Linear Algebra and its Applications 437 (2012) 692–712

Contents lists available at SciVerse ScienceDirect

Linear Algebra and its Applications

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

Transformed statistical distance measures and the Fisher information matrix

André Klein [a,*], Peter Spreij [b]

[a] Department of Quantitative Economics, University of Amsterdam, Valckenierstraat 65–67, 1018 XE Amsterdam, The Netherlands
[b] Korteweg-de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 9 January 2012
Accepted 6 March 2012
Available online 7 April 2012

Submitted by P. Šemrl

AMS classification:
15A18
15A23
15B99
62H30

Keywords:
Givens rotation matrix
Euclidean distance
Statistical distance measures
Fisher information matrix
Sylvester resultant matrix
Quantum information

ABSTRACT

Most multivariate statistical techniques are based upon the concept of distance. The purpose of this paper is to introduce statistical distance measures, which are normalized Euclidean distance measures, where the covariances of observed correlated measurements $x_1, \ldots, x_n$ and entries of the Fisher information matrix (FIM) are used as weighting coefficients. The measurements are subject to random fluctuations of different magnitudes and have therefore different variabilities. A rotation of the coordinate system through a chosen angle while keeping the scatter of points given by the data fixed, is therefore considered. It is shown that when the FIM is positive definite, the appropriate statistical distance measure is a metric. In case of a singular FIM, the metric property depends on the rotation angle. The introduced statistical distance measures, are matrix related, and are based on m parameters unlike a statistical distance measure in quantum information, which is also related to the Fisher information and where the information about one parameter in a particular measurement procedure is considered. A transformed FIM of a stationary process as well as the Sylvester resultant matrix are used to ensure the relevance of the appropriate statistical distance measure. The approach used in this paper is such that matrix properties are crucial for ensuring the relevance of the introduced statistical distance measures.

1. Introduction

In multivariate statistical analysis the concept of statistical distance is of fundamental importance because in order to produce a simple group structure from a complex data set, a measure of closeness

* Corresponding author. Tel.: +31 20 5254245; fax: +31 20 5254349.
E-mail address: A.A.B.Klein@uva.nl (A. Klein)

http://dx.doi.org/10.1016/j.laa.2012.03.002

or similarity is required. When units or cases are clustered, proximity is usually indicated by some distance or dissimilarity. This is also a key problem with image segmentation schemes when similarity or distance between two regions or pixels in an image have to be computed [25]. The most common metrics used are Euclidean.

In this paper we consider the statistical distance, which shall be specified below, between two vectors in a n-dimensional space. The straight-line or Euclidean distance between the stochastic vector $x = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix}^\top$ and fixed vector $y = \begin{pmatrix} y_1 & y_2 & \ldots & y_n \end{pmatrix}^\top$ where $x, y \in \mathbb{R}^n$, is given by

$$d_1(x, y) = \|x - y\| = \left(\sum_{j=1}^{n} (x_j - y_j)^2 \right)^{1/2}, \tag{1}$$

where the metric $d_1(x, y) := \|x - y\|$ is induced by the standard Euclidean norm $\|\cdot\|$ on $\mathbb{R}^n$.

All the components of the observations in (1) contribute equally to the Euclidean distance of x from y. However, in a statistical context the straight-line or Euclidean distance is unsatisfactory. This is because the coordinates or variables represent measurements that are subject to random fluctuations of different magnitudes. It is therefore important to consider a distance that takes the variability of these variables or measurements into account when determining its distance from a fix point. Components with a great deal of variability should receive less weight than components with low variability. We use the expression statistical distance, that accounts for differences in variation and will depend upon the sample variances and covariances, to distinguish it from ordinary Euclidean distance. A different measure of distance is obtained by rescaling the components or measurements, it is then a normalized Euclidean distance. This statistical distance is fundamental to multivariate statistical analysis, see e.g. [1,12]. In probability, statistics and more recently in quantum information, see [19,20,22], the statistical distance is used to quantify the similarity between two probability distributions in the space of probability distributions. Whereas in this paper, the statistical distance is based on observed measurements. However, variables are usually grouped on the basis of correlation coefficients or measure of association. The measure of association used in this paper is based on covariances and components of the FIM. In this study, we consider the statistical distance for the case that measurements x_j do not vary independently of the x_l measurements, for $j, l \in \{1, \ldots, n\}$, and they are correlated. The variability in the x_j direction is different from the variability in the x_l direction. Therefore, as suggested in [12], a rotation of the n-dimensional coordinate system through an angle ψ is considered while keeping the scatter of points given by the data fixed and label the rotated axes $\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_n$. A transformed distance measure is obtained and is determined entirely by the size of statistical fluctuations through the covariances and the rotation angle ψ. It is then identified with a suitable quadratic form. In the first part of this paper, Section 2, the covariances of the measurements $x_1, x_2, \ldots, x_n$, which are stochastic components, are used as weighting coefficients in the statistical distance measure, whereas in the second part, Sections 3 and 4, entries of the FIM $\mathcal{F}(\vartheta)$ are used for scaling the components. In this case the components considered are the estimated parameters $\vartheta_1, \vartheta_2, \ldots, \vartheta_m$ which are random and are computed from the measurements $x_1, x_2, \ldots, x_n$ according to appropriate statistical techniques, like the maximum likelihood method, see [11] for general results and see [5] for time series processes. The FIM $\mathcal{F}(\vartheta)$ measures the amount of information about the estimated parameters $\vartheta_1, \vartheta_2, \ldots, \vartheta_m$. Note, the $m \times 1$ estimated parameter vector is of the form

$$\vartheta = (\vartheta_1, \vartheta_2, \ldots, \vartheta_m)^\top \tag{2}$$

and $m < n$.

It is proved that the appropriate statistical distance measure is a metric when the corresponding FIM is positive definite, $\mathcal{F}(\vartheta) \succ 0$. In Section 3.2 we introduce the FIM $\mathcal{F}(\vartheta)$ of a stationary ARMA(p, q) process as well as a Sylvester resultant matrix, these matrices extend the sufficient condition for ensuring the relevance of the introduced statistical distance measure. The ARMA(p, q) process is set forth in Section 3.2. The presence or absence of common zeros between the ARMA(p, q) polynomials determines whether or not the property $\mathcal{F}(\vartheta) \succ 0$ holds and consequently, if the metric property of the introduced statistical distance measure is ensured. In Section 4 we discuss in detail metric conditions of the introduced statistical distance measure when the Fisher information matrices of ARMA(1, 1) and ARMA(2, 2) processes are singular or positive semidefinite. It is shown that some statistical distance measures fulfill the metric properties whereas some do not and this is determined by the choice of the rotation angle ϕ.

Consider the vectors x and y having n coordinates and suppose y is a fixed vector and the coordinate variables vary independently of one another. Let $s_{1,1}, s_{2,2}, \ldots, s_{n,n}$ be sample covariances constructed from n measurements $x_1, x_2, \ldots, x_n$ respectively. But since y is a fixed vector implies that $y_1, y_2, \ldots, y_n$ are not random components and this leads to corresponding covariances being zero. The statistical distance measure, as set forth in [12, p. 21], is of the form

$$d_2(x, y) = \left(\sum_{j=1}^{n} \left\{ \frac{(x_j - y_j)^2}{s_{j,j}} \right\} \right)^{1/2}. \tag{3}$$

However, in most cases the x_j measurements do not vary independently of the x_l measurements. The coordinates of the pairs (x_j, x_l) can exhibit a tendency to be large or small together so the variables x_j and x_l are correlated, for $j \neq l$ and $j, l \in \{1, 2, \ldots, n\}$. Consequently, the corresponding correlation coefficients are strongly positive or negative. Additionally, the variability in the x_j direction may differ from the variability in the x_l direction. We shall therefore consider a more general distance measure, this will be set forth in the next section.

2. A transformed statistical distance measure

As suggested in the multivariate statistical literature, see [12, p. 21–22], a more meaningful and general measure of distance is considered. We rotate the n-dimensional coordinate system through an angle ψ while keeping the scatter of points given by the data fixed and label the rotated axes $\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_n$. Note that in [12], the case displayed is limited to $n = 2$. The sample covariances are computed by using the $\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_n$ coordinate axes and the statistical distance measure is then formulated accordingly, to obtain

$$d(x, y) = \left(\sum_{j=1}^{n} \left\{ \frac{(\widetilde{x}_j - \widetilde{y}_j)^2}{\widetilde{s}_{j,j}} \right\} \right)^{1/2}, \tag{4}$$

where $\widetilde{s}_{1,1}, \ldots, \widetilde{s}_{i-1,i-1}, \widetilde{s}_{i,i}, \widetilde{s}_{i+1,i+1}, \widetilde{s}_{i+2,i+2}, \ldots, \widetilde{s}_{n,n}$ denote the sample covariances calculated with the $\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_n$ measurements. The newly obtained coordinates $\widetilde{x}_1 - \widetilde{y}_1, \widetilde{x}_2 - \widetilde{y}_2, \ldots, \widetilde{x}_n - \widetilde{y}_n$ are determined according to the following linear transformation which is the rotation through an angle $\psi \in \mathbb{R}$

$$(\widetilde{x} - \widetilde{y}) = \mathcal{R}_i(\psi)\,(x - y)\,, \tag{5}$$

where the block diagonal Givens rotation matrix $\mathcal{R}_i(\psi) \in \mathbb{R}^{n \times n}$ is of the form

$$\mathcal{R}_i(\psi) = \begin{pmatrix} I_{i-1} & 0 & 0 & 0 \\ 0 & (\cos(\psi))_{i,i} & (-\sin(\psi))_{i,i+1} & 0 \\ 0 & (\sin(\psi))_{i+1,i} & (\cos(\psi))_{i+1,i+1} & 0 \\ 0 & 0 & 0 & I_{n-i-1} \end{pmatrix}, \quad 0 \leqslant \psi \leqslant 2\pi, \tag{6}$$

and $i \in \{1, 2, \ldots, n-1\}$. As can be seen, the Givens matrix $\mathcal{R}_i(\psi)$ involves only two coordinates that are affected by the rotation angle ψ whereas the other directions, which correspond to eigenvalues 1, are unaffected by the rotation matrix. In dimension n there are $(n-1)$ Givens rotation matrices of type (6), these $(n-1)$ Givens rotations composed can generate a $n \times n$ matrix $\mathcal{R}(\psi)$ according to

$$\mathcal{R}(\psi) = \mathcal{R}_1(\psi)\mathcal{R}_2(\psi)\ldots\mathcal{R}_{n-1}(\psi). \tag{7}$$

It is clear that this choice of matrix $\mathcal{R}(\psi)$ is a special case in the sense that the rotation angles of the Givens rotation matrices $\mathcal{R}_i(\psi)$ are chosen to be equal. Note that matrix $\mathcal{R}(\psi)$ is used in [1] for the computation of the eigenvalues and eigenvectors of the covariance matrix of a random vector X of n principal components.

