Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 83-98, November 2018.

ESTIMATORS COMPARISON OF SEPARABLE COVARIANCE STRUCTURE WITH ONE COMPONENT AS COMPOUND SYMMETRY MATRIX*

KATARZYNA FILIPIAK†, DANIEL KLEIN‡, AND MONIKA MOKRZYCKA§

Abstract. The maximum likelihood estimation (MLE) of separable covariance structure with one component as compound symmetry matrix has been widely studied in the literature. Nevertheless, the proposed estimates are not given in explicit form and can be determined only numerically. In this paper we give an alternative form of MLE and we show that this new algorithm is much quicker than the algorithms given in the literature.
Another estimator of covariance structure can be found by minimizing the entropy loss function. In this paper we give three methods of finding the best approximation of separable covariance structure with one component as compound symmetry matrix and we compare the quickness of proposed algorithms.
We conduct simulation studies to compare statistical properties of MLEs and entropy loss estimators (ELEs), such us biasedness, variability and loss.

Key words. Separable covariance structure, Compound symmetry, Likelihood function, Entropy loss function, Estimation, Block trace operator.

AMS subject classifications. 15A30, 15B99, 62H12.

1. Preliminaries.

Let us consider experiments in which q characteristics at p time points are observed for each of n individuals. The data collected in this way are called doubly multivariate. Let $\mathbf{Y}_i$ for $i = 1, \ldots, n$ be independent and identically distributed $(q \times p)-$dimensional observation matrices. We assume $\mathbf{Y}_i \sim N_{q,p}(\mathbf{M}, \mathbf{\Omega})$, i.e., $\operatorname{vec} \mathbf{Y}_i \sim N_{pq}(\operatorname{vec} \mathbf{M}, \mathbf{\Omega})$, where $\operatorname{vec} \mathbf{M} \in \mathbb{R}^{pq}$, $\operatorname{vec}(\cdot)$ is the operator stacking the columns of a $q \times p$ matrix into a $pq-$dimensional vector, and $\mathbf{\Omega}$ is assumed to be an unstructured positive definite matrix of order pq. The vector of unknown parameters consists of $\operatorname{vec} \mathbf{M}$ and $\operatorname{vech} \mathbf{\Omega}$, where vec-half operator $\operatorname{vech}(\cdot)$ is the operator stacking the columns of a $(pq \times pq)-$dimensional symmetric matrix into a $pq(pq+1)/2-$dimensional vector by eliminating all of the supradiagonal elements. The number of unknown parameters to be estimated in $\mathbf{\Omega}$ is $pq(pq+1)/2$, which increases rapidly with an increase in either q or p. Estimation of $\mathbf{\Omega}$ is impossible when the sample size $n \leq pq$. Thus, researchers usually rely on a separable covariance matrix, expressed as the Kronecker product of two components:

$$\underset{pq \times pq}{\mathbf{\Omega}} = \underset{p \times p}{\mathbf{\Psi}} \otimes \underset{q \times q}{\mathbf{\Sigma}}.$$

Since usually matrices $\mathbf{\Psi}$ and $\mathbf{\Sigma}$ are unknown, the researcher is interested in their estimation and testing the hypothesis about the covariance structure; cf. e.g. Lu and Zimmerman [11], Roy and Khattree [16, 17], Mitchell et al. [14], Roy [15], Roy and Leiva [18], Srivastava et al. [21], Simpson [19], Manceur and Dutilleul [13], Filipiak et al. [6].

Filipiak et al. [6] have shown that the maximum likelihood estimators (MLEs) of unknown covariance

*Received by the editors on February 21, 2018. Accepted for publication on October 26, 2018. Handling Editor: Simo Puntanen.

†Institute of Mathematics, Poznań University of Technology, Poland (katarzyna.filipiak@put.poznan.pl).
‡Institute of Mathematics, Faculty of Science, P. J. Šafárik University, Košice, Slovakia (daniel.klein@upjs.sk).
§Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland (mmok@igr.poznan.pl).

