Linear Algebra and its Applications 432 (2010) 1975–1989

Contents lists available at ScienceDirect

Linear Algebra and its Applications

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

Tensor Sylvester matrices and the Fisher information matrix of VARMAX processes

André Klein [a,*], Peter Spreij [b]

[a] Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands
[b] Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 15 October 2008
Accepted 17 June 2009
Available online 12 July 2009

Submitted by B. Meini

AMS classification:
15A23
15A57
15A69
62B10

Keywords:
Multiple-tensor Sylvester matrix
Matrix polynomial
Fisher information matrix
VARMAX process

ABSTRACT

The purpose of this paper is to develop compact expressions for the Fisher information matrix (FIM) of a Gaussian stationary vector autoregressive and moving average process with exogenous or input variables, a vector ARMAX or VARMAX process. We develop a representation of the FIM based on multiple Sylvester matrices. An extension of this representation yields another one but in terms of tensor Sylvester matrices. In order to obtain the results presented in this paper, the approach used in [A. Klein, G. Mélard, P. Spreij, On the resultant property of the Fisher information matrix of a vector ARMA process, Linear Algebra Appl. 403 (2005) 291–313] is extended.

1. Introduction

The purpose of this paper is to develop compact representations of the Fisher information matrix of a Gaussian stationary vector autoregressive and moving average process with exogenous or input variables, a vector ARMAX or VARMAX process. These representations involve multiple and tensor Sylvester matrices. Especially the representation of the Fisher information matrix expressed in terms of

* Corresponding author. Tel.: +31 20 5254245; fax: +31 205254349.
E-mail address: A.A.B.Klein@uva.nl (A. Klein).

0024-3795/$ - see front matter
doi:10.1016/j.laa.2009.06.027

tensor Sylvester matrices leads to a better understanding of the underlying matrix structural properties of the Fisher information matrix. The latter representation shall be used for further study. In the next subsections the stationary process considered in this paper is introduced and the Fisher information matrix in this context is presented.

1.1. The VARMAX process

Consider the vector difference equation representation of a linear system ($\{y(t), t \in \mathbb{N}\}$, $\mathbb{N}$ the set of integers, of order (p, r, q)),

$$\sum_{j=0}^{p} \alpha_j y(t-j) = \sum_{j=0}^{r} \gamma_j x(t-j) + \sum_{j=0}^{q} \beta_j \varepsilon(t-j), \quad t \in \mathbb{N}, \tag{1}$$

where $y(t), x(t)$ and $\varepsilon(t)$ are the outputs, the observed inputs, and the errors, respectively, and where $\alpha_j \in \mathbb{R}^{n\times n}$, $\gamma_j \in \mathbb{R}^{n\times n}$, and $\beta_j \in \mathbb{R}^{n\times n}$ are the associate parameter matrices. The scalar version of process (1) is extensively studied in the system and control literature, where $x(t)$ assumes the role of a control variable, see e.g. [3,11,21]. In the statistical literature the process in (1) is extensively treated in [12].

We additionally have $\alpha_0 \equiv \beta_0 \equiv \gamma_0 \equiv I_n$, and starting the summation in the first sum of the right-hand side in (1) with 1 rather than with zero turns out to be more convenient and there is no loss in generality in the sense that $x(t)$ can always be redefined as $x(t+1)$. The error $\{\varepsilon(t),\ t \in \mathbb{N}\}$ is a collection of uncorrelated zero mean n-dimensional random variables each having positive definite covariance matrix Σ. We assume, for all s, t, $\mathbb{E}\{x(s)\varepsilon^\top(t)\} = 0$, where $\mathbb{E}$ is the expected value and $\top$ denotes the transposition.

We use L to denote the backward shift operator, for example $Lx(t) = x(t-1)$. Eq. (1) can be written as

$$\alpha(L)y(t) = \gamma(L)x(t) + \beta(L)\varepsilon(t), \tag{2}$$

and

$$\alpha(z) = \sum_{j=0}^{p} \alpha_j z^j; \quad \gamma(z) = \sum_{j=0}^{r} \gamma_j z^j; \quad \beta(z) = \sum_{j=0}^{q} \beta_j z^j. \tag{3}$$

The autoregressive matrix polynomial is given by $\alpha(z)$, the AR part, $\gamma(z)$ is the exogenous matrix polynomial, the X part and $\beta(z)$ is the moving average matrix polynomial, the MA part. Considering that matrix polynomials combined with vector processes are used, justifies the acronym VARMAX process. The assumption $\det(\alpha(z)) \neq 0$, $\det(\beta(z)) \neq 0$ and $\det(\gamma(z)) \neq 0$ for $|z| \leqslant 1$ or the determinants are different from zero in the closed unit disc will be imposed. Hence the zeros of the respective determinants, the eigenvalues, are outside the unit disc, so the elements of $\alpha^{-1}(z)$, $\beta^{-1}(z)$ and $\gamma^{-1}(z)$ can be written as power series in z with radius of convergence greater than 1.

Some more assumptions on the observed inputs $x(t)$ are given. The observed input variable $x(t)$ is assumed to be a stationary process with spectral density $R_x(\cdot)/2\pi$. If $x(t)$ is an n-dimensional VARMA process with $\eta(t)$ a white noise process satisfying $\mathbb{E}\{\eta(t)\eta^\top(t)\} = \Omega$,

$$a(L)x(t) = b(L)\eta(t), \tag{4}$$

then the spectral density of process $x(t)$ is

$$R_x(e^{i\omega}) = a^{-1}(e^{i\omega})b(e^{i\omega})\Omega b^*(e^{i\omega})a^{-*}(e^{i\omega}) \quad \omega \in [-\pi, \pi]. \tag{5}$$

We have $\mathbb{E}\{\varepsilon(t)\eta^\top(s)\} = 0$ for all s and t, the last property is a direct consequence of the fact that $x(t)$ and $\varepsilon(t)$ are orthogonal processes.

The parameter vector ϑ is defined by

$$\vartheta = \text{vec}\left\{\alpha_1, \alpha_2, \ldots, \alpha_p, \gamma_1, \gamma_2, \ldots, \gamma_r, \beta_1, \beta_2, \ldots, \beta_q\right\}, \tag{6}$$

where the ordering of the elements of the matrix polynomials $\alpha(z)$, $\beta(z)$ and $\gamma(z)$ into a vector ϑ is done according to (3). The vec operator transforms a matrix into a vector by stacking the columns of the matrix one underneath the other.

Estimation of the matrices $\alpha_1, \alpha_2, \ldots, \alpha_p, \gamma_1, \gamma_2, \ldots, \gamma_r, \beta_1, \beta_2, \ldots, \beta_q$ and Σ has received considerable attention in the time series and filtering theory literature [10,11]. In [12], the asymptotic properties of maximum likelihood estimates of the coefficients of VARMAX processes, stored in a $(\ell \times 1)$ vector ϑ, where $\ell = n^2(p+q+r)$ have been studied.

1.2. The Fisher information matrix

The Fisher information matrix prominently features in the asymptotic analysis of estimators. It is linked to the Cramér-Rao bound on the covariance of unbiased estimators, see e.g. [13] for general results and [1] for time series processes. Under mild assumptions but assuming that the estimators are asymptotically unbiased, the inverse of the asymptotic Fisher information matrix yields this bound, and provided that the estimators are asymptotically efficient, it equals the asymptotic covariance matrix. The inversion of the Fisher information matrix is thus of basic importance. In [22], an algorithm for the computation of asymptotic Fisher information matrix of a VARMA process is developed. It is based on a frequency domain representation of the Fisher information matrix, known as Whittle's formula, see [23]. In the pioneering paper [23], a scalar-level formula is developed for the asymptotic Fisher information matrix of a VARMA process, a stationary process that does not involve the input process $x(t)$ as in (2). In [16,19], the equivalence between a time and frequency domain representation of the asymptotic Fisher information matrix of VARMA-VARMAX processes has been established. The Fisher information matrix of a scalar version of (2) is described in [15]. The Fisher information matrix has also attracted much attention in the signal processing literature, see e.g. [7] and more recently in physics, see e.g. [5,6].

When the representation of the parameter vector ϑ as defined above is considered, the following expression may be taken as the definition of the $n^2(p+q+r) \times n^2(p+q+r)$ asymptotic Fisher information matrix of a VARMAX process

$$\mathcal{F}(\vartheta) = \mathbb{E}_\vartheta \left\{ \left(\frac{\partial \varepsilon}{\partial \vartheta} \right)^\top \Sigma^{-1} \left(\frac{\partial \varepsilon}{\partial \vartheta} \right) \right\}, \tag{7}$$

where $\mathbb{E}_\vartheta$ is the expected value under the parameter ϑ. A proof of an equivalent to (7) for a VARMA process is given in [18].

