Entropy **2014**, *16*, 2023-2055; doi:10.3390/e16042023

OPEN ACCESS

entropy

ISSN 1099-4300

www.mdpi.com/journal/entropy

Article

Matrix Algebraic Properties of the Fisher Information Matrix of Stationary Processes

André Klein

Rothschild Blv. 123 Apt.7, 65271 Tel Aviv, Israel;
E-Mail: A.A.B.Klein@uva.nl or klein@contact.uva.nl; Tel.: 972.5.25594723

Received: 12 February 2014; in revised form: 11 March 2014 / Accepted: 24 March 2014 / Published: 8 April 2014

Abstract: In this survey paper, a summary of results which are to be found in a series of papers, is presented. The subject of interest is focused on matrix algebraic properties of the Fisher information matrix (FIM) of stationary processes. The FIM is an ingredient of the Cramér-Rao inequality, and belongs to the basics of asymptotic estimation theory in mathematical statistics. The FIM is interconnected with the Sylvester, Bezout and tensor Sylvester matrices. Through these interconnections it is shown that the FIM of scalar and multiple stationary processes fulfill the resultant matrix property. A statistical distance measure involving entries of the FIM is presented. In quantum information, a different statistical distance measure is set forth. It is related to the Fisher information but where the information about one parameter in a particular measurement procedure is considered. The FIM of scalar stationary processes is also interconnected to the solutions of appropriate Stein equations, conditions for the FIM to verify certain Stein equations are formulated. The presence of Vandermonde matrices is also emphasized.

Keywords: Bezout matrix; Sylvester matrix; tensor Sylvester matrix; Stein equation; Vandermonde matrix; stationary process; matrix resultant; Fisher information matrix

MSC Classification: 15A23, 15A24, 15B99, 60G10, 62B10, 62M20.

---

1. Introduction

In this survey paper, a summary of results derived and described in a series of papers, is presented. It concerns some matrix algebraic properties of the Fisher information matrix (abbreviated as FIM)

of stationary processes. An essential property emphasized in this paper concerns the matrix resultant property of the FIM of stationary processes. To be more explicit, consider the coefficients of two monic polynomials $p(z)$ and $q(z)$ of finite degree, as the entries of a matrix such that the matrix becomes singular if and only if the polynomials $p(z)$ and $q(z)$ have at least one common root. Such a matrix is called a resultant matrix and its determinant is called the resultant. The Sylvester, Bezout and tensor Sylvester matrices have such a property and are extensively studied in the literature, see e.g., [1–3]. The FIM associated with various stationary processes will be expressed by these matrices. The derived interconnections are obtained by developing the necessary factorizations of the FIM in terms of the Sylvester, Bezout and tensor Sylvester matrices. These factored forms of the FIM enable us to show that the FIM of scalar and multiple stationary processes fulfill the resultant matrix property. Consequently, the singularity conditions of the appropriate Fisher information matrices and Sylvester, Bezout and tensor Sylvester matrices coincide, these results are described in [4–6].

A statistical distance measure involving entries of the FIM is presented and is based on [7]. In quantum information, a statistical distance measure is set forth, see [8–10], and is related to the Fisher information but where the information about one parameter in a particular measurement procedure is considered. This leads to a challenging question that can be presented as, can the existing distance measure in quantum information be developed at the matrix level ?

The matrix Stein equation, see e.g., [11], is associated with the Fisher information matrices of scalar stationary processes through the solutions of the appropriate Stein equations. Conditions for the Fisher information matrices or associated matrices to verify certain Stein equations are formulated and proved in this paper. The presence of Vandermonde matrices is also emphasized. The general and more detailed results are set forth in [12] and [13]. In this survey paper it is shown that the FIM of linear stationary processes form a class of structured matrices. Note that in [14], the authors emphasize that statistical problems related to stationary processes have been treated successfully with the aid of Toeplitz forms.
This paper is organized as follows. The various stationary processes, considered in this paper, are presented in Section 2, the Fisher information matrices of the stationary processes are displayed in Section 3. Section 3 sets forth the interconnections between the Fisher information matrices and the Sylvester, Bezout, tensor Sylvester matrices, and solutions to Stein equations. A statistical distance measure is expressed in terms of entries of a FIM.

2. The Linear Stationary Processes

In this section we display the class of linear stationary processes whose corresponding Fisher information matrix shall be investigated in a matrix algebraic context. But first some basic definitions are set forth, see e.g., [15].

If a random variable X is indexed to time, usually denoted by t, the observations $\{X_t, t \in \mathbb{T}\}$ is called a time series, where $\mathbb{T}$ is a time index set (for example, $\mathbb{T} = \mathbb{Z}$, the integer set).

Definition 2.1. *A stochastic process is a family of random variables* $\{X_t, t \in \mathbb{T}\}$ *defined on a probability space* $\{\Omega, \mathcal{F}, \mathcal{P}\}$.

Definition 2.2. *The Autocovariance function. If* $\{X_t, t \in \mathbb{T}\}$ *is a process such that Var*$(X_t) < \infty$ *(variance) for each* t*, then the autocovariance function* $\gamma_X(\cdot,\cdot)$ *of* $\{X_t\}$ *is defined by*

$\gamma_X(r,s) = Cov(X_r, X_s) = \mathbb{E}[(X_r - \mathbb{E}X_r)(X_s - \mathbb{E}X_s)]$, $r, s \in \mathbb{Z}$ *and* $\mathbb{E}$ *represents the expected value.*

Definition 2.3. *Stationarity. The time series* $\{X_t, t \in \mathbb{Z}\}$*, with the index set* $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$*, is said to be stationary if*

$$\begin{aligned} (i)\ \mathbb{E}\ |X_t|^2 &< \infty \\ (ii)\ \mathbb{E}\ (X_t) &= m \text{ for all } t \in \mathbb{Z},\ m \text{ is the constant average or mean} \\ (iii)\ \gamma_X(r,s) &= \gamma_X(r+t, s+t) \text{ for all } r, s, t \in \mathbb{Z}, \end{aligned}$$

From Definition 2.3 can be concluded that the joint probability distributions of the random variables $\{X_1, X_2, \ldots X_{t_n}\}$ and $\{X_{1+k}, X_{2+k}, \ldots X_{t_n+k}\}$ are the same for arbitrary times $t_1, t_2, \ldots, t_n$ for all n and all lags or leads $k = 0, \pm 1, \pm 2, \ldots$. The probability distribution of observations of a stationary process is invariant with respect to shifts in time. In the next section the linear stationary processes that will be considered throughout this paper are presented.

2.1. The Vector ARMAX or VARMAX Process

We display one of the most general linear stationary process called the multivariate autoregressive, moving average and exogenous process, the VARMAX process. To be more specific, consider the vector difference equation representation of a linear system $\{y(t), t \in \mathbb{Z}\}$, of order (p, r, q),

$$\sum_{j=0}^{p} A_j\, y(t-j) = \sum_{j=0}^{r} C_j\, x(t-j) + \sum_{j=0}^{q} B_j\, \varepsilon(t-j),\ t \in \mathbb{Z} \tag{1}$$

where $y(t)$ are the observable outputs, $x(t)$ the observable inputs and $\varepsilon(t)$ the unobservable errors, all are n-dimensional. The acronym VARMAX stands for vector autoregressive-moving average with exogenous variables. The left side of (1) is the autoregressive part the second term on the right is the moving average part and $x(t)$ is exogenous. If $x(t)$ does not occur the system is said to be (V)ARMA. Next to exogenous, the input $x(t)$ is also named the control variable, depending on the field of application, in econometrics and time series analysis, e.g., [15], and in signal processing and control, e.g., [16,17]. The matrix coefficients, $A_j \in \mathbb{R}^{n\times n}$, $C_j \in \mathbb{R}^{n\times n}$, and $B_j \in \mathbb{R}^{n\times n}$ are the associate parameter matrices. We have the property $A_0 \equiv B_0 \equiv C_0 \equiv I_n$.
Equation (1) can compactly be written as

$$A(z)\, y(t) = C(z)\, x(t) + B(z)\, \varepsilon(t) \tag{2}$$

where

$$A(z) = \sum_{j=0}^{p} A_j\, z^j;\ C(z) = \sum_{j=0}^{r} C_j\, z^j;\ B(z) = \sum_{j=0}^{q} B_j\, z^j$$

we use z to denote the backward shift operator, for example $z\, x_t = x_{t-1}$. The matrix polynomials $A(z)$, $B(z)$ and $C(z)$ are the associated autoregressive, moving average matrix polynomials, and the

exogenous matrix polynomial respectively of order p, q and r respectively. Hence the process described by Equation (2) is denoted as a VARMAX(p, r, q) process. Here $z \in \mathbb{C}$ with a duplicate use of z as an operator and as a complex variable, which is usual in the signal processing and time series literature, e.g., [15,16,18]. The assumptions $\text{Det}(A(z)) \neq 0$, such that $|z| \leq 1$ and $\text{Det}(B(z)) \neq 0$, such that $|z| < 1$ for all $z \in \mathbb{C}$, is imposed so that the VARMAX(p, r, q) process (2) has exactly one stationary solution and the condition $\text{Det}(B(z)) \neq 0$ implies the invertibility condition, see e.g., [15] for more details. Under these assumptions, the eigenvalues of the matrix polynomials $A(z)$ and $B(z)$ lie outside the unit circle. The eigenvalues of a matrix polynomial $Y(z)$ are the roots of the equation $\text{Det}(Y(z)) = 0$, $\text{Det}(X)$ is the determinant of X. The VARMAX(p, r, q) stationary process (2) is thoroughly discussed in [15,18,19].

The error $\{\varepsilon(t), t \in \mathbb{Z}\}$ is a collection of uncorrelated zero mean n-dimensional random variables each having positive definite covariance matrix Σ and we assume, for all s, t, $\mathbb{E}_\vartheta\{\ x(s)\ \varepsilon^\top(t)\} = 0$, where $X^\top$ denotes the transposition of matrix X and $\mathbb{E}_\vartheta$ represents the expected value under the parameter ϑ. The matrix ϑ represents all the VARMAX(p, r, q) parameters, with the total number of parameters being $n^2(p + q + r)$. For different purposes which will be specified in the next sections, two choices of the parameter structure are considered. First, the parameter vector $\vartheta \in \mathbb{R}^{n^2(p+q+r)\times 1}$ is defined by

$$\vartheta = \text{vec}\,\{A_1, A_2, ..., A_p, C_1, C_2, ..., C_r, B_1, B_2, ..., B_q\} \tag{3}$$

The vec operator transforms a matrix into a vector by stacking the columns of the matrix one underneath the other according to $\text{vec}\, X = \text{col}(\text{col}(X_{ij})_{i=1}^n)_{j=1}^n$, see e.g., [2,20]. A different choice is set forth, when the parameter matrix $\vartheta \in \mathbb{R}^{n\times n(p+q+r)}$ is of the form

$$\vartheta = (\vartheta_1\ \vartheta_2 \ldots \vartheta_p\ \vartheta_{p+1}\ \vartheta_{p+2} \ldots \vartheta_{p+r}\ \vartheta_{p+r+1}\ \vartheta_{p+r+2} \ldots \vartheta_{p+r+q}) \tag{4}$$

$$= (A_1\ A_2 \ldots A_p\ C_1\ C_2 \ldots C_r\ B_1\ B_2\ \ldots B_q) \tag{5}$$

Representation (5) of the parameter matrix has been used in [21]. The estimation of the matrices A_1, A_2,..., A_p, C_1, C_2,..., C_r, B_1, B_2, ..., B_q and Σ has received considerable attention in the time series and statistical signal processing literature, see e.g., [15,17,19]. In [19], the authors study the asymptotic properties of maximum likelihood estimates of the coefficients of VARMAX(p, r, q) processes, stored in a $(\ell \times 1)$ vector ϑ, where $\ell = n^2(p + q + r)$.

Before describing the control-exogenous variable $x(t)$ used in this survey paper, we shall present the different special cases of the model described in Equations (1) and (2).

2.2. The Vector ARMA or VARMA Process

When the process (2) does not contain the control process $x(t)$ it yields

$$A(z)y(t) = B(z)\varepsilon(t) \tag{6}$$

which is a vector autoregressive and moving average process, VARMA(p, q) process, see e.g., [15]. The matrix ϑ represents now all the VARMA parameters, with the total number of parameters being $n^2(p+q)$. The VARMA(p, q) version of the parameter vector ϑ defined in (3) is then given by

$$\vartheta = \text{vec}\,\{A_1, A_2, ..., A_p, B_1, B_2, ..., B_q\} \tag{7}$$

A VARMA process equivalent to the parameter matrix (4) is then the $n \times n(p+q)$ parameter matrix

$$\vartheta = (\vartheta_1 \; \vartheta_2 \ldots \vartheta_p \; \vartheta_{p+1} \; \vartheta_{p+2} \ldots \vartheta_{p+q}) = (A_1 \; A_2 \ldots A_p \; B_1 \; B_2 \; \ldots B_q) \tag{8}$$

A description of the input variable $x(t)$, in Equation (2) follows. Generally, one can assume either that $x(t)$ is non stochastic or that $x(t)$ is stochastic. In the latter case, we assume $\mathbb{E}_\vartheta\{ \; x(s) \; \varepsilon^\top(t)\} = 0$, for all s, t, and that statistical inference is performed conditionally on the values taken by $x(t)$. In this case it can be interpreted as constant, see [22] for a detailed exposition. However, in the papers referred in this survey, like in [21] and [23], the observed input variable $x(t)$, is assumed to be a stationary VARMA process, of the form

$$\alpha(z)x(t) = \beta(z)\eta(t) \tag{9}$$

where $\alpha(z)$ and $\beta(z)$ are the autoregressive and moving average polynomials of appropriate degree and $\{\eta(t), t \in \mathbb{Z}\}$ is a collection of uncorrelated zero mean n-dimensional random variables each having positive definite covariance matrix Ω. The spectral density of the VARMA process $x(t)$ is $R_x(\cdot)/2\pi$ and for a definition, see e.g., [15,16], to obtain

$$R_x(e^{\mathbf{i}\omega}) = \alpha^{-1}(e^{\mathbf{i}\omega})\beta(e^{\mathbf{i}\omega})\Omega\beta^*(e^{\mathbf{i}\omega})\alpha^{-*}(e^{\mathbf{i}\omega}) \quad \omega \in [-\pi, \pi] \tag{10}$$

where $\mathbf{i}$ is the imaginary unit with the property $\mathbf{i}^2 = -1$, ω is the frequency, the spectral density $R_x(e^{\mathbf{i}\omega})$ is Hermitian, and we further have, $R_x(e^{\mathbf{i}\omega}) \geq 0$ and $\int_{-\pi}^{\pi} R_x(e^{\mathbf{i}\omega})d\omega < \infty$. As mentioned above, the basic assumption, $x(t)$ and $\varepsilon(t)$ are independent or at least uncorrelated processes, which corresponds geometrically with orthogonal processes, holds and X^* is the complex conjugate transpose of matrix X.

2.3. The ARMAX and ARMA Processes

The scalar equivalent to the VARMAX(p, r, q) and VARMA(p, q) processes, given by Equations (2) and (6) respectively, shall now be displayed, to obtain for the ARMAX(p, r, q) process

$$a(z)y(t) = c(z)x(t) + b(z)\varepsilon(t) \tag{11}$$

and for the ARMA(p, q) process

$$a(z)y(t) = b(z)\varepsilon(t) \tag{12}$$

popularized in, among others, the Box-Jenkins type of time series analysis, see e.g., [15]. Where $a(z)$, $b(z)$ and $c(z)$ are respectively the scalar autoregressive, moving average polynomials and exogenous polynomial, with corresponding scalar coefficients a_j, b_j and c_j,

$$a(z) = \sum_{j=0}^{p} a_j \; z^j; \; c(z) = \sum_{j=0}^{r} c_j \; z^j; \; b(z) = \sum_{j=0}^{q} b_j \; z^j \tag{13}$$

Note that as in the multiple case, $a_0 = b_0 = 1$. The parameter vector, ϑ, for the processes, Equations (11) and (12) is then

$$\vartheta = \{a_1, a_2, ..., a_p, c_1, c_2, ..., c_r, b_1, b_2, ..., b_q\} \tag{14}$$

and

$$\vartheta = \{a_1, a_2, ..., a_p, b_1, b_2, ..., b_q\} \tag{15}$$

respectively.

In the next section the matrix algebraic properties of the Fisher information matrix of the stationary processes (2), (6), (11) and (12) will be verified. Interconnections with various known structured matrices like the Sylvester resultant matrix, the Bezout matrix and Vandermonde matrix are set forth. The Fisher information matrix of the various stationary processes is also expressed in terms of the unique solutions to the appropriate Stein equations.

