Applications of Mathematics

Malwina Janiszewska; Augustyn Markiewicz; Monika Mokrzycka

Block matrix approximation via entropy loss function

Applications of Mathematics, Vol. 65 (2020), No. 6, 829–844

Persistent URL: http://dml.cz/dmlcz/148399

65 (2020) APPLICATIONS OF MATHEMATICS No. 6, 829–844

BLOCK MATRIX APPROXIMATION VIA ENTROPY LOSS FUNCTION

Malwina Janiszewska, Augustyn Markiewicz,
Monika Mokrzycka, Poznań

Received January 29, 2020. Published online November 2, 2020.

Abstract. The aim of the paper is to present a procedure for the approximation of a symmetric positive definite matrix by symmetric block partitioned matrices with structured off-diagonal blocks. The entropy loss function is chosen as approximation criterion. This procedure is applied in a simulation study of the statistical problem of covariance structure identification.

Keywords: approximation; block covariance structure; entropy loss function

MSC 2020: 15A30, 15B99, 62H20, 65F99

1. Introduction

The problem of the best approximation of a given symmetric positive definite matrix by symmetric matrices with specified block partitioned structures is studied. These particular matrices are partitioned into four sub-blocks with square diagonal blocks, not necessarily of the same order. We assume that the diagonal blocks are arbitrary symmetric square matrices, while the off-diagonal blocks can get one of the following structures: all of the off-diagonal block elements are equal (this block can be also considered as a part of the compound symmetry structure) or the off-diagonal block is proportional to respective sub-block of the first-order autoregression matrix.

In the literature, the need for the estimation of covariance matrices is emphasized by many authors. It is worth noting that the number of unknown parameters to be estimated in an unstructured covariance matrix is usually much greater than in the structured one. Moreover, in case when the number of observations is not large enough, the sample covariance matrix is singular or ill-conditioned. One way to overcome this problem is to impose structural restrictions to the covariance matrix.

DOI: 10.21136/AM.2020.0023-20

There are several methods of the choice of the most relevant structure; for example graphical, like neural networks or mapping (cf. Gilson et al. (2019) [8]) or algebraic techniques (cf. Lin et al. (2014) [12], Cui et al. (2016) [1]). In mathematical statistics, structured covariance matrices are widely used in multivariate data analysis and applied in various areas of science, such as e.g. medicine, agronomy (Mieldzioc et al. (2019) [16]), economy, biology, geography, meteorology etc.

In classic multivariate models, usually covariance matrices with a specific structure, such as e.g. compound symmetry (Lin et al. (2014) [12]), autoregression of order one (Cui et al. (2016) [1]) or banded Toeplitz matrices (Filipiak et al. (2018b) [7], John, Mieldzioc (2019) [10]) are studied, without partitioning into blocks. Partitioned matrices are commonly used in doubly multivariate models. One of the most popular applications is that of separable structures of two unstructured matrices or two matrices with one component additionally structured, studied e.g. by Lu and Zimmerman (2005) [13], Srivastava et al. (2008) [18], Filipiak, Klein (2018) [4], Filipiak et al. (2018a) [6], Filipiak et al. (2020) [5]. In the literature, covariance structures partitioned into four submatrices (Szczepańska-Álvarez et al. (2017) [21]) or block compound symmetry covariance structures (Szatrowski et al. (1976) [19], Szatrowski et al. (1982) [20]) are also analyzed.

It is worth noting that block covariance structures can be also applied in modeling the covariance between two random vectors. However, usually the structure of covariance matrix is assumed a priori, without algebraic identification of the most relevant structure. Algebraic methods of covariance structure identification are often based on the approximation of a sample covariance matrix with respect to some criteria. The most commonly used criteria (discrepancy functions) are Frobenius norm (Cui et al. (2016) [1]) and the entropy loss function (Lin et al. (2014) [12]). In our studies the entropy loss function is used as the most relevant for statistical purposes (cf. Filipiak et al. (2020) [5]). The idea is to determine the minimum of the entropy loss function and to choose the structure for which this smallest discrepancy is attained. In this paper we perform simulation studies to verify whether the entropy loss function recognizes given covariance structure properly.

The paper is organized as follows. In Section 2 we present considered structures and the entropy loss function. In Section 3 the procedure of approximation via entropy loss function is given and a particular form of the entropy loss function for considered structures is derived. The description and results of simulation studies are presented in Section 4. Conclusions are given in Section 5.

2. Preliminaries

Let $\mathbf{\Omega} \in \mathbb{R}_m^>$, where $\mathbb{R}_m^>$ is a set of symmetric positive definite matrices of order m. This matrix can have $\frac{1}{2}m(m+1)$ different elements. To reduce the number of various elements in $\mathbf{\Omega}$, some structured matrices are used. Our aim is to determine the best approximation of $\mathbf{\Omega}$ by the block partitioned matrix

$$\mathbf{\Gamma} = \begin{pmatrix} \mathbf{\Gamma}_1 & \mathbf{\Psi} \\ \mathbf{\Psi}' & \mathbf{\Gamma}_2 \end{pmatrix}$$

with unstructured diagonal blocks of order p and q respectively, and structured off-diagonal blocks such that $\mathbf{\Gamma} \in \mathbb{R}_m^>$. We consider the following forms of matrix $\mathbf{\Psi}$

▷ the matrix with all elements equal to δ

$$\mathbf{\Psi} = \delta \mathbf{1}_p \mathbf{1}_q', \tag{2.1}$$

▷ the matrix of the form

$$\mathbf{\Psi} = \delta \mathbf{A} = \delta \begin{pmatrix} \varrho^p & \varrho^{p+1} & \dots & \varrho^{p+q-1} \\ \varrho^{p-1} & \dots & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots \\ \varrho^2 & \dots & \dots & \dots \\ \varrho & \varrho^2 & \dots & \varrho^q \end{pmatrix}. \tag{2.2}$$

Observe that such forms of off-diagonal blocks may follow from compound symmetry or first order autoregression structures. For example, for $p = 2$ and $q = 3$, structure (2.1) corresponds to the off-diagonal block of the compound symmetry structure, i.e.

$$\left(\begin{array}{cc|ccc} \alpha & \delta & \delta & \delta & \delta \\ \delta & \alpha & \delta & \delta & \delta \\ \hline \delta & \delta & \alpha & \delta & \delta \\ \delta & \delta & \delta & \alpha & \delta \\ \delta & \delta & \delta & \delta & \alpha \end{array}\right)$$

and structure (2.2) corresponds to the off-diagonal block of the autoregression of order one structure, i.e.