An explicit representation of $\mathcal{R}(\psi)$ is of the form

$$\begin{pmatrix} \cos(\psi) & -\cos(\psi)\sin(\psi) & \dots & \dots & (-1)^n\cos(\psi)\sin^{n-2}(\psi) & (-1)^{n+1}\sin^{n-1}(\psi) \\ \sin(\psi) & \cos^2(\psi) & -\cos^2(\psi)\sin(\psi) & & (-1)^{n-1}\cos^2(\psi)\sin^{n-3}(\psi) & (-1)^n\cos(\psi)\sin^{n-2}(\psi) \\ 0 & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -\cos^2(\psi)\sin(\psi) & \vdots \\ \vdots & & \ddots & \ddots & \cos^2(\psi) & -\cos(\psi)\sin(\psi) \\ 0 & \dots & \dots & 0 & \sin(\psi) & \cos(\psi) \end{pmatrix}. \tag{8}$$

The matrix (8) is almost an upper triangular matrix but with $\sin(\psi)$ on the first subdiagonal and the $(n-2)\times(n-2)$ submatrix starting on position (2, 2) is a Toeplitz upper triangular matrix. To summarize, the Givens rotations composed can transform the basis of the space to any other frame in the space. The matrix $\mathcal{R}(\psi)$ fulfills the property $\text{Det}(\mathcal{R}(\psi)) = 1$ and $\mathcal{R}^\top(\psi)\mathcal{R}(\psi) = I_n$, where $\text{Det}(X)$ is the determinant of square matrix X and I_n is the $n\times n$ identity matrix, and the property $\mathcal{R}(\psi = 0) = I_n$ holds. When n is odd, the matrix $\mathcal{R}(\psi)$ will have an eigenvalue 1 and the remaining eigenvalues are pairs of complex conjugates whose product is 1, the latter property also holds for n even. Consequently, the matrix $\mathcal{R}(\psi)$ is a rotation matrix and obviously orthogonal and it commutes with its transpose. We now conclude that every rotation matrix, when expressed in a suitable coordinate system, partitions into independent rotations of two-dimensional subspaces like in (7). All the transformed coordinates $\widetilde{x}_1 - \widetilde{y}_1, \widetilde{x}_2 - \widetilde{y}_2, \dots, \widetilde{x}_n - \widetilde{y}_n$ are then affected by the rotation matrix and angle ψ through the linear transformation

$$(\widetilde{x} - \widetilde{y}) = \mathcal{R}(\psi)\,(x - y)\,. \tag{9}$$

Considering the complexity of the rotation matrix $\mathcal{R}(\psi)$, we shall limit ourselves to the Givens rotation matrix $\mathcal{R}_i(\psi)$ present in (5) instead of using the computationally much more complex transformation displayed in (9). However, an appropriate approach to this problem can be a subject for future research.

Since the representation of the Givens rotation matrix $\mathcal{R}_i(\psi)$ is determined by subscript i, we set forth the Givens rotation matrix $\mathcal{G}(\psi)$ when $i = n-1$ in matrix $\mathcal{R}_i(\psi)$ and a $n\times n$ rotation matrix $\mathcal{D}(\psi)$ of the form

$$\begin{pmatrix} \cos(\psi) & \sin(\psi) & 0 \\ -\sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & I_{n-2} \end{pmatrix}.$$

In the next proposition we will show that the matrices $\mathcal{D}(\psi)$ and $\mathcal{G}(\psi)$ are similar.

Proposition 2.1. *The matrices $\mathcal{D}(\psi)$ and $\mathcal{G}(\psi)$ are similar.*

Proof. The following equality holds true

$$J_n\mathcal{D}(\psi)J_n = \mathcal{G}(\psi),$$

where

$$J_n = \begin{pmatrix} 0 & \cdots & 0 & 1 \\ \vdots & \cdot^{\cdot^{\cdot}} & 1 & 0 \\ 0 & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}$$

and taking into account the property $J_n^{-1} = J_n$ yields

$$\mathcal{D}(\psi)J_n = J_n\mathcal{G}(\psi),$$

which completes the proof. □

Similarity between other representations of Givens rotation matrices can be shown analogously.

Choice of the rotation matrix $\mathcal{R}_i(\psi)$ will yield equivalent results to the results derived throughout this paper. The subscript i will be subsequently used in this paper and refers to the ith position in matrix $\mathcal{R}_i(\psi)$ such that $i \in \{1, \ldots, n-1\}$ and is associated with the first coordinate that is affected by the rotation angle ψ. A possible change of the value i in the Givens rotation matrix $\mathcal{R}_i(\psi)$, will not result in structural changes. Equivalent results will be obtained for all i when $i \in \{1, \ldots, n-1\}$.

The relation between the original coordinates $(x - y) = (x_1 - y_1, x_2 - y_2, \ldots, x_n - y_n)^\top$ and the rotated coordinates $(\widetilde{x} - \widetilde{y}) = (\widetilde{x}_1 - \widetilde{y}_1, \widetilde{x}_2 - \widetilde{y}_2, \ldots, \widetilde{x}_n - \widetilde{y}_n)^\top$ is established through (5).

2.1. A transformed covariance matrix

We shall now use (5) to set forth the connection between the covariance matrices obtained in both coordinate systems. Considering the nonrandomness of $y_1, y_2, \ldots, y_n$, leads to $\text{Cov}(x-y) = \text{Cov}(x) = \mathcal{S}(x)$ and $\text{Cov}(\widetilde{x} - \widetilde{y}) = \text{Cov}(\widetilde{x}) = \mathcal{S}(\widetilde{x})$. Applying this property to (5) yields the following matrix equation

$$\mathcal{S}(\widetilde{x}) = \mathcal{R}_i(\psi)\mathcal{S}(x)\mathcal{R}_i^\top(\psi), \tag{10}$$

the symbol $\top$ is the transposition, an explicit representation is given by

$$\mathcal{S}(\widetilde{x}) = \begin{pmatrix} \mathcal{S}_{11}(\widetilde{x}) & \mathcal{S}_{12}(\psi) & \mathcal{S}_{13}(\widetilde{x}) \\ \mathcal{S}_{12}^\top(\psi) & \mathcal{S}_{22}(\psi) & \mathcal{S}_{23}(\psi) \\ \mathcal{S}_{13}^\top(\widetilde{x}) & \mathcal{S}_{23}^\top(\psi) & \mathcal{S}_{33}(\widetilde{x}) \end{pmatrix},$$

where

$$\mathcal{S}_{22}(\psi) = \begin{pmatrix} \widetilde{s}_{i,i}(\psi) & \widetilde{s}_{i,i+1}(\psi) \\ \widetilde{s}_{i+1,i}(\psi) & \widetilde{s}_{i+1,i+1}(\psi) \end{pmatrix}$$

and

$$\begin{cases} \widetilde{s}_{i,i}(\psi) = s_{i,i}\cos^2(\psi) - s_{i,i+1}\sin(2\psi) + s_{i+1,i+1}\sin^2(\psi), \\ \widetilde{s}_{i+1,i+1}(\psi) = s_{i+1,i+1}\cos^2(\psi) + s_{i,i+1}\sin(2\psi) + s_{i,i}\sin^2(\psi), \\ \widetilde{s}_{i,i+1}(\psi) = \widetilde{s}_{i+1,i}(\psi) = (1/2)\left(2s_{i,i+1}\cos(2\psi) + (s_{i,i} - s_{i+1,i+1})\sin(2\psi)\right). \end{cases}$$

Note that the submatrices of the transformed covariance matrix $\mathcal{S}(\widetilde{x})$ whose rows or columns are on the ith and $(i+1)$th position depend on the rotation angle ψ. Matrix $\mathcal{S}_{11}(\widetilde{x})$ is a $(i-1)\times(i-1)$ symmetric matrix with entries $s_{k,j}$, where $k, j \in \{1, \ldots, i-1\}$ and matrix $\mathcal{S}_{33}(\widetilde{x})$ is a $(n-i-1)\times(n-i-1)$ symmetric matrix with entries $s_{k,j}$, where $k, j \in \{i+2, \ldots, n\}$. The entries of the remaining submatrices shall not be specified explicitly since there are of no use in the statistical distance measure set forth in this paper.

Combining (5) and (10) will allow us to rewrite (4) in terms of the original data combined with its corresponding covariances and the rotation angle ψ, to obtain

$$\begin{aligned} d^2(x,y) = & \sum_{j=1, j\neq i,i+1}^{n} \left\{ \frac{(x_j - y_j)^2}{s_{j,j}} \right\} + \frac{\{(x_i - y_i)\cos(\psi) - (x_{i+1} - y_{i+1})\sin(\psi)\}^2}{\widetilde{s}_{i,i}(\psi)} \\ & + \frac{\{(x_{i+1} - y_{i+1})\cos(\psi) + (x_i - y_i)\sin(\psi)\}^2}{\widetilde{s}_{i+1,i+1}(\psi)}. \end{aligned} \tag{11}$$

A prerequisite for ensuring positivity of the covariances $\widetilde{s}_{i,i}(\psi)$ and $\widetilde{s}_{i+1,i+1}(\psi)$ in (11) is given in the next section. When no rotation is applied to the coordinate system, the distance measure (11) then coincides with (3).

2.2. A transformed statistical distance as quadratic form distance measures

The statistical distance of the stochastic vector x from the fixed vector y for situations in which the variables are correlated can also be expressed by the quadratic form

$$\begin{aligned} d^2(x,y) &= Q_A(x,y) = \langle A(x-y), (x-y)\rangle = (x-y)^\top A(x-y) \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n} a_{jk}(x_j-y_j)(x_k-y_k), \end{aligned} \tag{12}$$

where $\langle\cdot,\cdot\rangle$ denotes the standard scalar product on $\mathbb{R}^n$, and matrix A is symmetric with entries a_{ij}. The distance measures (12) can be seen as acting on a distorted Euclidean space. When A is diagonal with positive entries, the corresponding $Q_A(x,y)$ are weighted Euclidean distance measures. The family of quadratic form distances are defined for general positive semi-definite matrices. However, for positive definite matrices, the distance measure forms a metric. In Proposition 2.2 we show that the matrix A is positive definite.

Once the entries a_{ij} are known, the distance $d(x,y)$ can be computed. The coordinates of vector x are then a constant squared distance α^2 from y or $\sum_{j=1}^{n}\sum_{k=1}^{n} a_{jk}(x_j-y_j)(x_k-y_k) = \alpha^2$. The entries a_{ij} are numbers such that the distance or quadratic form is positive for all possible values of $x_1, x_2, \ldots, x_n$. The entries a_{ij} of the matrix A are determined by the angle ψ and the covariances $s_{1,1}, s_{2,2}, \ldots, s_{n,n}$ and $s_{n-1,n}$ computed by the original data. For that purpose we identify (12) with (11), the matrix A has then the representation

$$A = \begin{pmatrix} a_{1,1} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & & & \vdots \\ \vdots & \ddots & a_{i-1,i-1} & \ddots & & & \vdots \\ \vdots & & \ddots & A_{i,i+1} & \ddots & & \vdots \\ \vdots & & & \ddots & a_{i+2,i+2} & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & a_{n,n} \end{pmatrix},$$

where

$$A_{i,i+1} = \begin{pmatrix} a_{i,i} & a_{i,i+1} \\ a_{i,i+1} & a_{i+1,i+1} \end{pmatrix}$$

and

$$\begin{aligned} a_{i,i} &= \frac{\cos^2(\psi)}{\widetilde{s}_{i,i}(\psi)} + \frac{\sin^2(\psi)}{\widetilde{s}_{i+1,i+1}(\psi)}, \quad a_{i+1,i+1} = \frac{\sin^2(\psi)}{\widetilde{s}_{i,i}(\psi)} + \frac{\cos^2(\psi)}{\widetilde{s}_{i+1,i+1}(\psi)}, \\ a_{i,i+1} &= -\frac{\sin(\psi)\cos(\psi)}{\widetilde{s}_{i,i}(\psi)} + \frac{\sin(\psi)\cos(\psi)}{\widetilde{s}_{i+1,i+1}(\psi)}, \\ a_{1,1} &= \frac{1}{s_{1,1}}, \ldots, a_{i-1,i-1} = \frac{1}{s_{i-1,i-1}}, \quad a_{i+2,i+2} = \frac{1}{s_{i+2,i+2}}, \ldots, a_{n,n} = \frac{1}{s_{n,n}}. \end{aligned}$$

When no rotation is applied to the coordinate system, the distance measure (3) is then equivalent to expression (12) with the appropriate matrix A. To show that (12) is strict positive it suffices to show the positive definiteness of matrix A. For that purpose we remind the following property of symmetric matrices which is summarized in the next lemma, see e.g. [4].