The remainder of the paper is organized as follows: Differentiation of the error of the VARMAX process is described in Section 2.1. The method of differentiation applied in [17] is used. A convenient form for the derivative $\partial\varepsilon/\partial\vartheta$ is constructed in order to obtain appropriate representations of the Fisher information matrix. For that purpose we proceed in three stages. First an integral representation is constructed for the Fisher information matrix $\mathcal{F}(\vartheta)$ in (7). This is done in Sections 2.2 and 2.3. Second, the integral representation displayed in Section 2.3 is further developed and involves multiple Sylvester matrices. This is done in Section 2.4. Third, the results of Section 2.4 are used to construct a representation for the Fisher information matrix involving tensor Sylvester matrices. This is done in Section 2.5. In Section 2.6 an expression for the inverse of the matrix polynomials in terms of the corresponding parameters is given. In Section 2.7 a representation is derived from the Fisher information matrix, as a direct consequence of the results in Section 2.5.

2. Compact representations of the Fisher information matrix

In this section a representation of the Fisher information matrix $\mathcal{F}(\vartheta)$ expressed in terms of tensor Sylvester matrices is developed. For this purpose several steps set forth in Sections 2.1–2.5 are considered. We use a partitioned form of the Fisher information matrix, composed by the submatrices associated with the parameters $\alpha_1, \alpha_2, \ldots, \alpha_p, \gamma_1, \gamma_2, \ldots, \gamma_r, \beta_1, \beta_2, \ldots, \beta_q$, which is given by

$$\mathcal{F}(\vartheta) = \begin{pmatrix} \mathcal{F}_{\alpha\alpha}(\vartheta) & \mathcal{F}_{\alpha\gamma}(\vartheta) & \mathcal{F}_{\alpha\beta}(\vartheta) \\ \mathcal{F}_{\gamma\alpha}(\vartheta) & \mathcal{F}_{\gamma\gamma}(\vartheta) & \mathcal{F}_{\gamma\beta}(\vartheta) \\ \mathcal{F}_{\beta\alpha}(\vartheta) & \mathcal{F}_{\beta\gamma}(\vartheta) & \mathcal{F}_{\beta\beta}(\vartheta) \end{pmatrix}.$$

In a dynamic stationary stochastic context it has long been shown useful to use Fourier transform representations, or alternatively, circular integral representations, also called z-transform, to obtain

$$\mathcal{F}(\vartheta) = \frac{1}{2\pi i} \oint_{|z|=1} \begin{pmatrix} \mathcal{J}_{\alpha\alpha}(z) & \mathcal{J}_{\alpha\gamma}(z) & \mathcal{J}_{\alpha\beta}(z) \\ \mathcal{J}_{\gamma\alpha}(z) & \mathcal{J}_{\gamma\gamma}(z) & \mathcal{J}_{\gamma\beta}(z) \\ \mathcal{J}_{\beta\alpha}(z) & \mathcal{J}_{\beta\gamma}(z) & \mathcal{J}_{\beta\beta}(z) \end{pmatrix} \frac{dz}{z}. \tag{8}$$

The integration in (8) and elsewhere in the paper is counterclockwise around the unit circle. We then derive a compact representation of the Fisher information matrix of the form

$$\mathcal{F}(\vartheta) = \frac{1}{2\pi i} \oint_{|z|=1} \mathcal{J}(z) \frac{dz}{z}. \tag{9}$$

The integrand $\mathcal{J}(z)$ is Hermitian and shall eventually involve multiple and tensor Sylvester matrices. In order to transform expression (7) of $\mathcal{F}(\vartheta)$ into representation (9), appropriate matrix differential rules are first to be applied to the error process $\varepsilon(t)$. This is done in the following section where a representation of $\partial\varepsilon/\partial\vartheta$ is constructed. We apply the method used in [17].

2.1. Differentiation of the error process

In [19] differentiation is applied to the different parameter blocks of the Fisher information matrix whereas in this paper a global approach is considered. From (2) it can be seen that

$$\varepsilon(t) = \beta^{-1}(L)\alpha(L)y(t) - \beta^{-1}(L)\gamma(L)x(t).$$

Differentiation will be applied to this form of $\varepsilon(t)$. The following facts are used. If d represents any differential operator involving partial derivatives w.r.t. ϑ, then $dy(t) = 0$ and $dx(t) = 0$. These equalities hold because the realizations of $y(t)$ and $x(t)$ are independent of the parameters. For typographical brevity we omit the argument t, and write

$$\begin{aligned} d\varepsilon = {} & \beta^{-1}(L)d\alpha(L)\alpha^{-1}(L)\gamma(L)x - \beta^{-1}(L)d\gamma(L)x + \beta^{-1}(L)d\alpha(L)\alpha^{-1}(L)\beta(L)\varepsilon \\ & - \beta^{-1}(L)d\beta(L)\varepsilon. \end{aligned} \tag{10}$$

The rule

$$\mathrm{vec}\,(ABC) = \left(C^\top \otimes A\right) \mathrm{vec}\, B \quad \text{where } A \in \mathbb{R}^{m\times n},\ B \in \mathbb{R}^{n\times p} \text{ and } C \in \mathbb{R}^{p\times s},$$

is applied to (10), to obtain

$$\begin{aligned} d\varepsilon = {} & \left\{ \left(\alpha^{-1}(L)\gamma(L)x\right)^\top \otimes \beta^{-1}(L) \right\} \mathrm{vec}\, d\alpha(L) \\ & + \left\{ \left(\alpha^{-1}(L)\beta(L)\varepsilon\right)^\top \otimes \beta^{-1}(L) \right\} \mathrm{vec}\, d\alpha(L) \\ & - \left\{ x^\top \otimes \beta^{-1}(L) \right\} \mathrm{vec}\, d\gamma(L) \\ & - \left\{ \varepsilon^\top \otimes \beta^{-1}(L) \right\} \mathrm{vec}\, d\beta(L), \end{aligned}$$

where $\otimes$ denotes the Kronecker product.

An appropriate expression for the differentiation of the noise process $\partial\varepsilon/\partial\vartheta$ is then

$$\begin{aligned} \frac{\partial\varepsilon}{\partial\vartheta} = {} & \left\{ \left(\alpha^{-1}(L)\gamma(L)x\right)^\top \otimes \beta^{-1}(L) \right\} \frac{\partial \mathrm{vec}\, \alpha(L)}{\partial\vartheta} \\ & + \left\{ \left(\alpha^{-1}(L)\beta(L)\varepsilon\right)^\top \otimes \beta^{-1}(L) \right\} \frac{\partial \mathrm{vec}\, \alpha(L)}{\partial\vartheta} \end{aligned}$$

$$- \left\{ x^{\top} \otimes \beta^{-1}(L) \right\} \frac{\partial \mathrm{vec}\, \gamma(L)}{\partial \vartheta}$$
$$- \left(\varepsilon^{\top} \otimes \beta^{-1}(L) \right) \frac{\partial \mathrm{vec}\, \beta(L)}{\partial \vartheta}. \tag{11}$$

2.2. Representation with reordered factors and an integral representation

In the present and following sections, we use an approach similar to the method applied in [17]. However, the construction presented in the present paper is mainly concerned with the input part of process (2) that is not present in a VARMA context.

Substitution of (11) into (7) shall allow us to develop appropriate forms for the Fisher information matrix. For that purpose a useful equality is introduced. Consider the discrete-time stationary process $w(t)$ where $w(t) = H(L)u(t)$, $u(t)$ is the input process and $H(L)$ is an asymptotically stable filter. We apply Herglotz's theorem to express the covariance function of a stationary process $w(t)$ in terms of the spectral distribution function of $w(t)$, see e.g. [2,3]. For evaluating the covariance matrix of the output $w(t)$, we have the equality

$$\mathbb{E}_{\vartheta} \left\{ w(t) w^{\top}(t) \right\} = \int_{-\pi}^{\pi} \phi_w(\omega) d\omega, \quad \omega \in [-\pi, \pi] \tag{12}$$

where $\phi_w(\omega)$ is the spectral density of the proceses $w(t)$. It is defined as

$$\phi_w(\omega) = H(e^{i\omega}) \phi_u(\omega) H^*(e^{i\omega}). \tag{13}$$

Expression (13) is a Hermitian matrix, and $\phi_u(\omega)$ is the spectral density of the stationary process $u(t)$. Here Y^* denotes the complex conjugate transpose of the matrix Y. In order to use Herglotz's theorem given in equality (12), we rearrange the elements of the right-hand side of (7) so that a representation of the form $w(t)w^{\top}(t)$ is obtained. For that purpose the rule