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 83-98, November 2018.

matrices can be obtained as a solution of the following systems of matrix equations:

$$\left\{\begin{array}{rcl} q\ \operatorname{vec}\boldsymbol{\Psi} &=& (\mathbf{I}_{p^2}\otimes\operatorname{vec}^\top\boldsymbol{\Sigma}^{-1})(\mathbf{I}_p\otimes\mathbf{K}_{p,q}\otimes\mathbf{I}_q)\operatorname{vec}\mathbf{S}\\ p\ \operatorname{vec}\boldsymbol{\Sigma} &=& (\operatorname{vec}^\top\boldsymbol{\Psi}^{-1}\otimes\mathbf{I}_{q^2})(\mathbf{I}_p\otimes\mathbf{K}_{p,q}\otimes\mathbf{I}_q)\operatorname{vec}\mathbf{S}\end{array}\right. \tag{1.1}$$

with commutation matrix $\mathbf{K}_{p,q}$ (such that for an arbitrary $p\times q$ matrix $\mathbf{X}$, $\mathbf{K}_{p,q}\operatorname{vec}\mathbf{X}=\operatorname{vec}\mathbf{X}^\top$, cf. Kollo and von Rosen [9], and with

$$\mathbf{S}=n^{-1}\mathbf{Y}\mathbf{Q}_n\mathbf{Y}^\top \tag{1.2}$$

being the MLE of $\boldsymbol{\Omega}$ (cf. Filipiak et al. [6]), where $\mathbf{Y}=(\operatorname{vec}\mathbf{Y}_1,\operatorname{vec}\mathbf{Y}_2,\ldots,\operatorname{vec}\mathbf{Y}_n)$, and $\mathbf{Q}_n=\mathbf{I}_n-n^{-1}\mathbf{1}_n\mathbf{1}_n^\top$ is the orthogonal projector onto the orthocomplement of the column span of an $n-$dimensional vector of ones, $\mathbf{1}_n$. Note, that (1.1) follows from differentiation of the log-likelihood function

$$\ln L(\boldsymbol{\Psi},\boldsymbol{\Sigma};\mathbf{S}) = -\frac{npq}{2}\ln(2\pi)-\frac{n}{2}\ln|\boldsymbol{\Psi}\otimes\boldsymbol{\Sigma}|-\frac{n}{2}\operatorname{tr}\left[(\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{S}\right] \tag{1.3}$$

with respect to $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$.

To simplify (1.1) as well as to derive the main results of this paper we use the partial trace and block trace operators of square matrices (given below) and their properties (given in Appendix A). The following definition is a special case of partial trace and block trace operators defined by Filipiak et al. [7] for rectangular block matrices.

Definition 1.1. For an arbitrary $mp\times mp$ matrix $\mathbf{A}=(\mathbf{A}_{ij})$

(i) the **partial trace** operator, $\operatorname{PTr}_p:\mathbb{R}^{mp\times mp}\to\mathbb{R}^{m\times m}$, is the matrix of the traces of $p\times p$ blocks of $\mathbf{A}$, that is

$$\operatorname{PTr}_p\mathbf{A}=(\operatorname{tr}\mathbf{A}_{ij}),\quad i=1,\ldots,m,\ j=1,\ldots,n,$$

(ii) the **block trace** operator, $\operatorname{BTr}_m:\mathbb{R}^{mp\times mp}\to\mathbb{R}^{m\times m}$, is the sum of all diagonal $m\times m$ blocks of $\mathbf{A}$, that is

$$\operatorname{BTr}_m\mathbf{A}=\sum_{i=1}^{p}\mathbf{A}_{ii}.$$

From Lemma A.1 it can be easily seen that system (1.1) can be rewritten as

$$\left\{\begin{array}{rcl} q\boldsymbol{\Psi} &=& \operatorname{PTr}_q\left[(\mathbf{I}_p\otimes\boldsymbol{\Sigma}^{-1/2})\mathbf{S}(\mathbf{I}_p\otimes\boldsymbol{\Sigma}^{-1/2})\right]\\ p\boldsymbol{\Sigma} &=& \operatorname{BTr}_q\left[(\boldsymbol{\Psi}^{-1/2}\otimes\mathbf{I}_q)\mathbf{S}(\boldsymbol{\Psi}^{-1/2}\otimes\mathbf{I}_q)\right].\end{array}\right.$$

and it can be further reduced using Lemma A.2 to

$$\left\{\begin{array}{rcl} q\boldsymbol{\Psi} &=& \operatorname{PTr}_q\left[(\mathbf{I}_p\otimes\boldsymbol{\Sigma}^{-1})\mathbf{S}\right]\\ p\boldsymbol{\Sigma} &=& \operatorname{BTr}_q\left[(\boldsymbol{\Psi}^{-1}\otimes\mathbf{I}_q)\mathbf{S}\right].\end{array}\right. \tag{1.4}$$