Lemma 2.2. *Consider a symmetric $n \times n$ matrix C. For $r \in \{1, 2, \ldots, n\}$, let $C^{(r)}$ be the $r \times r$ matrix obtained by omitting all rows and columns of C past the rth. These matrices $C^{(r)}$ are called the principal submatrices of C. The matrix C is positive definite iff $\mathrm{Det}(C^{(r)}) > 0$, for all $r \in \{1, 2, \ldots, n\}$.*

Application of this lemma in the following proposition leads to:

Proposition 2.3. *The symmetric matrix A is positive definite $A \succ 0$ for all ψ when*

$$x_i \cos(\psi) - x_{i+1} \sin(\psi) = X(\psi) \quad \text{and} \quad x_i \sin(\psi) + x_{i+1} \cos(\psi) = Y(\psi), \tag{13}$$

where $X(\psi)$ and $Y(\psi)$ are random variables different from zero.

Proof. By virtue of Lemma 2.2, we first show that the matrix $A_{i,i+1}$ is positive definite, we therefore consider the principal submatrices of $A_{i,i+1}$. The first submatrix is entry $a_{i,i}$ which shall be represented accordingly by taking the equalities (13) into consideration, to obtain

$$a_{i,i} = \frac{\cos^2(\psi)}{\mathrm{Var}\,(X(\psi))} + \frac{\sin^2(\psi)}{\mathrm{Var}\,(Y(\psi))},$$

where $\mathrm{Var}(X)$ denotes variance of a random variable X and considering condition (13) implies that the variances

$$\mathrm{Var}\,(X(\psi)) > 0 \quad \text{and} \quad \mathrm{Var}\,(Y(\psi)) > 0.$$

It is now clear that the first principal submatrix $a_{i,i}$ is positive.

The determinant of the second principal submatrix of $A_{i,i+1}$ is

$$\mathrm{Det}(A_{i,i+1}) = \frac{1}{[\mathrm{Var}\,(X(\psi))]\,[\mathrm{Var}\,(Y(\psi))]} > 0.$$

Consequently, the matrix $A_{i,i+1}$ is positive definite considering the assumptions set forth in (13). The $(n-2)$ eigenvalues of the matrix A associated with the remaining components on the main diagonal next to the matrix $A_{i,i+1}$, are given by $(1/s_{1,1}), \ldots, (1/s_{i-1,i-1}), (1/s_{i+2,i+2}), \ldots, (1/s_{n,n})$, and are strict positive since $s_{1,1} > 0, \ldots, s_{i-1,i-1} > 0, s_{i+2,i+2} > 0, \ldots, s_{n,n} > 0$. It can be concluded that the matrix A is positive definite, this completes the proof. □

It follows from Proposition 2.3 that the distance measure $Q_A(x, y)$ forms a metric. Consequently, the statistical distance measure (4) is also a metric.

Imposing condition (13) in Proposition 2.3 is justified by the fact that the random variables x_1, $x_2, \ldots, x_n$ are stochastically dependent. Additionally, if $X(\psi)$ and $Y(\psi)$ in (13) are constant numbers instead of random variables leads to variances equal to 0 in the denominator of $a_{i,i}$ and $\mathrm{Det}(A_{i,i+1})$. Another aspect to consider, the rotation angles ψ should not fulfill the equations $x_i \cos(\psi) - x_{i+1} \sin(\psi) = 0$ and $x_i \sin(\psi) + x_{i+1} \cos(\psi) = 0$. A short illustration of the assumptions is now set forth. Assume the dependence between the random variables x_i and x_{i+1} to be of the form $x_{i+1} = \alpha x_i$, where $\alpha \in \mathbb{R}$ such that $X(\psi) = x_i\,(\cos(\psi) - \alpha \sin(\psi))$. The equation

$$\cos(\psi) - \alpha \sin(\psi) = 0 \tag{14}$$

has nontrivial solutions of the form

$$\psi_1 = \mathrm{Arc}\cos\left(\frac{\alpha}{\sqrt{1+\alpha^2}}\right), \quad \psi_2 = -\mathrm{Arc}\cos\left(-\frac{\alpha}{\sqrt{1+\alpha^2}}\right).$$

Consequently, the condition set forth in Proposition 2.3 implies that when the linear dependence between the random variables x_i and x_{i+1} as displayed above is considered, the rotation angles should fulfill the property, $\psi \neq \psi_1$ and $\psi \neq \psi_2$, in order to avoid $X(\psi) = 0$. In other words one does not expect the rotation angle ψ to fulfill equation (14). A similar approach holds for $Y(\psi)$.

The eigenvalues of matrix A are now displayed,

$$\lambda_i = \frac{1}{\widetilde{s}_{i,i}(\psi)} = \frac{1}{\mathrm{Var}\,(X(\psi))}, \quad \lambda_{i+1} = \frac{1}{\widetilde{s}_{i+1,i+1}(\psi)} = \frac{1}{\mathrm{Var}\,(Y(\psi))}, \tag{15}$$

$$\lambda_1 = (1/s_{1,1}), \ldots, \lambda_{i-1} = (1/s_{i-1,i-1}), \quad \lambda_{i+2} = (1/s_{i+2,i+2}), \ldots, \lambda_n = (1/s_{n,n}), \tag{16}$$

where λ_i and λ_{i+1} are the eigenvalues of matrix $A_{i,i+1}$, and by virtue of (13) we have $\lambda_i > 0$ and $\lambda_{i+1} > 0$. Consequently, all eigenvalues of matrix A are positive, this property can also serve as an alternative proof of Proposition 2.3. The eigenvectors of λ_i and λ_{i+1} are $v_i = (0, \ldots, 0, -\mathrm{cotan}(\psi), 1, 0, \ldots, 0)^\top$ and $v_{i+1} = (0, \ldots, 0, 1, \mathrm{cotan}(\psi), 0, \ldots, 0)^\top$, whereas the remaining $(n-2)$ eigenvectors are the first up to $(i-1)$th standard basic vectors in $\mathbb{R}^n$ as well as the $(i+2)$th up to the nth standard basic vectors in $\mathbb{R}^n$, which are labeled as $w_1, \ldots, w_{i-1}, w_{i+2}, \ldots, w_n$. By w_j we denote the jth coordinate vector, $w_j = (0, \ldots, 1, \ldots, 0)^\top$, with all its components equal to 0 except the jth component which equals 1. The orthonormal versions of v_i and v_{i+1} are set forth, to obtain $w_i = (0, \ldots, 0, -\cos(\psi), \sin(\psi), 0, \ldots, 0)^\top$ and $w_{i+1} = (0, \ldots, 0, \sin(\psi), \cos(\psi), 0, \ldots, 0)^\top$. Note that for $\psi = (\pi/2)$, matrix A is diagonal with $a_{i,i} = (1/s_{i,i})$ and $a_{i+1,i+1} = (1/s_{i+1,i+1})$, consequently, matrix A has only positive entries such that the corresponding quadratic form distance measure $Q_A(x, y)$ is a weighted Euclidean distance measure, cf. (3). When $\psi = (\pi/4)$, we have

$$a_{i,i} = a_{i+1,i+1} = \frac{2s_{i,i} + 2s_{i+1,i+1}}{(s_{i,i} - 2s_{i,i+1} + s_{i+1,i+1})\,(s_{i,i} + 2s_{i,i+1} + s_{i+1,i+1})}$$

and

$$a_{i,i+1} = \frac{-4s_{i,i+1}}{(s_{i,i} - 2s_{i,i+1} + s_{i+1,i+1})\,(s_{i,i} + 2s_{i,i+1} + s_{i+1,i+1})}.$$

As can be seen from (15), the corresponding eigenvalues are then given by

$$\lambda_i = \frac{2}{(s_{i,i} - 2s_{i,i+1} + s_{i+1,i+1})}, \quad \lambda_2 = \frac{2}{(s_{i,i} + 2s_{i,i+1} + s_{i+1,i+1})}.$$

Every quadratic form distance measure having a metric property is characterized by a square matrix A which is positive definite. However, every positive definite symmetric matrix A can be subjected to a Cholesky factorization of the form $A = L^\top L$, see for example [8,10]. Consequently, there is a unique upper triangular matrix L with positive diagonal entries. This yields for the vectors $x, y \in \mathbb{R}^n$ the quadratic form,

$$Q_A(x, y) = (x - y)^\top A\,(x - y) = (L\,(x - y))^\top L\,(x - y) = \|Lx - Ly\|^2, \tag{17}$$

which is the Euclidean distance between Lx and Ly. The matrix L of the Cholesky factorization associated with matrix A and for all ψ is given by

$$L = \begin{pmatrix}
\sqrt{a_{1,1}} & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & \ddots & \ddots & & & & & 0 \\
0 & 0 & \sqrt{a_{i-1,i-1}} & 0 & & & & \vdots \\
\vdots & \ddots & \ddots & \sqrt{a_{i,i}} & \frac{a_{i,i+1}}{\sqrt{a_{i,i}}} & & & \vdots \\
\vdots & & \ddots & \ddots & \sqrt{\frac{\mathrm{Det}(A_{i,i+1})}{a_{i,i}}} & 0 & & \vdots \\
\vdots & & & \ddots & \ddots & \sqrt{a_{i+2,i+2}} & \ddots & \vdots \\
\vdots & & & & \ddots & \ddots & \ddots & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 & \sqrt{a_{n,n}}
\end{pmatrix}.$$

Some geometry is now considered for the quadratic form (12) which is from $\mathbb{R}^n$ to $\mathbb{R}$. Let $\mathcal{B}$ be the eigenspace spanned by the orthonormal eigenvectors $w_1, w_2, \ldots, w_{n-2}, w_{n-1}, w_n$ of the matrix A that have associated distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{n-2}, \lambda_{n-1}, \lambda_n$, or $\mathcal{B}$ =span$\{w_1, w_2, \ldots, w_{n-2}, w_{n-1}, w_n\}$. The quadratic form distance measure (17) can be represented in terms of the eigenvalues of A, to obtain

$$1 = \frac{1}{\|Lx - Ly\|^2} \left\{ \sum_{j=1, j \neq i, i+1}^{n} \left(\frac{c_j^2}{s_{j,j}} \right) + \frac{c_i^2}{\widetilde{s}_{i,i}(\psi)} + \frac{c_{i+1}^2}{\widetilde{s}_{i+1,i+1}(\psi)} \right\}, \tag{18}$$

where the $c_j, j \in \{1, \ldots, n\}$ are the coordinates of $(x - y)$ with respect to the orthonormal eigenbasis $\mathcal{B}$. Equality (18) is the equation of an ellipsoid in a n-dimensional coordinate system and the corresponding eigenspaces (each of these eigenspaces are one-dimensional) of A are the principal axes of (18). This is justified by the fact that the covariances $s_{j,j}$, $\widetilde{s}_{i,i}(\psi)$ and $\widetilde{s}_{i+1,i+1}(\psi)$ are positive. Consequently, the statistical distance measure, which is now displayed by the representations (11) and (18), yields the equality

$$\|Lx - Ly\|^2 = d^2(x, y) \tag{19}$$

that describes the n-dimensional ellipsoid (18) where

$$\sqrt{k\, s_{1,1}}, \ldots, \sqrt{k\, s_{i-1,i-1}}, \sqrt{k\, \widetilde{s}_{i,i}(\psi)}, \sqrt{k\, \widetilde{s}_{i+1,i+1}(\psi)}, \sqrt{k\, s_{i+2,i+2}}, \ldots, \sqrt{k\, s_{n,n}}$$

are the lengths of the principal axes for $k = \|Lx - Ly\|^2$. Such that the principal axes of the n-dimensional ellipsoid given by (18) reflect the covariances or the associations between the random variables $x_1, x_2, \ldots, x_n$. In the next section we introduce a statistical distance measure where entries of the FIM are used as weighting coefficients.

3. Transformed statistical distance measures and the Fisher information matrix

The Fisher information is an ingredient of the Cramér–Rao inequality and belongs to the basics of asymptotic estimation theory in mathematical statistics. The Cramér–Rao theorem [23] is therefore considered. When assuming that the estimators (2) 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. Its quantum analog was introduced immediately after the foundation of mathematical quantum estimation theory in the 1960s, see [9,22] for a rigorous exposition of the subject. 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.