$$\delta \left(\begin{array}{cc|ccc} 1 & \varrho & \varrho^2 & \varrho^3 & \varrho^4 \\ \varrho & 1 & \varrho & \varrho^2 & \varrho^3 \\ \hline \varrho^2 & \varrho & 1 & \varrho & \varrho^2 \\ \varrho^3 & \varrho^2 & \varrho & 1 & \varrho \\ \varrho^4 & \varrho^3 & \varrho^2 & \varrho & 1 \end{array}\right).$$

We denote the sets of structured matrices defined above as

$$\begin{aligned} \mathcal{S}_1 &= \{\mathbf{\Gamma}_1 \in \mathbb{R}_p^{>},\ \mathbf{\Gamma}_2 \in \mathbb{R}_q^{>},\ \mathbf{\Psi} = \delta \mathbf{1}_p \mathbf{1}_q' : \mathbf{\Gamma} \in \mathbb{R}_{p+q}^{>}\}, \\ \mathcal{S}_2 &= \{\mathbf{\Gamma}_1 \in \mathbb{R}_p^{>},\ \mathbf{\Gamma}_2 \in \mathbb{R}_q^{>},\ \mathbf{\Psi} = \delta \mathbf{A} : \mathbf{\Gamma} \in \mathbb{R}_{p+q}^{>}\}. \end{aligned} \tag{2.3}$$

For a given matrix $\mathbf{\Omega}$, our aim is to determine a matrix from the set $\mathcal{S}_1$ and $\mathcal{S}_2$, which will be the closest to the unstructured matrix $\mathbf{\Omega}$ in the sense of some discrepancy function. The most relevant structure has the smallest discrepancy and the process of choosing the most appropriate structure is called by some authors, cf. Lin et al. (2014) [12], Cui et al. (2016) [1], regularization. However, in the literature regularization often has another meaning; that is, improvement of estimators by stating some additional requirements. Therefore, to avoid misunderstanding, the process described in this paper will be called covariance structure identification.

As a measure of discrepancy we use the entropy loss function

$$f(\mathbf{\Omega}, \mathbf{\Gamma}) = \operatorname{tr}(\mathbf{\Omega}^{-1}\mathbf{\Gamma}) - \ln|\mathbf{\Omega}^{-1}\mathbf{\Gamma}| - (p+q), \tag{2.4}$$

cf. James, Stein (1961) [9], Dey, Srinivasan (1985) [2]; which is also known as a Kullback-Leibler divergence between two probability distributions; cf. Pan, Fang (2002) [17]. The entropy loss function was considered in the approximation problem by Lin et al. (2014) [12] and Filipiak et al. (2018a) [6] in classic and doubly multivariate models, respectively. Since in the approximation process via the entropy loss function the inverses of matrices $\mathbf{\Omega}$ and $\mathbf{\Gamma}$ appear, their nonsingularity (positive definiteness) is required. The entropy loss function is convex and antisymmetric and has the important property that it is invariant with respect to the group of linear transformations (cf. James, Stein (1961) [9]), that is, for arbitrary nonsingular matrix $\mathbf{B}$

$$f(\mathbf{B}\mathbf{\Omega}\mathbf{B}', \mathbf{B}\mathbf{\Gamma}\mathbf{B}') = f(\mathbf{\Omega}, \mathbf{\Gamma}). \tag{2.5}$$

Let $\mathbf{\Omega} \in \mathbb{R}_m^{>}$ be partitioned as follows

$$\mathbf{\Omega} = \begin{pmatrix} \mathbf{\Omega}_{11} & \mathbf{\Omega}_{12} \\ \mathbf{\Omega}_{12}' & \mathbf{\Omega}_{22} \end{pmatrix},$$

where $\mathbf{\Omega}_{11} : p \times p$, $\mathbf{\Omega}_{12} : p \times q$, $\mathbf{\Omega}_{22} : q \times q$ and the inverse of matrix $\mathbf{\Omega}$ is a partitioned matrix, denoted by $\mathbf{V}$,

$$\mathbf{\Omega}^{-1} = \mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{12}' & \mathbf{V}_{22} \end{pmatrix},$$

where $\mathbf{V}_{11} : p \times p$, $\mathbf{V}_{12} : p \times q$, $\mathbf{V}_{22} : q \times q$.

Note that in the entropy loss function, we need a determinant of a partitioned nonsingular matrix $\mathbf{\Gamma}$ of the form

$$\mathbf{\Gamma} = \begin{pmatrix} \mathbf{\Gamma}_{11} & \mathbf{\Gamma}_{12} \\ \mathbf{\Gamma}_{21} & \mathbf{\Gamma}_{22} \end{pmatrix}$$

with non-singular diagonal blocks. That can be determined using the Schur complement (cf. Kollo, von Rosen (2005) [11]) as

$$|\mathbf{\Gamma}| = |\mathbf{\Gamma}_{22}| \cdot |\mathbf{\Gamma}_{11} - \mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Gamma}_{21}| = |\mathbf{\Gamma}_{11}| \cdot |\mathbf{\Gamma}_{22} - \mathbf{\Gamma}_{21}\mathbf{\Gamma}_{11}^{-1}\mathbf{\Gamma}_{12}|. \tag{2.6}$$

Then the entropy loss function of $\mathbf{\Omega}$ and $\mathbf{\Gamma}$ can be written in two forms

$$\begin{aligned} f(\mathbf{\Omega}, \mathbf{\Gamma}) = {} & \operatorname{tr}(\mathbf{V}_{11}\mathbf{\Gamma}_{11}) + 2\operatorname{tr}(\mathbf{V}_{12}\mathbf{\Gamma}_{21}) + \operatorname{tr}(\mathbf{V}_{22}\mathbf{\Gamma}_{22}) + \ln|\mathbf{\Omega}| \\ & - \ln|\mathbf{\Gamma}_{22}| - \ln|\mathbf{\Gamma}_{11} - \mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Gamma}_{21}| - (p+q) \\ = {} & \operatorname{tr}(\mathbf{V}_{11}\mathbf{\Gamma}_{11}) + 2\operatorname{tr}(\mathbf{V}_{12}\mathbf{\Gamma}_{21}) + \operatorname{tr}(\mathbf{V}_{22}\mathbf{\Gamma}_{22}) + \ln|\mathbf{\Omega}| \\ & - \ln|\mathbf{\Gamma}_{11}| - \ln|\mathbf{\Gamma}_{22} - \mathbf{\Gamma}_{21}\mathbf{\Gamma}_{11}^{-1}\mathbf{\Gamma}_{12}| - (p+q) \end{aligned} \tag{2.7}$$

with two variants of Schur complement (2.6) used.