3. Structured Matrix Properties of the Asymptotic Fisher Information Matrix of Stationary Processes

The Fisher information is an ingredient of the Cramér-Rao inequality, also called by some the Cauchy-Schwarz inequality in mathematical statistics, and belongs to the basics of asymptotic estimation theory in mathematical statistics. The Cramér-Rao theorem [24] is therefore considered. When assuming that the estimators of ϑ, defined in the previuos sections, are asymptotically unbiased, the inverse of the asymptotic information matrix yields the Cramér-Rao bound, and provided that the estimators are asymptotically efficient, the asymptotic covariance matrix then verifies the inequality

$$\mathrm{Cov}\left(\widehat{\vartheta}\right) \succeq \mathcal{I}^{-1}\left(\widehat{\vartheta}\right)$$

here $\mathcal{I}\left(\widehat{\vartheta}\right)$ is the FIM, $\mathrm{Cov}\left(\widehat{\vartheta}\right)$ is the covariance of $\widehat{\vartheta}$, the unbiased estimator of ϑ, for a detailed fundamental statistical analysis, see [25,26]. The FIM equals the Cramér-Rao lower bound, and the subject of the FIM is also of interest in the control theory and signal processing literature, see e.g., [27] . Its quantum analog was introduced immediately after the foundation of mathematical quantum estimation theory in the 1960's, see [28,29] for a rigorous exposition of the subject. More specifically, the Fisher information is also emphasized in the context of quantum information theory, see e.g., [30,31]. It is clear that the Cramér-Rao inequality takes a lot of attention because it is located on the highly exciting boundary of statistics, information and quantum theory and more recently matrix theory. In the next sections, the Fisher information matrices of linear stationary processes will be presented and its role as a new class of structured matrices will be the subject of study.

When time series models are the subject, using Equation (2) for all $t \in \mathbb{Z}$ to determine the residual $\varepsilon(t)$ or $\varepsilon_t(\vartheta)$, to emphasize the dependency on the parameter vector ϑ, and assuming that $x(t)$ is stochastic and that $(y(t), x(t))$ is a Gaussian stationary process, the asymptotic FIM $\mathcal{F}(\vartheta)$ is defined by the following $(\ell \times \ell)$ matrix which does not depend on t

$$\mathcal{F}(\vartheta) = \mathbb{E}\left\{\left(\frac{\partial \varepsilon_t(\vartheta)}{\partial \vartheta^\top}\right)^\top \Sigma^{-1}\left(\frac{\partial \varepsilon_t(\vartheta)}{\partial \vartheta^\top}\right)\right\} \tag{16}$$

where the $(\upsilon \times \ell)$ matrix $\partial(\cdot)/\partial\vartheta^\top$, the derivative with respect to $\vartheta^\top$, for any $(\upsilon \times 1)$ column vector $(\cdot)$ and ℓ is the total number of parameters. The derivative with respect to $\vartheta^\top$ is used for obtaining the appropriate dimensions. Equality (16) is used for computing the FIM of the various time series processes presented in the previous sections and appropriate definitions of the derivatives are used, especially for the multivariate processes (2) and (6), see [21,22].

3.1. The Fisher Information Matrix of an ARMA(p, q) Process

In this section, the focus is on the FIM of the ARMA process (12). When ϑ is given in Equation (15), the derivatives in Equation (16) are at the scalar level

$$\frac{\partial \varepsilon_t(\vartheta)}{\partial a_j} = \frac{1}{a(z)}\varepsilon_{t-j} \quad \text{for } j = 1, \ldots, p \quad \text{and} \quad \frac{\partial \varepsilon_t(\vartheta)}{\partial b_k} = -\frac{1}{b(z)}\varepsilon_{t-k} \quad \text{for } k = 1, \ldots, q$$

when combined for all j and k, the FIM of the ARMA process (12) with the variance of the noise process $\varepsilon_t(\vartheta)$ equal to one, yields the block decomposition, see [32]

$$\mathcal{F}(\vartheta) = \begin{pmatrix} \mathcal{F}_{aa}(\vartheta) & \mathcal{F}_{ab}(\vartheta) \\ \mathcal{F}_{ba}(\vartheta) & \mathcal{F}_{bb}(\vartheta) \end{pmatrix} \tag{17}$$

The expressions of the different blocks of the matrix $\mathcal{F}(\vartheta)$ are

$$\mathcal{F}_{aa}(\vartheta) = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{u_p(z)u_p^\top(z^{-1})}{a(z)a(z^{-1})}\frac{dz}{z} = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{u_p(z)v_p^\top(z)}{a(z)\widehat{a}(z)}dz \tag{18}$$

$$\mathcal{F}_{ab}(\vartheta) = -\frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{u_p(z)u_q^\top(z^{-1})}{a(z)b(z^{-1})}\frac{dz}{z} = -\frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{u_p(z)v_q^\top(z)}{a(z)\widehat{b}(z)}dz \tag{19}$$

$$\mathcal{F}_{ba}(\vartheta) = -\frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{u_q(z)u_p^\top(z^{-1})}{a(z^{-1})b(z)}\frac{dz}{z} = -\frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{u_q(z)v_p^\top(z)}{\widehat{a}(z)b(z)}dz \tag{20}$$

$$\mathcal{F}_{bb}(\vartheta) = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{u_q(z)u_q^\top(z^{-1})}{b(z)b(z^{-1})}\frac{dz}{z} = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{u_q(z)v_q^\top(z)}{b(z)\widehat{b}(z)}dz \tag{21}$$

where the integration above and everywhere below is counterclockwise around the unit circle. The reciprocal monic polynomials $\widehat{a}(z)$ and $\widehat{b}(z)$ are defined as $\widehat{a}(z) = z^p a(z^{-1})$ and $\widehat{b}(z) = z^q b(z^{-1})$ and $\vartheta = (a_1, \ldots, a_p, b_1, \ldots, b_q)^\top$ introduced in (15). For each positive integer k we have $u_k(z) = (1, z, z^2, \ldots, z^{k-1})^\top$ and $v_k(z) = z^{k-1}u_k(z^{-1})$. Considering the stability condition of the ARMA(p, q) process implies that all the roots of the monic polynomials $a(z)$ and $b(z)$ lie outside the unit circle. Consequently, the roots of the polynomials $\widehat{a}(z)$ and $\widehat{b}(z)$ lie within the unit circle and will be used as the poles for computing the integrals (18)-(21) when Cauchy's residue theorem is applied. Notice that the FIM $\mathcal{F}(\vartheta)$ is symmetric block Toeplitz so that $\mathcal{F}_{ab}(\vartheta) = \mathcal{F}_{ba}^\top(\vartheta)$ and the integrands in (18)-(21) are Hermitian. The computation of the integral expressions, (18)-(21) is easily implementable by using the standard residue theorem. The algorithms displayed in [33] and [22] are suited for numerical computations of among others the FIM of an ARMA(p, q) process.

3.2. The Sylvester Resultant Matrix - The Fisher Information Matrix

The resultant property of a matrix is considered, in order to show that the FIM $\mathcal{F}(\vartheta)$ has the matrix resultant property implies to show that the matrix $\mathcal{F}(\vartheta)$ becomes singular if and only if the appropriate scalar monic polynomials $\widehat{a}(z)$ and $\widehat{b}(z)$ have at least one common zero. To illustrate the subject, the

following known property of two polynomials is set forth. The greatest common divisor (frequently abbreviated as GCD) of two polynomials is a polynomial, of the highest possible degree, that is a factor of both the two original polynomials, the roots of the GCD of two polynomials are the common roots of the two polynomials. Consider the coefficients of two monic polynomials $p(z)$ and $q(z)$ of finite degree, as the entries of a matrix such that the matrix becomes singular if and only if the polynomials $p(z)$ and $q(z)$ have at least one common root. Such a matrix is called a resultant matrix and its determinant is called the resultant. Therefore we present the known $(p+q)\times(p+q)$ Sylvester resultant matrix of the polynomials a and b, see e.g., [2], to obtain

$$\mathcal{S}(a,b)=\begin{pmatrix} 1 & a_1 & \cdots & a_p & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 & a_1 & \cdots & a_p \\ 1 & b_1 & \cdots & b_q & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \ddots & 0_{n\times n} \\ 0 & \cdots & 0 & 1 & b_1 & \cdots & b_q \end{pmatrix} \tag{22}$$

Consider the $q\times(p+q)$ and $p\times(p+q)$ upper and lower submatrices $\mathcal{S}_p(b)$ and $\mathcal{S}_q(-a)$ of the Sylvester resultant matrix $\mathcal{S}(-b,a)$ such that

$$\mathcal{S}(b,-a)=\begin{pmatrix} \mathcal{S}_p(b) \\ -\mathcal{S}_q(a) \end{pmatrix} \tag{23}$$

The matrix $\mathcal{S}(a,b)$ becomes singular in the presence of one or more common zeros of the monic polynomials $\widehat{a}(z)$ and $\widehat{b}(z)$, this property is assessed by the following equalities

$$\mathcal{R}(a,b)=\prod_{\substack{i=1,\ldots,p\\ j=1,\ldots,q}}(\alpha_i-\beta_j),\ \mathcal{R}(b,a)=(-1)^{pq}\prod_{\substack{i=1,\ldots,p\\ j=1,\ldots,q}}(\alpha_i-\beta_j) \tag{24}$$

and

$$\mathcal{R}(b,-a)=(-1)^{q}\prod_{\substack{i=1,\ldots,p\\ j=1,\ldots,q}}(\beta_j-\alpha_i),\text{ and }\mathcal{R}(-b,a)=(-1)^{p}\prod_{\substack{i=1,\ldots,p\\ j=1,\ldots,q}}(\beta_j-\alpha_i) \tag{25}$$

where $\mathcal{R}(a,b)$ is the resultant of $\widehat{a}(z)$ and $\widehat{b}(z)$, and is equal to Det $\mathcal{S}(a,b)$. The string of equalities in (24) and (25) hold since $\mathcal{R}(b,a)=(-1)^{pq}\,\mathcal{R}(a,b)$, $\mathcal{R}(b,-a)=(-1)^{q}\,\mathcal{R}(b,a)$, and $\mathcal{R}(-b,a)=(-1)^{p}\,\mathcal{R}(b,a)$, see [34]. The zeros of the scalar monic polynomials $\widehat{a}(z)$ and $\widehat{b}(z)$ are α_i and β_j respectively and are assumed to be distinct. By this is meant, when we have $(z-\alpha_i)^{n_{\alpha_i}}$ and $\left(z-\beta_j\right)^{n_{\beta_j}}$ with the powers n_{α_i} and n_{β_j} both greater than one, that only the distinct roots will be considered free from the corresponding powers. The key property of the classical Sylvester resultant matrix $\mathcal{S}(a,b)$ is that its null space provides a complete description of the common zeros of the polynomials involved. In particular, in the scalar case the polynomials $\widehat{a}(z)$ and $\widehat{b}(z)$ are coprime if and only if $\mathcal{S}(a,b)$ is non-singular. The following key property of the classical Sylvester resultant matrix $\mathcal{S}(a,b)$, is given by the well known theorem on resultants, to obtain

$$\text{dim Ker}\,\mathcal{S}(a,b)=\nu(a,b) \tag{26}$$

where $\nu(a,b)$ is the number of common roots of the polynomials $\widehat{a}(z)$ and $\widehat{b}(z)$, with counting multiplicities, see e.g., [3]. The dimension of a subspace $\mathcal{V}$ is represented by dim $(\mathcal{V})$, Ker (X) is the null space or kernel of the matrix X, denoted by Null or Ker. The null space of an $n \times n$ matrix A with coefficients in a field K (typically the field of the real numbers or of the complex numbers) is the set Ker $A = \{x \in K^n : \ Ax = 0\}$, see e.g., [1,2,20].

In order to prove that the FIM $\mathcal{F}(\vartheta)$ fulfills the resultant matrix property, the following factorization is derived, Lemma 2.1 in [5],

$$\mathcal{F}(\vartheta) = \mathcal{S}(b,-a)\mathcal{P}(\vartheta)\mathcal{S}^{\top}(b,-a) \tag{27}$$

where the matrix $\mathcal{P}(\vartheta) \in \mathbb{R}^{(p+q)\times(p+q)}$ admits the form

$$\mathcal{P}(\vartheta) = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{u_{p+q}(z)u_{p+q}^{\top}(z^{-1})}{a(z)b(z)a(z^{-1})b(z^{-1})}\frac{dz}{z} = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{u_{p+q}(z)v_{p+q}^{\top}(z)}{a(z)b(z)\widehat{a}(z)\widehat{b}(z)}dz \tag{28}$$

It is proved in [5] that the symmetric matrix $\mathcal{P}(\vartheta)$ fulfills the property, $\mathcal{P}(\vartheta) \succ O$. The factorization (27) allows us to show the matrix resultant property of the FIM, Corollary 2.2 in [5] states.

The FIM of an ARMA(p,q) process with polynomials $a(z)$ and $b(z)$ of order p, q respectively becomes singular if and only if the polynomials $\widehat{a}(z)$ and $\widehat{b}(z)$ have at least one common root. From Corollary 2.2 in [5] can be concluded, the FIM of an ARMA(p,q) process and the Sylvester resultant matrix $\mathcal{S}(-b,a)$ have the same singularity property. By virtue of (26) and (27) we will specify the dimension of the null space of the FIM $\mathcal{F}(\vartheta)$, this is set forth in the following lemma.

Lemma 3.1. *Assume that the polynomials $\widehat{a}(z)$ and $\widehat{b}(z)$ have $\nu(a,b)$ common roots, counting multiplicities. The factorization (27) of the FIM and the property (26) enable us to prove the equality*

$$dim\ (Ker\ \mathcal{F}(\vartheta)) = dim\ (Ker\ \mathcal{S}(b,-a)) = \nu(a,b) \tag{29}$$

Proof. The matrix $\mathcal{P}(\vartheta) \in \mathbb{R}^{(p+q)\times(p+q)}$, given in (27), fulfills the property of positive definiteness, as proved in [5]. This implies that a Cholesky decomposition can be applied to $\mathcal{P}(\vartheta)$, see [35] for more details, to obtain $\mathcal{P}(\vartheta) = L^{\top}(\vartheta)L(\vartheta)$, where $L(\vartheta)$ is a $\mathbb{R}^{(p+q)\times(p+q)}$ upper triangular matrix that is unique if its diagonal elements are all positive. Consequently, all its eigenvalues are then positive so that the matrix $L(\vartheta)$ is also positive definite. Factorization of (27) now admits the representation

$$\mathcal{F}(\vartheta) = \mathcal{S}(b,-a)L^{\top}(\vartheta)L(\vartheta)\mathcal{S}^{\top}(b,-a) \tag{30}$$

and taking the property, if A is an $m \times n$ matrix, then Ker (A) = Ker $\left(A^{\top}A\right)$, into account, yields when applied to (30)

$$\text{Ker } \mathcal{F}(\vartheta) = \text{Ker } \mathcal{S}(b,-a)L^{\top}(\vartheta)L(\vartheta)\mathcal{S}^{\top}(b,-a) = \text{ Ker } L(\vartheta)\mathcal{S}^{\top}(b,-a)$$

Assume the vector $u \in$ Ker $L(\vartheta)\mathcal{S}^{\top}(b,-a)$, such that $L(\vartheta)\mathcal{S}^{\top}(b,-a)u = 0$ and set $\mathcal{S}^{\top}(b,-a)u = v \Longrightarrow L(\vartheta)v = 0$, since the matrix $L(\vartheta) \succ O \Longrightarrow v = 0$, this implies $\mathcal{S}^{\top}(b,-a)u = 0 \Longrightarrow u \in$ Ker $\mathcal{S}^{\top}(b,-a)$. Consequently,

$$\text{Ker } \mathcal{F}(\vartheta) = \text{Ker } \mathcal{S}^{\top}(b,-a) \tag{31}$$

We will now consider the Rank-Nullity Theorem, see e.g., [1], if A is an $m \times n$ matrix, then

$$\dim\ (\text{Ker}\ A) + \dim\ (\text{Im}\ A) = n$$

and the property $\dim\ (\text{Im}\ A) = \dim\ \left(\text{Im}\ A^{\top}\right)$. When applied to the $(p+q) \times (p+q)$ matrix $\mathcal{S}(b,-a)$, it yields

$$\dim\ (\text{Ker}\ \mathcal{S}(b,-a)) = \dim\ \left(\text{Ker}\ \mathcal{S}^{\top}(b,-a)\right) \Longrightarrow\ \dim\ (\text{Ker}\ \mathcal{F}(\vartheta))\ = \dim\ (\text{Ker}\ \mathcal{S}(b,-a))$$

which completes the proof. □

Notice that the dimension of the null space of matrix A is called the nullity of A and the dimension of the image of matrix A, $\dim\ (\text{Im}\ A)$, is termed the rank of matrix A. An alternative proof to the one developed in Corollary 2.2 in [5], is given in a corollary to Lemma 3.1, reconfirming the resultant matrix property of the FIM $\mathcal{F}(\vartheta)$.