$$(A_1 \otimes B_1)(A_2 \otimes B_2) \cdots (A_m \otimes B_m) = (A_1 A_2 \cdots A_m) \otimes (B_1 B_2 \ldots B_m) \tag{14}$$

is applied, for matrices $A_1, A_2, \ldots, A_m$ and $B_1, B_2, \ldots, B_m$ of appropriate dimensions, see e.g. [20]. We shall consider the first term in (11), which we denote by $(\partial \varepsilon / \partial \vartheta)_1$, to illustate the method used. The following representation is set forth

$$\begin{aligned}
\left(\frac{\partial \varepsilon}{\partial \vartheta} \right)_1^{\top} \Sigma^{-1} \left(\frac{\partial \varepsilon}{\partial \vartheta} \right)_1 &= \left(\frac{\partial \mathrm{vec}\, \alpha(L)}{\partial \vartheta} \right)^{\top} \left\{ \alpha^{-1}(L)\gamma(L)x \otimes \beta^{-\top}(L) \right\} \Sigma^{-1} \\
&\quad \times \left\{ \left(\alpha^{-1}(L)\gamma(L)x \right)^{\top} \otimes \beta^{-1}(L) \right\} \frac{\partial \mathrm{vec}\, \alpha(L)}{\partial \vartheta} \\
&= \left(\frac{\partial \mathrm{vec}\, \alpha(L)}{\partial \vartheta} \right)^{\top} \left\{ \alpha^{-1}(L)\gamma(L) \otimes \beta^{-\top}(L) \right\} (x \otimes I_n) \Sigma^{-1} \\
&\quad \times (x \otimes I_n)^{\top} \left\{ \left(\alpha^{-1}(L)\gamma(L) \right)^{\top} \otimes \beta^{-1}(L) \right\} \frac{\partial \mathrm{vec}\, \alpha(L)}{\partial \vartheta} \\
&= \left(\frac{\partial \mathrm{vec}\, \alpha(L)}{\partial \vartheta} \right)^{\top} \left\{ \alpha^{-1}(L)\gamma(L) \otimes \beta^{-\top}(L) \right\} (x \otimes \Sigma^{-1}) \\
&\quad \times (x \otimes I_n)^{\top} \left\{ \left(\alpha^{-1}(L)\gamma(L) \right)^{\top} \otimes \beta^{-1}(L) \right\} \frac{\partial \mathrm{vec}\, \alpha(L)}{\partial \vartheta}.
\end{aligned}$$

Our objective is to obtain a symmetric expression. We therefore apply a Cholesky factorization to Σ^{-1}, since the covariance matrix Σ is positive definite. Consequently, there is a unique lower triangular matrix Γ with positive diagonal entries such that $\Sigma^{-1} = \Gamma \Gamma^{\top}$. This yields the expression

$$\mathbb{E}_\vartheta \left\{ \left(\frac{\partial \varepsilon}{\partial \vartheta}\right)_1^\top \Sigma^{-1} \left(\frac{\partial \varepsilon}{\partial \vartheta}\right)_1 \right\}$$

$$= \mathbb{E}_\vartheta \left\{ \left(\frac{\partial \operatorname{vec} \alpha(L)}{\partial \vartheta}\right)^\top \left\{\alpha^{-1}(L)\gamma(L) \otimes \beta^{-\top}(L)\right\} (x \otimes \Gamma)\right.$$

$$\left. \times\ (x \otimes \Gamma)^\top \left\{\left(\alpha^{-1}(L)\gamma(L)\right)^\top \otimes \beta^{-1}(L)\right\} \frac{\partial \operatorname{vec} \alpha(L)}{\partial \vartheta} \right\}. \tag{15}$$

From (15) it can be seen that equality (12) can be used by setting

$$w(t) = \left(\frac{\partial \operatorname{vec} \alpha(L)}{\partial \vartheta}\right)^\top \left\{\alpha^{-1}(L)\gamma(L) \otimes \beta^{-\top}(L)\right\} (x \otimes \Gamma).$$

The spectral density of $(x \otimes \Gamma)$ is considered for obtaining an explicit expression of the covariance matrix. Since $\mathbb{E}_\vartheta x = 0$, this covariance matrix equals

$$\mathbb{E}_\vartheta\ (x \otimes \Gamma)\,(x \otimes \Gamma)^\top = \mathbb{E}_\vartheta \left(x\,x^\top \otimes \Gamma\Gamma^\top\right) = R_x(e^{i\omega}) \otimes \Sigma^{-1}.$$

It is straightforward to conclude that in view of (12) the values of the spectral density of $(x \otimes \Gamma)$ are $(1/2\pi)(R_x(e^{i\omega}) \otimes \Sigma^{-1})$, where $R_x(e^{i\omega})$ is defined in (5). By virtue of (12), expression (15) becomes

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\frac{\partial \operatorname{vec} \alpha(e^{i\omega})}{\partial \vartheta}\right)^\top \left\{\alpha^{-1}(e^{i\omega})\gamma(e^{i\omega}) \otimes \beta^{-\top}(e^{i\omega})\right\} \left(R_x(e^{i\omega}) \otimes \Sigma^{-1}\right)$$

$$\times \left\{\left(\alpha^{-1}(e^{-i\omega})\gamma(e^{-i\omega})\right)^\top \otimes \beta^{-1}(e^{-i\omega})\right\} \frac{\partial \operatorname{vec} \alpha(e^{-i\omega})}{\partial \vartheta}\, d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\frac{\partial \operatorname{vec} \alpha(e^{i\omega})}{\partial \vartheta}\right)^\top \left\{\alpha^{-1}(e^{i\omega})\gamma(e^{i\omega})R_x(e^{i\omega}) \otimes \beta^{-\top}(e^{i\omega})\Sigma^{-1}\right\}$$

$$\times \left\{\left(\alpha^{-1}(e^{-i\omega})\gamma(e^{-i\omega})\right)^\top \otimes \beta^{-1}(e^{-i\omega})\right\} \frac{\partial \operatorname{vec} \alpha(e^{-i\omega})}{\partial \vartheta}\, d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\frac{\partial \operatorname{vec} \alpha(e^{i\omega})}{\partial \vartheta}\right)^\top \left\{\alpha^{-1}(e^{i\omega})\gamma(e^{i\omega})R_x(e^{i\omega}) \left(\alpha^{-1}(e^{-i\omega})\gamma(e^{-i\omega})\right)^\top\right.$$

$$\left. \otimes\ \beta^{-\top}(e^{i\omega})\Sigma^{-1}\beta^{-1}(e^{-i\omega})\right\} \frac{\partial \operatorname{vec} \alpha(e^{-i\omega})}{\partial \vartheta}\, d\omega.$$

Equivalently for $z = e^{i\omega}$ we have

$$\mathbb{E}_\vartheta \left\{ \left(\frac{\partial \varepsilon}{\partial \vartheta}\right)_1^\top \Sigma^{-1} \left(\frac{\partial \varepsilon}{\partial \vartheta}\right)_1 \right\}$$

$$= \frac{1}{2\pi i}\oint_{|z|=1} \left(\frac{\partial \operatorname{vec} \alpha(z)}{\partial \vartheta}\right)^\top \left\{\alpha^{-1}(z)\gamma(z)R_x(z) \left(\alpha^{-1}(z^{-1})\gamma(z^{-1})\right)^\top\right.$$

$$\left. \otimes\ \beta^{-\top}(z)\Sigma^{-1}\beta^{-1}(z^{-1})\right\} \frac{\partial \operatorname{vec} \alpha(z^{-1})}{\partial \vartheta}\,\frac{dz}{z}. \tag{16}$$