More specifically, the Fisher information is extensively discussed in the quantum information theory literature, see e.g. [2,24]. An interconnection between the Fisher information and a statistical distance measure is established at the scalar level, see [3,13,20]. In [13], a review of the concept of statistical distance is given, both for classical probability distributions and for quantum states. The authors relate the statistical distance to the Fisher information, which measures the amount of information about a parameter obtained in a given measurement. The authors consider a statistical distance measure in the space of probability distributions, this measure quantifies the difference between two probability distributions or measures the length of a curve in the space of probability distributions. This curve is parameterized by a variable θ_1 and when more parameter variables $\theta_1, \ldots, \theta_N$, are involved they should be combined in a parameter vector $\varphi = (\theta_1, \ldots, \theta_N)^\top$ along the appropriate curve. However, in the quantum information literature mentioned above, the results are limited to one parameter variable. Whereas in this paper we consider statistical distance measures based on estimated parameters $\vartheta_1, \vartheta_2, \ldots, \vartheta_m$, which are random components and are derived from the observed measurements $x_1, x_2, \ldots, x_n$, where $m < n$. These parameters are collected in the stochastic parameter vector (2). In previous section, entries of the covariance matrix $\mathcal{S}(x)$ are the weighting coefficients of the statistical distance measure, whereas in this section, entries of the FIM are the appropriate weighting coefficients. This will enable us to formulate conditions of the FIM that ensure the metric properties of the appropriate statistical distance measure. The role of the FIM of stationary processes is also emphasized.

The approach used in this paper is such that matrix properties are crucial for ensuring the relevance of the introduced statistical distance measures.

3.1. The Fisher information matrix and statistical distance measures

An extension of the results set forth in the previous section shall be displayed. For that purpose we consider the set of maximum likelihood estimated parameters $\vartheta_1, \vartheta_2, \ldots, \vartheta_{,m}$ whose computation is based on observations $x_1, x_2, \ldots, x_n$ and where $m < n$. These estimated parameters are random variables, see e.g.[5]. We consider now the estimated random parameter vector ϑ given in (2) whose length is the distance that will be considered, for $\vartheta \in \mathbb{R}^m$. Consequently, an equivalent approach to the one used in Section 2 can be applied but instead of using the corresponding covariances as weighting coefficients, entries of the FIM will be inserted as the new weighting coefficients. The linear transformation $\widetilde{\vartheta} = \mathcal{L}_i(\phi)\vartheta$, which is equivalent to (5) is set forth and $\mathcal{L}_i(\phi)$ is an appropriate $m \times m$ Givens rotation matrix with rotation angle ϕ and having the same configuration as (6), with $0 \leqslant \phi \leqslant 2\pi$ and $i \in \{1, \ldots, m-1\}$. The covariance matrix of $\widetilde{\vartheta}$ is,

$$\mathrm{Cov}(\widetilde{\vartheta}) = \mathcal{L}_i(\phi)\mathrm{Cov}(\vartheta)\mathcal{L}_i^\top(\phi), \tag{20}$$

and is a fundamental matrix equation from which the FIM will be derived and is equivalent to (10). Considering the Cramér–Rao theorem mentioned in the former section allows us to assert that the covariance $\mathrm{Cov}(\vartheta)$ is equal to the inverse of the FIM. It suffices to assume that one of the covariance matrices $\mathrm{Cov}(\widetilde{\vartheta})$ and $\mathrm{Cov}(\vartheta)$ is nonsingular and taking the property of the rotation matrix $\mathcal{L}_i(\phi)$ into account, results in the following matrix equation when the matrix inversion of (20) is applied, to obtain

$$\mathcal{F}_\phi(\vartheta) = \mathcal{L}_i(\phi)\mathcal{F}(\vartheta)\mathcal{L}_i^\top(\phi), \tag{21}$$

where $\mathcal{F}_\phi(\vartheta)$ and $\mathcal{F}(\vartheta)$ are respectively the transformed and untransformed Fisher information matrices. Since the rotation matrix $\mathcal{L}_i(\phi)$ is orthogonal, it can be concluded from (21) that the transformed and untransformed Fisher information matrices $\mathcal{F}_\phi(\vartheta)$ and $\mathcal{F}(\vartheta)$ respectively, are similar. The FIM $\mathcal{F}(\vartheta)$ measures the amount of information about the parameter vector ϑ which is given in (2), and is obtained through observed measurements, see next to the statistical literature, [6,7] in the physics literature.

It is straightforward to conclude from (21) that a positive definite symmetric FIM $\mathcal{F}(\vartheta)$ yields a positive definite and symmetric transformed FIM $\mathcal{F}_\phi(\vartheta)$, this holds for all ϕ.

Note that when rotation matrix $\mathcal{L}(\phi) = \mathcal{L}_1(\phi)\mathcal{L}_2(\phi)\ldots\mathcal{L}_{m-1}(\phi)$ is considered, a variant of (20) and (21) is then of the form

$$\mathrm{Covar}(\widetilde{\vartheta}) = \mathcal{L}(\phi)\mathrm{Cov}(\vartheta)\mathcal{L}^\top(\phi) \quad \text{and} \quad \mathcal{J}_\phi(\vartheta) = \mathcal{L}(\phi)\mathcal{F}(\vartheta)\mathcal{L}^\top(\phi),$$

respectively and where $\mathrm{Covar}(\widetilde{\vartheta})$ and $\mathcal{J}_\phi(\vartheta)$ are the appropriate transformed covariance matrix and transformed Fisher information matrix.

By and large, when empirical estimation is applied, one computes first the FIM $\mathcal{F}(\vartheta)$ which is assumed to be nonsingular. This enables us to successfully perform the computation of a reliable corresponding covariance matrix of the estimated parameters, this is obtained after inversion of $\mathcal{F}(\vartheta)$. The derived covariance matrix is crucial for further statistical analysis. However, it is not uncommon that empirical statistical estimation results in a singular or near singular FIM. A near singular FIM $\mathcal{F}(\vartheta)$ yields very strongly inaccurate numerical results when an inversion is applied, consequently matrix equation (20) is then irrelevant. However, in Section 4, the example shows that even when the FIM $\mathcal{F}(\vartheta)$ is singular, the statistical distance measure (24) can fulfill the metric properties. This depends on the choice of the rotation angle ϕ. The invertibility condition of the matrix $\mathcal{F}(\vartheta)$ for stationary processes is formulated in the next section.

We shall introduce an appropriate variant of the distance measures (3) and (4), where the estimated parameters $\vartheta_1, \vartheta_2, \ldots, \vartheta_m$ are considered instead of the measurements $x_1 - y_1, x_2 - y_2, \ldots, x_n - y_n$ and some of the entries of the Fisher information matrices $\mathcal{F}(\vartheta)$ and $\mathcal{F}_\phi(\vartheta)$ are the weighting coefficients instead of the covariances $s_{1,1}, s_{2,2}, \ldots, s_{n,n}$, to obtain

$$d_{\mathcal{F}}(\vartheta) = \left(\sum_{j=1}^{m} \left\{ \frac{\vartheta_j^2}{f_{j,j}} \right\} \right)^{1/2} \tag{22}$$

and

$$d_{\mathcal{F}_\phi}(\vartheta) = \left(\sum_{j=1}^{m} \left\{ \frac{\widetilde{\vartheta}_j^2}{\widetilde{f}_{j,j}} \right\} \right)^{1/2} \tag{23}$$

respectively. As mentioned above, a positive definite FIM, $\mathcal{F}(\vartheta) \succ 0$, implies a positive definite transformed FIM, $\mathcal{F}_\phi(\vartheta) \succ 0$. This property holds because of the orthogonality property of rotation matrix $\mathcal{L}_i(\phi)$. 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, in Lemma 3.4 it is proved that the elements on the main diagonal of a singular FIM of a stationary ARMA process are also positive. It is clear that the entries of the FIM $\mathcal{F}(\vartheta)$ depend on ϑ but for typographical brevity we will henceforth represent the entries $f_{j,l}$ without argument ϑ. Note that matrix equation (21) leads to $\widetilde{f}_{j,j} = f_{j,j}$, for all $j \in \{1, \ldots, m\}$ when $j \neq i, i+1$. Combining (21), (22) and (23), yields an equation equivalent to (11) but with entries of the Fisher information matrices $\mathcal{F}(\vartheta)$ and $\mathcal{F}_\phi(\vartheta)$ as weighting coefficients, to obtain

$$d^2_{\mathcal{F}_\phi}(\vartheta) = \sum_{j=1, j\neq i, i+1}^{m} \left\{ \frac{\vartheta_j^2}{f_{j,j}} \right\} + \frac{\{\vartheta_i \cos(\phi) - \vartheta_{i+1} \sin(\phi)\}^2}{\widetilde{f}_{i,i}(\phi)} + \frac{\{\vartheta_{i+1} \cos(\phi) + \vartheta_i \sin(\phi)\}^2}{\widetilde{f}_{i+1,i+1}(\phi)}, \tag{24}$$

where

$$\widetilde{f}_{i,i}(\phi) = f_{i,i} \cos^2(\phi) - f_{i,i+1} \sin(2\phi) + f_{i+1,i+1} \sin^2(\phi), \tag{25}$$

$$\widetilde{f}_{i+1,i+1}(\phi) = f_{i+1,i+1} \cos^2(\phi) + f_{i,i+1} \sin(2\phi) + f_{i,i} \sin^2(\phi), \tag{26}$$

and $f_{j,l}$ are entries of the FIM $\mathcal{F}(\vartheta)$. The following inequalities are required for (24) to hold

$$\widetilde{f}_{i,i}(\phi) > 0 \tag{27}$$

and

$$\widetilde{f}_{i+1,i+1}(\phi) > 0. \tag{28}$$

It can be derived from (25) and (26) that $\mathrm{Tr}(\mathcal{F}(\vartheta)) = \mathrm{Tr}\big(\mathcal{F}_\phi(\vartheta)\big)$, where $\mathrm{Tr}(X)$ is the trace of a square matrix X. In case the rotation angle satisfies $\phi = \pi$, the properties $\widetilde{f}_{i,i}(\phi) = f_{i,i}$ and $\widetilde{f}_{i+1,i+1}(\phi) = f_{i+1,i+1}$ hold true. Despite the arguments mentioned above, we prove the inequalities (27) and (28) in the following proposition.

Proposition 3.1. *A positive definite FIM $\mathcal{F}(\vartheta) \succ 0$ implies that the inequalities (27) and (28) hold true for all values of ϕ.*

Proof. As mentioned above, a positive definite FIM, $\mathcal{F}(\vartheta) > 0$, implies $\mathcal{F}_\phi(\vartheta) > 0$ such that the determinant of the transformed FIM is positive $\mathrm{Det}\big(\mathcal{F}_\phi(\vartheta)\big) > 0$. Consequently, all submatrices on the main diagonal of the transformed FIM $\mathcal{F}_\phi(\vartheta)$ are positive definite. Therefore, when Lemma 2.2 is applied, it suffices to consider the ith 2×2 submatrix on the main diagonal of the symmetric matrix $\mathcal{F}_\phi(\vartheta)$ given in (21), cfr. $\mathcal{S}_{22}(\psi)$ in (10), to obtain

$$\Omega(\phi) = \begin{pmatrix} \widetilde{f}_{i,i}(\phi) & \widetilde{f}_{i,i+1}(\phi) \\ \widetilde{f}_{i,i+1}(\phi) & \widetilde{f}_{i+1,i+1}(\phi) \end{pmatrix}. \tag{29}$$

where

$$\widetilde{f}_{i,i+1}(\phi) = (1/2)\left(2 f_{i,i+1} \cos(2\phi) + (f_{i,i} - f_{i+1,i+1}) \sin(2\phi)\right).$$

The determinant of the first principal submatrix of (29) coincides with the left hand side of (27) given by $\widetilde{f}_{i,i}(\phi)$. By virtue of the property $\mathrm{Det}\big(\mathcal{F}_{\phi}(\vartheta)\big) > 0$, it should be positive. The determinant of the second principal submatrix of (21) is the determinant of (29), which is also positive by virtue of Lemma 2.2 and the assumption $\mathcal{F}(\vartheta) \succ 0$ which implies $\mathrm{Det}\big(\mathcal{F}_{\phi}(\vartheta)\big) > 0$. It is then straightforward to conclude that condition $\widetilde{f}_{i+1,i+1}(\phi) > 0$, holds. □

3.1.1. Statistical distance measure – Fisher information and quantum information

As mentioned above, in quantum information, see [3,13,20], the Fisher information, the information about a parameter θ in a particular measurement procedure, is expressed in terms of the statistical distance s which is defined as a measure to distinguish two probability distributions on the basis of measurement outcomes. The Fisher information and the statistical distance are statistical quantities, and generally refer to many measurements as it is the case in this paper. However, in the quantum information theory and quantum statistics context, the problem set up shall be formulated 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 in the quantum information literature mentioned above is of the form

$$\mathcal{F}\,(\theta) = \left(\frac{ds}{d\theta}\right)^2, \tag{30}$$

so the Fisher information is the square of the derivative to θ of the statistical distance. Whereas in (22) and (24), 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 associated with a certain process. So the statistical distance measures set forth in this paper are not limited to information based on one particular measurement and parameter. A challenging question could therefore be, how to generalize equality (30) to a case of many observations that lead to more parameters so that a Fisher information matrix instead of Fisher information could be interconnected to an appropriate statistical distance measure given by a distance matrix. This question could equally be a challenge in matrix theory and quantum information.