3. Approximation

In this section we present the best approximation of a given $\mathbf{\Omega} \in \mathbb{R}_m^{>}$ by the matrix respectively from the set $\mathcal{S}_1$ and $\mathcal{S}_2$ given by (2.3) via the entropy loss function (2.4). To determine the smallest value of the discrepancy function, the entropy loss function is minimized over the respective set of structures.

To shorten the notation, let us denote the set of relevant structures, $\mathcal{S}_1$ or $\mathcal{S}_2$, by $\mathcal{G}$ and for a given $\mathbf{\Omega}$, the entropy loss function as $f_\Omega(\mathbf{\Gamma})$. Then our aim is to minimize the entropy loss function with respect to $\mathbf{\Gamma} \in \mathcal{G}$, that is to determine

$$\zeta = \min_{\mathbf{\Gamma} \in \mathcal{G}} f_\Omega(\mathbf{\Gamma}).$$

Let $\mathcal{G} = \mathcal{S}_1$. Then the entropy loss function (2.7) with $\mathbf{\Gamma} \in \mathcal{S}_1$ can be written as

$$\begin{aligned} f_\Omega(\mathbf{\Gamma}) = {} & \operatorname{tr}(\mathbf{V}_{11}\mathbf{\Gamma}_1) + 2\delta\operatorname{tr}(\mathbf{V}_{12}\mathbf{1}_q\mathbf{1}_p') + \operatorname{tr}(\mathbf{V}_{22}\mathbf{\Gamma}_2) + \ln|\mathbf{\Omega}| \\ & - \ln|\mathbf{\Gamma}_2| - \ln|\mathbf{\Gamma}_1 - \delta^2\mathbf{1}_p\mathbf{1}_q'\mathbf{\Gamma}_2^{-1}\mathbf{1}_q\mathbf{1}_p'| - (p+q) \\ = {} & \operatorname{tr}(\mathbf{V}_{11}\mathbf{\Gamma}_1) + 2\delta\operatorname{tr}(\mathbf{V}_{12}\mathbf{1}_q\mathbf{1}_p') + \operatorname{tr}(\mathbf{V}_{22}\mathbf{\Gamma}_2) + \ln|\mathbf{\Omega}| \\ & - \ln|\mathbf{\Gamma}_1| - \ln|\mathbf{\Gamma}_2 - \delta^2\mathbf{1}_q\mathbf{1}_p'\mathbf{\Gamma}_1^{-1}\mathbf{1}_p\mathbf{1}_q'| - (p+q). \end{aligned}$$

To obtain the minimum of the entropy loss function over the set $\mathcal{S}_1$ we differentiate this function with respect to $\mathbf{\Gamma}_1$ and δ using the first form of Schur complement and with respect to $\mathbf{\Gamma}_2$ using the second form. Derivatives have the following forms:

$$\begin{aligned}
\frac{\partial f}{\partial \mathbf{\Gamma}_1} &= [\mathrm{vec}'(\mathbf{V}_{11}) - \mathrm{vec}'(\mathbf{\Gamma}_1 - \delta^2 \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p')^{-1}] \cdot \mathbf{D}_p, \\
\frac{\partial f}{\partial \mathbf{\Gamma}_2} &= [\mathrm{vec}'(\mathbf{V}_{22}) - \mathrm{vec}'(\mathbf{\Gamma}_2 - \delta^2 \mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}_q')^{-1}] \cdot \mathbf{D}_q, \\
\frac{\partial f}{\partial \delta} &= 2\,\mathrm{tr}(\mathbf{V}_{12} \mathbf{1}_q \mathbf{1}_p') + 2\delta\,\mathrm{vec}'(\mathbf{V}_{11})\,\mathrm{vec}(\mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p'),
\end{aligned}$$

where $\mathbf{D}_p$ is a $(p+q)^2 \times (p+q)(p+q+1)/2$ duplication matrix and $\mathrm{vec}(\cdot)$ is an operator stacking the columns of a given matrix one below another (cf. Magnus, Neudecker (1986) [14] or Magnus, Neudecker (1999) [15]). The formulas for matrix derivatives can be found in Fackler (2005) [3]. Equating the derivatives to zero and using some transformations, we obtain the stationary point as the solution of the system of equations given in Theorem 3.1. The theorem follows from convexity and differentiability of the entropy loss function over the convex set, and there exists the stationary point which is the global minimum (cf. Lin et al. (2014) [12]).

Theorem 3.1. *For a given $\mathbf{\Omega} \in \mathbb{R}_m^>$, there exists $\mathbf{\Gamma} \in \mathbb{R}_m^>$ that minimizes the entropy loss function* (2.4) *over $\mathcal{S}_1$ and this minimum is attained at $\mathbf{\Gamma_1}$, $\mathbf{\Gamma}_2$, δ satisfying the following system of equations:*

$$\begin{cases}
\mathbf{\Gamma}_1 = \mathbf{V}_{11}^{-1} + \delta^2 \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p', \\
\mathbf{\Gamma}_2 = \mathbf{V}_{22}^{-1} + \delta^2 \mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}_q', \\
\delta = -\dfrac{\mathrm{tr}(\mathbf{V}_{12} \mathbf{1}_q \mathbf{1}_p')}{\mathrm{tr}(\mathbf{V}_{11} \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p')}.
\end{cases}$$

Let $\mathcal{G} = \mathcal{S}_2$. Then the entropy loss function (2.7) with $\mathbf{\Gamma} \in \mathcal{S}_2$ can be written as

$$\begin{aligned}
f_\Omega(\mathbf{\Gamma}) = {} & \mathrm{tr}(\mathbf{V}_{11}\mathbf{\Gamma}_1) + \delta\,\mathrm{tr}(\mathbf{V}_{12}'\mathbf{A}) + \delta\,\mathrm{tr}(\mathbf{V}_{12}\mathbf{A}') \\
& + \mathrm{tr}(\mathbf{V}_{22}\mathbf{\Gamma}_2) + \ln|\mathbf{\Omega}| - \ln|\mathbf{\Gamma}| - (p+q),
\end{aligned}$$

where

$$\ln|\mathbf{\Gamma}| = \ln|\mathbf{\Gamma}_2| + \ln|\mathbf{\Gamma}_1 - \delta^2 \mathbf{A}\mathbf{\Gamma}_2^{-1}\mathbf{A}'| = \ln|\mathbf{\Gamma}_1| + \ln|\mathbf{\Gamma}_2 - \delta^2 \mathbf{A}'\mathbf{\Gamma}_1^{-1}\mathbf{A}|$$