A similar approach is applied to the remaining terms of representation (11). We denote the integrand of (16) by $\mathcal{J}^{(1)}(z)$. Taking into account the fact that $\mathbb{E}\{\varepsilon(t)\eta^\top(s)\} = 0$ for all s and t, we obtain for the integral representation given in (8)

$$\mathcal{F}(\theta) = \mathbb{E}_\vartheta \left\{ \left(\frac{\partial \varepsilon}{\partial \vartheta}\right)^\top \Sigma^{-1} \left(\frac{\partial \varepsilon}{\partial \vartheta}\right) \right\} = \frac{1}{2\pi i} \oint_{|z|=1} \sum_{j=1}^{8} \mathcal{J}^{(j)}(z) \frac{dz}{z}, \tag{17}$$

where

$$\mathcal{J}^{(2)}(z) = -\left(\frac{\partial \mathrm{vec}\,\alpha(z)}{\partial \vartheta}\right)^\top \left\{\alpha^{-1}(z)\gamma(z)R_x(z) \otimes \sigma(z)\right\} \left(\frac{\partial \mathrm{vec}\,\gamma(z^{-1})}{\partial \vartheta}\right),$$

$$\mathcal{J}^{(3)}(z) = \left(\frac{\partial \mathrm{vec}\,\alpha(z)}{\partial \vartheta}\right)^\top \left\{\alpha^{-1}(z)\beta(z)\Sigma\left(\alpha^{-1}(z^{-1})\beta(z^{-1})\right)^\top \otimes \sigma(z)\right\} \left(\frac{\partial \mathrm{vec}\,\alpha(z^{-1})}{\partial \vartheta}\right),$$

$$\mathcal{J}^{(4)}(z) = -\left(\frac{\partial \mathrm{vec}\,\alpha(z)}{\partial \vartheta}\right)^\top \left\{\alpha^{-1}(z)\beta(z)\Sigma \otimes \sigma(z)\right\} \left(\frac{\partial \mathrm{vec}\,\beta(z^{-1})}{\partial \vartheta}\right),$$

$$\mathcal{J}^{(5)}(z) = -\left(\frac{\partial \mathrm{vec}\,\gamma(z)}{\partial \vartheta}\right)^\top \left\{R_x(z)\left(\alpha^{-1}(z^{-1})\gamma(z^{-1})\right)^\top \otimes \sigma(z)\right\} \left(\frac{\partial \mathrm{vec}\,\alpha(z^{-1})}{\partial \vartheta}\right),$$

$$\mathcal{J}^{(6)}(z) = \left(\frac{\partial \mathrm{vec}\,\gamma(z)}{\partial \vartheta}\right)^\top \left\{R_x(z) \otimes \sigma(z)\right\} \left(\frac{\partial \mathrm{vec}\,\gamma(z^{-1})}{\partial \vartheta}\right),$$

$$\mathcal{J}^{(7)}(z) = -\left(\frac{\partial \mathrm{vec}\,\beta(z)}{\partial \vartheta}\right)^\top \left\{\Sigma\left(\alpha^{-1}(z^{-1})\beta(z^{-1})\right)^\top \otimes \sigma(z)\right\} \left(\frac{\partial \mathrm{vec}\,\alpha(z^{-1})}{\partial \vartheta}\right),$$

$$\mathcal{J}^{(8)}(z) = \left(\frac{\partial \mathrm{vec}\,\beta(z)}{\partial \vartheta}\right)^\top \left\{\Sigma \otimes \sigma(z)\right\} \left(\frac{\partial \mathrm{vec}\,\beta(z^{-1})}{\partial \vartheta}\right),$$

where $\sigma(z) = \beta^{-\top}(z)\Sigma^{-1}\beta^{-1}(z^{-1})$. The representation of the parameter vector ϑ as displayed in (6) yields

$$\frac{\partial \mathrm{vec}\,\alpha(z)}{\partial \vartheta^\top} = z\left\{u_p^\top(z) \otimes I_{n^2}, 0_r^\top \otimes I_{n^2}, 0_q^\top \otimes I_{n^2}\right\}, \tag{18}$$

$$\frac{\partial \mathrm{vec}\,\gamma(z)}{\partial \vartheta^\top} = z\left\{0_p^\top \otimes I_{n^2}, u_r^\top(z) \otimes I_{n^2}, 0_q^\top \otimes I_{n^2}\right\}, \tag{19}$$

$$\frac{\partial \mathrm{vec}\,\beta(z)}{\partial \vartheta^\top} = z\left\{0_p^\top \otimes I_{n^2}, 0_r^\top \otimes I_{n^2}, u_q^\top(z) \otimes I_{n^2}\right\}, \tag{20}$$

where $u_x^\top(z) = (1, z, z^2, \ldots, z^{x-1})$.

2.3. Second integral representation

To analyze (17) further it requires an additional matrix property. If a matrix A is decomposed as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

then the Kronecker product $A \otimes B$ takes the form

$$A \otimes B = \begin{pmatrix} A_{11} \otimes B & A_{12} \otimes B \\ A_{21} \otimes B & A_{22} \otimes B \end{pmatrix}. \tag{21}$$

Combining (18), (19), (20) and (21) yields the following formulas for the elements of (17)

$$\mathcal{J}^{(1)}(z)=\begin{pmatrix} u_p(z)u_p^\top(z^{-1}) & 0_{p\times r} & 0_{p\times q}\\ 0_{r\times p} & 0_{r\times r} & 0_{r\times q}\\ 0_{q\times p} & 0_{q\times r} & 0_{q\times q}\end{pmatrix} \otimes \left\{\alpha^{-1}(z)\gamma(z)R_x(z)\left(\alpha^{-1}(z^{-1})\gamma(z^{-1})\right)^\top \otimes \sigma(z)\right\},$$

$$\mathcal{J}^{(2)}(z)=-\begin{pmatrix} 0_{p\times p} & u_p(z)u_r^\top(z^{-1}) & 0_{p\times q}\\ 0_{r\times p} & 0_{r\times r} & 0_{r\times q}\\ 0_{q\times p} & 0_{q\times r} & 0_{q\times q}\end{pmatrix} \otimes \left\{\alpha^{-1}(z)\gamma(z)R_x(z)\otimes \sigma(z)\right\},$$

$$\mathcal{J}^{(3)}(z)=\begin{pmatrix} u_p(z)u_p^\top(z^{-1}) & 0_{p\times r} & 0_{p\times q}\\ 0_{r\times p} & 0_{r\times r} & 0_{r\times q}\\ 0_{q\times p} & 0_{q\times r} & 0_{q\times q}\end{pmatrix} \otimes \left\{\alpha^{-1}(z)\beta(z)\Sigma\left(\alpha^{-1}(z^{-1})\beta(z^{-1})\right)^\top \otimes \sigma(z)\right\},$$

$$\mathcal{J}^{(4)}(z)=-\begin{pmatrix} 0_{p\times p} & 0_{p\times r} & u_p(z)u_q^\top(z^{-1})\\ 0_{r\times p} & 0_{r\times r} & 0_{r\times q}\\ 0_{q\times p} & 0_{q\times r} & 0_{q\times q}\end{pmatrix} \otimes \left\{\alpha^{-1}(z)\beta(z)\Sigma\otimes \sigma(z)\right\},$$

$$\mathcal{J}^{(5)}(z)=-\begin{pmatrix} 0_{p\times p} & 0_{p\times r} & 0_{p\times q}\\ u_r(z)u_p^\top(z^{-1}) & 0_{r\times r} & 0_{r\times q}\\ 0_{q\times p} & 0_{q\times r} & 0_{q\times q}\end{pmatrix} \otimes \left\{R_x(z)\left(\alpha^{-1}(z^{-1})\gamma(z^{-1})\right)^\top \otimes \sigma(z)\right\},$$

$$\mathcal{J}^{(6)}(z)=\begin{pmatrix} 0_{p\times p} & 0_{p\times r} & 0_{p\times q}\\ 0_{r\times p} & u_r(z)u_r^\top(z^{-1}) & 0_{r\times q}\\ 0_{q\times p} & 0_{q\times r} & 0_{q\times q}\end{pmatrix} \otimes \left\{R_x(z)\otimes \sigma(z)\right\},$$

$$\mathcal{J}^{(7)}(z)=-\begin{pmatrix} 0_{p\times p} & 0_{p\times r} & 0_{p\times q}\\ 0_{r\times p} & 0_{r\times r} & 0_{r\times q}\\ u_q(z)u_p^\top(z^{-1}) & 0_{q\times r} & 0_{q\times q}\end{pmatrix} \otimes \left\{\Sigma\left(\alpha^{-1}(z^{-1})\beta(z^{-1})\right)^\top \otimes \sigma(z)\right\},$$

$$\mathcal{J}^{(8)}(z)=\begin{pmatrix} 0_{p\times p} & 0_{p\times r} & 0_{p\times q}\\ 0_{r\times p} & 0_{r\times r} & 0_{r\times q}\\ 0_{q\times p} & 0_{q\times r} & u_q(z)u_q^\top(z^{-1})\end{pmatrix} \otimes \left\{\Sigma\otimes \sigma(z)\right\}.$$

The submatrix $\mathcal{J}_{\alpha\alpha}(z)$ in (8) is equal to the sum of the non-zero blocks of $\mathcal{J}^{(1)}(z)$ and $\mathcal{J}^{(3)}(z)$. Similarly, the submatrices $\mathcal{J}_{\alpha\gamma}(z)$, $\mathcal{J}_{\alpha\beta}(z)$, $\mathcal{J}_{\gamma\alpha}(z)$, $\mathcal{J}_{\gamma\gamma}(z)$, $\mathcal{J}_{\beta\alpha}(z)$ and $\mathcal{J}_{\beta\beta}(z)$ are equal to the non-zero blocks of $\mathcal{J}^{(2)}(z)$, $\mathcal{J}^{(4)}(z)$, $\mathcal{J}^{(5)}(z)$, $\mathcal{J}^{(6)}(z)$, $\mathcal{J}^{(7)}(z)$ and $\mathcal{J}^{(8)}(z)$ respectively. Since the input process $x(t)$ and the noise process $\varepsilon(t)$ are orthogonal, we have $\mathcal{F}_{\gamma\beta}(\vartheta)=\mathcal{F}_{\beta\gamma}(\vartheta)=0$. Inserting representations $\mathcal{J}^{(1)}(z)$ through $\mathcal{J}^{(8)}(z)$ in (17) yields a compact expression for the Fisher information matrix which is summarized in Proposition 2.1.

