$,

$$\sum_{\ell=0}^{\infty}v^\top\mathbf{p}_\ell\beta_{n+j+\ell+1}=v^\top\left(\sum_{\ell=0}^{\infty}\mathbf{p}_\ell\mathbf{p}_\ell^*\right)(\Pi_n\Theta)^*\overline{\mathbf{p}}_j=v^\top\Lambda(\Pi_n\Theta)^*\overline{\mathbf{p}}_j.$$

This and the second equality in (D.3) give (D.5). □

Here is a key lemma.

Lemma 19. *For $n\geq m_0$, both $I_{dM}-\tilde{G}_nG_n$ and $I_{dM}-G_n\tilde{G}_n$ are invertible and we have $\sum_{k=0}^{\infty}(\tilde{G}_nG_n)^k=(I_{dM}-\tilde{G}_nG_n)^{-1}$ and $\sum_{k=0}^{\infty}(G_n\tilde{G}_n)^k=(I_{dM}-G_n\tilde{G}_n)^{-1}$, where $(\tilde{G}_nG_n)^0=(G_n\tilde{G}_n)^0=I_{dM}$.*

Proof. We assume $n \geq m_0$. It is enough for us to show that both $\sum_{k=0}^{\infty}(\tilde{G}_n G_n)^k$ and $\sum_{k=0}^{\infty}(G_n\tilde{G}_n)^k$ converge. We see from Proposition 18 that, for $k \in \mathbb{N}$ and $v \in \mathbb{C}^{dM\times d}$,

$$(H_n\tilde{H}_n)^k\left(\sum_{\ell=0}^{\infty}(v^\top \mathbf{p}_\ell)\varepsilon_\ell\right) = \sum_{j=0}^{\infty}(v^\top \Lambda(\tilde{G}_n G_n)^{k-1}\tilde{G}_n \Pi_n \Theta \mathbf{p}_j)\varepsilon_j,$$

hence, for $k \in \mathbb{N}$ and $u, v \in \mathbb{C}^{dM\times d}$,

$$\left\langle (H_n\tilde{H}_n)^k\left(\sum_{\ell=0}^{\infty}(v^\top \mathbf{p}_\ell)\varepsilon_\ell\right), \sum_{j=0}^{\infty}(u^\top \mathbf{p}_j)\varepsilon_j\right\rangle = v^\top \Lambda(\tilde{G}_n G_n)^{k-1}\tilde{G}_n \Pi_n \Theta\left(\sum_{j=0}^{\infty}\mathbf{p}_j\mathbf{p}_j^*\right)\overline{u} = v^\top \Lambda(\tilde{G}_n G_n)^k \overline{u},$$

and similarly for $k = 0$. Since $(H_n\tilde{H}_n)^k x = ((\mathcal{H}_n\tilde{\mathcal{H}}_n)^k x^1, \ldots, (\mathcal{H}_n\tilde{\mathcal{H}}_n)^k x^d)^\top$ for $x = (x^1, \ldots, x^d)^\top \in \{(V^X_{(-\infty,-1]})^\perp\}^d$, it follows from Proposition 16 that

$$\sum_{k=0}^{N} v^\top \Lambda(\tilde{G}_n G_n)^k \overline{u} = \left\langle \sum_{k=0}^{N}(H_n\tilde{H}_n)^k\left(\sum_{\ell=0}^{\infty}(v^\top \mathbf{p}_\ell)\varepsilon_\ell\right), \sum_{j=0}^{\infty}(u^\top \mathbf{p}_j)\varepsilon_j\right\rangle$$

converges as $N \to \infty$, for any $u, v \in \mathbb{C}^{dM\times d}$. By choosing $u_i, v_i \in \mathbb{C}^{dM\times d}$ $(i = 1, \ldots, d)$ so that $(u_1, \ldots, u_d) = (v_1, \ldots, v_d) = I_{dM}$, we find that $\sum_{k=0}^{\infty}\Lambda(\tilde{G}_n G_n)^k$ converges. Since Λ is invertible by Proposition 15, $\sum_{k=0}^{\infty}(\tilde{G}_n G_n)^k$ also converges. Finally, from $\sum_{k=1}^{N}(G_n\tilde{G}_n)^k = G_n\left\{\sum_{k=0}^{N-1}(\tilde{G}_n G_n)^k\right\}\tilde{G}_n$ for $N \in \mathbb{N}$, $\sum_{k=0}^{\infty}(G_n\tilde{G}_n)^k$ converges, too. $\square$