follows from Schur complement (2.6). To obtain the minimum of the entropy loss function over the set $\mathcal{S}_2$, we differentiate this function with respect to $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$, δ

and ϱ. Derivatives have the following forms:

$$\begin{aligned}
\frac{\partial f}{\partial \boldsymbol{\Gamma}_1} &= [\operatorname{vec}'(\mathbf{V}_{11}) - \operatorname{vec}'(\boldsymbol{\Gamma}_1 - \delta^2 \mathbf{A}\boldsymbol{\Gamma}_2^{-1}\mathbf{A}')^{-1}] \cdot \mathbf{D}_p, \\
\frac{\partial f}{\partial \boldsymbol{\Gamma}_2} &= [\operatorname{vec}'(\mathbf{V}_{22}) - \operatorname{vec}'(\boldsymbol{\Gamma}_2 - \delta^2 \mathbf{A}'\boldsymbol{\Gamma}_1^{-1}\mathbf{A})^{-1}] \cdot \mathbf{D}_q, \\
\frac{\partial f}{\partial \delta} &= 2\operatorname{tr}(\mathbf{V}_{12}'\mathbf{A}) + 2\delta \operatorname{vec}'(\mathbf{V}_{11}) \operatorname{vec}(\mathbf{A}\boldsymbol{\Gamma}_2^{-1}\mathbf{A}'), \\
\frac{\partial f}{\partial \varrho} &= [2\operatorname{vec}'(\mathbf{V}_{12}) - \operatorname{vec}'(\boldsymbol{\Gamma}_1 - \delta^2 \mathbf{A}\boldsymbol{\Gamma}_2^{-1}\mathbf{A}')^{-1}(\mathbf{I}_{p^2} + \mathbf{K}_{p,p})(\mathbf{A}\boldsymbol{\Gamma}_2^{-1} \otimes \mathbf{I}_p)]\operatorname{vec}\mathbf{F}
\end{aligned}$$

with $\mathbf{F}$ being a derivative of $\mathbf{A}$ with respect to ϱ, that is,

$$\mathbf{F} = \begin{pmatrix}
p\varrho^{p-1} & (p+1)\varrho^{p} & \dots & (p+q-1)\varrho^{p+q-2} \\
(p-1)\varrho^{p-2} & \dots & \dots & \dots \\
\vdots & \vdots & \vdots & \vdots \\
2\varrho & \dots & \dots & \dots \\
1 & 2\varrho & \dots & q\varrho^{q-1}
\end{pmatrix}$$

and $\mathbf{K}_{s,t}$ being an $st \times st$ commutation matrix (cf. Magnus, Neudecker (1986) [14]), such that for any $\mathbf{G} \in \mathbb{R}^{s,t}$ we have the following

$$\mathbf{K}_{s,t} \operatorname{vec}\mathbf{G} = \operatorname{vec}\mathbf{G}'.$$

Equating the derivatives to zero and using some transformations gives the stationary point as the solution of the system of equations given in Theorem 3.2 given below. The entropy loss function is a convex function, but the set $\mathcal{S}_2$ is not convex. Therefore, there may exist more than one stationary point and thus $\boldsymbol{\Gamma}$ from Theorem 3.2 can be only a local minimum (cf. Lin et al. (2014) [12]).

Theorem 3.2. *For a given $\boldsymbol{\Omega} \in \mathbb{R}_m^{\geq}$, there exists $\boldsymbol{\Gamma} \in \mathbb{R}_m^{\geq}$ that minimizes the entropy loss function* (2.4) *over $\mathcal{S}_2$ and this minimum is attained at $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$, δ, ϱ satisfying the following system of equations:*

$$\begin{cases}
\boldsymbol{\Gamma}_1 = \mathbf{V}_{11}^{-1} + \delta^2 \mathbf{A}\boldsymbol{\Gamma}_2^{-1}\mathbf{A}', \\
\boldsymbol{\Gamma}_2 = \mathbf{V}_{22}^{-1} + \delta^2 \mathbf{A}'\boldsymbol{\Gamma}_1^{-1}\mathbf{A}, \\
\delta = -\dfrac{\operatorname{tr}(\mathbf{V}_{12}'\mathbf{A})}{\operatorname{tr}(\mathbf{V}_{11}\mathbf{A}\boldsymbol{\Gamma}_2^{-1}\mathbf{A}')}, \\
2\operatorname{tr}(\mathbf{V}_{12}'\mathbf{F}) + \delta \operatorname{tr}[\mathbf{V}_{11}(\mathbf{F}\boldsymbol{\Gamma}_2^{-1}\mathbf{A}' + \mathbf{A}\boldsymbol{\Gamma}_2^{-1}\mathbf{F}')] = 0.
\end{cases}$$

Observe that the equations received by differentiating the entropy loss function with respect to $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$ and δ are comparable for $\mathcal{S}_1$ and $\mathcal{S}_2$ sets. In the second case we consider the structure with one more parameter (ϱ); thus we receive one more equation, which is a polynomial of order $3(p+q)-4$. It is worth noting that the systems of equations for minimizing the entropy loss function over both sets have no explicit solutions and can be solved only numerically. For this purpose we use an iterative procedure, where we start with some initial assumption and determine the sequence of improving approximate solutions such that the next result is derived from the previous one. The calculations are continued until the stopping rule is fulfilled. The details for initial assumptions and stopping rule are given in Section 4.

4. Simulation studies

Let us assume an experiment where two groups of characteristics are observed for n objects. The measurements are collected in vectors $\mathbf{x}_i$ for each object, $i = 1, \ldots, n$. We can denote a vector of observations as $\mathbf{x}_i = (\mathbf{x}_a', \mathbf{x}_b')'$, where the first group, with p elements, has mean vector $\boldsymbol{\mu}_a$, and the second one, with q elements, $\boldsymbol{\mu}_b$. Assume that each of the observation vectors $\mathbf{x}_i$ has the normal distribution with mean vector $\boldsymbol{\mu} = (\boldsymbol{\mu}_a', \boldsymbol{\mu}_b')'$ and covariance matrix $\mathbf{\Omega}$, i.e. $\mathbf{x}_i \sim N_m(\boldsymbol{\mu}, \mathbf{\Omega})$ with $m = p+q$. Using vectors of observations for each object, we constuct the observation matrix as $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ having matrix normal distribution $N_{m,n}(\boldsymbol{\mu}\mathbf{1}_n', \mathbf{\Omega}, \mathbf{I}_n)$. We are interested in relations between characterisitcs from the first and second group described in the covariance matrix $\mathbf{\Omega}_{12}$, especially $\delta\mathbf{1}_p\mathbf{1}_q'$ and $\delta\mathbf{A}$ described in Section 2.