3.1.2. Some additional properties

It is worth observing that an equivalent to Proposition 2.3, when applied to the statistical distance measure (24), can be proved in a straightforward manner when Proposition 3.1 is taken into account. Consider therefore the quadratic form distance measure

$$d^2_{\mathcal{F}_{\phi}}(\vartheta) = Q_B(\vartheta) = \langle B\vartheta, \vartheta\rangle = \vartheta^{\top} B\vartheta, \tag{31}$$

where the $m \times m$ symmetric matrix B has an equivalent representation to matrix A in (12) and with $s_{jl} \to f_{jl}$ and $\psi \to \phi$. Condition (13) set forth in Proposition 2.3 is replaced by the condition, positive definite FIM $\mathcal{F}(\vartheta) \succ 0$, so by virtue of Proposition 3.1 we have $\widetilde{f}_{i,i}(\phi) > 0, \widetilde{f}_{i+1,i+1}(\phi) > 0$. Note that the equivalent to the determinant of submatrix $A_{i,i+1}$ of A in (12), is $\mathrm{Det}\,(B_{i,i+1}) = \Big(1/\Big(\widetilde{f}_{i,i}(\phi)\widetilde{f}_{i+1,i+1}(\phi)\Big)\Big)$. A corollary to Proposition 2.3 reads now as follows.

Corollary 3.2. *When the FIM is positive definite $\mathcal{F}(\vartheta) \succ 0$, then the matrix B is positive definite for all values of the rotation angle ϕ.*

Proof. As mentioned above, a positive definite FIM has the property, $\mathcal{F}(\vartheta) \succ 0 \iff \mathcal{F}_{\phi}(\vartheta) \succ 0$, a direct consequence of (21). Proposition 3.1 and the property shown in Lemma 3.4, yield $f_{1,1} > 0, \ldots, f_{i-1,i-1} > 0, \widetilde{f}_{i,i}(\phi) > 0, \widetilde{f}_{i+1,i+1}(\phi) > 0, f_{i+2,i+2} > 0, \ldots, f_{m,m} > 0$. It can be deduced from (15) and (16) for $s_{jl} \to f_{jl}$ and $\psi \to \phi$, that the eigenvalues of matrix B are

$$\mu_1 = \frac{1}{f_{1,1}}, \ldots, \ \mu_{i-1} = \frac{1}{f_{i-1,i-1}}, \mu_i = \frac{1}{\tilde{f}_{i,i}(\phi)}, \mu_{i+1} = \frac{1}{\tilde{f}_{i+1,i+1}(\phi)},$$
$$\mu_{i+2} = \frac{1}{f_{i+2,i+2}}, \ldots, \mu_n = \frac{1}{f_{m,m}}.$$

Consequently, $\mu_1 > 0, \ldots, \mu_{i-1} > 0$, $\mu_i > 0$, $\mu_{i+1} > 0$, $\mu_{i+2} > 0, \ldots, \mu_m > 0$, from which the positive definiteness of the symmetric matrix B can be concluded. This completes the proof. □

When $\mathcal{F}(\vartheta) \succ 0$, we get $d^2_{\mathcal{F}_\phi}(\vartheta) = Q_B(\vartheta) > 0$, such that the quadratic form distance measure (31) and consequently the statistical distance measure (24) are metrics.

Similar equalities to (17) and (19) can be established through the substitutions $s_{jl} \to f_{jl}$ and $\psi \to \phi$, as in

$$Q_B(\vartheta) = d^2_{\mathcal{F}_\phi}(\vartheta) = \|M\vartheta\|^2, \tag{32}$$

which involves the Euclidean norm of $M\vartheta$, and M is the appropriate $m \times m$ Cholesky factorization matrix associated with matrix B, such that $B = M^\top M$. Matrix M has the same configuration as matrix L in (17) but where $s_{jl} \to f_{jl}$ and $\psi \to \phi$. When the entries of the FIM are the weighting coefficients for the statistical distance measure, the QR factorization of matrix B is set forth in addition to the Cholesky factorization, it is of the form

$$B = QR, \tag{33}$$

where

$$Q = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & & & \vdots \\ \vdots & \ddots & 1 & \ddots & & & \vdots \\ \vdots & & \ddots & Q_{i,i+1} & \ddots & & \vdots \\ \vdots & & & \ddots & 1 & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix}, \quad R = \begin{pmatrix} 1/f_{1,1} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & & & \vdots \\ \vdots & \ddots & 1/f_{i-1,i-1} & \ddots & & & \vdots \\ \vdots & & \ddots & R_{i,i+1} & \ddots & & \vdots \\ \vdots & & & \ddots & 1/f_{i+2,i+2} & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1/f_{m,m} \end{pmatrix}$$

and

$$Q_{i,i+1} = \frac{1}{\gamma(\phi)} \begin{pmatrix} \delta(\phi) & \rho(\phi) \\ -\rho(\phi) & \delta(\phi) \end{pmatrix},$$

$$R_{i,i+1} = \begin{pmatrix} \gamma(\phi)/\left(\tilde{f}_{i,i}(\phi)\tilde{f}_{i+1,i+1}(\phi)\right) & \chi(\phi)/\left(\tilde{f}_{i,i}(\phi)\tilde{f}_{i+1,i+1}(\phi)\gamma(\phi)\right) \\ 0 & 1/\gamma(\phi) \end{pmatrix},$$

whereby the Gram–Schmidt orthonormalization process ensures that the matrix R is an upper triangular matrix and Q is a $m \times m$ orthogonal matrix, where

$$\gamma(\phi) = \sqrt{\tilde{f}^2_{i,i}(\phi)\sin^2(\phi) + \tilde{f}^2_{i+1,i+1}(\phi)\cos^2(\phi)}, \quad \delta(\phi) = \tilde{f}_{i,i}(\phi)\sin^2(\phi) + \tilde{f}_{i+1,i+1}(\phi)\cos^2(\phi). \tag{34}$$

$$\chi(\phi) = \left(\tilde{f}^2_{i,i}(\phi) - \tilde{f}^2_{i+1,i+1}(\phi)\right)\sin(\phi)\cos(\phi), \quad \rho(\phi) = \left(\tilde{f}_{i+1,i+1}(\phi) - \tilde{f}_{i,i}(\phi)\right)\sin(\phi)\cos(\phi). \tag{35}$$

Next to the quadratic form distance measure representations (31) or (32), the form

$$Q_B(\vartheta) = d^2_{\mathcal{F}_\phi}(\vartheta) = \left\langle R\vartheta, Q^\top\vartheta \right\rangle \tag{36}$$

is introduced. Interconnections between the Euclidean norm $\|\vartheta\|$ and the matrices M, R and Q are set forth by inequalities displayed in the next proposition.

Proposition 3.3. *When the FIM is positive definite, $\mathcal{F}(\vartheta) \succ 0$, the inequalities*

$$\|\vartheta\|^2 \geqslant \frac{1}{(\|M\|^2 + \|R\|)} \left\{\|M\vartheta\|^2 + \left\langle R\vartheta, Q^\top\vartheta\right\rangle\right\} \tag{37}$$

$$\|\vartheta\|^2 \geqslant \frac{1}{2} \left\{\frac{\|M\vartheta\|^2}{\|M\|^2} + \frac{\left\langle R\vartheta, Q^\top\vartheta\right\rangle}{\|R\|}\right\} \tag{38}$$

hold true.

Proof. By virtue of Corollary 3.2, when $\mathcal{F}(\vartheta) \succ 0 \Longrightarrow B \succ 0$, consequently there exists an appropriate matrix M that satisfies the Cholesky factorization of matrix B iff distance measure (24) fulfills the metric requirements. Consider then the inequality

$$\|M\vartheta\|^2 \leqslant \|M\|^2\|\vartheta\|^2. \tag{39}$$

Considering the equality (36) and the positive definiteness of the FIM $\mathcal{F}(\vartheta)$ implies that the scalar product $\left\langle R\vartheta, Q^\top\vartheta\right\rangle$ is positive. Taking the property into consideration that the columns of the matrix Q are orthonormal, results in $\|Q\| \leqslant 1$ and $\|Q^\top\| \leqslant 1$. Applying the Cauchy–Schwarz inequality to $\left\langle R\vartheta, Q^\top\vartheta\right\rangle$, then yields the inequality

$$\left\langle R\vartheta, Q^\top\vartheta\right\rangle \leqslant \|R\|\|\vartheta\|^2. \tag{40}$$

Combining (39) and (40) yields (37) and (38). □

We measure the magnitude of a matrix A with the operator norm, $\|A\| = \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|}$. Note that the inequalities (39) and (40) result in a connection between the square of the Euclidean norm $\|\vartheta\|^2$ and the transformed statistical distance measure (24).

Equivalently to (18), the string of equalities formulated in (32) yields an ellipsoid in a m-dimensional coordinate system, given by

$$1 = \frac{1}{\|M\vartheta\|^2} \left\{\sum_{j=1, j\neq i, i+1}^{m} \left(\frac{k_j^2}{f_{j,j}}\right) + \frac{k_i^2}{\widetilde{f}_{i,i}(\phi)} + \frac{k_{i+1}^2}{\widetilde{f}_{i+,i+1}(\phi)}\right\}. \tag{41}$$

This is justified by the fact that the entries $f_{j,j}$, $\widetilde{f}_{i,i}(\phi)$ and $\widetilde{f}_{i+1,i+1}(\phi)$ of the Fisher information matrices $\mathcal{F}(\vartheta)$ and $\mathcal{F}_\phi(\vartheta)$, are positive. The dimension of the eigenspace $\mathcal{M}$ =span$\{w_1, w_2, \ldots, w_i, w_{i+1}, \ldots, w_m\}$, where $w_1, w_2, \ldots, w_i, w_{i+1}, \ldots, w_m$ are the normalized eigenvectors of matrix B, is smaller than the dimension of the eigenspace $\mathcal{B}$ used in (18), so $\mathcal{M} \subset \mathcal{B}$. But to some extent, the eigenspace $\mathcal{M}$ is spanned by the same basis orthonormal eigenvectors as eigenspace $\mathcal{B}$. The principal axes of ellipsoid (41) are labeled $k_1, \ldots, k_{i-1}, k_i, k_{i+1}, k_{i+2}, \ldots, k_m$ and are the coordinates of ϑ with respect to the eigenspace $\mathcal{M}$. Consequently, the lengths of these principal axes are given by

$$\sqrt{h f_{1,1}}, \ldots, \sqrt{h f_{i-1,i-1}}, \sqrt{h \widetilde{f}_{i,i}(\phi)}, \sqrt{h \widetilde{f}_{i+,i+1}(\phi)}, \sqrt{h f_{i+2,i+2}}, \ldots, \sqrt{h f_{m,m}}, \tag{42}$$

for $h = \|M\vartheta\|^2$. When $\mathcal{F}(\vartheta) \succeq 0$ combined with $\mathrm{Det}(\mathcal{F}(\vartheta)) = 0$, at least one eigenvalue of matrix B can be zero for some value of ϕ, such that the length of at least one of the main axes of the m-dimensional ellipsoid in (42), $f_{1,1}, \ldots, f_{i-1,i-1}, \widetilde{f}_{i,i}(\phi), \widetilde{f}_{i+,i+1}(\phi), f_{i+2,i+2}, \ldots, f_{m,m}$, will be arbitrarily large. In this case the statistical distance measure (24) does not fulfill the metric requirements. As shall be seen in Section 4 where stationary processes are considered, the case $\mathcal{F}(\vartheta) \succeq 0$ combined with $\mathrm{Det}(\mathcal{F}(\vartheta)) = 0$ does not necessarily lead to a statistical distance measure (24) that is not a metric. This is determined by the rotation angle ϕ. However, the condition $\mathcal{F}(\vartheta) \succ 0$, is of fundamental importance for ensuring the metric properties of statistical distance measure (24) for all rotation angles ϕ.