For $n \in \mathbb{N}$ and $k \in \mathbb{N}\cup\{0\}$, the two sequences $\{b^k_{n,j}\}_{j=0}^{\infty} \in \ell_{2+}^{d\times d}$ and $\{\tilde{b}^k_{n,j}\}_{j=0}^{\infty} \in \ell_{2+}^{d\times d}$ are defined by the recursions

$$b^0_{n,j} = \delta_{0,j}I_d, \qquad b^{2k+1}_{n,j} = \sum_{\ell=0}^{\infty} b^{2k}_{n,\ell}\beta_{n+j+\ell+1}, \qquad b^{2k+2}_{n,j} = \sum_{\ell=0}^{\infty} b^{2k+1}_{n,\ell}\beta^*_{n+j+\ell+1}$$

and

$$\tilde{b}^0_{n,j} = \delta_{0,j}I_d, \qquad \tilde{b}^{2k+1}_{n,j} = \sum_{\ell=0}^{\infty} \tilde{b}^{2k}_{n,\ell}\beta^*_{n+j+\ell+1}, \qquad \tilde{b}^{2k+2}_{n,j} = \sum_{\ell=0}^{\infty} \tilde{b}^{2k+1}_{n,\ell}\beta_{n+j+\ell+1},$$

respectively (see Section 4 in [11]).

Lemma 20. *For $n \geq \max(m_0, 1)$, $k \in \mathbb{N}$ and $j \in \mathbb{N}\cup\{0\}$, we have*

$$b^{2k-1}_{n,j} = \mathbf{p}_0^\top(\tilde{G}_n G_n)^{k-1}(\Pi_n\Theta)^*\overline{\mathbf{p}}_j, \tag{D.6}$$
$$b^{2k}_{n,j} = \mathbf{p}_0^\top(\tilde{G}_n G_n)^{k-1}\tilde{G}_n\Pi_n\Theta\mathbf{p}_j, \tag{D.7}$$
$$\tilde{b}^{2k-1}_{n,j} = \mathbf{p}_0^\top(G_n\tilde{G}_n)^{k-1}\Pi_n\Theta\mathbf{p}_j, \tag{D.8}$$
$$\tilde{b}^{2k}_{n,j} = \mathbf{p}_0^\top(G_n\tilde{G}_n)^{k-1}G_n(\Pi_n\Theta)^*\overline{\mathbf{p}}_j. \tag{D.9}$$

Proof. We assume $n \geq \max(m_0, 1)$, and prove (D.6) and (D.7) by induction. First, from Lemma 17, $b^1_{n,j} = \beta_{n+j+1} = \mathbf{p}_0^\top(\Pi_n\Theta)^*\overline{\mathbf{p}}_j$. Next, for $k \in \mathbb{N}$, we assume (D.6). Then, by Lemma 17,

$$\begin{aligned}b^{2k}_{n,j} &= \sum_{\ell=0}^{\infty} b^{2k-1}_{n,\ell}\beta^*_{n+j+\ell+1} = \sum_{\ell=0}^{\infty}\mathbf{p}_0^\top(\tilde{G}_n G_n)^{k-1}(\Pi_n\Theta)^*\overline{\mathbf{p}}_\ell\mathbf{p}_\ell^\top\Pi_n\Theta\mathbf{p}_j = \mathbf{p}_0^\top(\tilde{G}_n G_n)^{k-1}(\Pi_n\Theta)^*\left(\sum_{\ell=0}^{\infty}\overline{\mathbf{p}}_\ell\mathbf{p}_\ell^\top\right)\Pi_n\Theta\mathbf{p}_j\\ &= \mathbf{p}_0^\top(\tilde{G}_n G_n)^{k-1}(\Pi_n\Theta)^*\Lambda^\top\Pi_n\Theta\mathbf{p}_j = \mathbf{p}_0^\top(\tilde{G}_n G_n)^{k-1}\tilde{G}_n\Pi_n\Theta\mathbf{p}_j\end{aligned}$$

or (D.7). From this as well as Lemma 17,

$$\begin{aligned}b^{2k+1}_{n,j} &= \sum_{\ell=0}^{\infty} b^{2k}_{n,\ell}\beta_{n+j+\ell+1} = \mathbf{p}_0^\top(\tilde{G}_n G_n)^{k-1}\tilde{G}_n\Pi_n\Theta\left(\sum_{\ell=0}^{\infty}\mathbf{p}_\ell\mathbf{p}_\ell^*\right)(\Pi_n\Theta)^*\overline{\mathbf{p}}_j = \mathbf{p}_0^\top(\tilde{G}_n G_n)^{k-1}\tilde{G}_n\Pi_n\Theta\Lambda(\Pi_n\Theta)^*\overline{\mathbf{p}}_j\\ &= \mathbf{p}_0^\top(\tilde{G}_n G_n)^{k}(\Pi_n\Theta)^*\overline{\mathbf{p}}_j\end{aligned}$$

or (D.6) with k replaced by $k+1$. Thus (D.6) and (D.7) follow. We can prove (D.8) and (D.9) by induction similarly; we omit the details. $\square$

We are now ready to prove Theorem 6.