In this section we apply the formulas from Theorems 3.1 and 3.2 in a simulation study. Our aim is to verify whether the considered discrepancy function recognizes the true structure properly. For this purpose we take sample size $n = 100$ and number of parameters $(p, q) = \{(2,2), (2,3), (2,5), (2,10), (3,2), (3,3), (3,5), (3,10)\}$, for which we determine matrix $\mathbf{\Omega}$ as a well-conditioned matrix from the set $\mathcal{S}_1$ or $\mathcal{S}_2$. Under these assumptions we generate the data from the normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{\Omega}$, i.e. $\mathbf{X} \sim N_{p+q,n}(\mathbf{0}, \mathbf{\Omega}, \mathbf{I}_n)$.

The choice of the most appropriate structure is based on the estimation of matrix $\mathbf{\Omega}$ and determination of the nearest structure in the sense of some discrepancy function. In real life applications, the covariance matrix $\mathbf{\Omega}$ is usually unknown; thus we replace it with a maximum likelihood estimator of $\mathbf{\Omega}$, i.e. $\mathbf{S} = \frac{1}{n}\mathbf{X}(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n')\mathbf{X}'$. By minimazing the entropy loss function over the set $\mathcal{S}_1$ and $\mathcal{S}_2$, we obtain respective discrepancies that we denote as ζ_1 and ζ_2, that is

$$\zeta_1 = \min_{\mathbf{\Gamma} \in \mathcal{S}_1} f_{\mathbf{S}}(\mathbf{\Gamma}) \quad \text{and} \quad \zeta_2 = \min_{\mathbf{\Gamma} \in \mathcal{S}_2} f_{\mathbf{S}}(\mathbf{\Gamma}).$$

The discrepancies ζ_1 and ζ_2 are computed using iterative algorithm based on a system of equations given in Theorems 3.1 and 3.2, respectively. In the first considered structure the initial assumption of the algorithm solving the system of equations given in Theorem 3.1 is $\mathbf{\Gamma}_2 = \mathbf{I}_q$, while in Theorem 3.2: $\mathbf{\Gamma}_2 = \mathbf{I}_q$ and $\varrho = 0.5$. The other variables, δ and $\mathbf{\Gamma}_1$, are computed from their respective equations (Theorems 3.1 and 3.2) and we receive the first step results of appropriate variables. Then the second step results are computed based on the first step. The iteration process goes on until some criterion is fulfilled, e.g. the difference between the solutions of two steps (or trace of the matrix difference) is smaller than a given threshold, here 10^{-5}. The last step results can be viewed as the estimators of respectively structured covariances. Obviously, in both algorithms the initial conditions can be chosen arbitrarily, for example instead of the identity matrix a block of $\mathbf{S}$ can be also assumed.

All simulations are repeated 1000 times, i.e. for the chosen true matrix $\mathbf{\Omega}$ we generate 1000 matrices $\mathbf{X}$ and for each of them we determine the maximum likelihood estimator of $\mathbf{\Omega}$, i.e. the matrix $\mathbf{S}$ and discrepancies ζ_1 and ζ_2. Then the averaged results of discrepancies, i.e. $\overline{\zeta}_1$ and $\overline{\zeta}_2$, are computed.

We expect that when the true matrix $\mathbf{\Omega} \in \mathcal{S}_1$ the discrepancy between $\mathbf{S}$ and the $\mathcal{S}_1$ set of structures is smaller than the discrepancy between $\mathbf{S}$ and the $\mathcal{S}_2$ set of structures, and when the true matrix $\mathbf{\Omega} \in \mathcal{S}_2$ we expect the opposite rule, that is

$$\mathbf{\Omega} \in \mathcal{S}_1 \Rightarrow \overline{\zeta}_1 \leqslant \overline{\zeta}_2,$$
$$\mathbf{\Omega} \in \mathcal{S}_2 \Rightarrow \overline{\zeta}_2 \leqslant \overline{\zeta}_1.$$

The simulations results are presented in Tables 1 and 2.

It can be seen from Table 1 that the entropy loss function identifies the true $\mathbf{\Omega} \in \mathcal{S}_1$ very rarely and that there are only a few cases with the proportion of proper structure identification higher than 50%. Moreover, for $q = 2$ there are some cases where none of the structures has been identified properly ($p = 2$ and $\delta = -3$ or $p = 3$ and $\delta \in \{-3, 3\}$). However, the best proportion (93.6%) of correct identification is also obtained for $q = 2$ ($p = 2$ and $\delta = -2$). We can also observe that bigger values of the parameters p and q usually provide to higher values of discrepancy.

In Table 2 the results for the true $\mathbf{\Omega} \in \mathcal{S}_2$ are presented. We can see that the averaged discrepancies satisfy $\overline{\zeta}_2 \leqslant \overline{\zeta}_1$ for each value of parameters p, q and δ (except one case). Moreover, the true $\mathbf{\Omega} \in \mathcal{S}_2$ proportion of proper structure identification is usually higher than 80% and for many cases the proportion is very high, larger than 95%. The results obtained for $\delta = 1$ and $\delta = 2$ are comparable. Similarly as in the first case, bigger values of parameters p and q provide higher values of discrepancy.

The small proportion of proper structure identification (small π values) for true $\mathbf{\Omega} \in \mathcal{S}_1$ is caused by the inclusion $\mathcal{S}_1 \subset \mathcal{S}_2$. It follows from the fact that the discrepancy from the larger set is always smaller. It can be also caused by the high bias of the matrix $\mathbf{S}$, which does not need to be a good estimator of $\mathbf{\Omega}$, as we will show below. However, if ζ_1 is close to ζ_2, then $\mathcal{S}_1$ can be recommended as the relevant structure, since it has smaller number of parameters.