Corollary 3.2. *The FIM $\mathcal{F}(\vartheta)$ of an ARMA(p,q) process becomes singular if and only if the autoregressive and moving average polynomials $\widehat{a}(z)$ and $\widehat{b}(z)$ have at least one common root.*

Proof. By virtue of the equality (31) combining with the property $\text{Det}\ \mathcal{S}^{\top}(b,-a) = \text{Det}\ \mathcal{S}(b,-a)$ and the matrix resultant property of the Sylvester matrix $\mathcal{S}(b,-a)$ yields, $\text{Det}\ \mathcal{S}^{\top}(b,-a) = 0 \Longleftrightarrow \text{Ker}\ \mathcal{S}^{\top}(b,-a) \neq \{0\}$ if and only if the ARMA(p,q) polynomials $\widehat{a}(z)$ and $\widehat{b}(z)$ have at least one common root. Equivalently, $\text{Det}\ \mathcal{S}^{\top}(b,-a) \neq 0 \Longleftrightarrow \text{Ker}\ \mathcal{S}^{\top}(b,-a) = \{0\}$ if and only if the ARMA(p,q) polynomials $\widehat{a}(z)$ and $\widehat{b}(z)$ have no common roots. Consequently, by virtue of the equality $\text{Ker}\ \mathcal{F}(\vartheta) = \text{Ker}\ \mathcal{S}^{\top}(b,-a)$ can be concluded, the FIM $\mathcal{F}(\vartheta)$ becomes singular if and only if the ARMA(p,q) polynomials $\widehat{a}(z)$ and $\widehat{b}(z)$ have at least one common root. This completes the proof. □

3.3. The Statistical Distance Measure and the Fisher Information Matrix

In [7] statistical distance measures are studied. Most multivariate statistical techniques are based upon the concept of distance. For that purpose a statistical distance measure is considered that is a normalized Euclidean distance measure with entries of the FIM as weighting coefficients. The measurements x_1, x_2,..., x_n are subject to random fluctuations of different magnitudes and have therefore different variabilities. It is then important to consider a distance that takes the variability of these variables or measurements into account when determining its distance from a fix point. A rotation of the coordinate system through a chosen angle while keeping the scatter of points given by the data fixed, is also applied, see [7] for more details. It is shown that when the FIM is positive definite, the appropriate statistical distance measure is a metric. In case of a singular FIM of an ARMA stationary process, the metric property depends on the rotation angle. The statistical distance measure, is based on m parameters unlike a statistical distance measure introduced in quantum information, see e.g., [8,9], that is also related to the Fisher information but where the information about one parameter in a particular measurement procedure is considered.

The straight-line or Euclidean distance between the stochastic vector $x = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}^{\top}$ and fixed vector $y = \begin{pmatrix} y_1 & y_2 & \dots & y_n \end{pmatrix}^{\top}$ where $x, y \in \mathbb{R}^n$, is given by

$$d(x,y) = \|x-y\| = \left(\sum\nolimits_{j=1}^{n} (x_j - y_j)^2\right)^{1/2} \tag{32}$$

where the metric $d(x,y) := \|x-y\|$ is induced by the standard Euclidean norm $\|\cdot\|$ on $\mathbb{R}^n$, see e.g., [2] for the metric conditions.

The observations $x_1, x_2, \ldots, x_n$ are used to compute maximum likelihood estimated of the parameters $\vartheta_1, \vartheta_2, \ldots, \vartheta_m$ and where $m < n$. These estimated parameters are random variables, see e.g., [15]. The distance of the estimated vector $\vartheta \in \mathbb{R}^m$, given in (15), is studied. Entries of the FIM are inserted in the distance measure as weighting coefficients. The linear transformation

$$\widetilde{\vartheta} = \mathcal{L}_i(\phi)\vartheta \tag{33}$$

is applied, where $\mathcal{L}_i(\phi) \in \mathbb{R}^{m\times n}$ is the Givens rotation matrix with rotation angle ϕ, with $0 \le \phi \le 2\pi$ and $i \in \{1, \ldots, m-1\}$, see e.g., [36], and is given by

$$\mathcal{L}_i(\phi) = \begin{pmatrix} I_{i-1} & 0 & 0 & 0 \\ 0 & (\cos(\phi))_{i,i} & (-\sin(\phi))_{i,i+1} & 0 \\ 0 & (\sin(\phi))_{i+1,i} & (\cos(\phi))_{i+1,i+1} & 0 \\ 0 & 0 & 0 & I_{m-i-1} \end{pmatrix}, \quad 0 \le \phi \le 2\pi \tag{34}$$

The following matrix decomposition is applied in order to obtain a transformed FIM

$$\mathcal{F}_\phi(\vartheta) = \mathcal{L}_i(\phi)\mathcal{F}(\vartheta)\mathcal{L}_i^\top(\phi) \tag{35}$$

where $\mathcal{F}_\phi(\vartheta)$ and $\mathcal{F}(\vartheta)$ are respectively the transformed and untransformed Fisher information matrices. It is straightforward to conclude that by virtue of (35), the transformed and untransformed Fisher information matrices $\mathcal{F}_\phi(\vartheta)$ and $\mathcal{F}(\vartheta)$, are similar since the rotation matrix $\mathcal{L}_i(\phi)$ is orthogonal. Two matrices A and B are similar if there exists an invertible matrix X such that the equality $AX = XB$ holds. As can be seen, the Givens matrix $\mathcal{L}_i(\phi)$ involves only two coordinates that are affected by the rotation angle ϕ whereas the other directions, which correspond to eigenvalues of one, are unaffected by the rotation matrix.

By virtue of (35) can be concluded that a positive definite FIM, $\mathcal{F}(\vartheta) \succ 0$, implies a positive definite transformed FIM, $\mathcal{F}_\phi(\vartheta) \succ 0$. Consequently, the elements on the main diagonal of $\mathcal{F}(\vartheta)$, $f_{1,1}, f_{2,2}, \ldots, f_{m,m}$, as well as the elements on the main diagonal of $\mathcal{F}_\phi(\vartheta)$, $\widetilde{f}_{1,1}, \widetilde{f}_{2,2}, \ldots, \widetilde{f}_{m,m}$ are all positive. However, the elements on the main diagonal of a singular FIM of a stationary ARMA process are also positive.

As developed in [7], combining (33) and (35) yields the distance measure of the estimated parameters $\vartheta_1, \vartheta_2, \ldots, \vartheta_m$ accordingly, to obtain

$$d^2_{\mathcal{F}_\phi}(\vartheta) = \sum_{j=1, j\neq i, i+1}^{m} \left\{ \frac{\vartheta_j^2}{f_{j,j}} \right\} + \frac{\{\vartheta_i \cos(\phi) - \vartheta_{i+1}\sin(\phi)\}^2}{\widetilde{f}_{i,i}(\phi)} + \frac{\{\vartheta_{i+1}\cos(\phi) + \vartheta_i \sin(\phi)\}^2}{\widetilde{f}_{i+1,i+1}(\phi)} \tag{36}$$

where

$$\widetilde{f}_{i,i}(\phi) = f_{i,i}\cos^2(\phi) - f_{i,i+1}\sin(2\phi) + f_{i+1,i+1}\sin^2(\phi) \tag{37}$$

$$\widetilde{f}_{i+1,i+1}(\phi) = f_{i+1,i+1}\cos^2(\phi) + f_{i,i+1}\sin(2\phi) + f_{i,i}\sin^2(\phi) \tag{38}$$

and $f_{j,l}$ are entries of the FIM $\mathcal{F}(\vartheta)$ whereas $\widetilde{f}_{i,i}(\phi)$ and $\widetilde{f}_{i+1,i+1}(\phi)$ are the transformed components since the rotation affects only the entries i, i and $i+1, i+1$ as can be seen in matrix $\mathcal{L}_i(\phi)$. In [7], the existence of the following inequalities is proved

$$\widetilde{f}_{i,i}(\phi) > 0 \quad \text{and} \quad \widetilde{f}_{i+1,i+1}(\phi) > 0$$

this guaratees the metric property of (36). When the FIM of an ARMA(p,q) process is the case, a combination of (27) and (35) for the ARMA(p,q) parameters, given in (15) yields for the transformed FIM,

$$\mathcal{F}_\phi(\vartheta) = \mathcal{S}_\phi(-b,a)\mathcal{P}(\vartheta)\mathcal{S}_\phi^\top(-b,a) \tag{39}$$

where $\mathcal{P}(\vartheta)$ is given by (28) and the transformed Sylvester resultant matrix is of the form

$$\mathcal{S}_\phi(-b,a) = \mathcal{L}_i(\phi)\mathcal{S}(-b,a) \tag{40}$$

Proposition 3.5 in [7], proves that the transformed FIM $\mathcal{F}_\phi(\vartheta)$ and the transformed Sylvester matrix $\mathcal{S}_\phi(-b,a)$ fulfill the resultant matrix property by using the equalities (40) and (39). The following property is then set forth.

Proposition 3.3. *The properties*

$$Ker\ \mathcal{F}_\phi(\vartheta) = Ker\ \mathcal{S}_\phi^\top(-b,a)\ \ and\ \ Ker\ \mathcal{S}_\phi(-b,a) = Ker\ \mathcal{S}(-b,a)$$

hold true.

Proof. By virtue of the equalities (39), (40) and the orthogonality property of the rotation matrix $\mathcal{L}_i(\phi)$ which implies that Ker $\mathcal{L}_i(\phi) = \{0\}$ combined with the same approach as in Lemma 3.1 completes the proof. □

A straightforward conclusion from Proposition 3.3 is then

$$\dim \text{Ker}\ \mathcal{F}_\phi(\vartheta) = \dim \text{Ker}\ \mathcal{S}_\phi(-b,a),\ \dim \text{Ker}\ \mathcal{S}_\phi(-b,a) = \dim \text{Ker}\ \mathcal{S}(-b,a)$$

In the next section a distance measure introduced in quantum information is discussed.

Statistical Distance Measure - Fisher Information and Quantum Information

In quantum information, the Fisher information, the information about a parameter θ in a particular measurement procedure, is expressed in terms of the statistical distance s, see [8–10]. The statistical distance used is defined as a measure to distinguish two probability distributions on the basis of measurement outcomes, see [37]. The Fisher information and the statistical distance are statistical quantities, and generally refer to many measurements as it is the case in this survey. However, in the quantum information theory and quantum statistics context, the problem set up is presented as follows. There may or may not be a small phase change θ, and the question is whether it is there. In that case you can design quantum experiments that will tell you the answer unambiguously in a single measurement. The equality derived is of the form

$$\mathcal{F}(\theta) = \left(\frac{ds}{d\theta}\right)^2 \tag{41}$$

the Fisher information is the square of the derivative of the statistical distance s with respect to θ. Contrary to (36), where the square of the statistical distance measure is expressed in terms of entries of a FIM $\mathcal{F}(\vartheta)$ which is based on information about m parameters estimated from n measurements, for $m < n$. A challenging question could therefore be formulated as follows, can a generalization of equality

(41) be developed in a quantum information context but at the matrix level ? To be more specific, many observations or measurements that lead to more than one parameter such that the corresponding Fisher information matrix is interconnected to an appropriate statistical distance matrix, a matrix where entries are scalar distance measures. This question could equally be a challenge to algebraic matrix theory and to quantum information.

3.4. The Bezoutian - The Fisher Information Matrix

In this section an additional resultant matrix is presented, it concerns the Bezout matrix or Bezoutian. The notation of Lancaster and Tismenetsky [2] shall be used and the results presented are extracted from [38]. Assume the polynomials a and b given by $a(z) = \sum_{j=0}^{n} a_j\, z^j$ and $b(z) = \sum_{j=0}^{n} b_j\, z^j$, cfr. (13) but where $p = q = n$, and we further assume $a_0 = b_0 = 1$. The Bezout matrix $B(a, b)$ of the polynomials a and b is defined by the relation

$$a(z)b(w) - a(w)b(z) = (z - w)u_n^{\top}(z)B(a, b)u_n(z)$$

This matrix is often referred as the Bezoutian. We will display a decomposition of the Bezout matrix $B(a, b)$ developed in [38]. For that purpose the matrix U_ϕ and its inverse T_ϕ are presented, where ϕ is a given complex number, to obtain

$$U_\phi = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -\phi & 1 & \cdots & \cdots & 0 \\ 0 & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & 0 & -\phi & 1 \end{pmatrix}, T_\phi = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ \phi & 1 & \cdots & \cdots & 0 \\ \phi^2 & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \phi^{n-1} & \cdots & \phi^2 & \phi & 1 \end{pmatrix}$$

Let $(1 - \alpha_1 z)$ and $(1 - \beta_1 z)$ be a factor of $a(z)$ and $b(z)$ respectively and α_1 and β_1 are zeros of $\widehat{a}(z)$ and $\widehat{b}(z)$. Consider the factored form of the nth order polynomials $a(z)$ and $b(z)$ of the form $a(z) = (1 - \alpha_1 z)a_{-1}(z)$ and $b(z) = (1 - \beta_1 z)b_{-1}(z)$ respectively. Proceeding this way, for $\alpha_2, \ldots, \alpha_n$ yields the recursion $a_{-(k-1)}(z) = (1 - \alpha_k z)a_{-k}(z)$, equivalently for the polynomials $b_{-k}(z)$ and $a_0(z) = a(z)$ and $b_0(z) = b(z)$. Proposition 3.1 in [38] is presented.

The following non-symmetric decomposition of the Bezoutian is derived, considering the notations above

$$B(a, b) = U_{\alpha_1} \begin{pmatrix} B(a_{-1}, b_{-1}) & 0 \\ 0 & 0 \end{pmatrix} U_{\beta_1}^{\top} + (\beta_1 - \alpha_1) b_{\beta_1} a_{\alpha_1}^{\top} \tag{42}$$

with a_{α_1} such that $a_{\alpha_1}^{\top} u_n(z) = a_{-1}$ similarly for b_{β_1}. Iteration gives the following expansion for the Bezout matrix

$$B(a, b) = \sum_{k=1}^{n} (\beta_k - \alpha_k) U_{\alpha_1} \ldots U_{\alpha_{k-1}} U_{\beta_{k+1}} \ldots U_{\beta_n} e_1^n \left(e_1^n\right)^{\top} U_{\beta_1}^{\top} \ldots U_{\beta_{k-1}}^{\top} U_{\alpha_{k+1}}^{\top} \ldots U_{\alpha_n}^{\top}$$

where e_1^n is the first unit standard basis column vector in $\mathbb{R}^n$, by e_j we denote the jth coordinate vector, $e_j = (0, \ldots, 1, \ldots, 0)^{\top}$, with all its components equal to 0 except the jth component which equals 1. The following corollarys to Proposition 3.1 in [38] are now presented.

Corollary 3.2 in [38] states. Let ϕ be a common zero of the polynomials $\widehat{a}(z)$ and $\widehat{b}(z)$. Then $a(z) = (1-\phi z)a_{-1}(z)$ and $b(z) = (1-\phi z)b_{-1}(z)$ and

$$B(a,b) = U_\phi \begin{pmatrix} B(a_{-1},b_{-1}) & 0 \\ 0 & 0 \end{pmatrix} U_\phi^\top$$

This a direct consequence of (42) and from which can be concluded that the Bezoutian $B(a,b)$ is non-singular if and only if the polynomials $a(z)$ and $b(z)$ have no common factors. A similar conclusion is drawn for the FIM in (27) so that matrices $\mathcal{F}(\vartheta)$ and $B(a,b)$ have the same singularity property.

Related to Corollary 3.2 in[38], this is where we give a description of the kernel or nullspace of the Bezout matrix.