It is worth to mention that to solve (1.1) much more operations are required than to solve (1.4). Thus, the use of the partial trace and block trace operators allows to express the estimators in shorter form as well as it makes the algorithm faster.

The problems which arise here are the convergence of the proposed algorithm and the uniqueness of its solution. These issues were studied e.g. by Lu and Zimmerman [11] or Soloveychik and Trushin [20]. In this

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 83-98, November 2018.

last paper the authors proved that the sufficient condition for the convergence of the so called "flip-flop" algorithm used to solve (1.4) and for the uniqueness almost surely of the solution of (1.4) is the sample size n greater than $p/q + q/p + 1$.

In a variety of applications there is no prior information about the possible structure of the covariance matrix. In such a case the structure may be prespecified from the range of potential covariance structures using some criteria based on e.g. Frobenius norm (Cui et al. [1], Filipiak and Klein [4]) or entropy loss function (Lin et al. [10]). In the latter situation the problem is to find a minimum of the entropy loss function defined in the case of separable structure as

$$f(\mathbf{\Psi}, \mathbf{\Sigma}; \mathbf{\Omega}) = \operatorname{tr}\left[\mathbf{\Omega}^{-1}(\mathbf{\Psi} \otimes \mathbf{\Sigma})\right] - \ln\left|\mathbf{\Omega}^{-1}(\mathbf{\Psi} \otimes \mathbf{\Sigma})\right| - pq \tag{1.5}$$

(cf. Dey and Srinivasan [2], James and Stein [8], Lin et al. [10]) with respect to unknown covariance structure parameters, $\mathbf{\Psi}$ and $\mathbf{\Sigma}$. Since the true $\mathbf{\Omega}$ is unknown, to find the best approximation of $\mathbf{\Omega}$ by a separable matrix $\mathbf{\Psi} \otimes \mathbf{\Sigma}$ in (1.5) we use $\mathbf{S}$ defined in (1.2) being the function of sufficient statistic for $\mathbf{\Omega}$, and hence our goal is to minimize

$$f(\mathbf{\Psi}, \mathbf{\Sigma}; \mathbf{S}) = \operatorname{tr}\left[\mathbf{S}^{-1}(\mathbf{\Psi} \otimes \mathbf{\Sigma})\right] - \ln\left|\mathbf{S}^{-1}(\mathbf{\Psi} \otimes \mathbf{\Sigma})\right| - pq \tag{1.6}$$

with respect to $\mathbf{\Psi}$ and $\mathbf{\Sigma}$. It can be easily seen that since the inverse of $\mathbf{S}$ is used in (1.6), the sample size should exceed pq. Therefore we assume $n > pq$ $(p, q \geq 2)$ throughout this article. It is worth noting that the condition of Soloveychik and Trushin [20] is then satisfied.

Observe that the solution $\widetilde{\mathbf{\Psi}}$ and $\widetilde{\mathbf{\Sigma}}$ in which the minimum of (1.6) is attained can be treated as an estimator of unknown structured covariance matrix in the doubly multivariate data. The estimators $\widetilde{\mathbf{\Psi}}$ and $\widetilde{\mathbf{\Sigma}}$ will be called in this paper the *entropy loss estimators* (ELEs).