3.2. The Fisher information matrix of stationary processes

In [15], an interconnection between the FIM of a stationary ARMA process, a symmetric block Toeplitz matrix, and the Sylvester resultant matrix, is set forth. For a detailed exposition of ARMA time series processes, see [5] and for Sylvester resultant matrices, see [21]. The ARMA(p, q) time series process is defined as follows:

$$a(\mathcal{Q})\, y(t) = b(\mathcal{Q})\, \varepsilon(t), \quad a(\mathcal{Q}) = \sum_{j=0}^{p} a_j\, \mathcal{Q}^j; \quad b(\mathcal{Q}) = \sum_{l=0}^{q} b_l\, \mathcal{Q}^l, \tag{43}$$

where $\mathcal{Q}$ denotes the backward shift operator, e.g. $\mathcal{Q}x(t) = x(t-1)$. The polynomials $a(\mathcal{Q})$ and $b(\mathcal{Q})$ are the autoregressive and moving average polynomials of degree p and q respectively with corresponding coefficient parameters $a_1, \ldots, a_p$ and $b_1, \ldots, b_q$ and $a_0 = b_0 = 1$. The stability condition of the ARMA(p, q) process is imposed, this implies that the absolute values of the roots of the polynomials $a(\mathcal{Q})$ and $b(\mathcal{Q})$ are outside and not on the unit circle. The process $y(t)$ is driven by a white noise process $\varepsilon(t)$. The error $\{\varepsilon(t), t \in \mathbb{N}\}$ is a collection of uncorrelated zero mean random variables with constant variance. The shift operator $\mathcal{Q}$ is for further analysis substituted by the z transform. Note that $p+q = m$ and the associated parameter vector is $\vartheta = \left(a_1, \ldots, a_p, b_1, \ldots, b_q\right)^\top$. In [18], representations of the submatrices of the FIM $\mathcal{F}(\vartheta)$ of a multivariate ARMA(p, q) process are developed, the corresponding autoregressive and moving average polynomials are then matrix polynomials. The scalar version of these representations will be applied to an ARMA(p, q) process, this yields explicit expressions for the entries $f_{i,i}$ and $f_{i,i+1}$ of the FIM that appear in (25) and (26), to obtain

$$f_{i,i} = \begin{cases} \frac{1}{2\pi\mathbf{i}} \oint\limits_{|z|=1} \frac{1}{a(z)a(\frac{1}{z})} \frac{dz}{z} & i \in \{1, 2, \ldots, p\} \\ \frac{1}{2\pi\mathbf{i}} \oint\limits_{|z|=1} \frac{1}{b(z)b(\frac{1}{z})} \frac{dz}{z} & i \in \{p+1, \ldots, p+q\} \end{cases} \tag{44}$$

and

$$f_{i,i+1} = \begin{cases} \frac{1}{2\pi\mathbf{i}} \oint\limits_{|z|=1} \frac{1}{za(z)a(\frac{1}{z})} \frac{dz}{z} & i \in \{1, \ldots, p-1\} \\ -\frac{1}{2\pi\mathbf{i}} \oint\limits_{|z|=1} \frac{1}{za(z)b(\frac{1}{z})} \frac{dz}{z} & i = p \\ \frac{1}{2\pi\mathbf{i}} \oint\limits_{|z|=1} \frac{1}{zb(z)b(\frac{1}{z})} \frac{dz}{z} & i \in \{p+1, \ldots, p+q-1\}, \end{cases} \tag{45}$$

where $\mathbf{i}$ is the standard imaginary unit with the property $\mathbf{i}^2 = -1$. The integrals are counterclockwise and the contour is the unit circle $|z| = 1$ that encloses all roots of $z^p a(\frac{1}{z})$ and $z^q b(\frac{1}{z})$. The poles are the solutions of $z^p a(\frac{1}{z}) = 0$ and $z^q b(\frac{1}{z}) = 0$ and are used for the standard residue theorem. Remark that (44) can also be used for computing all remaining values $f_{j,j}$, where $j \in \{1, \ldots, m\}$ and with $j \neq i, i+1$, that appear in statistical distance measure (24).

We shall use (44) to prove that even for singular Fisher information matrices of stationary ARMA processes, the entries on the main diagonal are positive.

Lemma 3.4. *For singular as well as nonsingular Fisher information matrices of stationary ARMA processes, the components given in (44) satisfy the property $f_{i,i} > 0$ for $i = 1, \ldots, p+q$.*

Proof. We consider $i \in \{1, 2, \ldots, p\}$ and introduce the $p \times p$ matrix $F(\vartheta) =$Diag$\{f_{1,1}, f_{2,2}, \ldots, f_{p,p}\}$ and suppose there is a fixed vector x such that $F(\vartheta)\, x = 0$, so that

$$0 = \frac{1}{2\pi\mathbf{i}} \oint\limits_{|z|=1} \mathcal{K}(z)\, x \frac{dz}{z},$$

and $\mathcal{K}(z)$ is a $p \times p$ diagonal matrix consisting of the integrands of (44). Take $z = e^{\mathbf{i}\omega}$, we then get

$$0 = \frac{1}{2\pi}\int_0^{2\pi} x^* \mathcal{K}(e^{\mathbf{i}\omega})\, x\, d\omega,$$

and with

$$f_{i,i} = \frac{1}{2\pi}\int_0^{2\pi} \frac{d\omega}{a(e^{\mathbf{i}\omega})a(e^{-\mathbf{i}\omega})} = \frac{1}{2\pi}\int_0^{2\pi} \frac{d\omega}{|a(e^{\mathbf{i}\omega})|^2},$$

where x^* is the complex conjugate of x. Then we must have $x^*\mathcal{K}(e^{\mathbf{i}\omega})x \equiv 0$ and note that $\mathcal{K}(e^{\mathbf{i}\omega}) > 0$, it then follows that $x_i^* \frac{1}{a(e^{\mathbf{i}\omega})} = 0$ for almost all ω. This is only possible if $x_i^* = 0$ and consequently $x_i = 0$ for all $i \in \{1, 2, \ldots, p\}$, so $f_{i,i} > 0$. It is clear that $f_{i,i} > 0$ also holds for $i \in \{p+1, \ldots, p+q\}$. $\square$

From Lemma 3.4 it can be concluded that statistical distance measure (22) is always a metric, even for singular Fisher information matrices, when applied to stationary ARMA processes. Whereas statistical distance measure (24), under these conditions, does not always fulfill the metric condition, as shall be seen in Section 4. However, equality (20) can in no way be considered when a singular FIM is the case since inversion of the FIM is then necessary. This property does not only hold for stationary processes.

The following decomposition of the asymptotic FIM of a stationary ARMA process, is displayed accordingly, see [15]

$$\mathcal{F}(\vartheta) = \mathcal{S}(-b, a)\mathcal{P}(\vartheta)\mathcal{S}^\top(-b, a), \tag{46}$$

where the $(p+q) \times (p+q)$ Sylvester resultant matrix of the polynomials $a(z)$ and $b(z)$ is defined as

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

It is shown that the matrix $\mathcal{P}(\vartheta)$ is positive definite and is given by

$$\mathcal{P}(\vartheta) = \frac{1}{2\pi\mathbf{i}} \oint_{|z|=1} \frac{u_{p+q}(z)u_{p+q}^\top(\frac{1}{z})}{a(z)a(\frac{1}{z})b(z)b(\frac{1}{z})}\frac{dz}{z}, \tag{47}$$

where $u_k(z) = (z^{k-1}, z, \ldots, 1)^\top$, $a(z)$ and $b(z)$ are the ARMA(p, q) process monic polynomials set forth in (43). A fundamental property of the Sylvester matrix $\mathcal{S}(a, b)$ is that it becomes singular iff the polynomials $a(z)$ and $b(z)$ have at least one common root. This is called the resultant property. In [15], matrix equality (46) is used to prove the resultant property of the FIM of an ARMA process. Remark that in [14], the resultant property is proved for the FIM of an ARMAX process. This is also proved through Sylvester resultant matrices. An ARMAX process is an extension of the ARMA process where an input-control-exogenous variable is added to the ARMA part.

An alternative representation of the transformed FIM is now obtained by a combination of (21) and (46),

$$\mathcal{F}_\phi(\vartheta) = \mathcal{S}_\phi(-b, a)\mathcal{P}(\vartheta)\mathcal{S}_\phi^\top(-b, a), \tag{48}$$

where the transformed Sylvester resultant matrix is given by

$$\mathcal{S}_\phi(-b, a) = \mathcal{L}_i(\phi)\mathcal{S}(-b, a). \quad (49)$$

In the next proposition we shall formulate the condition under which the transformed FIM $\mathcal{F}_\phi(\vartheta)$ fulfills the resultant property condition.

Proposition 3.5. *The transformed FIM $\mathcal{F}_\phi(\vartheta)$ in (48) is a resultant matrix for all values of the rotation angle ϕ when the FIM is positive definite, $\mathcal{F}(\vartheta) \succ 0$.*

Proof. It is straightforward to conclude that for all values of the rotation angle ϕ, the rotation matrix $\mathcal{L}_i(\phi)$ is invertible since it is an orthogonal matrix. Consequently, from (49) can be seen that the transformed Sylvester matrix $\mathcal{S}_\phi(-b, a)$ has the same invertibility condition as matrix $\mathcal{S}(-b, a)$ and is therefore a resultant matrix. It can be concluded, the transformed FIM $\mathcal{F}_\phi(\vartheta)$, given the property $\mathcal{P}(\vartheta) > 0$ and matrix equality (48), has the resultant matrix property. □

It is clear that when a stationary ARMA process is considered, a corollary to Proposition 3.5 can be proved equivalently when the nonsingularity condition of the Sylvester resultant matrix $\text{Det}(\mathcal{S}(-b, a)) \neq 0$ instead of $\mathcal{F}(\vartheta) \succ 0$, is imposed. Equation (46) can be interpreted as $\text{Det}(\mathcal{S}(-b, a)) \neq 0 \Longleftrightarrow \mathcal{F}(\vartheta) \succ 0$ since the Sylvester resultant matrix $\mathcal{S}(-b, a)$ does not have to fulfill the condition $\mathcal{S}(-b, a) \succ 0$ and can be $\mathcal{S}(-b, a) \prec 0$, negative definite instead.

The distance measures (22) and (23) are partially expressed by the original and transformed Sylvester resultant matrices $\mathcal{S}(-b, a)$ and $\mathcal{S}_\phi(-b, a)$ through the corresponding values $f_{1,1}, \ldots, f_{i-1,i-1}, \widetilde{f}_{i,i}(\phi), \widetilde{f}_{i+,i+1}(\phi), f_{i+2,i+2}, \ldots, f_{m,m}$. These values are computed by means of the equalities (46) and (48) respectively.