Corollary 3.3 in [38] is now presented. Let $\phi_1, \ldots, \phi_m$ be all the common zeros of the polynomials $\widehat{a}(z)$ and $\widehat{b}(z)$, with multiplicities $n_1, \ldots, n_m$. Let ℓ be the last unit standard basis column vector in $\mathbb{R}^n$ and put

$$w_k^j = \left(T_{\phi_k}^j J^{j-1}\right)^\top \ell$$

for $k = 1, \ldots, m$ and $j = 1, \ldots, n_k$ and by J we denote the forward $n \times n$ shift matrix, $J_{ij} = 1$ if $i = j+1$. Consequently, the subspace Ker $B(a,b)$ is the linear span of the vectors w_k^j.
An alternative representation to (27) but involving the Bezoutian $B(b,a)$ and derived in Proposition 5.1 in [38] is of the form

$$\mathcal{F}(\vartheta) = \mathcal{M}^{-1}(b,a)\mathcal{H}(\vartheta)\mathcal{M}^{-\top}(b,a) \tag{43}$$

where

$$\mathcal{H}(\vartheta) = \begin{pmatrix} I & 0 \\ 0 & B(b,a) \end{pmatrix} Q(\vartheta) \begin{pmatrix} I & 0 \\ 0 & B(b,a) \end{pmatrix} \text{ and } \mathcal{M}(b,a) = \begin{pmatrix} P & 0 \\ PS(\widehat{a})P & PS(\widehat{b})P \end{pmatrix} \tag{44}$$

and

$$P = \begin{pmatrix} 0 & \cdots & 0 & 1 \\ \vdots & \iddots & 1 & 0 \\ 0 & \iddots & \iddots & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}, S(\widehat{a}) = \begin{pmatrix} a_{n-1} & a_{n-2} & \cdots & a_0 \\ a_{n-2} & \iddots & a_0 & 0 \\ \vdots & \iddots & \iddots & \vdots \\ a_0 & 0 & \cdots & 0 \end{pmatrix} \text{ and } Q(\vartheta) \succ 0$$

The matrix $S(\widehat{a})$ is the symmetrizer of the polynomial $\widehat{a}(z)$, in this paper $a_0 = 1$, see [2] and P is a permutation matrix. In [38] it is shown that the matrix $Q(\vartheta)$ is the unique solution to an appropriate Stein equation and is strictly positive definite. However, in the next section an explicit form of the Stein solution $Q(\vartheta)$ is developed. Some comments concerning the property summarized in Corollary 5.2 in [38] follow.

The matrix $\mathcal{H}(\vartheta)$ is non-singular if and only if the polynomials $a(z)$ and $b(z)$ have no common factors. The proof is straightforward since the matrix $Q(\vartheta)$ is non-singular which implies that the matrix $\mathcal{H}(\vartheta)$ is only non-singular when the Bezoutian $B(b,a)$ is non-singular and this is fulfilled if and only if the polynomials $a(z)$ and $b(z)$ have no common factors.
The matrix $\mathcal{M}(b,a)$ is non-singular if $a_0 \neq 0$ and $b_0 \neq 0$, which is the case since we have $a_0 = b_0 = 1$. From (43) can be concluded that the FIM $\mathcal{F}(\vartheta)$ is non-singular only when the matrix $\mathcal{H}(\vartheta)$ is non-singular or by virtue of (44) when the Bezoutian $B(b,a)$ is non-singular. Consequently,

the singularity conditions of the Bezoutian $B(b,a)$, the FIM $\mathcal{F}(\vartheta)$ and the Sylvester resultant matrix $\mathcal{S}(b,-a)$ are therefore equivalent. Can be concluded, by virtue of (29) proved in Lemma 3.1 and the equality dim (Ker $\mathcal{S}(a,b)$) = dim (Ker $B(a,b)$) proved in Theorem 21.11 in [1], yields

$$\text{dim } (\text{Ker } \mathcal{S}(b,-a)) = \text{dim } (\text{Ker } \mathcal{F}(\vartheta)) = \text{dim } (\text{Ker } B(b,a)) = \nu(a,b)$$

3.5. The Stein Equation - The Fisher Information Matrix of an ARMA(p,q) Process

In [12], a link between the FIM of an ARMA process and an appropriate solution of a Stein equation is set forth. In this survey paper we shall present some of the results and confront some results displayed in the previous sections. However, alternative proofs will be given to some results obtained in [12,38].

The Stein matrix equation is now set forth. Let $A \in \mathbb{C}^{m\times m}$, $B \in \mathbb{C}^{n\times n}$ and $\Gamma \in \mathbb{C}^{n\times m}$ and consider the Stein equation

$$S - BSA^{\top} = \Gamma \tag{45}$$

It has a unique solution if and only if $\lambda\mu \neq 1$ for any $\lambda \in \sigma(A)$ and $\mu \in \sigma(B)$, the spectrum of D is $\sigma(D) = \{\lambda \in \mathbb{C}\text{: } \det(\lambda I_m - D) = 0\}$, the set of eigenvalues of D. The unique solution will be given in the next theorem [11].

Theorem 3.4. *Let A and B be, such that there is a single closed contour C with $\sigma(B)$ inside C and for each non-zero $w \in \sigma(A)$, w^{-1} is outside C. Then for an arbitrary Γ the Stein Equation (45) has a unique solution S*

$$S = \frac{1}{2\pi \mathbf{i}} \oint_C (\lambda I_n - B)^{-1}\Gamma(I_m - \lambda A)^{-\top} d\lambda \tag{46}$$

In this section an interconnection between the representation (27) of the FIM $\mathcal{F}(\vartheta)$ and an appropriate solution to a Stein equation of the form (45) as developed in [12] is set forth. The distinct roots of the polynomials $\widehat{a}(z)$ and $\widehat{b}(z)$ are denoted by $\alpha_1, \alpha_2, \ldots, \alpha_p$ and $\beta_1, \beta_2, \ldots, \beta_q$ respectively such that the non-singularity of the FIM $\mathcal{F}(\vartheta)$ is guaranteed. The following representation of the integral expression (28) is given when Cauchy's residue theorem is applied, equation (4.8) in [12]

$$\mathcal{P}(\vartheta) = \mathcal{U}(\vartheta)\mathcal{D}(\vartheta)\widehat{\mathcal{U}}(\vartheta) \tag{47}$$

where

$$\mathcal{U}(\vartheta) = \left\{u_{p+q}(\alpha_1), u_{p+q}(\alpha_2), \ldots, u_{p+q}(\alpha_p), u_{p+q}(\beta_1), u_{p+q}(\beta_2), \ldots, u_{p+q}(\beta_q)\right\}$$

$$\mathcal{D}(\vartheta) = \text{diag}\left\{\left(\frac{1}{\widehat{a}(z;\alpha_i)\widehat{b}(\alpha_i)a(\alpha_i)b(\alpha_i)}\right), \left(\frac{1}{\widehat{a}(\beta_j)\widehat{b}(z;\beta_j)a(\beta_j)b(\beta_j)}\right)\right\}, i = 1,\ldots,p \text{ and } j = 1,\ldots,q$$

and

$$\widehat{\mathcal{U}}(\vartheta) = \left\{v_{p+q}(\alpha_1), v_{p+q}(\alpha_2), \ldots, v_{p+q}(\alpha_p), v_{p+q}(\beta_1), v_{p+q}(\beta_2), \ldots, v_{p+q}(\beta_q)\right\}^{\top}$$

the polynomial $p(\cdot;\beta)$ is defined accordingly, $p(z;\beta) = \dfrac{p(z)}{(z-\beta)}$ and $\mathcal{D}(\vartheta)$ is the $(p+q)\times(p+q)$ diagonal matrix. The matrices $\mathcal{U}(\vartheta)$ and $\widehat{\mathcal{U}}(\vartheta)$ in (47) are the $(p+q)\times(p+q)$ Vandermonde matrices $V_{\alpha\beta}$ and $\widehat{V}_{\alpha\beta}$ respectively, given by

$$V_{\alpha\beta} = \begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^{p+q-1} \\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^{p+q-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \alpha_p & \alpha_p^2 & \cdots & \alpha_p^{p+q-1} \\ 1 & \beta_1 & \beta_1^2 & \cdots & \beta_1^{p+q-1} \\ 1 & \beta_2 & \beta_2^2 & \cdots & \beta_2^{p+q-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \beta_q & \beta_q^2 & \cdots & \beta_q^{p+q-1} \end{pmatrix}^{\top} \text{ and } \widehat{V}_{\alpha\beta} = \begin{pmatrix} \alpha_1^{p+q-1} & \alpha_1^{p+q-2} & \cdots & \alpha_1 & 1 \\ \alpha_2^{p+q-1} & \alpha_2^{p+q-2} & \cdots & \alpha_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_p^{p+q-1} & \alpha_p^{p+q-2} & \cdots & \alpha_p & 1 \\ \beta_1^{p+q-1} & \beta_1^{p+q-2} & \cdots & \beta_1 & 1 \\ \beta_2^{p+q-1} & \beta_2^{p+q-2} & \cdots & \beta_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \beta_q^{p+q-1} & \beta_q^{p+q-2} & \cdots & \beta_q & 1 \end{pmatrix}$$

It is clear that the $(p+q)\times(p+q)$ Vandermonde matrices $V_{\alpha\beta}$ and $\widehat{V}_{\alpha\beta}$ are nonsingular when $\alpha_i \neq \alpha_j$, $\beta_k \neq \beta_h$ and $\alpha_i \neq \beta_k$ for all $i,j = 1,\ldots,p$ and $k,h = 1,\ldots,q$. A rigorous systematic evaluation of the Vandermonde determinants $\mathrm{Det}V_{\alpha\beta}$ and $\mathrm{Det}\widehat{V}_{\alpha\beta}$, yields

$$\mathrm{Det}V_{\alpha\beta} = (-1)^{(p+q)(p+q-1)/2}\,\Phi\left(\alpha_i,\beta_k\right)$$

where

$$\Phi\left(\alpha_i,\beta_k\right) = \prod_{1\leq i<j\leq p}(\alpha_i-\alpha_j)\prod_{1\leq k<h\leq q}(\beta_k-\beta_h)\prod_{\substack{m=1,\ldots p\\ n=1,\ldots q}}(\alpha_m-\beta_n)$$

Since $V_{\alpha\beta} = P\widehat{V}_{\alpha\beta}^{\top}$ and given the configuration of the permutation matrix, P, this leads to the equalities $\mathrm{Det}\widehat{V}_{\alpha\beta}^{\top}$ =$\mathrm{Det}P\ \mathrm{Det}V_{\alpha\beta}$ and $\mathrm{Det}P = (-1)^{(p+q)(p+q-1)/2}$ so that

$$\mathrm{Det}\widehat{V}_{\alpha\beta} = (-1)^{(p+q)(p+q-1)}\,\Phi\left(\alpha_i,\beta_k\right) \Longrightarrow \left|\mathrm{Det}V_{\alpha\beta}\right| = \left|\mathrm{Det}\widehat{V}_{\alpha\beta}\right|$$

We shall now introduce an appropriate Stein equation of the form (45) such that an interconnection with $\mathcal{P}(\vartheta)$ in (47) can be verified. Therefore the following $(p+q)\times(p+q)$ companion matrix is introduced,

$$\mathcal{C}_g = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \\ -g_{p+q} & -g_{p+q-1} & \cdots & -g_1 \end{pmatrix} \tag{48}$$

where the entries g_i are given by $z^{p+q} + \sum_{i=1}^{p+q} g_i(\vartheta)z^{p+q-i} = \widehat{a}(z)\widehat{b}(z) = \widehat{g}(z,\vartheta)$ and $\widehat{g}(\vartheta)$ is the vector $\widehat{g}(\vartheta) = \left(g_{p+q}(\vartheta), g_{p+q-1}(\vartheta),\ldots,g_1(\vartheta)\right)^{\top}$. Likewise is the vector $g(z,\vartheta) = a(z)b(z)$ and $g(\vartheta) = \left(g_1(\vartheta),g_1(\vartheta),\ldots,g_{p+q}(\vartheta)\right)^{\top}$, for investigating the properties of a companion matrix see e.g., [36], [2]. Since all the roots of the polynomials $\widehat{a}(z)$ and $\widehat{b}(z)$ are distinct and lie within the unit circle implies that the products $\alpha_i\beta_j \neq 1$, $\alpha_i\alpha_j \neq 1$ and $\beta_i\beta_j \neq 1$ hold for all $i = 1,2,\ldots,p$ and $j = 1,2,\ldots,q$. Consequently, the uniqueness condition of the solution of an appropriate Stein equation is verified. The following Stein equation and its solution, according to (45) and (46), are now presented

$$S - \mathcal{C}_g S \mathcal{C}_g^{\top} = \Gamma \text{ and } S = \frac{1}{2\pi\mathbf{i}}\oint_{|z|=1}(zI_{p+q}-\mathcal{C}_g)^{-1}\Gamma(I_{p+q}-z\mathcal{C}_g)^{-\top}dz$$

where the closed contour is now the unit circle $|z| = 1$ and the matrix Γ is of size $(p+q) \times (p+q)$. A more explicit expression of the solution S is of the form

$$S = \frac{1}{2\pi \mathbf{i}} \oint_{|z|=1} \frac{\text{adj}(zI_{p+q} - \mathcal{C}_g)\Gamma \, \text{adj}(I_{p+q} - z\mathcal{C}_g)^\top}{a(z)b(z)\widehat{a}(z)\widehat{b}(z)} dz \tag{49}$$

where $\text{adj}(X) = X^{-1}\text{Det}(X)$, the adjoint of matrix X. When Cauchy's residue theorem is applied to the solution S in (49), the following factored form of S is derived, equation (4.9) in [12]

$$S = (\mathcal{C}_1, \mathcal{C}_2) \, (I_{p+q} \otimes \Gamma) \, (\mathcal{D}(\vartheta) \otimes I_{p+q}) \, (\mathcal{C}_3, \mathcal{C}_4)^\top \tag{50}$$

where

$$\begin{aligned}
\mathcal{C}_1 &= \text{adj}(\alpha_1 I_{p+q} - \mathcal{C}_g), \text{adj}(\alpha_2 I_{p+q} - \mathcal{C}_g), \ldots, \text{adj}(\alpha_p I_{p+q} - \mathcal{C}_g) \\
\mathcal{C}_2 &= \text{adj}(\beta_1 I_{p+q} - \mathcal{C}_g), \text{adj}(\beta_2 I_{p+q} - \mathcal{C}_g), \ldots, \text{adj}(\beta_p I_{p+q} - \mathcal{C}_g) \\
\mathcal{C}_3 &= \text{adj}(I_{p+q} - \alpha_1 \mathcal{C}_g), \text{adj}(I_{p+q} - \alpha_2 \mathcal{C}_g), \ldots, \text{adj}(I_{p+q} - \alpha_p \mathcal{C}_g) \\
\mathcal{C}_4 &= \text{adj}(I_{p+q} - \beta_1 \mathcal{C}_g), \text{adj}(I_{p+q} - \beta_2 \mathcal{C}_g), \ldots, \text{adj}(I_{p+q} - \beta_p \mathcal{C}_g)
\end{aligned}$$

and $\mathcal{D}(\vartheta)$ is given in (47), the following matrix rule is applied

$$(A \otimes B)(C \otimes D) = AC \otimes BD$$

and the operator $\otimes$ is the tensor (Kronecker) product of two matrices, see e.g., [2], [20].

Combining (47) and (50) and taking the assumption, $\alpha_i \neq \alpha_j$, $\beta_k \neq \beta_h$ and $\alpha_i \neq \beta_k$, into account implies that the inverse of the $(p+q) \times (p+q)$ Vandermonde matrices $V_{\alpha\beta}$ and $\widehat{V}_{\alpha\beta}$ exist, as Lemma 4.2 [12] states.

The following equality holds true

$$S = (\mathcal{C}_1, \mathcal{C}_2) \left(V_{\alpha\beta}^{-1} \mathcal{P}(\vartheta) \widehat{V}_{\alpha\beta}^{-1} \otimes \Gamma \right) (\mathcal{C}_3, \mathcal{C}_4)^\top$$

or

$$S = (\mathcal{C}_1, \mathcal{C}_2) \left(V_{\alpha\beta}^{-1} \mathcal{S}^{-1}(b, -a) \mathcal{F}(\vartheta) \mathcal{S}^{-\top}(b, -a) \widehat{V}_{\alpha\beta}^{-1} \otimes \Gamma \right) (\mathcal{C}_3, \mathcal{C}_4)^\top \tag{51}$$

Consequently, under the condition $\alpha_i \neq \alpha_j$, $\beta_k \neq \beta_h$ and $\alpha_i \neq \beta_k$, and by virtue of (27) and (51), an interconnection involving the FIM $\mathcal{F}(\vartheta)$, a solution to an appropriate Stein equation S, the Sylvester matrix $\mathcal{S}(b, -a)$ and the Vandermonde matrices $V_{\alpha\beta}$ and $\widehat{V}_{\alpha\beta}$ is established. It is clear that by using the expression (43), the Bezoutian $B(a, b)$ can be inserted in equality (51).

We will formulate a Stein equation when the matrix $\Gamma = e_{p+q} e_{p+q}^\top$,

$$S - \mathcal{C}_g S \mathcal{C}_g^\top = e_{p+q} e_{p+q}^\top \tag{52}$$

where e_{p+q} is the last standard basis column vector in $\mathbb{R}^{p+q}$, e_i^m is the i-th unit standard basis column vector in $\mathbb{R}^m$, with all its components equal to 0 except the i-th component which equals 1. The next lemma is formulated.