Note, that the entropy loss function of the inverses of $\mathbf{S}$ and $\mathbf{\Psi} \otimes \mathbf{\Sigma}$ can be expressed using the log-likelihood function defined in (1.3), that is

$$\begin{aligned}
\ln f\left(\mathbf{\Psi}^{-1}, \mathbf{\Sigma}^{-1}; \mathbf{S}^{-1}\right) &= \operatorname{tr}\left[\mathbf{S}\left(\mathbf{\Psi}^{-1} \otimes \mathbf{\Sigma}^{-1}\right)\right] - \ln\left|\mathbf{S}\left(\mathbf{\Psi}^{-1} \otimes \mathbf{\Sigma}^{-1}\right)\right| - pq \\
&= \operatorname{tr}\left[\left(\mathbf{\Psi}^{-1} \otimes \mathbf{\Sigma}^{-1}\right)\mathbf{S}\right] + \ln|\mathbf{\Psi} \otimes \mathbf{\Sigma}| + pq\ln(2\pi) \\
&\quad -pq\ln(2\pi) - \ln|\mathbf{S}| - pq \\
&= -\tfrac{2}{n}\ln L(\mathbf{\Psi}, \mathbf{\Sigma}; \mathbf{S}) - \ln|\mathbf{S}| - pq[1 + \ln(2\pi)]
\end{aligned}$$

Therefore, the minimum of the entropy loss function can be found in the same way as the maximum of the log-likelihood function, with $\mathbf{S}$ and $\mathbf{\Psi} \otimes \mathbf{\Sigma}$ replaced by the inverses. Thus, using the same technique as in Filipiak et al. [6] and Lemmas A.1 and A.2 from Appendix A, we obtain the following system of equations:

$$\begin{cases} q\mathbf{\Psi}^{-1} &= \mathrm{PTr}_q\left[(\mathbf{I}_p \otimes \mathbf{\Sigma})\mathbf{S}^{-1}\right] \\ p\mathbf{\Sigma}^{-1} &= \mathrm{BTr}_q\left[(\mathbf{\Psi} \otimes \mathbf{I}_q)\mathbf{S}^{-1}\right]. \end{cases} \tag{1.7}$$

More detailed procedure based on the differentiation of (1.6) with respect to $\mathbf{\Psi}$ and $\mathbf{\Sigma}$ can be found in Appendix B.

Note that both functions, (1.3) and (1.6), are invariant with respect to the linear transformation of the data matrix. Moreover, since there is a strong relation between the MLEs (1.4) and ELEs (1.7) of $\mathbf{\Psi} \otimes \mathbf{\Sigma}$, one would expect that this relation still holds if one of the components of Kronecker product is additionally

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 83-98, November 2018.

structured. One of the structure considered in the literature is compound symmetry (CS): see e.g. Cui et al. [1], Filipiak and Klein [4] or Lin et al. [10] for specification of the structure, or Roy and Khatree [16], Roy and Leiva [18], Filipiak et al. [5, 6] for estimation and testing hypotheses about the covariance structure. Without loss of generality let us assume that

$$\mathbf{\Psi} = \mathbf{\Psi}_{CS} \quad = \quad (1-\varrho)\mathbf{I}_p + \varrho \mathbf{1}_p \mathbf{1}_p^\top,$$

with $\varrho \in \left(-\frac{1}{p-1}, 1\right)$ and $\mathbf{\Sigma}$ is arbitrary unstructured positive definite matrix. Observe, that if the second component of Kronecker product is assumed to be compound symmetric, i.e., $\mathbf{\Sigma} = \mathbf{\Sigma}_{CS}$, and $\mathbf{\Psi}$ is arbitrary unstructured positive definite matrix, it is enough to change the order of Kronecker product components with respect to the rule

$$\mathbf{\Psi} \otimes \mathbf{\Sigma} = \mathbf{K}_{q,p}(\mathbf{\Sigma} \otimes \mathbf{\Psi})\mathbf{K}_{p,q};$$

cf. Magnus and Neudecker [12].

The main goal of this paper will be to verify whether similar relation as between the MLEs and ELEs of separable structure with both components unstructured holds also in the case of $\mathbf{\Psi} = \mathbf{\Psi}_{CS}$ and to compare the properties of MLEs $\widehat{\mathbf{\Psi}}_{CS}$ and $\widehat{\mathbf{\Sigma}}$ with the properties of ELEs, $\widetilde{\mathbf{\Psi}}_{CS}$ and $\widetilde{\mathbf{\Psi}}_{CS}$.