It can be concluded, when a stationary ARMA process is considered, an equivalent to Proposition 3.1 and Corollary 3.2 can be proved in a similar manner when the nonsingularity condition of the Sylvester resultant matrix $\text{Det}(\mathcal{S}(-b, a)) \neq 0$ instead of $\mathcal{F}(\vartheta) \succ 0$, is imposed.

In [15], it is proved that the FIM is nonsingular when the corresponding ARMA polynomials have no common zeros. Under these conditions it can be concluded that the FIM of an ARMA process is strict positive, $\mathcal{F}(\vartheta) \succ 0$ or $\text{Det}(\mathcal{F}(\vartheta)) > 0$. Consequently, the properties shown in Propositions 3.1 and 3.3 automatically hold for ARMA processes when no common roots are detected between the ARMA polynomials. This yields, by virtue of Corollary 2.2 proved in [15, p. 276], that the nonsingularity condition of the Sylvester resultant matrix, $\text{Det}(\mathcal{S}(-b, a)) \neq 0$ instead of $\mathcal{F}(\vartheta) \succ 0$, can alternatively be imposed as a sufficient condition for Propositions 3.1 and 3.3 and Corollary 3.2 to hold. The results shown in Propositions 3.5, 3.1 and 3.3 and Corollary 3.2 can also be proved for ARMAX processes since the resultant property of the corresponding FIM has been confirmed in [14]. In [17], an interconnection between the Bezout matrix, which has the resultant property, and the FIM of an ARMA process is set forth. Introducing an appropriate Bezout matrix in equalities which are equivalent to (46) and (48) can be applied in the same way as with the Sylvester resultant matrices. This would lead to an alternative version of the statistical distance measure (24), whereby the appropriate weighting coefficients are a combination of entries of the Bezout matrix.

In the next section, cases based on ARMA(1, 1) and ARMA(2, 2) processes are set forth to show that when there is a common zero between the ARMA polynomials, which implies a singular FIM, the distance measure (24) can fulfill the metric properties depending on the choice of the rotation angle ϕ.

4. Singular Fisher information matrices of ARMA(2, 2) and ARMA(1, 1) processes

In this section we study the metric conditions of statistical distance measure (24) when the FIM $\mathcal{F}(\vartheta)$ is singular. This will be analyzed by means of ARMA(2, 2) and ARMA(1, 1) processes.

Consider the FIM of an ARMA(2, 2) process, the autoregressive and moving average polynomials are of degree two, and described by, $y(t)a(z) = b(z)\varepsilon(t)$, where $y(t)$ is the stationary process driven by white noise $\varepsilon(t)$, $a(z) = (1 + a_1z + a_2z^2)$ and $b(z) = (1 + b_1z + b_2z^2)$. The condition, zeros of the polynomials

$$p(z) = z^2 a(z^{-1}) = z^2 + a_1 z + a_2 \quad \text{and} \quad q(z) = z^2 b(z^{-1}) = z^2 + b_1 z + b_2 \tag{50}$$

are in absolute value smaller than one, is considered. The Sylvester resultant matrix $\mathcal{S}(-b, a)$ corresponding to the polynomials $a(z)$ and $b(z)$ as well as the FIM of the ARMA(2, 2) process are set forth. Computation of the FIM $\mathcal{F}(\vartheta)$ at a matrix level implies the evaluation of (46), this can be done through a combination of the appropriate matrix $\mathcal{P}(\vartheta)$ displayed in (47), with $p = q = 2$, and the Sylvester matrix of the form (52). However, the scalar version of the representations in [18] will be applied to the ARMA(2, 2) process where $\vartheta = (a_1, a_2, b_1, b_2)^\top$, to obtain

$$\mathcal{F}(\vartheta) = \begin{pmatrix} \mathcal{F}_{aa}(\vartheta) & \mathcal{F}_{ab}(\vartheta) \\ \mathcal{F}_{ab}^\top(\vartheta) & \mathcal{F}_{bb}(\vartheta) \end{pmatrix} = \begin{pmatrix} f_{11} & f_{12} & f_{13} & f_{14} \\ f_{12} & f_{22} & f_{23} & f_{24} \\ f_{13} & f_{23} & f_{33} & f_{34} \\ f_{14} & f_{24} & f_{34} & f_{44} \end{pmatrix}, \tag{51}$$

and

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

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

$$\mathcal{F}_{ab}(\vartheta) = \frac{1}{(a_2 b_2 - 1)^2 + (a_2 b_1 - a_1)(b_1 - a_1 b_2)} \begin{pmatrix} a_2 b_2 - 1 & a_1 - a_2 b_1 \\ b_1 - a_1 b_2 & a_2 b_2 - 1 \end{pmatrix}.$$

The submatrices $\mathcal{F}_{aa}(\vartheta)$ and $\mathcal{F}_{bb}(\vartheta)$ are symmetric and Toeplitz whereas $\mathcal{F}_{ab}(\vartheta)$ is Toeplitz. This property 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 Sylvester resultant matrix associated with the stationary ARMA(2, 2) process is

$$\mathcal{S}(-b, a) = \begin{pmatrix} -1 & -b_1 & -b_2 & 0 \\ 0 & -1 & -b_1 & -b_2 \\ 1 & a_1 & a_2 & 0 \\ 0 & 1 & a_1 & a_2 \end{pmatrix}. \tag{52}$$

The zeros of the autoregressive and moving average polynomials $p(z)$ and $q(z)$ are $z_{1,2} = \frac{1}{2}\left(-a_1 \pm \sqrt{a_1^2 - 4a_2}\right)$ and $z_{3,4} = \frac{1}{2}\left(-b_1 \pm \sqrt{b_1^2 - 4b_2}\right)$ respectively, the conditions $\mid z_{1,2} \mid < 1$ and $\mid z_{3,4} \mid < 1$ are imposed and we assume $z_1 = z_3$. Apply the substitution $z_1 \to z_3$ to polynomial $p(z)$, the appropriate coefficients are then

$$a_1 \to \frac{1}{2}\left(a_1 + b_1 + \sqrt{a_1^2 - 4a_2} - \sqrt{b_1^2 - 4b_2}\right) \tag{53}$$

and

$$a_2 \to \frac{1}{4}\left(a_1 b_1 + b_1\sqrt{a_1^2 - 4a_2} - a_1\sqrt{b_1^2 - 4b_2} - \sqrt{a_1^2 - 4a_2}\sqrt{b_1^2 - 4b_2}\right). \tag{54}$$

By inserting representations (53) and (54) in the matrices $\mathcal{S}(-b, a)$ displayed in (52) and the FIM $\mathcal{F}(\vartheta)$ given in (51) and leaving the coefficients of polynomial $q(z)$ unchanged, results in the singularity of these matrices.

The choice of i, in the 4×4 Givens rotation matrix $\mathcal{L}_i(\phi)$ of type (6), that shall first be considered, is 2. The entries of the Fisher information matrix $\mathcal{F}(\vartheta)$, f_{22}, f_{23} and f_{33}, are used for the computation of the appropriate values of $\widetilde{f}_{2,2}(\phi)$ and $\widetilde{f}_{3,3}(\phi)$ according to (25) and (26), to obtain

$$\widetilde{f}_{2,2}(\phi) = f_{22}\cos^2(\phi) - f_{23}\sin(2\phi) + f_{33}\sin^2(\phi) \quad \text{and}$$
$$\widetilde{f}_{3,3}(\phi) = f_{33}\cos^2(\phi) + f_{23}\sin(2\phi) + f_{22}\sin^2(\phi). \tag{55}$$

From (55) can be seen that for $\phi = k\pi$, with $k = 0, 1, 2, \ldots$, we have $\widetilde{f}_{2,2}(\phi) = f_{22}$ and $\widetilde{f}_{3,3}(\phi) = f_{33}$, by virtue of Lemma 3.4 we have $\widetilde{f}_{2,2}(\phi) > 0$ and $\widetilde{f}_{3,3}(\phi) > 0$. Equivalently, when $\phi = \pi/2$ and $\phi = 3\pi/2$, it yields $\widetilde{f}_{2,2}(\phi) = f_{33}$ and $\widetilde{f}_{3,3}(\phi) = f_{22}$. The structure of the matrix $\Omega(\phi)$, introduced in (29) and when (55) is taken into account, results in Det $(\Omega(\phi)) = f_{22}f_{33} - f_{23}^2$, which is the determinant of an appropriate 2×2 submatrix on the main diagonal of the FIM $\mathcal{F}(\vartheta)$ in (51). Despite the property $\mathcal{F}(\vartheta) \succeq 0 \Longrightarrow \mathcal{F}_\phi(\vartheta) \succeq 0$, we have Det $(\Omega(\phi)) \neq 0$, due to the fact $f_{2,2}(\phi) > 0$ and $f_{3,3}(\phi) > 0$, at least for some values of ϕ. This case shows that when $\mathcal{F}(\vartheta) \succeq 0$, the case of a singular FIM, the statistical distance measure $d^2_{\mathcal{F}_\phi}(P, Q)$ given in (24) can fulfill the metric requirements. By virtue of Lemma 3.4, the entries of $\mathcal{F}(\vartheta)$, $f_{11} > 0$ and $f_{44} > 0$, such that all the eigenvalues of the corresponding matrix B, introduced in (31), are positive. The corresponding Cholesky matrix M, and corresponding QR decomposition (33) are under these conditions relevant. Considering that usually $f_{2,2} \neq f_{3,3}$, implies that the resulting inequalities in Proposition 3.3, also hold since the appropriate values of $\gamma(\phi)$, $\delta(\phi)$, $\chi(\phi)$ and $\rho(\phi)$, given in (34) and (35), are then different from zero.

When $i = 1$ and $i = 3$ in the Givens rotation matrix (6), we have the property $f_{1,1} = f_{2,2}$ and $f_{3,3} = f_{4,4}$ respectively, see the FIM (51). The case $i = 1$ combined with rotation angle $\phi = (k\pi/2)$, where $k = 0, 1, 2, \ldots$, results in $\widetilde{f}_{1,1}(\phi) = \widetilde{f}_{2,2}(\phi) = f_{1,1}$. We further have $\gamma(\phi) = \delta(\phi) = f_{1,1}$ and $\chi(\phi) = \rho(\phi) = 0$, so that $Q = I_4$, $R = B = \text{diag}\,\{1/f_{1,1}, 1/f_{1,1}, 1/f_{3,3}, 1/f_{4,4}\}$ and the Cholesky matrix is then $M = \text{diag}\left\{1/\sqrt{f_{1,1}}, 1/\sqrt{f_{1,1}}, 1/\sqrt{f_{3,3}}, 1/\sqrt{f_{4,4}}\right\}$, this confirms the positive definiteness of B. The obtained inequalities in Proposition 3.3, are then relevant. The resulting statistical distance measure (24) or equivalently representations (31) and (32) are of the form

$$d^2_{\mathcal{F}_\phi}(\vartheta) = \frac{\vartheta_1^2}{f_{1,1}} + \frac{\vartheta_2^2}{f_{1,1}} + \frac{\vartheta_3^2}{f_{3,3}} + \frac{\vartheta_4^2}{f_{4,4}}, \tag{56}$$

where ϑ_1 and ϑ_2 are given by representations (53) and (54) respectively, $\vartheta_3 = b_1$ and $\vartheta_4 = b_2$. Equivalently, when $i = 3$ in the Givens rotation matrix (6) and $\phi = (k\pi/2)$, with $k = 0, 1, 2, \ldots$, the appropropriate matrices are then $Q = I_4$ and $R = B = \text{diag}\,\{1/f_{1,1}, 1/f_{2,2}, 1/f_{3,3}, 1/f_{3,3}\}$ respectively and we have $B \succ 0$. An equivalent to (56) is then

$$d^2_{\mathcal{F}_\phi}(\vartheta) = \frac{\vartheta_1^2}{f_{1,1}} + \frac{\vartheta_2^2}{f_{2,2}} + \frac{\vartheta_3^2}{f_{3,3}} + \frac{\vartheta_4^2}{f_{3,3}}.$$

We have considered some rotation angles ϕ that confirm the metric properties of statistical distance measure (24) despite the singularity of the FIM (51).