Lemma 3.5. *The symmetric matrix $\mathcal{P}(\vartheta)$ defined in (28) fulfills the Stein Equation (52).*

Proof. The unique solution of (52) is according to (46)

$$S = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} (zI_{p+q} - \mathcal{C}_g)^{-1} e_{p+q} e_{p+q}^{\top} (I_{p+q} - z\mathcal{C}_g)^{-\top} dz$$

more explictely written,

$$S = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{\mathrm{adj}(zI_{p+q} - \mathcal{C}_g) e_{p+q} e_{p+q}^{\top} \mathrm{adj}(I_{p+q} - z\mathcal{C}_g)^{\top}}{a(z)b(z)\widehat{a}(z)\widehat{b}(z)} dz$$

Using the property of the companion matrix $\mathcal{C}_g$, standard computation shows that the last column of $\mathrm{adj}(zI_{p+q} - \mathcal{C}_g)$ is the basic vector $u_{p+q}(z)$ and consequently the last column of $\mathrm{adj}(I_{p+q} - z\mathcal{C}_g)$ is the basic vector $v_{p+q}(z) = z^{p+q-1} u_{p+q}(z^{-1})$. This implies that $\mathrm{adj}(zI_{p+q} - \mathcal{C}_g) e_{p+q} = u_{p+q}(z)$ and $e_{p+q}^{\top} \mathrm{adj}(I_{p+q} - z\mathcal{C}_g)^{\top} = v_{p+q}^{\top}(z)$ or

$$S = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{u_{p+q}(z) v_{p+q}^{\top}(z)}{a(z)b(z)\widehat{a}(z)\widehat{b}(z)} dz = \mathcal{P}(\vartheta)$$

Consequently, the solution S to the Stein Equation (52) coincides with the matrix $\mathcal{P}(\vartheta)$ defined in (28). □

The Stein equation that is verified by the FIM $\mathcal{F}(\vartheta)$ will be considered. For that purpose we display the following $p \times p$ and $q \times q$ companion matrices $\mathcal{C}_a$ and $\mathcal{C}_b$ of the form,

$$\mathcal{C}_a = \begin{pmatrix} -a_1 & -a_2 & \cdots & \cdots & -a_p \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}, \mathcal{C}_b = \begin{pmatrix} -b_1 & -b_2 & \cdots & \cdots & -b_q \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

respectively. Introduce the $(p+q) \times (p+q)$ matrix $\mathcal{K}(\vartheta) = \begin{pmatrix} \mathcal{C}_a & O \\ O & \mathcal{C}_b \end{pmatrix}$ and the $(p+q) \times 1$ vector $\mathcal{B} = \begin{pmatrix} e_p^1 \\ -e_q^1 \end{pmatrix}$, where e_p^1 and e_q^1 are the first standard basis column vectors in $\mathbb{R}^p$and $\mathbb{R}^q$ respectively. Consider the Stein equation

$$S - \mathcal{K}(\vartheta) S \mathcal{K}^{\top}(\vartheta) = \mathcal{B}\mathcal{B}^{\top} \tag{53}$$

followed by the theorem.

Theorem 3.6. *The Fisher information matrix $\mathcal{F}(\vartheta)$ (17) coincides with the solution to the Stein equation (53).*

Proof. The eigenvalues of the companion matrices $\mathcal{C}_a$ and $\mathcal{C}_b$ are respectively the zeros of the polynomials $\widehat{a}(z)$ and $\widehat{b}(z)$ which are in absolute value smaller than one. This implies that the unique solution of the Stein Equation (53) exists and is given by

$$S = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} (zI_{p+q} - \mathcal{K}(\vartheta))^{-1} \mathcal{B}\mathcal{B}^{\top} (I_{p+q} - z\mathcal{K}(\vartheta))^{-\top} dz$$

developing this integral expression in a more explicit form yields

$$S = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \begin{pmatrix} \frac{\operatorname{adj}(zI_p - \mathcal{C}_a)}{\widehat{a}(z)} & O \\ O & \frac{\operatorname{adj}(zI_q - \mathcal{C}_b)}{\widehat{b}(z)} \end{pmatrix} \begin{pmatrix} e_p^1 \\ -e_q^1 \end{pmatrix} \left\{ \begin{pmatrix} \frac{\operatorname{adj}(I_p - z\mathcal{C}_a)}{a(z)} & O \\ O & \frac{\operatorname{adj}(I_q - z\mathcal{C}_b)}{b(z)} \end{pmatrix} \begin{pmatrix} e_p^1 \\ -e_q^1 \end{pmatrix} \right\}^\top dz$$

Considering the form of the companion matrices $\mathcal{C}_a$ and $\mathcal{C}_b$ leads through straightforward computation to the conclusion, the first column of $\operatorname{adj}(zI_p - \mathcal{C}_a)$ is the basic vector $v_p(z)$ and consequently the first column of $\operatorname{adj}(I_p - z\mathcal{C}_a)$ is the basic vector $u_p(z)$. Equivalently for the companion matrix $\mathcal{C}_b$, this yields

$$S = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \begin{pmatrix} \frac{v_p(z)}{\widehat{a}(z)} \\ -\frac{v_q(z)}{\widehat{b}(z)} \end{pmatrix} \begin{pmatrix} \frac{u_p^\top(z)}{a(z)} & -\frac{u_q^\top(z)}{b(z)} \end{pmatrix} dz \tag{54}$$

Representation (54) is such that in order to obtain an equivalent representation to the FIM $\mathcal{F}(\vartheta)$ in (17), the transpose of the solution to the Stein Equation (53) is therefore required, to obtain

$$S^\top = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \begin{pmatrix} \frac{u_p(z)v_p^\top(z)}{a(z)\widehat{a}(z)} & -\frac{u_p(z)v_q^\top(z)}{a(z)\widehat{b}(z)} \\ -\frac{u_q(z)v_p^\top(z)}{\widehat{a}(z)b(z)} & \frac{u_q(z)v_q^\top(z)}{b(z)\widehat{b}(z)} \end{pmatrix} dz = \mathcal{F}(\vartheta) \tag{55}$$

or

$$S^\top = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} (I_{p+q} - z\mathcal{K}(\vartheta))^{-1} \mathcal{B}\mathcal{B}^\top (zI_{p+q} - \mathcal{K}(\vartheta))^{-\top} dz = \mathcal{F}(\vartheta)$$

The symmetry property of the FIM $\mathcal{F}(\vartheta)$, leads to $S = \mathcal{F}(\vartheta)$. From the representation (55) can be concluded that the solution S of the Stein Equation (53) coincides with the symmetric block Toeplitz FIM $\mathcal{F}(\vartheta)$ given in (17). This completes the proof. □

It is straightforward to verify that the submatrix (1,2) in (55) is the complex conjugate transpose of the submatrix (2,1), whereas each submatrix on the main diagonal is Hermitian, consequently, the integrand is Hermitian. This implies that when the standard residue theorem is applied, it yields $\mathcal{F}(\vartheta) = \mathcal{F}^\top(\vartheta)$.

An Illustrative Example of Theorem 3.6

To illustrate Theorem 3.6, the case of an ARMA$(2,2)$ process is considered. We will use the representation (17) for computing the FIM $\mathcal{F}(\vartheta)$ of an ARMA$(2,2)$ process. The autoregressive and moving average polynomials are of degree two or $p = q = 2$ and the ARMA$(2,2)$ process is described by,

$$y(t)a(z) = b(z)\varepsilon(t) \tag{56}$$

where $y(t)$ is the stationary process driven by white noise $\varepsilon(t)$, $a(z) = (1 + a_1 z + a_2 z^2)$ and $b(z) = (1 + b_1 z + b_2 z^2)$ and the parameter vector is $\vartheta = (a_1, a_2, b_1, b_2)^\top$. The condition, the zeros of the polynomials

$$\widehat{a}(z) = z^2 a(z^{-1}) = z^2 + a_1 z + a_2 \text{ and } \widehat{b}(z) = z^2 b(z^{-1}) = z^2 + b_1 z + b_2$$

are in absolute value smaller than one, is imposed. The FIM $\mathcal{F}(\vartheta)$ of the ARMA$(2,2)$ process (56) is of the form

$$\mathcal{F}(\vartheta) = \begin{pmatrix} \mathcal{F}_{aa}(\vartheta) & \mathcal{F}_{ab}(\vartheta) \\ \mathcal{F}_{ab}^{\top}(\vartheta) & \mathcal{F}_{bb}(\vartheta) \end{pmatrix} \tag{57}$$

where

$$\mathcal{F}_{aa}(\vartheta) = \frac{1}{(1-a_2)\left[(1+a_2)^2 - a_1^2\right]} \begin{pmatrix} 1+a_2 & -a_1 \\ -a_1 & 1+a_2 \end{pmatrix}$$

$$\mathcal{F}_{bb}(\vartheta) = \frac{1}{(1-b_2)\left[(1+b_2)^2 - b_1^2\right]} \begin{pmatrix} 1+b_2 & -b_1 \\ -b_1 & 1+b_2 \end{pmatrix}$$

$$\mathcal{F}_{ab}(\vartheta) = \frac{1}{(a_2b_2-1)^2 + (a_2b_1 - a_1)(b_1 - a_1b_2)} \begin{pmatrix} a_2b_2 - 1 & a_1 - a_2b_1 \\ b_1 - a_1b_2 & a_2b_2 - 1 \end{pmatrix}$$

The submatrices $\mathcal{F}_{aa}(\vartheta)$ and $\mathcal{F}_{bb}(\vartheta)$ are symmetric and Toeplitz whereas $\mathcal{F}_{ab}(\vartheta)$ is Toeplitz. One can assert that without any loss of generality, the property, symmetric block Toeplitz, holds for the class of Fisher information matrices of stationary ARMA(p,q) processes, where p and q are arbitrary, finite integers that represent the degrees of the autoregressive and moving average polynomials, respectively. The appropriate companion matrices $\mathcal{C}_a$, $\mathcal{C}_b$, the 4×4 matrices $\mathcal{K}(\vartheta)$ and $\mathcal{B}\mathcal{B}^{\top}$ are

$$\mathcal{C}_a = \begin{pmatrix} -a_1 & -a_2 \\ 1 & 0 \end{pmatrix}, \mathcal{C}_b = \begin{pmatrix} -b_1 & -b_2 \\ 1 & 0 \end{pmatrix}, \mathcal{K}(\vartheta) = \begin{pmatrix} -a_1 & -a_2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & -b_1 & -b_2 \\ 0 & 0 & 1 & 0 \end{pmatrix} \text{ and } \mathcal{B}\mathcal{B}^{\top} = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{58}$$

where $\mathcal{B} = \begin{pmatrix} 1 & 0 & -1 & 0 \end{pmatrix}^{\top}$. It can be verified that the Stein equation

$$\mathcal{F}(\vartheta) - \mathcal{K}(\vartheta)\mathcal{F}(\vartheta)\mathcal{K}^{\top}(\vartheta) = \mathcal{B}\mathcal{B}^{\top}$$

holds true, when $\mathcal{F}(\vartheta)$ is of the form (57) and the matrices $\mathcal{K}(\vartheta)$ and $\mathcal{B}\mathcal{B}^{\top}$ are given in (58).

3.5.1. Some Additional Results

In Proposition 5.1 in [38], the matrix $Q(\vartheta)$ in (44) fulfills the Stein Equation (59) and the property $Q(\vartheta) \succ 0$ is proved. It states that when $e_P^{\top} = \left(e_1^{\top}P, 0\right)^{\top} = (e_n, 0_n)^{\top} \in \mathbb{R}^{2n}$, where e_1 is the first unit standard basis column vector in $\mathbb{R}^n$ and e_n is the last or n-th unit standard basis column vector in $\mathbb{R}^n$, the following Stein equation admits the form

$$Q(\vartheta) = F_N(\vartheta)Q(\vartheta)F_N^{\top}(\vartheta) + e_P e_P^{\top} \tag{59}$$

where

$$F_N(\vartheta) = \begin{pmatrix} \widehat{\mathcal{C}}_a & 0 \\ e_1 e_1^{\top} & \mathcal{C}_b \end{pmatrix}, \widehat{\mathcal{C}}_a = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & & \ddots & \vdots \\ 0 & \ddots & \ddots & & 1 \\ -a_p & -a_{p-1} & \cdots & \cdots & -a_1 \end{pmatrix}$$

A corollary to Proposition 5.1, [38] will be set forth, the involvement of various Vandermonde matrices in the explicit solution to equation (59) is confirmed. For that purpose the following Vandermonde matrices are displayed,

$$V_\alpha = \begin{pmatrix} 1 & 1 & 1 \\ \alpha_1 & \alpha_2 & \alpha_n \\ \alpha_1^2 & \alpha_2^2 & \alpha_n^2 \\ \vdots & \vdots & \vdots \\ \alpha_1^{n-1} & \alpha_2^{n-1} & \alpha_n^{n-1} \end{pmatrix}, \widehat{V}_\alpha = \begin{pmatrix} \alpha_1^{n-1} & \alpha_1^{n-2} & 1 \\ \alpha_2^{n-1} & \alpha_2^{n-2} & 1 \\ \alpha_3^{n-1} & \alpha_3^{n-2} & 1 \\ \vdots & \vdots & \vdots \\ \alpha_n^{n-1} & \alpha_n^{n-2} & 1 \end{pmatrix}, \widehat{V}_{\alpha\beta} = \begin{pmatrix} \widehat{V}_\alpha \\ \widehat{V}_\beta \end{pmatrix}, \text{and } V_{\alpha\beta} = \begin{pmatrix} V_\alpha & V_\beta \end{pmatrix} \tag{60}$$

where $\widehat{V}_\beta$ and V_β have the same configuration as $\widehat{V}_\alpha$ and V_α respectively. A corollary to Proposition 5.1 in [38] is now formulated.

Corollary 3.7. *An explicit expression of the solution to the Stein equation (59) is of the form*

$$Q(\vartheta) = \begin{pmatrix} V_\alpha \mathcal{D}_{11}(\vartheta) \widehat{V}_\alpha & V_\alpha \mathcal{D}_{12}(\vartheta) V_\alpha^\top \\ \widehat{V}_{\alpha\beta}^\top \mathcal{D}_{21}(\vartheta) \widehat{V}_{\alpha\beta} & \widehat{V}_{\alpha\beta}^\top \mathcal{D}_{22}(\vartheta) V_{\alpha\beta}^\top \end{pmatrix} \tag{61}$$

where the $n \times n$ and $2n \times 2n$ diagonal matrices $\mathcal{D}_{kl}(\vartheta)$ shall be specified in the proof.