The MLEs of $\mathbf{\Psi}_{CS}$ and $\mathbf{\Sigma}$ has been widely studied in the literature (see e.g. Roy and Khatree [16], Filipiak et al. [6]). Nevertheless, the proposed estimators are not given in explicit form and are determined iteratively. In this paper we give an alternative technique of determination of the MLEs of $\mathbf{\Psi}_{CS}$ and $\mathbf{\Sigma}$ directly, and we show that this new algorithm is much quicker than the algorithms given in the literature.

In the case of ELEs of $\mathbf{\Psi}_{CS}$ and $\mathbf{\Sigma}$ in this paper we give three methods of their determination: iterative, direct, and the third one (also iterative), but with the use of spectral decomposition of CS structure. Similarly as in the case of MLEs we compare the time necessary to obtain ELEs using these three different approaches.

It is worth noting that assuming CS structure in both cases (MLE and ELE) we can avoid iterative procedure, which was necessary to determine the estimators under general separable structure.

To compare statistical properties of MLEs and ELEs, such us the biasedness or the mean square error and expected loss, we conduct simulation studies.

2. MLEs of $\mathbf{\Psi}_{CS} \otimes \mathbf{\Sigma}$. Filipiak et al. [6] have shown that the MLEs of unknown separable covariance matrix with one component structured as compound symmetry matrix can be obtained as a solution of the following systems of matrix equations:

$$\begin{cases} (p-1)k_0\varrho^3 + \{k_0 - (p-1)k_0 + (p-1)^2 a - (p-1)b\}\varrho^2 + \{2(p-1)a - k_0\}\varrho + (a-b) = 0 \\ p\mathbf{\Sigma} = \mathrm{BTr}_q\left[(\mathbf{\Psi}_{CS}^{-1} \otimes \mathbf{I}_q)\mathbf{S}\right] \end{cases} \tag{2.8}$$

with

$$a = \mathrm{tr}\left[(\mathbf{I}_p \otimes \mathbf{\Sigma}^{-1})\mathbf{S}\right], \quad b = \mathrm{tr}\left[(\mathbf{1}_p\mathbf{1}_p^\top \otimes \mathbf{\Sigma}^{-1})\mathbf{S}\right], \quad \text{and} \quad k_0 = nq(p-1)p.$$

Since the first equation in (2.8) depends on $\mathbf{\Sigma}$ through a and b, and since the second one depends on ϱ through $\mathbf{\Psi}_{CS}$, one possibility of finding the solution could be the iterative flip-flop algorithm. It should be mentioned that the MLEs given as numerical solution of (2.8) are equivalent to the MLEs given by Roy and

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 83-98, November 2018.

Khattree [16, Equations (3) and (6)], who have pointed out that the MLE of ϱ is always between $-1/(p-1)$ and 1, and thus the resulting estimators give the positive definite matrices. Moreover, the MLE of $\mathbf{\Sigma}$ is positive definite, as the block trace operator of positive definite matrix preserves positive definiteness (cf. Filipiak et al. [7, Lemma 2.4]). Recall, that we assume $n > pq$ to ensure positive definiteness of $\mathbf{S}$.

Let us present now a different approach to determine the MLEs of $\mathbf{\Psi}$ and $\mathbf{\Sigma}$.

For every matrix $\mathbf{X}$ of order pq,

$$\operatorname{tr}\mathbf{X} = \operatorname{tr}(\operatorname{BTr}_q\mathbf{X}) = \operatorname{tr}(\operatorname{BTr}_p\mathbf{X}). \tag{2.9}$$

Thus, (1.3) with $\mathbf{\Psi} = \mathbf{\Psi}_{CS}$ can be expressed as

$$\ln L(\mathbf{\Psi}_{CS},\mathbf{\Sigma};\mathbf{S}) = -\frac{npq}{2}\ln(2\pi) - \frac{n}{2}\ln|\mathbf{\Psi}_{CS}\otimes\mathbf{\Sigma}| - \frac{n}{2}\operatorname{tr}\left\{\operatorname{BTr}_q\left[(\mathbf{\Psi}_{CS}^{-1}\otimes\mathbf{\Sigma}^{-1})\mathbf{S}\right]\right\},$$

and further, using Lemma A.3,

$$\ln L(\mathbf{\Psi}_{CS},\mathbf{\Sigma};\mathbf{S}) = -\frac{npq}{2}\ln(2\pi) - \frac{n}{2}\ln|\mathbf{\Psi}_{CS}\otimes\mathbf{\Sigma}| - \frac{n}{2}\operatorname{tr}\left\{\mathbf{\Sigma}^{-1}\operatorname{BTr}_q\left[(\mathbf{\Psi}_{CS}^{-1}\otimes\mathbf{I}_q)\mathbf{S}\right]\right\}. \tag{2.10}$$