Proposition 2.1. *The following integral expression for the Fisher information matrix of a VARMAX process holds true*

$$\mathcal{F}(\theta)=\frac{1}{2\pi i}\oint_{|z|=1}\left(\mathcal{P}(z)\otimes\sigma(z)\right)\frac{dz}{z}+\frac{1}{2\pi i}\oint_{|z|=1}\left(\mathcal{Q}(z)\otimes\sigma(z)\right)\frac{dz}{z}, \tag{22}$$

where the $(p+r+q)n\times(p+r+q)n$ matrices $\mathcal{P}(z)$ and $\mathcal{Q}(z)$ are

$$\mathcal{P}(z)=\mathcal{G}(z)\Sigma\mathcal{G}(z)^*$$

and

$$\mathcal{Q}(z) = \mathcal{K}(z)R_x(z)\mathcal{K}(z)^*,$$

where

$$\mathcal{G}(z) = \begin{pmatrix} u_p(z) \otimes \alpha^{-1}(z)\,(-\beta(z)) \\ 0_{rn\times n} \\ u_q(z) \otimes I_n \end{pmatrix} \quad \text{and} \quad \mathcal{K}(z) = \begin{pmatrix} u_p(z) \otimes \alpha^{-1}(z)\,(-\gamma(z)) \\ u_r(z) \otimes I_n \\ 0_{qn\times n} \end{pmatrix}.$$

2.4. Representation based on multiple Sylvester matrices

The representations developed in the previous section are such that a multiple Sylvester matrix can be used to rewrite (22). For that purpose we apply a convenient factorization to the matrix polynomials $\mathcal{G}(z)$ and $\mathcal{K}(z)$, to obtain

$$\mathcal{G}(z) = \begin{pmatrix} I_p \otimes \alpha^{-1}(z) & 0_{pn\times rn} & 0_{pn\times qn} \\ 0_{rn\times pn} & 0_{rn\times rn} & 0_{rn\times qn} \\ 0_{qn\times pn} & 0_{qn\times rn} & I_q \otimes \alpha^{-1}(z) \end{pmatrix} \begin{pmatrix} u_p(z) \otimes (-\beta(z)) \\ 0_{rn\times n} \\ u_q(z) \otimes \alpha(z) \end{pmatrix} \tag{23}$$

and

$$\mathcal{K}(z) = \begin{pmatrix} I_p \otimes \alpha^{-1}(z) & 0_{pn\times rn} & 0_{pn\times qn} \\ 0_{rn\times pn} & I_r \otimes \alpha^{-1}(z) & 0_{rn\times qn} \\ 0_{qn\times pn} & 0_{qn\times rn} & 0_{qn\times qn} \end{pmatrix} \begin{pmatrix} u_p(z) \otimes (-\gamma(z)) \\ u_r(z) \otimes \alpha(z) \\ 0_{qn\times n} \end{pmatrix}. \tag{24}$$

We now introduce the $n(p+q) \times n(p+q)$ multiple Sylvester matrix involving the coefficients of the matrix polynomials $\alpha(z)$ and $\beta(z)$. It is given by

$$\mathcal{S}(-\beta,\alpha) = \begin{pmatrix} -I_n & -\beta_1 & \cdots & -\beta_q & 0_{n\times n} & \cdots & 0_{n\times n} \\ 0_{n\times n} & \ddots & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \ddots & 0_{n\times n} \\ 0_{n\times n} & \cdots & 0_{n\times n} & -I_n & -\beta_1 & \cdots & -\beta_q \\ I_n & \alpha_1 & \cdots & \alpha_p & 0_{n\times n} & \cdots & 0_{n\times n} \\ 0_{n\times n} & \ddots & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \ddots & 0_{n\times n} \\ 0_{n\times n} & \cdots & 0_{n\times n} & I_n & \alpha_1 & \cdots & \alpha_p \end{pmatrix}.$$

Consider the $pn \times (p+q)n$ and $qn \times (p+q)n$ upper and lower submatrices $\mathcal{S}_p(-\beta)$ and $\mathcal{S}_q(\alpha)$ of the multiple Sylvester matrix $\mathcal{S}(-\beta,\alpha)$ such that

$$\mathcal{S}(-\beta,\alpha) = \begin{pmatrix} \mathcal{S}_p(-\beta) \\ \mathcal{S}_q(\alpha) \end{pmatrix}.$$

It is straightforward to verify that the following equalities hold true,

$$\mathcal{S}_p(-\beta)\left(u_{p+q}(z) \otimes I_n\right) = u_p(z) \otimes (-\beta(z)) \tag{25}$$

and

$$\mathcal{S}_q(\alpha)\left(u_{p+q}(z) \otimes I_n\right) = u_q(z) \otimes \alpha(z). \tag{26}$$

Similarly for the submatrices $\mathcal{S}_p(-\gamma)$ and $\mathcal{S}_r(\alpha)$ of the $(p+r)n \times (p+r)n$ multiple Sylvester matrix $\mathcal{S}(-\gamma,\alpha)$, we have

$$\mathcal{S}_p(-\gamma)\left(u_{p+r}(z) \otimes I_n\right) = u_p(z) \otimes (-\gamma(z)) \tag{27}$$

and

$$\mathcal{S}_r(\alpha)\left(u_{p+r}(z)\otimes I_n\right)=u_r(z)\otimes\alpha(z). \tag{28}$$

Insertion of (25) and (26) in (23) and (27) and (28) in (24) respectively yields a representation of the Fisher information matrix in terms of multiple Sylvester matrices. These matrices are represented by the submatrices associated with $\mathcal{S}(-\beta,\alpha)$ and $\mathcal{S}(-\gamma,\alpha)$, respectively. This is summarized in Proposition 2.2.

Proposition 2.2. *The Fisher information matrix of a VARMAX process expressed in terms of multiple Sylvester matrices is given by*

$$\mathcal{F}(\theta)=\frac{1}{2\pi i}\oint\limits_{|z|=1}\left(\Omega(z)\Sigma\Omega^{*}(z)\otimes\sigma(z)\right)\frac{dz}{z}+\frac{1}{2\pi i}\oint\limits_{|z|=1}\left(\Gamma(z)R_x(z)\Gamma^{*}(z)\otimes\sigma(z)\right)\frac{dz}{z}, \tag{29}$$

where

$$\Omega(z)=\begin{pmatrix}I_p\otimes\alpha^{-1}(z) & 0_{pn\times rn} & 0_{pn\times qn}\\ 0_{rn\times pn} & 0_{rn\times rn} & 0_{rn\times qn}\\ 0_{qn\times pn} & 0_{qn\times rn} & I_q\otimes\alpha^{-1}(z)\end{pmatrix}\begin{pmatrix}\mathcal{S}_p(-\beta)\\ 0_{rn\times n(p+q)}\\ \mathcal{S}_q(\alpha)\end{pmatrix}\left(u_{p+q}(z)\otimes I_n\right)$$

and

$$\Gamma(z)=\begin{pmatrix}I_p\otimes\alpha^{-1}(z) & 0_{pn\times rn} & 0_{pn\times qn}\\ 0_{rn\times pn} & I_r\otimes\alpha^{-1}(z) & 0_{rn\times qn}\\ 0_{qn\times pn} & 0_{qn\times rn} & 0_{qn\times qn}\end{pmatrix}\begin{pmatrix}\mathcal{S}_p(-\gamma)\\ \mathcal{S}_r(\alpha)\\ 0_{qn\times n(p+r)}\end{pmatrix}\left(u_{p+r}(z)\otimes I_n\right),$$

where $\mathcal{S}_p(-\beta)$ and $\mathcal{S}_q(\alpha)$ are submatrices of the multiple Sylvester matrix $\mathcal{S}(-\beta,\alpha)$ whereas $\mathcal{S}_p(-\gamma)$ and $\mathcal{S}_r(\alpha)$ are submatrices of the multiple Sylvester matrix $\mathcal{S}(-\gamma,\alpha)$.