Proof. The condition of a unique solution of the Stein Equation (59) is guaranteed since the eigenvalues of the companions matrices $\widehat{\mathcal{C}}_a$ and $\mathcal{C}_b$ given respectively by the zeros of the polynomials $\widehat{a}(z)$ and $\widehat{b}(z)$ are in absolute value smaller than one. Consequently, the unique solution to the Stein Equation (59) exists and is given by

$$Q(\vartheta) = \frac{1}{2\pi \mathbf{i}} \oint_{|z|=1} (zI_{2n} - F_N(\vartheta))^{-1} e_P e_P^\top (I_{2n} - zF_N(\vartheta))^{-\top} dz \tag{62}$$

in order to proceed successfully, the following matrix property is displayed, to obtain

$$\begin{pmatrix} A & O \\ B & C \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} & O \\ -C^{-1}BA^{-1} & C^{-1} \end{pmatrix}$$

When applied to the Equation (62), it yields

$$Q(\vartheta) = \frac{1}{2\pi \mathbf{i}} \oint_{|z|=1} \begin{pmatrix} \dfrac{\operatorname{adj}(zI_p - \widehat{\mathcal{C}}_a)}{\widehat{a}(z)} & O \\ \dfrac{\operatorname{adj}(zI_q - \mathcal{C}_b) e_1 e_1^\top \operatorname{adj}(zI_p - \widehat{\mathcal{C}}_a)}{\widehat{a}(z)\widehat{b}(z)} & \dfrac{\operatorname{adj}(zI_q - \mathcal{C}_b)}{\widehat{b}(z)} \end{pmatrix} \begin{pmatrix} e_n \\ 0 \end{pmatrix} \times$$
$$\left\{ \begin{pmatrix} \dfrac{\operatorname{adj}(I_n - z\widehat{\mathcal{C}}_a)}{a(z)} & O \\ \dfrac{\operatorname{adj}(I_n - z\mathcal{C}_b) e_1 e_1^\top \operatorname{adj}(I_p - z\widehat{\mathcal{C}}_a)}{a(z)b(z)} & \dfrac{\operatorname{adj}(I_n - z\mathcal{C}_b)}{b(z)} \end{pmatrix} \begin{pmatrix} e_n \\ 0 \end{pmatrix} \right\}^\top dz$$

Considering that the last column vector of the matrices $\mathrm{adj}(zI_p - \widehat{\mathcal{C}}_a)$ and $\mathrm{adj}(I_n - z\widehat{\mathcal{C}}_a)$ are the vectors $u_n(z)$ and $v_n(z)$ respectively, it then yields

$$\begin{aligned} Q(\vartheta) &= \frac{1}{2\pi\mathbf{i}} \oint_{|z|=1} \begin{pmatrix} \dfrac{u_n(z)}{\widehat{a}(z)} \\ \dfrac{v_n(z)}{\widehat{a}(z)\widehat{b}(z)} \end{pmatrix} \begin{pmatrix} \dfrac{v_n^\top(z)}{a(z)} & \dfrac{z^{n-1}u_n^\top(z)}{a(z)b(z)} \end{pmatrix} dz \\ &= \frac{1}{2\pi\mathbf{i}} \oint_{|z|=1} \begin{pmatrix} \dfrac{u_n(z)v_n^\top(z)}{a(z)\widehat{a}(z)} & \dfrac{z^{n-1}u_n(z)u_n^\top(z)}{\widehat{a}(z)a(z)b(z)} \\ \dfrac{v_n(z)v_n^\top(z)}{\widehat{a}(z)\widehat{b}(z)a(z)} & \dfrac{z^{n-1}v_n(z)u_n^\top(z)}{\widehat{a}(z)\widehat{b}(z)a(z)b(z)} \end{pmatrix} dz = \begin{pmatrix} Q_{11}(\vartheta) & Q_{12}(\vartheta) \\ Q_{21}(\vartheta) & Q_{22}(\vartheta) \end{pmatrix} \end{aligned}$$

Applying the standard residue theorem leads for the respective submatrices

$$Q_{11}(\vartheta) = \{u_n(\alpha_1), \ldots, u_n(\alpha_n)\}\, \mathcal{D}_{11}(\vartheta)\, \{v_n(\alpha_1), \ldots, v_n(\alpha_n)\}^\top$$

$$Q_{12}(\vartheta) = \{u_n(\alpha_1), \ldots, u_n(\alpha_n)\}\, \mathcal{D}_{12}(\vartheta)\, \{u_n(\alpha_1), \ldots, u_n(\alpha_n)\}^\top$$

$$Q_{21}(\vartheta) = \{v_n(\alpha_1), \ldots, v_n(\alpha_n), v_n(\beta_1), \ldots, v_n(\beta_n)\}\, \mathcal{D}_{21}(\vartheta)\, \{v_n(\alpha_1), \ldots, v_n(\alpha_n), v_n(\beta_1), \ldots, v_n(\beta_n)\}^\top$$

$$Q_{22}(\vartheta) = \{v_n(\alpha_1), \ldots, v_n(\alpha_n), v_n(\beta_1), \ldots, v_n(\beta_n)\}\, \mathcal{D}_{22}(\vartheta)\, \{u_n(\alpha_1), \ldots, u_n(\alpha_n), u_n(\beta_1), \ldots, u_n(\beta_n)\}^\top$$

where the $n \times n$ diagonal matrices are

$$\mathcal{D}_{11}(\vartheta) = \mathrm{diag}\left\{1/\left(a(\alpha_i)\widehat{a}(z;\alpha_i)\right)\right\}, \mathcal{D}_{12}(\vartheta) = \mathrm{diag}\left\{\alpha_i^{n-1}/\left(a(\alpha_i)b(\alpha_i)\widehat{a}(z;\alpha_i)\right)\right\} \text{ for } i = 1, \ldots, n$$

and the $2n \times 2n$ diagonal matrices are

$$\mathcal{D}_{21}(\vartheta) = \mathrm{diag}\left\{1/\left(a(\alpha_i)\widehat{b}(\alpha_i)\widehat{a}(z;\alpha_i)\right), 1/\left(\widehat{a}(\beta_j)a(\beta_j)\widehat{b}(z;\beta_j)\right)\right\}, \text{ for } i, j = 1, \ldots, n$$

$$\mathcal{D}_{22}(\vartheta) = \mathrm{diag}\left\{\alpha_i^{n-1}/\left(a(\alpha_i)b(\alpha_i)\widehat{b}(\alpha_i)\widehat{a}(z;\alpha_i)\right), \beta_j^{n-1}/\left(\widehat{a}(\beta_j)a(\beta_j)b(\beta_j)\widehat{b}(z;\beta_j)\right)\right\}, \text{ for } i, j = 1, \ldots, n$$

It is clear that the first and third matrices in $Q_{11}(\vartheta)$, $Q_{12}(\vartheta)$, $Q_{21}(\vartheta)$ and $Q_{22}(\vartheta)$ are the appropriate Vandermonde matrices displayed in (60), it can be concluded that the representation (61) is verified. This completes the proof. □

In this section an explicit form of the solution $Q(\vartheta)$, expressed in terms of various Vandermonde matrices, is displayed. Also, an interconnection between the Fisher information $\mathcal{F}(\vartheta)$ and appropriate solutions to Stein equations and related matrices is presented. Proofs are given when the Stein equations are verified by the FIM $\mathcal{F}(\vartheta)$ and the associated matrix $\mathcal{P}(\vartheta)$. These are alternative to the proofs developed in [38]. The presence of various forms of Vandermonde matrices is also emphasized. In the next section some matrix properties of the FIM $\mathcal{F}(\vartheta)$ of an ARMAX process is presented.

3.6. The Fisher Information Matrix of an ARMAX(p, r, q) Process

The FIM of the ARMAX process (11) is set forth according to [4]. The derivatives in the corresponding representation (16) are

$$\frac{\partial \varepsilon_t(\vartheta)}{\partial a_j} = \frac{c(z)}{a(z)b(z)}x(t-j) + \frac{1}{a(z)}\varepsilon(t-j),\ \frac{\partial \varepsilon_t(\vartheta)}{\partial c_l} = -\frac{1}{b(z)}\varepsilon(t-l) \text{ and } \frac{\partial \varepsilon_t(\vartheta)}{\partial b_k} = -\frac{1}{b(z)}\varepsilon_{t-k}$$

where $j = 1, \ldots, p$, $l = 1, \ldots, r$ and $k = 1, \ldots, q$. Combining all j, l and k yields the $(p + r + q) \times (p + r + q)$ FIM

$$\mathcal{G}(\vartheta) = \begin{pmatrix} \mathcal{G}_{aa}(\vartheta) & \mathcal{G}_{ac}(\vartheta) & \mathcal{G}_{ab}(\vartheta) \\ \mathcal{G}_{ac}^{\top}(\vartheta) & \mathcal{G}_{cc}(\vartheta) & \mathcal{G}_{cb}(\vartheta) \\ \mathcal{G}_{ab}^{\top}(\vartheta) & \mathcal{G}_{cb}^{\top}(\vartheta) & \mathcal{G}_{bb}(\vartheta) \end{pmatrix} \tag{63}$$

where the submatrices of $\mathcal{G}(\vartheta)$ are given by

$$\begin{aligned} \mathcal{G}_{aa}(\vartheta) &= \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} R_x(z) \frac{u_p(z)u_p^{\top}(z^{-1})c(z)c(z^{-1})}{a(z)a(z^{-1})b(z)b(z^{-1})} \frac{dz}{z} + \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{u_p(z)u_p^{\top}(z^{-1})}{a(z)a(z^{-1})} \frac{dz}{z} \\ &= \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} R_x(z) \frac{u_p(z)v_p^{\top}(z)c(z)\widehat{c}(z)}{a(z)\widehat{a}(z)b(z)\widehat{b}(z)z^{r-q}} dz + \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{u_p(z)v_p^{\top}(z)}{a(z)\widehat{a}(z)} dz \end{aligned}$$

$$\mathcal{G}_{ab}(\vartheta) = -\frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{u_p(z)u_q^{\top}(z^{-1})}{a(z)b(z^{-1})} \frac{dz}{z} = -\frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{u_p(z)v_q^{\top}(z)}{a(z)\widehat{b}(z)} dz$$

$$\mathcal{G}_{ac}(\vartheta) = -\frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} R_x(z) \frac{u_p(z)u_r^{\top}(z^{-1})c(z)}{a(z)b(z)b(z^{-1})} \frac{dz}{z} = -\frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} R_x(z) \frac{u_p(z)v_r^{\top}(z)c(z)}{a(z)b(z)\widehat{b}(z)z^{r-q}} dz$$

$$\mathcal{G}_{cc}(\vartheta) = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} R_x(z) \frac{u_r(z)u_r^{\top}(z^{-1})}{b(z)b(z^{-1})} \frac{dz}{z} = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} R_x(z) \frac{u_r(z)v_r^{\top}(z)}{b(z)\widehat{b}(z)z^{r-q}} dz$$

$$\mathcal{G}_{bb}(\vartheta) = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{u_q(z)u_q^{\top}(z^{-1})}{b(z)b(z^{-1})} \frac{dz}{z} = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{u_q(z)v_q^{\top}(z)}{b(z)\widehat{b}(z)} dz, \text{ and } \mathcal{G}_{cb}(\vartheta) = O$$

where $R_x(z)$ is the spectral density of the process $x(t)$ and is defined in (10). Let $K(z) = a(z)a(z^{-1})b(z)b(z^{-1})$, combining all the expressions in (63) leads to the following representation of $\mathcal{G}(\vartheta)$ as the sum of two matrices

$$\frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{R_x(z)}{K(z)} \begin{pmatrix} c(z)u_p(z) \\ -a(z)u_r(z) \\ O \end{pmatrix} \begin{pmatrix} c(z)u_p(z) \\ -a(z)u_r(z) \\ O \end{pmatrix}^* \frac{dz}{z} + \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{1}{K(z)} \begin{pmatrix} b(z)u_p(z) \\ O \\ -a(z)u_q(z) \end{pmatrix} \begin{pmatrix} b(z)u_p(z) \\ O \\ -a(z)u_q(z) \end{pmatrix}^* \frac{dz}{z} \tag{64}$$

where $(X)^*$ is the complex conjugate transpose of the matrix $X \in \mathbb{C}^{m \times n}$. Like in (23) we set forth

$$\mathcal{S}(-c, a) = \begin{pmatrix} -\mathcal{S}_p(c) \\ \mathcal{S}_r(a) \end{pmatrix}$$

here $\mathcal{S}_p(c)$ is formed by the top p rows of $\mathcal{S}(-c,a)$. In a similar way we decompose

$$\mathcal{S}(-b,a)=\begin{pmatrix}-\mathcal{S}_p(b)\\ \mathcal{S}_q(a)\end{pmatrix}$$

The representation (64) can be expressed by the appropriate block representations of the Sylvester resultant matrices, to obtain

$$\mathcal{G}(\vartheta)=\begin{pmatrix}-\mathcal{S}_p(c)\\ \mathcal{S}_r(a)\\ O\end{pmatrix}\mathcal{W}(\vartheta)\begin{pmatrix}-\mathcal{S}_p(c)\\ \mathcal{S}_r(a)\\ O\end{pmatrix}^{\top}+\begin{pmatrix}-\mathcal{S}_p(b)\\ O\\ \mathcal{S}_q(a)\end{pmatrix}\mathcal{P}(\vartheta)\begin{pmatrix}-\mathcal{S}_p(b)\\ O\\ \mathcal{S}_q(a)\end{pmatrix}^{\top} \tag{65}$$

where the matrix $\mathcal{P}(\vartheta)$ is given in (28) and the matrix $\mathcal{W}(\vartheta)\in\mathbb{R}^{(p+r)\times(p+r)}$ is of the form

$$\mathcal{W}(\vartheta)=\frac{1}{2\pi\mathbf{i}}\oint_{|z|=1}R_x(z)\frac{u_{p+r}(z)u_{p+r}^{\top}(z^{-1})}{a(z)a(z^{-1})b(z)b(z^{-1})}\frac{dz}{z}=\frac{1}{2\pi\mathbf{i}}\oint_{|z|=1}R_x(z)\frac{u_{p+r}(z)v_{p+r}^{\top}(z)}{a(z)b(z)\widehat{a}(z)\widehat{b}(z)}dz \tag{66}$$

It is shown in [4] that $\mathcal{W}(\vartheta)\succ O$. As can be seen in (65), the ARMAX part is explained by the first term, whereas the ARMA part is described by the second term, the combination of both terms is a summary of the Fisher information of a ARMAX(p,r,q) process. The FIM $\mathcal{G}(\vartheta)$ under form (65) allows us to prove the following property, Theorem 3.1 in [4]. The FIM $\mathcal{G}(\vartheta)$ of the ARMAX(p,r,q) process with polynomials $a(z)$, $c(z)$ and $b(z)$ of order p,r,q respectively becomes singular if and only if these polynomials have at least one common root. Consequently, the class of resultant matrices is extended by the FIM $\mathcal{G}(\vartheta)$.

3.7. The Stein Equation - The Fisher Information Matrix of an ARMAX(p,r,q) Process

In Lemma 3.5 it is proved that the matrix $\mathcal{P}(\vartheta)$ (28) fulfills the Stein Equation (52). We will now consider the conditions under which the matrix $\mathcal{W}(\vartheta)$ (66) verifies an appropriate Stein equation. For that purpose we consider the spectral density to be of the form $R_x(z)=(1/h(z)h(z^{-1}))$. The degree of the polynomial $h(z)$ is ℓ and we assume the distinct roots of the polynomial $h(z)$ to lie outside the unit circle, consequently, the roots of the polynomial $\widehat{h}(z)$ lie within the unit circle. We therefore rewrite $\mathcal{W}(\vartheta)$ accordingly

$$\mathcal{W}(\vartheta)=\frac{1}{2\pi\mathbf{i}}\oint_{|z|=1}\frac{u_{p+r}(z)u_{p+r}^{\top}(z^{-1})}{h(z)h(z^{-1})a(z)a(z^{-1})b(z)b(z^{-1})}\frac{dz}{z}$$

We consider a companion matrix of the form (48) and with size $p+q+\ell$, it is denoted by $\mathcal{C}_f$ and the entries f_i are given by $z^{p+q+\ell}+\sum_{i=1}^{p+q+\ell}f_i(\vartheta)z^{p+q+\ell-i}=\widehat{a}(z)\widehat{b}(z)\widehat{h}(z)=\widehat{f}(z,\vartheta)$ and $\widehat{f}(\vartheta)$ is the vector $\widehat{f}(\vartheta)=(f_{p+q+\ell}(\vartheta),f_{p+q+\ell-1}(\vartheta),\ldots,f_1(\vartheta))^{\top}$. Likewise for the vector $f(z,\vartheta)=a(z)b(z)h(z)$ and $f(\vartheta)=(f_1(\vartheta),f_1(\vartheta),\ldots,f_{p+q+\ell}(\vartheta))^{\top}$. The property $\mathrm{Det}(zI_{p+q+\ell}-\mathcal{C}_f)=\widehat{a}(z)\widehat{b}(z)\widehat{h}(z)$ and $\mathrm{Det}(I_{p+q+\ell}-z\mathcal{C}_f)=a(z)b(z)h(z)$ holds and assume

$$r=q+\ell \text{ or } p+q+\ell=p+r \text{ and } r>q \tag{67}$$

$\mathcal{W}(\vartheta)$ is then of the form

$$\mathcal{W}(\vartheta) = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{u_{p+r}(z) v_{p+r}^{\top}(z)}{h(z)\widehat{h}(z)a(z)\widehat{a}(z)b(z)\widehat{b}(z)} dz \tag{68}$$

We will formulate a Stein equation when the matrix $\Gamma = e_{p+r} e_{p+r}^{\top}$ and which is of the form

$$S - \mathcal{C}_f S \mathcal{C}_f^{\top} = e_{p+r} e_{p+r}^{\top} \tag{69}$$

where e_{p+r} is the last standard basis column vector in $\mathbb{R}^{p+r}$. The next lemma is formulated.