Proof of Theorem 6. By Theorem 5.4 in [11], we have $\phi_{n,j} = \sum_{k=0}^{\infty}\{\phi^{2k}_{n,j} + \phi^{2k+1}_{n,n-j+1}\}$ for $n \in \mathbb{N}$, $j \in \{1, \ldots, n\}$, where $\phi^{2k}_{n,j} := c_0\sum_{\ell=0}^{\infty} b^{2k}_{n,\ell}a_{j+\ell}$ and $\phi^{2k+1}_{n,j} := c_0\sum_{\ell=0}^{\infty} b^{2k+1}_{n,\ell}\tilde{a}_{j+\ell}$ for $n \in \mathbb{N}$ and $k, j \in \mathbb{N}\cup\{0\}$. Since $b^0_{n,j} = \delta_{0,j}I_d$, we have $\phi^0_{n,j} = c_0 a_j$,

$\phi_{n,j} = c_0 a_j + \sum_{k=1}^{\infty}\{\phi_{n,j}^{2k} + \phi_{n,n-j+1}^{2k-1}\}$. By Lemma 20, we have, for $n \geq \max(m_0, 1)$, $k \in \mathbb{N}$ and $j \in \{1, \ldots, n\}$,

$$\phi_{n,j}^{2k} = c_0 \mathbf{p}_0^\top (\tilde{G}_n G_n)^{k-1} \tilde{G}_n \Pi_n \Theta v_j = c_0 \mathbf{p}_0^\top (\tilde{G}_n G_n)^{k-1} (\Pi_n \Theta)^* \Lambda^\top \Pi_n \Theta v_j,$$

$$\phi_{n,n-j+1}^{2k-1} = c_0 \mathbf{p}_0^\top (\tilde{G}_n G_n)^{k-1} (\Pi_n \Theta)^* \tilde{v}_{n-j+1}.$$

Therefore, thanks to Lemma 19, we obtain the theorem. □

References

[1] D.Z. Arov, H. Dym, *J*-Contractive Matrix Valued Functions and Related Topics, Cambridge University Press, 2008.
[2] G. Baxter, An asymptotic result for the finite predictor, Math. Scand. 10 (1962) 137–144.
[3] K.N. Berk, Consistent autoregressive spectral estimates, Ann. Statist. 2 (1974) 489–502.
[4] N.H. Bingham, Multivariate prediction and matrix Szegö theory, Probab. Surv. 9 (2012) 325–339.
[5] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, second ed., Springer, 1991.
[6] P. Bühlmann, Sieve bootstrap for time series, Bernoulli 3 (1997) 123–148.
[7] A. Inoue, Linear-time algorithms for block Toeplitz systems with rational symbols, 2019, in preparation.
[8] A. Inoue, Y. Kasahara, Explicit representation of finite predictor coefficients and its applications, Ann. Statist. 34 (2006) 973–993.
[9] A. Inoue, Y. Kasahara, Simple matrix representations of the orthogonal polynomials for a rational spectral density on the unit circle, J. Math. Anal. Appl. 464 (2018) 1366–1374.
[10] A. Inoue, Y. Kasahara, M. Pourahmadi, The intersection of past and future for multivariate stationary processes, Proc. Amer. Math. Soc. 144 (2016) 1779–1786.
[11] A. Inoue, Y. Kasahara, M. Pourahmadi, Baxter's inequality for finite predictor coefficients of multivariate long-memory stationary processes, Bernoulli 24 (2018) 1202–1232.
[12] V.E. Katsnelson, B. Kirstein, On the theory of matrix-valued functions belonging to the Smirnov class, in: H. Dym, B. Fritzsche, V. Katsnelson, B. Kirstein (Eds.), Topics in Interpolation Theory, Leipzig, 1994, in: Operator Theory: Advances and Applications, vol. 95, Birkhäuser, 1997, pp. 299–350.
[13] J.-P. Kreiss, E. Paparoditis, D.N. Politis, On the range of validity of the autoregressive sieve bootstrap, Ann. Statist. 39 (2011) 2103–2130.
[14] P. Masani, The prediction theory of multivariate stochastic processes. III, Acta Math. 104 (1960) 141–162.
[15] V.V. Peller, Hankel Operators and Their Applications, Springer, 2003.
[16] M. Pourahmadi, Foundations of Time Series Analysis and Prediction Theory, Wiley, 2001.
[17] S. Treil, A. Volberg, Wavelets and the angle between past and future, J. Funct. Anal. 143 (1997) 269–308.
[18] S. Treil, A. Volberg, Completely regular multivariate stationary processes and the Muckenhoupt condition, Pacific J. Math. 190 (1999) 361–382.
[19] Y. Xi, J. Xia, S. Cauley, V. Balakrishnan, Superfast and stable structured solvers for Toeplitz least squares via randomized sampling, SIAM J. Matrix Anal. Appl. 35 (2014) 44–72.