We have derived a compact representation of $\mathcal{F}(\vartheta)$ at the vector-matrix level and expressed in terms of multiple Sylvester matrices. It is known that the scalar version of $\mathcal{S}(-\beta,\alpha)$ has the resultant property, the matrix $\mathcal{S}(-\beta,\alpha)$ becomes singular if and only if the scalar polynomials $\alpha(z)$ and $\beta(z)$ have at least one common zero. However, the multiple resultant property does not hold for multiple Sylvester matrices, see e.g. [8,17]. If the two matrix polynomials $\alpha(z)$ and $\beta(z)$ have a common eigenvalue the det $\mathcal{S}(-\beta,\alpha)$ will not necessarily be equal to zero. In the next section we represent $\mathcal{F}(\theta)$ in terms of tensor Sylvester matrices, for which the multiple resultant property does hold.

2.5. Representation based on tensor Sylvester matrices

In this section we shall further exploit the approach used in Section 2.4. We will develop a representation of the Fisher information matrix in terms of tensor Sylvester matrices. We shall therefore apply another factorization to the matrices in (22). We have

$$\mathcal{P}(z)\otimes\sigma(z)=\mathcal{D}(z)\left(\Sigma\otimes\sigma(z)\right)\mathcal{D}(z)^{*} \tag{30}$$

and

$$\mathcal{Q}(z)\otimes\sigma(z)=\mathcal{N}(z)\left(R_x(z)\otimes\sigma(z)\right)\mathcal{N}(z)^{*}, \tag{31}$$

where

$$\mathcal{D}(z)=\begin{pmatrix}u_p(z)\otimes\alpha^{-1}(z)\left(-\beta(z)\right)\otimes I_n\\ 0_{rn^2\times n^2}\\ u_q(z)\otimes\alpha^{-1}(z)\alpha(z)\otimes I_n\end{pmatrix}$$

and

$$\mathcal{N}(z)=\begin{pmatrix}u_p(z)\otimes\alpha^{-1}(z)\left(-\gamma(z)\right)\otimes I_n\\ u_r(z)\otimes\alpha^{-1}(z)\alpha(z)\otimes I_n\\ 0_{qn^2\times n^2}\end{pmatrix}.$$

For matrix polynomials $\alpha(z) = \sum_{i=0}^{p} \alpha_i z^i$ and $\beta(z) = \sum_{j=0}^{q} \beta_j z^j$ the $n^2(p+q) \times n^2(p+q)$ tensor Sylvester matrix is defined as

$$\mathcal{S}^{\otimes}(-\beta, \alpha) = \begin{pmatrix} (-I_n)\otimes I_n & (-\beta_1)\otimes I_n & \cdots & (-\beta_q)\otimes I_n & 0_{n^2\times n^2} & \cdots & 0_{n^2\times n^2} \\ 0_{n^2\times n^2} & \ddots & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \ddots & 0_{n^2\times n^2} \\ 0_{n^2\times n^2} & \cdots & 0_{n^2\times n^2} & (-I_n)\otimes I_n & (-\beta_1)\otimes I_n & \cdots & (-\beta_q)\otimes I_n \\ I_n\otimes I_n & I_n\otimes\alpha_1 & \cdots & I_n\otimes\alpha_p & 0_{n^2\times n^2} & \cdots & 0_{n^2\times n^2} \\ 0_{n^2\times n^2} & \ddots & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \ddots & 0_{n^2\times n^2} \\ 0_{n^2\times n^2} & \cdots & 0_{n^2\times n^2} & I_n\otimes I_n & I_n\otimes\alpha_1 & \cdots & I_n\otimes\alpha_p \end{pmatrix}.$$

Gohberg and Lerer (1976), in [8], have proved that the matrix polynomials $\alpha(z)$ and $\beta(z)$ have at least one common eigenvalue if and only if $\det \mathcal{S}^{\otimes}(-\beta, \alpha) = 0$ or when the matrix $\mathcal{S}^{\otimes}(-\beta, \alpha)$ is singular. In other words the tensor Sylvester matrix $\mathcal{S}^{\otimes}(-\beta, \alpha)$ has the multiple resultant property contrary to the multiple Sylvester matrix $\mathcal{S}(-\beta, \alpha)$.
We need the $pn^2 \times (p+q)n^2$ and $qn^2 \times (p+q)n^2$ submatrices $\mathcal{S}_p^{\otimes}(-\beta)$ and $\mathcal{S}_q^{\otimes}(\alpha)$ of the tensor Sylvester matrix $\mathcal{S}^{\otimes}(-\beta, \alpha)$ such that

$$\mathcal{S}^{\otimes}(-\beta, \alpha) = \begin{pmatrix} \mathcal{S}_p^{\otimes}(-\beta) \\ \mathcal{S}_q^{\otimes}(\alpha) \end{pmatrix}.$$

The following two properties are easily verified. We have

$$\mathcal{S}_p^{\otimes}(-\beta)\left(u_{p+q}(z) \otimes I_{n^2}\right) = u_p(z) \otimes (-\beta(z)) \otimes I_n \tag{32}$$

and

$$\mathcal{S}_q^{\otimes}(\alpha)\left(u_{p+q}(z) \otimes I_{n^2}\right) = u_q(z) \otimes \alpha(z) \otimes I_n. \tag{33}$$

Similarly, for the submatrices $\mathcal{S}_p^{\otimes}(-\gamma)$ and $\mathcal{S}_r^{\otimes}(\alpha)$ of the $(p+r)n^2 \times (p+r)n^2$ tensor Sylvester matrix $\mathcal{S}^{\otimes}(-\gamma, \alpha)$, it holds that

$$\mathcal{S}_p^{\otimes}(-\gamma)\left(u_{p+r}(z) \otimes I_{n^2}\right) = u_p(z) \otimes (-\gamma(z)) \otimes I_n \tag{34}$$

and

$$\mathcal{S}_r^{\otimes}(\alpha)\left(u_{p+r}(z) \otimes I_{n^2}\right) = u_r(z) \otimes \alpha(z) \otimes I_n. \tag{35}$$

A factorization of the matrices in (30) and (31) will be applied in order to express the Fisher information matrix in terms of the tensor Sylvester matrices $\mathcal{S}^{\otimes}(-\beta, \alpha)$ and $\mathcal{S}^{\otimes}(-\gamma, \alpha)$. For that purpose we first factorize the matrix polynomials $\mathcal{D}(z)$ and $\mathcal{N}(z)$ accordingly, to obtain

$$\mathcal{D}(z) = \begin{pmatrix} I_p \otimes \alpha^{-1}(z) \otimes I_n & 0_{pn^2\times rn^2} & 0_{pn^2\times qn^2} \\ 0_{rn^2\times pn^2} & 0_{rn^2\times rn^2} & 0_{rn^2\times qn^2} \\ 0_{qn^2\times pn^2} & 0_{qn^2\times rn^2} & I_q \otimes \alpha^{-1}(z) \otimes I_n \end{pmatrix} \begin{pmatrix} u_p(z) \otimes (-\beta(z)) \otimes I_n \\ 0_{rn^2\times n^2} \\ u_q(z) \otimes \alpha(z) \otimes I_n \end{pmatrix} \tag{36}$$

and

$$\mathcal{N}(z) = \begin{pmatrix} I_p \otimes \alpha^{-1}(z) \otimes I_n & 0_{pn^2\times rn^2} & 0_{pn^2\times qn^2} \\ 0_{rn^2\times pn^2} & I_r \otimes \alpha^{-1}(z) \otimes I_n & 0_{rn^2\times qn^2} \\ 0_{qn^2\times pn^2} & 0_{qn^2\times rn^2} & 0_{qn^2\times qn^2} \end{pmatrix} \begin{pmatrix} u_p(z) \otimes (-\gamma(z)) \otimes I_n \\ u_r(z) \otimes \alpha(z) \otimes I_n \\ 0_{qn^2\times n^2} \end{pmatrix}. \tag{37}$$