Lemma 3.8. *The matrix $\mathcal{W}(\vartheta)$ given in (68) fulfills the Stein Equation (69).*

Proof. The unique solution of (69) is assured since the product of all the eigenvalues of $\mathcal{C}_f$ are different from one, the solution is of the form

$$S = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} (zI_{p+r} - \mathcal{C}_f)^{-1} e_{p+r} e_{p+r}^{\top} (I_{p+r} - z\mathcal{C}_f)^{-\top} dz$$

or

$$S = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{\text{adj}(zI_{p+r} - \mathcal{C}_f) e_{p+r} e_{p+r}^{\top} \text{adj}(I_{p+r} - z\mathcal{C}_f)^{\top}}{\widehat{a}(z)\widehat{b}(z)\widehat{h}(z)a(z)b(z)h(z)} dz$$

taking the property of the companion matrix $\mathcal{C}_f$ into account implies that the last column vector of $\text{adj}(zI_{p+r} - \mathcal{C}_f)$ is the basic vector $u_{p+r}(z)$, consequently the last column of $\text{adj}(I_{p+r} - z\mathcal{C}_f)$ is the basic vector $v_{p+r}(z)$, this yields

$$S = \frac{1}{2\pi \mathbf{i}} \oint\limits_{|z|=1} \frac{u_{p+r}(z) v_{p+r}^{\top}(z)}{\widehat{a}(z)\widehat{b}(z)\widehat{h}(z)a(z)b(z)h(z)} dz = \mathcal{W}(\vartheta)$$

Consequently, the matrix $\mathcal{W}(\vartheta)$ defined in (68) verifies the Stein Equation (69). This completes the proof. □

The matrices, $\mathcal{P}(\vartheta)$ and $\mathcal{W}(\vartheta)$, in (65), verify under specific conditions appropriate Stein equations, as has been shown in Lemma 3.5 and Lemma 3.8, respectively. We will now confirm the presence of Vandermonde matrices by applying the standard residue theorem to $\mathcal{W}(\vartheta)$ in (68), to obtain

$$\mathcal{W}(\vartheta) = V_{\alpha\beta\xi} \mathcal{R}(\vartheta) \widehat{V}_{\alpha\beta\xi} \tag{70}$$

The $(p+r) \times (p+r)$ diagonal matrix $\mathcal{R}(\vartheta)$ is of the form

$$\mathcal{R}(\vartheta) = \text{diag}\left\{ \left(1/\widehat{a}(z;\alpha_i)\widehat{b}(\alpha_i)\widehat{h}(\alpha_i)\phi(\alpha_i)\right), \left(1/\widehat{a}(\beta_j)\widehat{b}(z;\beta_j)\widehat{h}(\beta_j)\phi(\beta_j)\right), \left(1/\widehat{a}(\xi_k)\widehat{b}(\xi_k)\widehat{h}(z;\xi_k)\phi(\xi_k)\right) \right\}$$

where $\phi(z) = a(z)b(z)h(z)$ and $i = 1,\ldots,p$, $j = 1,\ldots,q$ and $k = 1,\ldots,\ell$. Whereas the $(p+r) \times (p+r)$ matrices $V_{\alpha\beta\xi}$ and $\widehat{V}_{\alpha\beta\xi}$ are of the form

$$V_{\alpha\beta\xi} = \begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^{p+r-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \alpha_p & \alpha_p^2 & \cdots & \alpha_p^{p+r-1} \\ 1 & \beta_1 & \beta_1^2 & \cdots & \beta_1^{p+r-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \beta_q & \beta_q^2 & \cdots & \beta_q^{p+r-1} \\ 1 & \xi_1 & \xi_1^2 & \cdots & \xi_1^{p+r-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \xi_\ell & \xi_\ell^2 & \cdots & \xi_\ell^{p+r-1} \end{pmatrix}^\top, \quad \widehat{V}_{\alpha\beta\xi} = \begin{pmatrix} \alpha_1^{p+r-1} & \alpha_1^{p+r-2} & \cdots & \alpha_1 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ \alpha_p^{p+r-1} & \alpha_p^{p+r-2} & \cdots & \alpha_p & 1 \\ \beta_1^{p+r-1} & \beta_1^{p+r-2} & \cdots & \beta_1 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ \beta_q^{p+r-1} & \beta_q^{p+r-2} & \cdots & \beta_q & 1 \\ \xi_1^{p+r-1} & \xi_1^{p+r-2} & \cdots & \xi_1 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ \xi_\ell^{p+r-1} & \xi_\ell^{p+r-2} & \cdots & \xi_\ell & 1 \end{pmatrix}$$

The $(p+r) \times (p+r)$ Vandermonde matrices $V_{\alpha\beta\xi}$ and $\widehat{V}_{\alpha\beta\xi}$ are nonsingular when $\alpha_i \neq \alpha_j$, $\beta_k \neq \beta_h$, $\xi_m \neq \xi_n$, $\alpha_i \neq \beta_k$, $\alpha_i \neq \xi_m$, $\beta_k \neq \xi_m$ for all $i,j = 1,\ldots,p$, $k,h = 1,\ldots,q$ and $m,n = 1,\ldots,\ell$. The Vandermonde determinants $\mathrm{Det}V_{\alpha\beta\xi}$ and $\mathrm{Det}\widehat{V}_{\alpha\beta\xi}$, are

$$\mathrm{Det}V_{\alpha\beta\xi} = (-1)^{(p+r)(p+r-1)/2}\, \Psi\left(\alpha_i, \beta_k, \xi_m\right)$$

where

$$\Psi\left(\alpha_i, \beta_k, \xi_m\right) =$$

$$\prod_{1\leq i<j\leq p} (\alpha_i - \alpha_j) \prod_{1\leq k<h\leq q} (\beta_k - \beta_h) \prod_{1\leq m<n\leq \ell} (\xi_m - \xi_n) \prod_{\substack{r=1,\ldots,p \\ s=1,\ldots,q}} (\alpha_r - \beta_s) \prod_{\substack{r=1,\ldots,p \\ w=1,\ldots,\ell}} (\alpha_r - \xi_w) \prod_{\substack{s=1,\ldots,q \\ w=1,\ldots,\ell}} (\beta_s - \xi_w)$$

Like for the Vandermonde matrices $V_{\alpha\beta}$ and $\widehat{V}_{\alpha\beta}^\top$,

$$\mathrm{Det}\widehat{V}_{\alpha\beta\xi} = (-1)^{(p+r)(p+r-1)}\, \Psi\left(\alpha_i, \beta_k, \xi_m\right) \Longrightarrow \left|\mathrm{Det}V_{\alpha\beta\xi}\right| = \left|\mathrm{Det}\widehat{V}_{\alpha\beta\xi}\right|$$

Equation (70) is the ARMAX equivalent to (47). A combination of both equations generates a new representation of the FIM $\mathcal{G}(\vartheta)$, this is set forth in the following lemma.

Lemma 3.9. *Assume the conditions (67) to hold and consider the representations of $\mathcal{P}(\vartheta)$ and $\mathcal{W}(\vartheta)$ in (47) and (70) respectively, leads to an alternative form to (65), it is given by*

$$\mathcal{G}(\vartheta) = \begin{pmatrix} -\mathcal{S}_p(c) \\ \mathcal{S}_r(a) \\ O \end{pmatrix} V_{\alpha\beta\xi} \mathcal{R}\left(\vartheta\right) \widehat{V}_{\alpha\beta\xi} \begin{pmatrix} -\mathcal{S}_p(c) \\ \mathcal{S}_r(a) \\ O \end{pmatrix}^\top + \begin{pmatrix} -\mathcal{S}_p(b) \\ O \\ \mathcal{S}_q(a) \end{pmatrix} V_{\alpha\beta} \mathcal{D}(\vartheta) \widehat{V}_{\alpha\beta} \begin{pmatrix} -\mathcal{S}_p(b) \\ O \\ \mathcal{S}_q(a) \end{pmatrix}^\top$$

In Lemma 3.9, the FIM $\mathcal{G}(\vartheta)$ is expressed by submatrices of two Sylvester matrices and various Vandermonde matrices, both type of matrices become singular if and only if the appropriate polynomials have at least one common root.

3.8. The Fisher Information Matrix of a Vector ARMA(p, q) Process

The process (6) is summarized as,

$$A(z)y(t) = B(z)\varepsilon(t)$$

and we assume that $\{y(t), t \in \mathbb{N}\}$, is a zero mean Gaussian time series and $\{\varepsilon(t), t \in \mathbb{N}\}$ is a n-dimensional vector random variable, such that $\mathbb{E}_\vartheta\{\varepsilon(t)\} = 0$ and $\mathbb{E}_\vartheta\{\varepsilon(t)\varepsilon^\top(t)\} = \Sigma$ and the parameter vector ϑ is of the form (7). In [6] it is shown that representation (16) for the $n^2(p+q)\times n^2(p+q)$ asymptotic FIM of the VARMA process (6) is

$$\boldsymbol{F}(\vartheta) = \mathbb{E}_\vartheta\left\{\left(\frac{\partial\varepsilon}{\partial\vartheta^\top}\right)^\top \Sigma^{-1}\left(\frac{\partial\varepsilon}{\partial\vartheta^\top}\right)\right\} \tag{71}$$

where $\partial\varepsilon/\partial\vartheta^\top$ is of size $n \times n^2(p+q)$ and for convenience t is omitted from $\varepsilon(t)$. Using the differential rules outlined in [6], yields

$$\frac{\partial\varepsilon}{\partial\vartheta^\top} = \left\{\left(A^{-1}(z)B(z)\varepsilon\right)^\top \otimes B^{-1}(z)\right\}\frac{\partial \text{vec } A(z)}{\partial\vartheta^\top} - \left(\varepsilon^\top \otimes B^{-1}(z)\right)\frac{\partial \text{vec } B(z)}{\partial\vartheta^\top} \tag{72}$$

The substitution of representation (72) of $\partial\varepsilon/\partial\vartheta^\top$ in (71) yields the FIM of a VARMA process. The purpose is to construct a factorization of the FIM $\boldsymbol{F}(\vartheta)$ that should be a multiple variant of the factorization (27), so that a multiple resultant matrix property can be proved for $\boldsymbol{F}(\vartheta)$. As illustrated in [6], the multiple version of the Sylvester resultant matrix (22) does not fulfill the multiple resultant matrix property. In that case even when the matrix polynomials $A(z)$ and $B(z)$ have a common zero or a common eigenvalue, the multiple Sylvester matrix is not neccessarily singular. This has also been illustrated in [3]. In order to consider a multiple equivalent to the resultant matrix $\mathcal{S}(-b, a)$, Gohberg and Lerer set forth the $n^2(p+q) \times n^2(p+q)$ tensor Sylvester matrix

$$\mathcal{S}^{\otimes}(-B, A) := \begin{pmatrix}
(-I_n)\otimes I_n & (-B_1)\otimes I_n & \cdots & (-B_q)\otimes I_n & O_{n^2\times n^2} & \cdots & O_{n^2\times n^2} \\
O_{n^2\times n^2} & \ddots & \ddots & & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & & \ddots & O_{n^2\times n^2} \\
O_{n^2\times n^2} & \cdots & O_{n^2\times n^2} & (-I_n)\otimes I_n & (-B_1)\otimes I_n & \cdots & (-B_q)\otimes I_n \\
I_n\otimes I_n & I_n\otimes A_1 & \cdots & I_n\otimes A_p & O_{n^2\times n^2} & \cdots & O_{n^2\times n^2} \\
O_{n^2\times n^2} & \ddots & \ddots & & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & & \ddots & O_{n^2\times n^2} \\
O_{n^2\times n^2} & \cdots & O_{n^2\times n^2} & I_n\otimes I_n & I_n\otimes A_1 & \cdots & I_n\otimes A_p
\end{pmatrix} \tag{73}$$

In [3], the authors prove that the tensor Sylvester matrix $\mathcal{S}^{\otimes}(-B, A)$ fulfills the multiple resultant property, it becomes singular if and only if the appropriate matrix polynomials $A(z)$ and $B(z)$ have at least one common zero. In Proposition 2.2 in [6], the following factorized form of the Fisher information $\boldsymbol{F}(\vartheta)$ is developed

$$\boldsymbol{F}(\vartheta) = \frac{1}{2\pi\mathbf{i}}\oint_{|z|=1}\Phi(z)\Theta(z)\Phi^*(z)\frac{dz}{z} \tag{74}$$

where

$$\Phi(z) = \begin{pmatrix} I_p \otimes A^{-1}(z) \otimes I_n & O_{pn^2 \times qn^2} \\ O_{qn^2 \times pn^2} & I_q \otimes I_n \otimes A^{-1}(z) \end{pmatrix} \mathcal{S}^{\otimes}(-B, A) \left(u_{p+q}(z) \otimes I_{n^2}\right)$$

and

$$\Theta(z) = \Sigma \otimes \sigma(z), \sigma(z) = B^{-\top}(z)\Sigma^{-1}B^{-1}(z^{-1}) \tag{75}$$

In order to obtain a multiple variant of (27), the following matrix is introduced,

$$\boldsymbol{M}(\vartheta) = \frac{1}{2\pi \mathbf{i}} \oint_{|z|=1} \Lambda(z)\mathcal{J}(z)\Lambda^*(z)\frac{dz}{z} = \mathcal{S}^{\otimes}(-B, A)\boldsymbol{P}(\vartheta)\left(\mathcal{S}^{\otimes}(-B, A)\right)^{\top} \tag{76}$$

where

$$\mathcal{J}(z) = \Phi(z)\Theta(z)\Phi^*(z) \text{ and } \Lambda(z) = \begin{pmatrix} I_p \otimes A(z) \otimes I_n & O_{pn^2 \times qn^2} \\ O_{qn^2 \times pn^2} & I_q \otimes I_n \otimes A(z) \end{pmatrix}$$

and the matrix $\boldsymbol{P}(\vartheta)$ is a multiple variant of the matrix $\mathcal{P}(\vartheta)$ in (28), it is of the form

$$\boldsymbol{P}(\vartheta) = \frac{1}{2\pi \mathbf{i}} \oint_{|z|=1} \left(u_{p+q}(z) \otimes I_{n^2}\right) \Theta(z) \left(u_{p+q}(z) \otimes I_{n^2}\right)^* \frac{dz}{z} \tag{77}$$

In Lemma 2.3 in [6], it is proved that the matrix $\boldsymbol{M}(\vartheta)$ in (76) becomes singular if and only if the matrix polynomials $A(z)$ and $B(z)$ have at least one common eigenvalue-zero. The proof is a multiple equivalent of the proof of Corollary 2.2 in [5], since the equality (76) is a multiple version of (27). Consequently, the matrix $\boldsymbol{M}(\vartheta)$ like the tensor Sylvester matrix $\mathcal{S}^{\otimes}(-B, A)$, fulfills the multiple resultant matrix property. Since the matrix $\boldsymbol{M}(\vartheta)$ is derived from the FIM $\boldsymbol{F}(\vartheta)$, this enables us to prove that the matrix $\boldsymbol{F}(\vartheta)$ fulfills the multiple resultant matrix property by showing that it becomes singular if and only if the matrix $\boldsymbol{M}(\vartheta)$ is singular, this is done in Proposition 2.4 in [6]. Consequently, it can be concluded from [6] that the FIM of a VARMA process $\boldsymbol{F}(\vartheta)$ and the tensor Sylvester matrix $\mathcal{S}^{\otimes}(-B, A)$ have the same singularity conditions. The FIM of a VARMA process $\boldsymbol{F}(\vartheta)$ can therefore be added to the class of multiple resultant matrices.

A brief summary of the contribution of [6] follows, in order to show that the FIM of a VARMA process $\boldsymbol{F}(\vartheta)$ is a multiple resultant matrix two new representations of the FIM are derived. To construct such representations appropriate matrix differential rules are applied. The newly obtained representations are expressed in terms of the multiple Sylvester matrix and the tensor Sylvester matrix. The representation of the FIM expressed by the tensor Sylvester matrix is used to prove that the FIM becomes singular if and only if the autoregressive and moving average matrix polynomials have at least one common eigenvalue. It then follows that the FIM and the tensor Sylvester matrix have equivalent singularity conditions. In a numerical example it is shown, however, that the FIM fails to detect common eigenvalues due to some kind of numerical instability. The tensor Sylvester matrix reveals it clearly, proving the usefulness of the results derived in this paper.