However, for singular Fisher information matrices, there exists nontrivial solutions ϕ to the equations

$$f_{i,i}\cos^2(\phi) - f_{i,i+1}\sin(2\phi) + f_{i+1,i+1}\sin^2(\phi) = 0 \tag{57}$$

and

$$f_{i+1,i+1}\cos^2(\phi) + f_{i,i+1}\sin(2\phi) + f_{i,i}\sin^2(\phi) = 0, \tag{58}$$

which are of the form

$$\pm\text{Arccos}\left(\pm\sqrt{\frac{f_{j,j}^2 - f_{i,i}f_{i+1,i+1} + 2f_{i,i+1}^2 \pm 2\sqrt{f_{i,i+1}^4 - f_{i,i}f_{i+1,i+1}f_{i,i+1}^2}}{f_{i,i}^2 - 2f_{i,i}f_{i+1,i+1} + f_{i+1,i+1}^2 + 4f_{i,i+1}^2}}\right), \tag{59}$$

where $j = i + 1$ and $j = i$ for (57) and (58) respectively and are evaluated through the Mathematica 6.0 version. The equalities (57) and (58) have different solutions that are variants of representation

(59), this implies different rotation angles ϕ. When $i = 2$ in the 4×4 Givens rotation matrix $\mathcal{L}_i(\phi)$, introduce the representations (53) and (54) in $f_{2,2}$ and $f_{2,3}$ and eventually insert the corresponding values of $f_{2,2}$ and $f_{2,3}$ in ϕ given by (59) results in $\widetilde{f}_{2,2}(\phi) = \widetilde{f}_{3,3}(\phi) = 0$. Consequently, in the case of a singular FIM, there exists a set of values of the rotation angle ϕ that yields a statistical distance measure (24) that does not satisfy the metric conditions. Equivalently, for any stationary ARMA(p, q) process with a singular FIM, there exists rotation angles ϕ that satisfy equations (57) and (58). Note that the equations (57) and (58) hold for all $i = 1, \ldots, m-1$ and for any process or statistical model that differs from a stationary ARMA(p, q) process when a singular FIM is the case, see Proposition 3.1. By virtue of Proposition 3.1, can also be concluded that there are no solutions to the equations (57) and (58) when the Fisher information matrices are nonsingular.

When the FIM is singular and the rotation angle ϕ is given by the representation (59), the appropriate matrix B, its associated Cholesky matrix M, and the corresponding QR decomposition (33) are then irrelevant. Consequently, the resulting inequalities in Proposition 3.3, are also irrelevant since the appropriate values of $\gamma(\phi)$, $\delta(\phi)$, $\chi(\phi)$ and $\rho(\phi)$ are under these conditions zero.

Note that when a_2 is close to the unit circle $|z| = 1$, f_{22} will be arbitrarily large, the corresponding eigenvalue $\left(1/\widetilde{f}_{2,2}(\phi)\right)$ of the matrix B is then close to zero. Under these conditions, the property $B \succ 0$ does not hold such that $d^2_{\mathcal{F}\phi}(P, Q)$ in (31) is not a metric. This could be the case for any f_{ii}.

An ARMA(1, 1) process is now considered, this is obtained by choosing $a_2 = b_2 = 0$ in the autoregressive and moving average polynomials $a(z)$ and $b(z)$ respectively, and displayed in (50). The corresponding FIM and Sylvester matrix are then 2×2, and are of the form

$$\mathcal{F}(\vartheta) = \begin{pmatrix} f_{11} & f_{12} \\ f_{12} & f_{22} \end{pmatrix} = \begin{pmatrix} 1/(1-a^2) & -1/(1-ab) \\ -1/(1-ab) & 1/(1-b^2) \end{pmatrix} \quad \text{and} \quad \mathcal{S}(-b, a) = \begin{pmatrix} -1 & -b \\ 1 & a \end{pmatrix}. \tag{60}$$

Assume that a and b coincide, it is then straightforward to see that $\mathcal{F}(\vartheta)$ and $\mathcal{S}(-b, a)$ in (60) are singular matrices. The presence of a common root results in, $f_{11} = f_{22} = -f_{12} > 0$, such that Det $(\mathcal{F}(\vartheta)) =$ Det $(\Omega(\phi)) = f_{11}^2 - f_{12}^2 = 0$, this implies that matrix $\Omega(\phi)$ is singular for all values of ϕ, contrary to the ARMA(2, 2) case. This is equivalent with a semipositive transformed FIM $\mathcal{F}_\phi(\vartheta)$. The corresponding values of representations (59) are then $\pm\text{Arccos}\left(1/\sqrt{2}\right)$ and $\pm\text{Arccos}\left(-1/\sqrt{2}\right)$ which yield the following rotation angles for $0 \leqslant \phi \leqslant 2\pi$ and when taken counterclockwise, $\phi_1 = (\pi/4)$, $\phi_2 = (3\pi/4)$, $\phi_3 = (5\pi/4)$ and $\phi_4 = (7\pi/4)$. The appropriate entries of the transformed FIM $\mathcal{F}_\phi(\vartheta)$, $\widetilde{f}_{1,1}(\phi)$ and $\widetilde{f}_{2,2}(\phi)$, are zero and the equations (57) and (58) are then

$$f_{11}\,(1 - \sin(2\phi_1)) = f_{11}\,(1 - \sin(2\phi_3)) = 0 \quad \text{and} \quad f_{11}\,(1 + \sin(2\phi_2)) = f_{11}\,(1 + \sin(2\phi_4)) = 0.$$

In this case one of the eigenvalues of the appropriate matrix B is arbitrarily large such that the length of one principal ax of the corresponding ellips is zero. Under these conditions, Proposition 3.1 does not hold and consequently, the statistical distance measure (24) does not satisfy the metric requirements. It can be concluded for an ARMA(1, 1) process which has a singular FIM, that there exist values of ϕ for which Proposition 3.1 does not hold, and where conditions (27) and (28) are not feasible, consequently, the statistical distance measure (24) does not satisfy the metric requirements. However, for the values of ϕ equal to π and $(\pi/2)$ we have $\widetilde{f}_{1,1}(\phi) = \widetilde{f}_{2,2}(\phi) = f_{11}$ such that (27) and (28) hold, since $f_{11} = f_{22} > 0$. The statistical distance measure (24) is then a metric.

The cases considered in this section show that when the matrix $\mathcal{F}(\vartheta)$ is singular, the choice of the rotation angle ϕ determines whether or not the statistical distance measure (24) fulfills the metric requirements. It is therefore not possible to draw a coherent conclusion concerning the metric properties of the statistical distance measure (24) when singular Fisher information matrices of stationary ARMA processes are considered, contrary to statistical distance measure (22), this by virtue of Lemma 3.4.

It is straightforward to conclude that a singular FIM $\mathcal{F}(\vartheta)$ and not only for stationary ARMA(p, q) processes, and when a numerical inversion is still feasible, see e.g. [16] for vector ARMA processes, will result in a trivial covariance matrix $\mathrm{Cov}(\vartheta)$. This implies that matrix equation (20) is under this condition irrelevant.

5. Conclusions

In this paper it is proved that when the Fisher information matrix $\mathcal{F}(\vartheta)$ is positive definite, the statistical distance measure (24) fulfills the metric requirements. The metric properties of statistical distance measure (22) are always ensured, both for singular and nonsingular Fisher information matrices of stationary ARMA processes, whereas the metric properties of statistical distance measure (24) hold only for some values of the rotation angle ϕ when a singular FIM is the case.

Based on the results obtained for ARMA$(2, 2)$ and ARMA$(1, 1)$ processes which have singular Fisher information matrices, one can assert that without any loss of generality, similar conclusions can be drawn for any ARMA(p, q) process when the ARMA polynomials have at least one common zero.

The statistical distance measures derived in this paper, are matrix related through the entries of the FIM. These distance measures can be a challenge to its quantum information counterpart (30). This because (22) and (24) involve information about m parameters estimated from n measurements associated with a certain process. Whereas in quantum information, the information about one parameter in a particular measurement procedure is considered for establishing an interconnection with the appropriate statistical distance measure.

Acknowledgements

The first author would like to thank Pieter Kok for insightful comments on the subject of statistical distance and Fisher information in quantum information theory. The authors thank the perceptive reviewer for his comments which significantly improved the presentation.

References

[1] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, third ed., John Wiley, New York, 2003.
[2] O.E. Barndorff-Nielsen, R.D. Gill, Fisher information in quantum statistics, J. Phys. A 30 (2000) 4481–4490.
[3] S.L. Braunstein, C.M. Caves, Statistical Distance and the Geometry of Quantum States, Phys. Rev. Lett. 72 (1994) 3439–3443.
[4] O. Bretscher, Linear Algebra with Applications, first ed., Prentice-Hall, 1997.
[5] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, second ed., Springer-Verlag, Berlin, New York, 1991.
[6] B.R. Frieden, Physics from Fisher Information: A Unification, Cambridge University Press, New York, 1998.
[7] B.R. Frieden, Science from Fisher Information: A Unification, Cambridge University Press, New York, 2004.
[8] G.H. Golub, C.F. Van Loan, Matrix Computations, third ed., John Hopkins University Press, 1996.
[9] A.S. Holevo, Probabilistic and Statistical Aspects of Quantum Theory, second ed., Edizioni Della Normale, SNS Pisa, 2011.
[10] R.A. Horn, C.R. Johnson, Matrix Analysis, North-Holland, Cambridge University Press, 1985.
[11] I.A. Ibragimov, R.Z. Has'minskiĭ, Statistical Estimation, Asymptotic Theory, Springer-Verlag, New York, 1981.
[12] R.A. Johnson, D.W. Wichern, Applied Multivariate Statistical Analysis, second ed., Prentice-Hall, 1988.
[13] P.J. Jones, P. Kok, Geometric derivation of the quantum speed limit, Phys. Rev. A 82 (2010) 022107
[14] A. Klein, P. Spreij, On Fisher's information matrix of an ARMAX process and Sylvester's resultant matrices, Linear Algebra Appl. 237/238 (1996) 579–590.
[15] A. Klein, P. Spreij, On Fisher's information matrix of an ARMA process, in: I. Csiszar, Gy. Michaletzky (Eds.), Stochastic Differential and Difference Equations, Progress in Systems and Control Theory, vol. 23, Birkhäuser, Boston, 1997, pp. 273–284.
[16] A. Klein, G. Mélard, P. Spreij, On the resultant property of the Fisher information matrix of a vector ARMA process, Linear Algebra Appl. 403 (2005) 291–313.
[17] A. Klein, P. Spreij, The Bezoutian, state space realizations and Fisher's information matrix of an ARMA process, Linear Algebra Appl. 416 (2006) 160–174.
[18] A. Klein, P. Spreij, An explicit expression for the Fisher information matrix of a multiple time series process, Linear Algebra Appl. 417 (2006) 140–149.
[19] S. Kullback, Information Theory and Statistics, John Wiley & Sons, New York, 1959.
[20] P. Kok, Tutorial: statistical distance and Fisher information, 8 August 2006.
[21] P. Lancaster, M. Tismenetsky, The Theory of Matrices with Applications, second ed., Academic Press, Orlando, 1985.
[22] D. Petz, Quantum Information Theory and Quantum Statistics, Springer-Verlag, Berlin, Heidelberg, 2008.
[23] C.R. Rao, Information and the accuracy attainable in the estimation of statistical parameters, Bull. Calcutta Math. Soc. 37 (1945) 81–91.
[24] Shunlong Luo, Wigner–Yanase skew information vs. quantum Fisher information, Proc. Amer. Math. Soc. 132 (3) (2003) 885–890.
[25] S. Wesolkowski, P. Fieguth, A probabilistic shading invariant color distance measure, in: 15th European Signal Processing Conference, Poznan, Poland, September 2007, pp. 1907—1911.