Expression (30), when equalities (32) and (33) are used in (36), becomes

$$\mathcal{P}(z) \otimes \sigma(z) = \Phi(z)\Theta(z)\Phi^*(z), \tag{38}$$

where

$$\Phi(z) = \begin{pmatrix} I_p \otimes \alpha^{-1}(z) \otimes I_n & 0_{pn^2 \times rn^2} & 0_{pn^2 \times qn^2} \\ 0_{rn^2 \times pn^2} & 0_{rn^2 \times rn^2} & 0_{rn^2 \times qn^2} \\ 0_{qn^2 \times pn^2} & 0_{qn^2 \times rn^2} & I_q \otimes \alpha^{-1}(z) \otimes I_n \end{pmatrix} \begin{pmatrix} -\mathcal{S}_p^{\otimes}(\beta) \\ 0_{rn^2 \times n^2(p+q)} \\ \mathcal{S}_q^{\otimes}(\alpha) \end{pmatrix} \left(u_{p+q}(z) \otimes I_{n^2}\right)$$

and

$$\Theta(z) = \Sigma \otimes \sigma(z).$$

When equalities (34) and (35) are used in (37), we obtain

$$\mathcal{Q}(z) \otimes \sigma(z) = \Lambda(z)\Psi(z)\Lambda^*(z), \tag{39}$$

where

$$\Lambda(z) = \begin{pmatrix} I_p \otimes \alpha^{-1}(z) \otimes I_n & 0_{pn^2 \times rn^2} & 0_{pn^2 \times qn^2} \\ 0_{rn^2 \times pn^2} & I_r \otimes \alpha^{-1}(z) \otimes I_n & 0_{rn^2 \times qn^2} \\ 0_{qn^2 \times pn^2} & 0_{qn^2 \times rn^2} & 0_{qn^2 \times qn^2} \end{pmatrix} \begin{pmatrix} -\mathcal{S}_p^{\otimes}(\gamma) \\ \mathcal{S}_r^{\otimes}(\alpha) \\ 0_{qn^2 \times n^2(p+r)} \end{pmatrix} \left(u_{p+r}(z) \otimes I_{n^2}\right)$$

and

$$\Psi(z) = R_x(z) \otimes \sigma(z).$$

Combining (38) and (39) in (22) yields a representation for the Fisher information matrix $\mathcal{F}(\vartheta)$ in terms of submatrices of the tensor Sylvester matrices $\mathcal{S}^{\otimes}(-\beta, \alpha)$ and $\mathcal{S}^{\otimes}(-\gamma, \alpha)$. This is given in Proposition 2.3.

Proposition 2.3. *The following representation of the Fisher information matrix of a VARMAX process expressed in terms of tensor Sylvester matrices holds true*

$$\mathcal{F}(\vartheta) = \frac{1}{2\pi i} \oint_{|z|=1} \Phi(z)\Theta(z)\Phi^*(z)\frac{dz}{z} + \frac{1}{2\pi i} \oint_{|z|=1} \Lambda(z)\Psi(z)\Lambda^*(z)\frac{dz}{z}. \tag{40}$$

As mentioned before, the multiple Sylvester matrix has no resultant property whereas the tensor Sylvester matrix does have this property. This implies that representation (40) is more appropriate than (29) to prove a possible resultant property of the Fisher information matrix of a VARMAX process. This will be a subject for further research.

The resultant property of the Fisher information matrix of a VARMA process is proved in [17]. In [19] an elementwise representation of the Fisher information matrix of a VARMAX process is developed. The obtained expressions are easily implementable circular integrals which are convenient for computational purposes. Therefore, a numerical computation of the Fisher information matrix of a VARMAX process is a subject for further study. An efficient and fast algorithm described in [4] can be used to compute the circular integrals.

A representation of the inverses $(\alpha(z))^{-1}$, $(\beta(z))^{-1}$, $(\alpha(z^{-1}))^{-1}$ and $(\beta(z^{-1}))^{-1}$, which appear in (29) and (40), is expressed in terms of the coefficients. More details are given in the next section according to a property proved in [9].

2.6. Inversion of matrix polynomials

Let $\tilde{\alpha}(z) = z^p\alpha(z^{-1})$ and $\tilde{\beta}(z) = z^q\beta(z^{-1})$. Companion matrices which shall be associated with the matrix polynomials $\tilde{\alpha}(z)$ and $\tilde{\beta}(z)$ are defined by the $np \times np$ and $nq \times nq$ matrices accordingly, to have

$$C_\alpha = \begin{pmatrix} 0 & I & 0 & \dots & 0 \\ 0 & 0 & I & \dots & 0 \\ \vdots & & & \ddots & \\ 0 & & & & I \\ -\alpha_p & -\alpha_{p-1} & \cdots & \cdots & -\alpha_1 \end{pmatrix} \quad \text{and}$$

$$C_\beta = \begin{pmatrix} 0 & I & 0 & \dots & 0 \\ 0 & 0 & I & \dots & 0 \\ \vdots & & & \ddots & \\ 0 & & & & I \\ -\beta_q & -\beta_{q-1} & \dots & \dots & -\beta_1 \end{pmatrix}.$$

respectively. As in the scalar case, the properties

$$\det(Iz - C_\alpha) = \det \tilde{\alpha}(z) \text{ and } \det(Iz - C_\beta) = \det \tilde{\beta}(z)$$

and

$$\det(I - zC_\alpha) = \det \alpha(z) \text{ and } \det(I - zC_\beta) = \det \beta(z)$$

hold, see [9].

The following equalities hold for every $z \in \mathbb{C}$ which is not an eigenvalue of the matrix polynomials $\tilde{\alpha}(z)$ and $\tilde{\beta}(z)$,

$$(\tilde{\alpha}(z))^{-1} = P_\alpha (Iz - C_\alpha)^{-1} R_\alpha \text{ and } \left(\tilde{\beta}(z)\right)^{-1} = P_\beta (Iz - C_\beta)^{-1} R_\beta,$$

where

$$\text{the } n \times np \text{ matrix } P_\alpha = \begin{pmatrix} I & 0 & \dots & 0 \end{pmatrix} \text{ and } np \times n \text{ matrix } R_\alpha = \begin{pmatrix} 0 & \dots & 0 & I \end{pmatrix}^\top$$

and

$$\text{the } n \times nq \text{ matrix } P_\beta = \begin{pmatrix} I & 0 & \dots & 0 \end{pmatrix} \text{ and } nq \times n \text{ matrix } R_\beta = \begin{pmatrix} 0 & \dots & 0 & I \end{pmatrix}^\top.$$

Since $(\alpha(z^{-1}))^{-1} = z^p \ (\tilde{\alpha}(z))^{-1}$ and $(\beta(z^{-1}))^{-1} = z^q \ (\tilde{\beta}(z))^{-1}$ the relations

$$\left(\alpha(z^{-1})\right)^{-1} = z^p P_\alpha (Iz - C_\alpha)^{-1} R_\alpha \text{ and } \left(\beta(z^{-1})\right)^{-1} = z^q P_\beta (Iz - C_\beta)^{-1} R_\beta$$

and

$$(\alpha(z))^{-1} = z^{-p+1} P_\alpha (I - zC_\alpha)^{-1} R_\alpha \text{ and } (\beta(z))^{-1} = z^{-q+1} P_\beta (I - zC_\beta)^{-1} R_\beta$$

hold true. However, the computation of the values that the inverses of the matrix polynomials take at some points can be of interest; i.e. when we are not interested in the computation of the matrix coefficients of the inverses.

2.7. A representation derived from the Fisher information matrix

The purpose of this section consists of displaying a setting which is similar to the representation of the Fisher information matrix of a scalar ARMAX process set forth in [14]. For that purpose we start from representation (40). The integrand of the first term is considered,

$$\mathcal{A}(z) := \Phi(z)\Theta(z)\Phi^*(z).$$

We obtain

$$\mathcal{L}(z)\mathcal{A}(z)\mathcal{L}^*(z) = \begin{pmatrix} -\mathcal{S}_p^{\otimes}(\beta) \\ 0_{rn^2 \times n^2(p+q)} \\ \mathcal{S}_q^{\otimes}(\alpha) \end{pmatrix} \left(u_{p+q}(z) \otimes I_{n^2}\right) \Theta(z) \left(u_{p+q}(z) \otimes I_{n^2}\right)^* \begin{pmatrix} -\mathcal{S}_p^{\otimes}(\beta) \\ 0_{rn^2 \times n^2(p+q)} \\ \mathcal{S}_q^{\otimes}(\alpha) \end{pmatrix}^\top, \tag{41}$$

where

$$\mathcal{L}(z) = \begin{pmatrix} I_p \otimes \alpha(z) \otimes I_n & 0_{pn^2 \times rn^2} & 0_{pn^2 \times qn^2} \\ 0_{rn^2 \times pn^2} & 0_{rn^2 \times rn^2} & 0_{rn^2 \times qn^2} \\ 0_{qn^2 \times pn^2} & 0_{qn^2 \times rn^2} & I_q \otimes \alpha(z) \otimes I_n \end{pmatrix}.$$