3.9. The Fisher Information Matrix of a Vector ARMAX(p, r, q) Process

The $n^2(p+q+r) \times n^2(p+q+r)$ asymptotic FIM of the VARMAX(p, r, q) process (2)

$$A(z)y(t) = C(z)x(t) + B(z)\varepsilon(t)$$

is displayed according to [23] and is an extension of the FIM of the VARMA(p, q) process (6). Representation (16) of the FIM of the VARMAX(p, r, q) process is then

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

where

$$\begin{aligned} \frac{\partial \varepsilon}{\partial \vartheta^\top} &= \left\{ \left(A^{-1}(z)C(z)x \right)^\top \otimes B^{-1}(z) \right\} \frac{\partial \text{vec } A(z)}{\partial \vartheta^\top} \\ &+ \left\{ \left(A^{-1}(z)B(z)\varepsilon \right)^\top \otimes B^{-1}(z) \right\} \frac{\partial \text{vec } A(z)}{\partial \vartheta^\top} \\ &- \left\{ x^\top \otimes B^{-1}(z) \right\} \frac{\partial \text{vec } C(z)}{\partial \vartheta^\top} \\ &- \left(\varepsilon^\top \otimes B^{-1}(z) \right) \frac{\partial \text{vec } B(z)}{\partial \vartheta^\top} \end{aligned} \tag{78}$$

To obtain the term $\partial \varepsilon / \partial \vartheta^\top$, of size $n \times n^2(p+q+r)$, the same differential rules are applied as for the VARMA(p, q) process. In Proposition 2.3 in [23], the representation of the FIM of a VARMAX process is expressed in terms of tensor Sylvester matrices, this obtained when $\partial \varepsilon / \partial \vartheta^\top$ in (78) is substituted in (16), to obtain

$$\boldsymbol{G}(\vartheta) = \frac{1}{2\pi i} \oint_{|z|=1} \Phi_x(z)\Theta(z)\Phi_x^*(z) \frac{dz}{z} + \frac{1}{2\pi i} \oint_{|z|=1} \Lambda_x(z)\Psi(z)\Lambda_x^*(z) \frac{dz}{z} \tag{79}$$

The matrices in (79) are of the form

$$\Phi_x(z) = \begin{pmatrix} I_p \otimes A^{-1}(z) \otimes I_n & O_{pn^2 \times rn^2} & O_{pn^2 \times qn^2} \\ O_{rn^2 \times pn^2} & O_{rn^2 \times rn^2} & O_{rn^2 \times qn^2} \\ O_{qn^2 \times pn^2} & O_{qn^2 \times rn^2} & I_q \otimes I_n \otimes A^{-1}(z) \end{pmatrix} \begin{pmatrix} -\mathcal{S}_p^\otimes(B) \\ O_{rn^2 \times n^2(p+q)} \\ \mathcal{S}_q^\otimes(A) \end{pmatrix} (u_{p+q}(z) \otimes I_{n^2})$$

$$\Lambda_x(z) = \begin{pmatrix} I_p \otimes A^{-1}(z) \otimes I_n & O_{pn^2 \times rn^2} & O_{pn^2 \times qn^2} \\ O_{rn^2 \times pn^2} & I_r \otimes I_n \otimes A^{-1}(z) & O_{rn^2 \times qn^2} \\ O_{qn^2 \times pn^2} & O_{qn^2 \times rn^2} & O_{qn^2 \times qn^2} \end{pmatrix} \begin{pmatrix} -\mathcal{S}_p^\otimes(C) \\ \mathcal{S}_r^\otimes(A) \\ O_{qn^2 \times n^2(p+r)} \end{pmatrix} (u_{p+r}(z) \otimes I_{n^2})$$

$$\mathcal{S}_{p,q}^\otimes(-B, A) = \begin{pmatrix} -\mathcal{S}_p^\otimes(B) \\ \mathcal{S}_q^\otimes(A) \end{pmatrix}, \ \mathcal{S}_{p,r}^\otimes(-C, A) = \begin{pmatrix} -\mathcal{S}_p^\otimes(C) \\ \mathcal{S}_r^\otimes(A) \end{pmatrix} \tag{80}$$

additionally we have $\Psi(z) = R_x(z) \otimes \sigma(z)$ and the Hermitian spectral density matrix $R_x(z)$ is defined in (10), whereas the matrix polynomials $\Theta(z)$ and $\sigma(z)$ are presented in (75). In (80), we have the $pn^2 \times (p+q)n^2$ and $qn^2 \times (p+q)n^2$ submatrices $\mathcal{S}_p^\otimes(-B)$ and $\mathcal{S}_q^\otimes(A)$ of the tensor Sylvester resultant matrix $\mathcal{S}_{p,q}^\otimes(-B, A)$. Whereas the matrices $\mathcal{S}_p^\otimes(-C)$ and $\mathcal{S}_r^\otimes(A)$ are the upper and lower blocks of the

$(p+r)n^2 \times (p+r)n^2$ tensor Sylvester resultant matrix $\mathcal{S}_{p,r}^{\otimes}(-C, A)$. As for the FIM of the VARMA(p, q) process, the objective is to construct a multiple version of (65), this done in [23], to obtain

$$\boldsymbol{M}_x(\vartheta) = \frac{1}{2\pi i}\oint_{|z|=1} \mathcal{L}(z)\mathcal{A}(z)\mathcal{L}^*(z)\frac{dz}{z} + \frac{1}{2\pi i}\oint_{|z|=1} \mathcal{W}(z)\mathcal{B}(z)\mathcal{W}^*(z)\frac{dz}{z}$$

$$= \begin{pmatrix} -\mathcal{S}_p^{\otimes}(B) \\ O_{rn^2\times n^2(p+q)} \\ \mathcal{S}_q^{\otimes}(A) \end{pmatrix} \boldsymbol{P}(\vartheta) \begin{pmatrix} -\mathcal{S}_p^{\otimes}(B) \\ O_{rn^2\times n^2(p+q)} \\ \mathcal{S}_q^{\otimes}(A) \end{pmatrix}^{\top} + \begin{pmatrix} -\mathcal{S}_p^{\otimes}(C) \\ \mathcal{S}_r^{\otimes}(A) \\ O_{qn^2\times n^2(p+r)} \end{pmatrix} \boldsymbol{T}(\vartheta) \begin{pmatrix} -\mathcal{S}_p^{\otimes}(C) \\ \mathcal{S}_r^{\otimes}(A) \\ O_{qn^2\times n^2(p+r)} \end{pmatrix}^{\top} \quad (81)$$

The matrices involved are of the form

$$\mathcal{L}(z) = \begin{pmatrix} I_p \otimes A(z) \otimes I_n & O_{pn^2\times rn^2} & O_{pn^2\times qn^2} \\ O_{rn^2\times pn^2} & O_{rn^2\times rn^2} & O_{rn^2\times qn^2} \\ O_{qn^2\times pn^2} & O_{qn^2\times rn^2} & I_q \otimes I_n \otimes A(z) \end{pmatrix} \text{ and } \mathcal{A}(z) := \Phi_x(z)\Theta(z)\Phi_x^*(z)$$

$$\mathcal{W}(z) = \begin{pmatrix} I_p \otimes A(z) \otimes I_n & O_{pn^2\times rn^2} & O_{pn^2\times qn^2} \\ O_{rn^2\times pn^2} & I_r \otimes I_n \otimes A(z) & O_{rn^2\times qn^2} \\ O_{qn^2\times pn^2} & O_{qn^2\times rn^2} & O_{qn^2\times qn^2} \end{pmatrix} \text{ and } \mathcal{B}(z) := \Lambda_x(z)\Psi(z)\Lambda_x^*(z)$$

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

and $\boldsymbol{P}(\vartheta)$ is given in (77). Note, the matrices $\Phi_x(z)$, $\Lambda_x(z)$, $\mathcal{L}(z)$ and $\mathcal{W}(z)$ are the corrected versions of the corresponding matrices in [23].

A parallel between the scalar and multiple structures is straightforward. This is best illustrated by comparing the representations (27) and (28) with (76) and (77) respectively, confronting the FIM for scalar and vector ARMA(p, q) processes. The FIM of the scalar ARMAX(p, r, q) process contains an ARMA(p, q) part, this is confirmed by (65), through the presence of the matrix $\mathcal{P}(\vartheta)$ which is originally displayed in (28). The multiple resultant matrices $\boldsymbol{M}(\vartheta)$ and $\boldsymbol{M}_x(\vartheta)$ derived from the FIM of the VARMA(p, q) and VARMAX(p, r, q) processes respectively both contain $\boldsymbol{P}(\vartheta)$, whereas the first matrix term of the matrices $\Phi(z)$ and $\Phi_x(z)$, which are of different size, consist of the same nonzero submatrices. To summarize, in [23] compact forms of the FIM of a VARMAX process expressed in terms of multiple and tensor Sylvester matrices are developed. The tensor Sylvester matrices allow us to investigate the multiple resultant matrix property of the FIM of VARMAX(p, r, q) processes. However, since no proof of the multiple resultant matrix property of the FIM $\boldsymbol{G}(\vartheta)$ has been done yet, justifies the consideration of a conjecture. A conjecture that states, the FIM $\boldsymbol{G}(\vartheta)$ of a VARMAX(p, r, q) process becomes singular if and only if the matrix polynomials $A(z)$, $B(z)$ and $C(z)$ have at least one common eigenvalue. A multiple equivalent to Theorem 3.1 in [4] and combined with Proposition 2.4 in [6], but based on the representations (79) and (81), can be envisaged to formulate a proof which will be a subject for future study.

4. Conclusions

In this survey paper, matrix algebraic properties of the FIM of stationary processes are discussed. The presented material is a summary of papers where several matrix structural aspects of the FIM

are investigated. The FIM of scalar and multiple processes like the (V)ARMA(X) are set forth with appropriate factorized forms involving (tensor) Sylvester matrices. These representations enable us to prove the resultant matrix property of the corresponding FIM. This has been done for (V)ARMA(p, q) and ARMAX(p, r, q) processes in the papers [4–6]. The development of the stages that lead to the appropriate factorized form of the FIM $\boldsymbol{G}(\vartheta)$ (79) is set forth in [23]. However, there is no proof done yet that confirms the multiple resultant matrix property of the FIM $\boldsymbol{G}(\vartheta)$ of a VARMAX(p, r, q) process. This justifies the consideration of a conjecture which is formulated in the former section, this can be a subject for future study.

The statistical distance measure derived in [7], involves entries of the FIM. This distance measure can be a challenge to its quantum information counterpart (41). Because (36) involves information about m parameters estimated from n measurements. Whereas in quantum information, like in e.g., [8–10], the information about one parameter in a particular measurement procedure is considered for establishing an interconnection with the appropriate statistical distance measure. A possible approach, by combining matrix algebra and quantum information, for developing a statistical distance measure in quantum information or quantum statistics but at the matrix level, can be a subject of future research.
Some results concerning interconnections between the FIM of ARMA(X) models and appropriate solutions to Stein matrix equations are discussed, the material is extracted from the papers, [12] and [13]. However, in this paper, some alternative and new proofs that emphasize the conditions under which the FIM fulfills appropriate Stein equations, are set forth. The presence of various types of Vandermonde matrices is also emphasized when an explicit expansion of the FIM is computed. These Vandermonde matrices are inserted in interconnections with appropriate solutions to Stein equations. This explains, when the matrix algebraic structures of the FIM of stationary processes are investigated, the involvement of structured matrices like the (tensor) Sylvester, Bezoutian and Vandermonde matrices is essential.

Acknowledgements

The author thanks a perceptive reviewer for his comments which significantly improved the quality and presentation of the paper.

Conflicts of Interest

The authors have declared no conflict of interest.

References

1. Dym, H. *Linear Algebra in Action*; American Mathematical Society: Providence, RI, USA, 2006; Volume 78 .
2. Lancaster, P.; Tismenetsky, M. *The Theory of Matrices with Applications*, 2nd ed.; Academic Press: Orlando, FL, USA, 1985.
3. Gohberg, I.; Lerer, L. Resultants of matrix polynomials. *Bull. Am. Math. Soc.* **1976**, *82*, 565–567.
4. Klein, A.; Spreij, P. On Fisher's information matrix of an ARMAX process and Sylvester's resultant matrices. *Linear Algebra Appl.* **1996**, *237/238*, 579–590.

5. Klein, A.; Spreij, P. On Fisher's information matrix of an ARMA process. *Stochastic Differential and Difference Equations*; Csiszar, I., Michaletzky, Gy., Eds; Birkhäuser Boston: Boston, USA, 1997; Progress in Systems and Control Theory; Volume 23, pp. 273–284.
6. Klein, A.; Mélard, G.; Spreij, P. On the Resultant Property of the Fisher Information Matrix of a Vector ARMA process. *Linear Algebra Appl.* **2005**, *403*, 291–313.
7. Klein, A.; Spreij, P. Transformed Statistical Distance Measures and the Fisher Information Matrix. *Linear Algebra Appl.* **2012**, *437*, 692–712.
8. Braunstein, S.L.; Caves, C.M. Statistical Distance and the Geometry of Quantum States. *Phys. Rev. Lett.* **1994**, *72*, 3439–3443.
9. Jones, P.J.; Kok, P. Geometric derivation of the quantum speed limit. *Phys. Rev. A* **2010**, *82*, 022107.
10. Kok, P. Tutorial: Statistical distance and Fisher information. Oxford, UK, 2006.
11. Lancaster, P.; Rodman, L. *Algebraic Riccati Equations*; Clarendon Press: Oxford, UK, 1995.
12. Klein, A.; Spreij, P. On Stein's equation, Vandermonde matrices and Fisher's information matrix of time series processes. Part I: The autoregressive moving average process. *Linear Algebra Appl.* **2001**, *329*, 9–47.
13. Klein, A.; Spreij, P. On the solution of Stein's equation and Fisher's information matrix of an ARMAX process. *Linear Algebra Appl.* **2005**, *396*, 1–34.
14. Grenander, U.; Szegő, G.P. *Toeplitz Forms and Their Applications*; University of California Press: New York, NY, USA, 1958.
15. Brockwell, P.J.; Davis, R.A. *Time Series: Theory and Methods*, 2nd ed.; Springer Verlag: Berlin, Germany; New York, NY, USA, 1991.
16. Caines, P. *Linear Stochastic Systems*; John Wiley and Sons: New York, NY, USA, 1988.
17. Ljung, L.; Söderström, T. *Theory and Practice of Recursive Identification*; M.I.T. Press: Cambridge, MA, USA, 1983.
18. Hannan, E.J.; Deistler, M. *The Statistical Theory of Linear Systems*; John Wiley and Sons: New York, NY, USA, 1988.
19. Hannan, E.J.; Dunsmuir, W.T.M.; Deistler, M. Estimation of vector Armax models. *J. Multivar. Anal.* **1980**, *10*, 275–295.
20. Horn, R.A.; Johnson, C.R. *Topics in Matrix Analysis*; Cambridge University Press: New York, NY, USA, 1995.
21. Klein, A.; Spreij, P. Matrix differential calculus applied to multiple stationary time series and an extended Whittle formula for information matrices. *Linear Algebra Appl.* **2009**, *430*, 674–691.
22. Klein, A.; Mélard, G. An algorithm for the exact Fisher information matrix of vector ARMAX time series. *Linear Algebra Its Appl.* **2014**, *446*, 1–24.
23. Klein, A.; Spreij, P. Tensor Sylvester matrices and the Fisher information matrix of VARMAX processes. *Linear Algebra Appl.* **2010**, *432*, 1975–1989.
24. Rao, C.R. Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **1945**, *37*, 81–91.
25. Ibragimov, I.A.; Has'minskiĭ, R.Z. Statistical Estimation. *Asymptotic Theory*; Springer-Verlag: New York, NY, USA, 1981.
26. Lehmann, E.L. *Theory of Point Estimation*; Wiley: New York, NY, USA, 1983.

27. Friedlander, B. On the computation of the Cramér-Rao bound for ARMA parameter estimation. *IEEE Trans. Acoust. Speech Signal Process* **1984**, *32* 721–727.
28. Holevo, A.S. *Probabilistic and Statistical Aspects of Quantum Theory*, 2nd ed.; Edizioni Della Normale, SNS Pisa: Pisa, Italy, 2011.
29. Petz, T. *Quantum Information Theory and Quantum Statistics*; Springer-Verlag: Berlin Heidelberg, Germany, 2008.
30. Barndorff-Nielsen, O.E.; Gill, R.D. Fisher Information in quantum statistics. *J. Phys. A* **2000**, *30*, 4481–4490.
31. Luo, S. Wigner-Yanase skew information *vs*. quantum Fisher information. *Proc. Amer. Math. Soc.* **2004**, Volume 132, pp. 885–890.
32. Klein, A.; Mélard, G. On algorithms for computing the covariance matrix of estimates in autoregressive moving average processes. *Comput. Stat. Q.* **1989**, *5*, 1–9.
33. Klein, A.; Mélard, G. An algorithm for computing the asymptotic Fisher information matrix for seasonal SISO models. *J. Time Ser. Anal. 2004*, *25*, 627–648.
34. Bistritz, Y.; Lifshitz, A. Bounds for resultants of univariate and bivariate polynomials. *Linear Algebra Appl.* **2010**, *432*, 1995–2005.
35. Horn, R.A.;Johnson, C.R. *Matrix Analysis*; Cambridge University Press: New York, NY, USA, 1996.
36. Golub, G.H.; van Loan, C.F. *Matrix Computations*, 3rd ed.; John Hopkins University Press: Baltimore, USA, 1996.
37. Kullback, S. *Information Theory and Statistics*; John Wiley and Sons: New York, NY, USA, 1959.
38. Klein, A.; Spreij, P. The Bezoutian, state space realizations and Fisher's information matrix of an ARMA process. *Linear Algebra Appl.* **2006**, *416*, 160–174.

© 2014 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).