Matrix $\mathbf{S}$ does not have structure from $\mathcal{S}_1$ with probability one. One would expect that $\zeta_1 < \zeta_2$, however this is not true in general. For example, in one of the cases where only 7.9% of structures were identified properly ($p = 2$, $q = 2$, $\delta = -5$), the true matrix $\mathbf{\Omega}$ has the form

$$\mathbf{\Omega} = \begin{pmatrix} 40.4805 & -3.2857 & -5 & -5 \\ -3.2857 & 30.4662 & -5 & -5 \\ -5 & -5 & 13.7867 & 7.0661 \\ -5 & -5 & 7.0661 & 29.8369 \end{pmatrix}$$

and its maximum likelihood estimate has the form

$$\mathbf{S} = \begin{pmatrix} 43.1327 & -3.2283 & -2.0342 & -4.0651 \\ -3.2283 & 30.2062 & -3.5415 & -4.5745 \\ -2.0342 & -3.5415 & 10.9243 & 7.4380 \\ -4.0651 & -4.5745 & 7.4380 & 29.4835 \end{pmatrix},$$

while

$$\widehat{\mathbf{\Gamma}}_{CS} = \begin{pmatrix} 43.3859 & -3.1991 & -3.1071 & -3.1071 \\ -3.1991 & 29.7926 & -3.1071 & -3.1071 \\ -3.1071 & -3.1071 & 10.9721 & 7.2440 \\ -3.1071 & -3.1071 & 7.2440 & 28.9051 \end{pmatrix}$$

and

$$\widehat{\mathbf{\Gamma}}_{AR} = \begin{pmatrix} 43.6253 & -3.1640 & -3.3547 & -4.0682 \\ -3.1640 & 29.6590 & -2.7663 & -3.3547 \\ -3.3547 & -2.7663 & 10.9387 & 7.3309 \\ -4.0682 & -3.3547 & 7.3309 & 29.1380 \end{pmatrix}$$

are the minimum of the entropy loss function over $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively, that is

$$\min_{\mathbf{\Gamma} \in \mathcal{S}_1} f_{\mathbf{S}}(\mathbf{\Gamma}) = f_{\mathbf{S}}(\widehat{\mathbf{\Gamma}}_{CS}) = \zeta_1 \quad \text{and} \quad \min_{\mathbf{\Gamma} \in \mathcal{S}_2} f_{\mathbf{S}}(\mathbf{\Gamma}) = f_{\mathbf{S}}(\widehat{\mathbf{\Gamma}}_{AR}) = \zeta_2.$$

In this case $\zeta_1 = 0.0079 > 0.0067 = \zeta_2$. It can follow from the fact that ζ_1 presents a discrepancy between the structure from the set $\mathcal{S}_1$ and a matrix $\mathbf{S}$ like ζ_2 gives the discrepancy between the structure from the set $\mathcal{S}_2$ and also the matrix $\mathbf{S}$, which is not a good estimator of the true structure.

Similarly, for $p = 3$, $q = 2$, $\delta = 3$ the true matrix has the following form:

$$\mathbf{\Omega} = \begin{pmatrix} 13.6376 & 7.0020 & 7.4252 & 3 & 3 \\ 7.0020 & 13.2476 & -5.1712 & 3 & 3 \\ 7.4252 & -5.1712 & 19.5732 & 3 & 3 \\ 3 & 3 & 3 & 7.9498 & 3.8762 \\ 3 & 3 & 3 & 3.8762 & 14.1895 \end{pmatrix}$$

and its maximum likelihood estimate is

$$\mathbf{S} = \begin{pmatrix} 12.8821 & 5.4486 & 8.9021 & 2.4804 & 3.5079 \\ 5.4486 & 11.4081 & -3.4473 & 4.0197 & 2.5917 \\ 8.9021 & -3.4473 & 19.7483 & 3.3223 & 4.1575 \\ 2.4804 & 4.0197 & 3.3223 & 9.4947 & 4.5119 \\ 3.5079 & 2.5917 & 4.1575 & 4.5119 & 10.9852 \end{pmatrix}.$$

As previously, matrix $\mathbf{S}$ has no structure from $\mathcal{S}_1$ anymore and the estimates obtained over respective sets are

$$\widehat{\mathbf{\Gamma}}_{CS} = \begin{pmatrix} 13.4928 & 6.0912 & 9.2271 & 3.6822 & 3.6822 \\ 6.0912 & 11.5260 & -3.1804 & 3.6822 & 3.6822 \\ 9.2271 & -3.1804 & 19.7059 & 3.6822 & 3.6822 \\ 3.6822 & 3.6822 & 3.6822 & 8.6508 & 4.7567 \\ 3.6822 & 3.6822 & 3.6822 & 4.7567 & 11.6244 \end{pmatrix}$$

and

$$\widehat{\mathbf{\Gamma}}_{AR} = \begin{pmatrix} 12.5147 & 5.3189 & 8.7078 & 2.7464 & 2.2326 \\ 5.3189 & 11.0068 & -3.3883 & 3.3783 & 2.7464 \\ 8.7078 & -3.3883 & 19.8810 & 4.1557 & 3.3783 \\ 2.7464 & 3.3783 & 4.1557 & 9.2361 & 4.7472 \\ 2.2326 & 2.7464 & 3.3783 & 4.7472 & 11.1314 \end{pmatrix}.$$

Similarly as in the first example, the $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ estimates are quite similar and the discrepancy between $\mathbf{S}$ and its estimate is smaller for $\widehat{\mathbf{\Gamma}}_{AR}$, $\zeta_1 = 0.07429 > 0.0447 = \zeta_2$.

On the other hand, the discrepancy values obtained in simulation studies are rather small. This means that the considered approximations ($\widehat{\mathbf{\Gamma}}_{CS}$ and $\widehat{\mathbf{\Gamma}}_{AR}$) are close to the matrix $\mathbf{S}$ and close to each other; thus they are very difficult to distinguish. Sometimes, it is better to choose a slightly more distant structure (in the sense of discrepancy function) and have a smaller number of parameters to estimate.