Integrating expression (41) yields

$$\frac{1}{2\pi i}\oint_{|z|=1}\mathcal{L}(z)\mathcal{A}(z)\mathcal{L}^*(z)\frac{dz}{z}=\begin{pmatrix}-\mathcal{S}_p^{\otimes}(\beta)\\ 0_{rn^2\times n^2(p+q)}\\ \mathcal{S}_q^{\otimes}(\alpha)\end{pmatrix}\mathcal{R}(\vartheta)\begin{pmatrix}-\mathcal{S}_p^{\otimes}(\beta)\\ 0_{rn^2\times n^2(p+q)}\\ \mathcal{S}_q^{\otimes}(\alpha)\end{pmatrix}^{\top}, \tag{42}$$

where

$$\mathcal{R}(\vartheta)=\frac{1}{2\pi i}\oint_{|z|=1}\left(u_{p+q}(z)\otimes I_{n^2}\right)\Theta(z)\left(u_{p+q}(z)\otimes I_{n^2}\right)^*\frac{dz}{z}.$$

The integrand of the second term of (40) is now considered,

$$\mathcal{B}(z):=\Lambda(z)\,\Psi(z)\Lambda^*(z).$$

A similar procedure as above is applied, to obtain

$$\mathcal{W}(z)\mathcal{B}(z)\mathcal{W}^*(z)=\begin{pmatrix}-\mathcal{S}_p^{\otimes}(\gamma)\\ \mathcal{S}_r^{\otimes}(\alpha)\\ 0_{qn^2\times n^2(p+r)}\end{pmatrix}\left(u_{p+r}(z)\otimes I_{n^2}\right)\Psi(z)\left(u_{p+r}(z)\otimes I_{n^2}\right)^*\begin{pmatrix}-\mathcal{S}_p^{\otimes}(\gamma)\\ \mathcal{S}_r^{\otimes}(\alpha)\\ 0_{qn^2\times n^2(p+r)}\end{pmatrix}^{\top}, \tag{43}$$

with

$$\mathcal{W}(z)=\begin{pmatrix}I_p\otimes\alpha(z)\otimes I_n & 0_{pn^2\times rn^2} & 0_{pn^2\times qn^2}\\ 0_{rn^2\times pn^2} & I_r\otimes\alpha(z)\otimes I_n & 0_{rn^2\times qn^2}\\ 0_{qn^2\times pn^2} & 0_{qn^2\times rn^2} & 0_{qn^2\times qn^2}\end{pmatrix}.$$

Integration of (43) is applied, to obtain

$$\frac{1}{2\pi i}\oint_{|z|=1}\mathcal{W}(z)\mathcal{B}(z)\mathcal{W}^*(z)\frac{dz}{z}=\begin{pmatrix}-\mathcal{S}_p^{\otimes}(\gamma)\\ \mathcal{S}_r^{\otimes}(\alpha)\\ 0_{qn^2\times n^2(p+r)}\end{pmatrix}\mathcal{T}(\vartheta)\begin{pmatrix}-\mathcal{S}_p^{\otimes}(\gamma)\\ \mathcal{S}_r^{\otimes}(\alpha)\\ 0_{qn^2\times n^2(p+r)}\end{pmatrix}^{\top}, \tag{44}$$

where

$$\mathcal{T}(\vartheta)=\frac{1}{2\pi i}\oint_{|z|=1}\left(u_{p+r}(z)\otimes I_{n^2}\right)\Psi(z)\left(u_{p+r}(z)\otimes I_{n^2}\right)^*\frac{dz}{z}.$$

A matrix, which is a combination of (42) and (44), is set forth

$$\begin{aligned}\mathcal{M}(\vartheta)&=\frac{1}{2\pi i}\oint_{|z|=1}\mathcal{L}(z)\mathcal{A}(z)\mathcal{L}^*(z)\frac{dz}{z}+\frac{1}{2\pi i}\oint_{|z|=1}\mathcal{W}(z)\mathcal{B}(z)\mathcal{W}^*(z)\frac{dz}{z}\\ &=\begin{pmatrix}-\mathcal{S}_p^{\otimes}(\beta)\\ 0_{rn^2\times n^2(p+q)}\\ \mathcal{S}_q^{\otimes}(\alpha)\end{pmatrix}\mathcal{R}(\vartheta)\begin{pmatrix}-\mathcal{S}_p^{\otimes}(\beta)\\ 0_{rn^2\times n^2(p+q)}\\ \mathcal{S}_q^{\otimes}(\alpha)\end{pmatrix}^{\top}\\ &\quad+\begin{pmatrix}-\mathcal{S}_p^{\otimes}(\gamma)\\ \mathcal{S}_r^{\otimes}(\alpha)\\ 0_{qn^2\times n^2(p+r)}\end{pmatrix}\mathcal{T}(\vartheta)\begin{pmatrix}-\mathcal{S}_p^{\otimes}(\gamma)\\ \mathcal{S}_r^{\otimes}(\alpha)\\ 0_{qn^2\times n^2(p+r)}\end{pmatrix}^{\top}.\end{aligned} \tag{45}$$

In [14], the Fisher information matrix of a ARMAX process has a similar representation to (45) with appropriate submatrices. This expression is used to prove that the Fisher information matrix of a ARMAX time series process is singular iff the three polynomials have at least one common zero.

3. Conclusion

Compact forms of the Fisher information matrix of a VARMAX process expressed in terms of multiple and tensor Sylvester matrices have been established. Especially the representation expressed by tensor

Sylvester matrices will allow us to study matrix structural properties of the Fisher information matrix of VARMAX processes. This will be a subject of further study.

Acknowledgments

The authors would like to thank two anonymous referees and the associate editor Beatrice Meini for their valuable comments and suggestions that have improved the quality of the paper.

References

[1] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, Time Series Analysis, third ed., Prentice-Hall Inc., Englewood Cliffs, NJ, 1994.
[2] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, second ed., Springer-Verlag, Berlin, New York, 1991.
[3] P. Caines, Linear Stochastic Systems, John Wiley and Sons, New York, 1988.
[4] C.J. Demeure, C.T. Mullis, The Euclid algorithm and the fast computation of cross-covariance and autocovariance sequences, IEEE Trans. Acoust. Speech Signal Process. 37 (1989) 545–552.
[5] B.R. Frieden, Physics from Fisher Information: A Unification, Cambridge University Press, New York, 1998.
[6] B.R. Frieden, Science from Fisher Information: A Unification, Cambridge University Press, New York, 2004.
[7] B. Friedlander, On the computation of the Cramér-Rao bound for ARMA parameter estimation, IEEE Trans. Acoust. Speech Signal Process. 32 (4) (1984) 721–727.
[8] I. Gohberg, L. Lerer, Resultants of matrix polynomials, Bull. Amer. Math. Soc. 82 (1976) 565–567.
[9] I. Gohberg, P. Lancaster, P.L. Rodman, Matrix Polynomials, Academic Press, New York, 1982.
[10] E.J. Hannan, Multiple Time Series, John Wiley, New York, 1970.
[11] E.J. Hannan, M. Deistler, The Statistical Theory of Linear Systems, John Wiley and Sons, New York, 1988.
[12] E.J. Hannan, W.T.M. Dunsmuir, M. Deistler, Estimation of vector ARMAX models, J. Multivariate Anal. 10 (1980) 275–295.
[13] I.A. Ibragimov, R.Z. Has'minskiĭ, Statistical Estimation, Asymptotic Theory, Springer-Verlag, New York, 1981.
[14] A. Klein, P. Spreij, On Fisher's information matrix of an ARMAX process and Sylvester's resultant matrices, Linear Algebra Appl. 237/238 (1996) 579–590.
[15] A. Klein, G. Mélard, Computation of the Fisher information matrix for SISO models, IEEE Trans. Signal Process. 42 (1994) 684–688.
[16] A. Klein, A generalization of Whittle's formula for the information matrix of vector mixed time series, Linear Algebra Appl. 321 (2000) 197–208.
[17] A. Klein, G. Mélard, P. Spreij, On the resultant property of the Fisher information matrix of a vector ARMA process, Linear Algebra Appl. 403 (2005) 291–313.
[18] A. Klein, G. Mélard, A. Saidi, The asymptotic and exact Fisher information matrices of a vector ARMA process, Statist. Probab. Lett. 78 (2008) 1430–1433.
[19] A. Klein, P. Spreij, Matrix differential calculus applied to multiple stationary time series and an extended Whittle formula for information matrices, Linear Algebra Appl. 430 (2009) 674–691.
[20] P. Lancaster, M. Tismenetsky, The Theory of Matrices with Applications, second ed., Academic Press, Orlando, 1985.
[21] L. Ljung, T. Söderström, Theory and Practice of Recursive Identification, MIT Press, Cambridge, Mass, 1983.
[22] H.J. Newton, The information matrices of multiple time series, J. Multivariate Anal. 8 (1978) 317–323.
[23] P. Whittle, The analysis of multiple stationary time series, J. Roy. Statist. Soc. B 15 (1953) 125–139.