		$p=2$			$p=3$		
q	δ	$\overline{\zeta}_1$	$\overline{\zeta}_2$	π	$\overline{\zeta}_1$	$\overline{\zeta}_2$	π
2	−5	0.0299	0.0351	7.9	0.0491	0.1507	38.3
2	−3	0.0301	0.0206	0.0	0.0509	0.0404	0.0
2	−2	0.0303	0.1864	93.6	0.0496	0.0442	4.4
2	−1	0.0287	0.0587	46.4	0.0501	0.0399	5.5
2	0	0.0285	0.0209	6.4	0.0503	0.0390	8.7
2	1	0.0275	0.0271	18.5	0.0479	0.0408	3.0
2	2	0.0293	0.1013	70.8	0.0519	0.0512	7.1
2	3	0.0300	0.0694	36.7	0.0528	0.0417	0.0
2	5	0.0317	0.1434	53.9	0.0827	0.4040	49.4
3	−5	0.0564	0.0908	13.4	0.0954	0.3421	42.0
3	−3	0.0538	0.1174	32.3	0.0846	0.2178	44.5
3	−2	0.0489	0.0403	4.4	0.0833	0.1931	54.3
3	−1	0.0519	0.0417	2.3	0.0802	0.0719	7.1
3	0	0.0504	0.0464	2.3	0.0809	0.0727	8.2
3	1	0.0521	0.1318	55.9	0.0828	0.0737	13.8
3	2	0.0579	0.0974	26.9	0.0827	0.1956	84.8
3	3	0.0495	0.0452	5.6	0.0840	0.0914	13.2
3	5	0.0495	0.0866	14.9	0.0826	0.1674	39.4
5	−5	0.0941	0.1354	32.2	0.1494	0.1511	5.5
5	−3	0.1013	0.1458	24.7	0.1485	0.2032	24.6
5	−2	0.0946	0.0945	7.8	0.1487	0.1546	3.4
5	−1	0.0916	0.0831	14.5	0.1472	0.1423	21.1
5	0	0.0907	0.0795	4.7	0.1445	0.1410	15.8
5	1	0.0922	0.1181	27.5	0.1483	0.1444	10.3
5	2	0.0925	0.0906	7.5	0.1469	0.1524	25.7
5	3	0.0945	0.1971	50.9	0.1447	0.2356	25.2
5	5	0.0906	0.2403	46.6	0.1524	0.1445	0.8
10	−5	0.2113	0.4297	49.6	0.3195	0.3959	29.8
10	−3	0.2032	0.3481	65.1	0.3165	0.3187	18.9
10	−2	0.2074	0.2144	18.0	0.3102	0.3214	15.3
10	−1	0.2026	0.1980	22.0	0.3165	0.3154	27.7
10	0	0.2053	0.1964	17.2	0.3140	0.3027	18.9
10	1	0.2076	0.2017	20.0	0.3131	0.3088	14.8
10	2	0.2069	0.2144	17.7	0.3176	0.3183	9.1
10	3	0.2046	0.2104	11.3	0.3144	0.3338	29.5
10	5	0.2018	0.3070	40.1	0.3137	0.4553	42.8

Table 1. The averaged discrepancies $\overline{\zeta}_1$, $\overline{\zeta}_2$ and proportion π (in %) of proper structure identification among 1000 simulation runs in the case of $\mathbf{\Omega} \in \mathcal{S}_1$ with $n = 100$ for various values of δ.

			$\delta = 1$			$\delta = 2$		
p	q	ϱ	$\overline{\zeta}_1$	$\overline{\zeta}_2$	π	$\overline{\zeta}_1$	$\overline{\zeta}_2$	π
2	2	−0.9	0.0430	0.0293	89.1	0.1037	0.0633	94.1
2	2	−0.5	0.0542	0.0385	77.6	0.0424	0.0283	87.8
2	2	−0.1	0.0266	0.0220	84.6	0.0273	0.0230	88.6
2	2	0	0.0289	0.0222	83.2	0.0285	0.0234	90.5
2	2	0.1	0.0310	0.0202	93.5	0.0277	0.0211	97.0
2	2	0.5	0.0282	0.0238	92.8	0.0262	0.0643	73.8
2	2	0.9	0.0286	0.0201	99.8	0.0309	0.0303	87.4
2	3	−0.9	0.0630	0.0482	90.8	0.0889	0.0524	95.1
2	3	−0.5	0.0515	0.0416	91.8	0.1011	0.0448	100.0
2	3	−0.1	0.0512	0.0403	92.7	0.0483	0.0373	97.9
2	3	0	0.0524	0.0462	92.4	0.0493	0.0407	98.9
2	3	0.1	0.0520	0.0445	90.0	0.0494	0.0398	97.7
2	3	0.5	0.0504	0.0434	92.9	0.0568	0.0421	96.9
2	3	0.9	0.0511	0.0453	95.2	0.0543	0.0707	50.1
2	5	−0.9	0.1036	0.0837	87.3	0.1384	0.0911	94.8
2	5	−0.5	0.1008	0.0867	96.8	0.0936	0.0893	88.4
2	5	−0.1	0.0960	0.0873	91.1	0.0945	0.0843	93.3
2	5	0	0.0910	0.0854	90.3	0.0917	0.0851	95.6
2	5	0.1	0.0937	0.0876	89.2	0.0936	0.0871	87.8
2	5	0.5	0.0937	0.0865	97.8	0.0979	0.0802	95.5
2	5	0.9	0.0913	0.0814	99.8	0.0939	0.0840	95.3
2	10	−0.9	0.2925	0.2273	97.4	0.2224	0.2040	89.5
2	10	−0.5	0.2092	0.1974	80.3	0.2208	0.2114	93.2
2	10	−0.1	0.2017	0.1881	90.7	0.1995	0.1891	89.5
2	10	0	0.2065	0.1969	84.8	0.2041	0.1977	85.7
2	10	0.1	0.1984	0.1871	88.8	0.2052	0.1961	86.6
2	10	0.5	0.2037	0.1898	80.7	0.2098	0.1951	87.8
2	10	0.9	0.2208	0.1999	82.8	0.2065	0.1932	86.3
3	5	−0.9	0.5050	0.4646	88.9	0.1738	0.1520	97.5
3	5	−0.5	0.1627	0.1401	88.7	0.1500	0.1409	88.2
3	5	−0.1	0.1482	0.1343	88.1	0.1483	0.1405	88.2
3	5	0	0.1440	0.1348	89.1	0.1467	0.1355	88.7
3	5	0.1	0.1498	0.1377	90.7	0.1486	0.1380	89.9
3	5	0.5	0.1518	0.1397	92.6	0.1586	0.1359	95.1
3	5	0.9	0.1485	0.1418	92.9	0.1651	0.1396	99.8

			$\delta = 1$			$\delta = 2$		
p	q	ϱ	$\overline{\zeta}_1$	$\overline{\zeta}_2$	π	$\overline{\zeta}_1$	$\overline{\zeta}_2$	π
3	10	−0.9	0.3873	0.3470	89.4	0.6324	0.5133	97.7
3	10	−0.5	0.3108	0.2995	90.1	0.3148	0.3060	89.5
3	10	−0.1	0.3121	0.2973	76.0	0.3164	0.3027	82.3
3	10	0	0.3152	0.3040	78.8	0.3179	0.3073	81.4
3	10	0.1	0.3143	0.3083	88.4	0.3195	0.3044	87.2
3	10	0.5	0.3150	0.2999	86.2	0.3190	0.3036	86.8
3	10	0.9	0.3126	0.3047	84.4	0.3321	0.4113	15.6

Table 2. The averaged discrepancies $\overline{\zeta}_1$, $\overline{\zeta}_2$ and proportion π (in %) of proper structure identification among 1000 simulation runs in the case of $\mathbf{\Omega} \in \mathcal{S}_2$ with $n = 100$ for various values of ϱ.

5. Conclusions

We have proposed a method to find the best approximation of a given positive definite matrix by matrices with specified block structures, based on minimizing the entropy loss function, between a given $\mathbf{\Omega} \in \mathbb{R}^{>}_{m}$ and the set of matrices that has a structure under consideration. The simulation study showed that the $\mathcal{S}_2$ structure is better identified by the entropy loss function even when the assumed structure is $\mathcal{S}_1$. Moreover, with the increase of the parameters p and q, there is no apparent change in the proportion of proper structure identification. To sum up, this method can be used in statistics to approximate the covariance matrix.

Finally, it is worth mentioning that the problems of estimation of the partitioned covariance matrix with known structure of diagonal blocks of $\mathbf{\Gamma}$ can be considered, in particular, $\mathbf{\Gamma}_1 = \delta_1\, \mathbf{\Theta}_1$ and $\mathbf{\Gamma}_2 = \delta_2 \mathbf{\Theta}_2$, where $\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$ are known positive definite matrices. It can be related to the situation when the covariance structure for each of the observation vectors $\mathbf{x}_a$ and $\mathbf{x}_b$ has been already studied while the covariance between vectors $\mathbf{x}_a$ and $\mathbf{x}_b$ remains unknown. In such a case the formulas given in Theorems 3.1 and 3.2 are less complicated, and for $\mathbf{\Theta}_1 = \mathbf{I}_p$ and $\mathbf{\Theta}_2 = \mathbf{I}_q$ we have three or four variables instead of $\frac{1}{2}p(p+1)+\frac{1}{2}q(q+1)+1$ and $\frac{1}{2}p(p+1)+\frac{1}{2}q(q+1)+2$ for the structure from $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively.

In this paper the approximation by only two special structures is given. Observe however that there is a wide range of potential candidate structures; therefore, this topic will be a subject of our future research.

References

[1] *X. Cui, C. Li, J. Zhao, L. Zeng, D. Zhang, J. Pan*: Covariance structure regularization via Frobenius-norm discrepancy. Linear Algebra Appl. *510* (2016), 124–145. zbl MR doi

[2] *D. K. Dey, C. Srinivasan*: Estimation of a covariance matrix under Stein's loss. Ann. Stat. *13* (1985), 1581–1591. zbl MR doi

[3] *P. L. Fackler*: Notes on matrix calculus. Available at `https://studyres.com/doc/15927579/notes-on-matrix-calculus` (2005), 14 pages.

[4] *K. Filipiak, D. Klein*: Approximation with a Kronecker product structure with one component as compound symmetry or autoregression. Linear Algebra Appl. *559* (2018), 11–33. zbl MR doi

[5] *K. Filipiak, D. Klein, A. Markiewicz, M. Mokrzycka*: Approximation with a Kronecker product structure with one component as compound symmetry or autoregression via entropy loss function. To appear in Linear Algebra Appl. (2020). doi

[6] *K. Filipiak, D. Klein, M. Mokrzycka*: Estimators comparison of separable covariance structure with one component as compound symmetry matrix. Electron. J. Linear Algebra *33* (2018), 83–98. zbl MR doi

[7] *K. Filipiak, A. Markiewicz, A. Mieldzioc, A. Sawikowska*: On projection of a positive definite matrix on a cone of nonnegative definite Toeplitz matrices. Electron. J. Linear Algebra *33* (2018), 74–82. zbl MR doi

[8] *M. Gilson, D. Dahmen, R. Moreno-Bote, A. Insabato, M. Helias*: The covariance perceptron: A new paradigm for classication and processing of time series in recurrent neuronal networks. Available at `https://www.biorxiv.org/content/10.1101/562546v4` (2020), 39 pages. doi

[9] *W. James, C. Stein*: Estimation with quadratic loss. Proc. 4th Berkeley Symp. Math. Stat. Probab. Vol. 1. University of California Press, Berkeley, 1961, pp. 361–379. zbl MR

[10] *M. John, A. Mieldzioc*: The comparison of the estimators of banded Toeplitz covariance structure under the high-dimensional multivariate model. Commun. Stat., Simulation Comput. *49* (2020), 734–752. MR doi

[11] *T. Kollo, D. von Rosen*: Advanced Multivariate Statistics with Matrices. Mathematics and Its Applications 579. Springer, Dordrecht, 2005. zbl MR doi

[12] *L. Lin, N. J. Higham, J. Pan*: Covariance structure regularization via entropy loss function. Comput. Stat. Data Anal. *72* (2014), 315–327. zbl MR doi

[13] *N. Lu, D. L. Zimmerman*: The likelihood ratio test for a separable covariance matrix. Stat. Probab. Lett. *73* (2005), 449–457. zbl MR doi

[14] *J. R. Magnus, H. Neudecker*: Symmetry, 0-1 matrices and Jacobians: A review. Econom. Theory *2* (1986), 157–190. doi

[15] *J. R. Magnus, H. Neudecker*: Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley Series in Probability and Statistics. Wiley, Chichester, 1999. zbl MR doi

[16] *A. Mieldzioc, M. Mokrzycka, A. Sawikowska*: Covariance regularization for metabolomic data on the drought resistance of barley. Biometrical Lett. *56* (2019), 165–181. doi

[17] *J.-X. Pan, K.-T. Fang*: Growth Curve Models and Statistical Diagnostics. Springer Series in Statistics. Springer, New York, 2002. zbl MR doi

[18] *M. S. Srivastava, T. von Rosen, D. von Rosen*: Models with a Kronecker product covariance structure: Estimation and testing. Math. Methods Stat. *17* (2008), 357–370. zbl MR doi

[19] *T. H. Szatrowski*: Estimation and testing for block compound symmetry and other patterned covariance matrices with linear and non-linear structure. Technical Report OLK NSF 107, Stanford University, Stanford (1976), 187 pages.

[20] *T. H. Szatrowski*: Testing and estimation in the block compound symmetry problem. J. Educ. Behavioral Stat. *7* (1982), 3–18. doi

[21] *A. Szczepańska-Álvarez, C. Hao, Y. Liang, D. von Rosen*: Estimation equations for multivariate linear models with Kronecker structured covariance matrices. Commun. Stat., Theory Methods *46* (2017), 7902–7915. zbl MR doi

Authors' addresses: *Malwina Janiszewska*, *Augustyn Markiewicz* (corresponding author), Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Wojska Polskiego 28, 60-637 Poznań, Poland, e-mail: amark@up.poznan.pl; *Monika Mokrzycka* (http://orcid.org/0000-0002-4512-845X), Institute of Plant Genetics, Polish Academy of Sciences, Strzeszyńska 34, 60-479 Poznań, Poland.