Using the spectral decomposition of $\mathbf{\Psi}_{CS}$, that is

$$\mathbf{\Psi}_{CS} = c_1\mathbf{P}_p + c_2\mathbf{Q}_p \tag{2.11}$$

with $\mathbf{P}_p = \mathbf{I}_p - \mathbf{Q}_p$ being an orthogonal projector onto the vector of ones, and $c_1 = 1+(p-1)\varrho$, $c_2 = 1-\varrho$, obviously $\mathbf{\Psi}_{CS}^{-1} = \frac{1}{c_1}\mathbf{P}_p + \frac{1}{c_2}\mathbf{Q}_p$, and we can express the block trace term of (2.10) as

$$\operatorname{BTr}_q\left[(\mathbf{\Psi}_{CS}^{-1}\otimes\mathbf{I}_q)\mathbf{S}\right] = \frac{1}{c_1}\operatorname{BTr}_q\left[(\mathbf{P}_p\otimes\mathbf{I}_q)\mathbf{S}\right] + \frac{1}{c_2}\operatorname{BTr}_q\left[(\mathbf{Q}_p\otimes\mathbf{I}_q)\mathbf{S}\right].$$

Let us denote

$$\mathbf{A} = \operatorname{BTr}_q\left[(\mathbf{P}_p\otimes\mathbf{I}_q)\mathbf{S}\right], \qquad \mathbf{B} = \operatorname{BTr}_q\left[(\mathbf{Q}_p\otimes\mathbf{I}_q)\mathbf{S}\right].$$

Then, since $|\mathbf{\Psi}_{CS}| = c_1c_2^{(p-1)}$,

$$\begin{aligned}\ln L(\mathbf{\Psi}_{CS},\mathbf{\Sigma};\mathbf{S}) = {} & -\frac{npq}{2}\ln(2\pi) - \frac{nq}{2}\ln c_1 - \frac{nq(p-1)}{2}\ln c_2 - \frac{np}{2}\ln|\mathbf{\Sigma}| \\ & -\frac{n}{2c_1}\operatorname{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{A}\right) - \frac{n}{2c_2}\operatorname{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{B}\right).\end{aligned} \tag{2.12}$$

Differentiating the above function with respect to $\mathbf{\Sigma}$ and equating the result to zero we get

$$p\widehat{\mathbf{\Sigma}} = \frac{1}{c_1}\mathbf{A} + \frac{1}{c_2}\mathbf{B} \tag{2.13}$$

or

$$\widehat{\mathbf{\Sigma}} = \frac{1}{p}\operatorname{BTr}_q\left[\left(\mathbf{\Psi}_{CS}^{-1}\otimes\mathbf{I}_q\right)\mathbf{S}\right]. \tag{2.14}$$

Differentiating now (2.12) with respect to ϱ and equating the result to zero we obtain

$$-\frac{nq(p-1)}{2c_1} + \frac{nq(p-1)}{2c_2} + \frac{n(p-1)}{2c_1^2}\operatorname{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{A}\right) - \frac{n}{2c_2^2}\operatorname{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{B}\right) = 0.$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 83-98, November 2018.

Plugging (2.14) into the above equality we get

$$q(p-1)\varrho + (p-1)c_2^2 \operatorname{tr}\left[(c_2\mathbf{A}+c_1\mathbf{B})^{-1}\mathbf{A}\right] - c_1^2 \operatorname{tr}\left[(c_2\mathbf{A}+c_1\mathbf{B})^{-1}\mathbf{B}\right] = 0. \tag{2.15}$$

Since (2.15) does not depend on $\boldsymbol{\Sigma}$, the solution $\widehat{\varrho}$ of (2.15) gives us the MLE of ϱ directly. It can be observed, that $\widehat{\varrho} \in (-1/(p-1), 1)$. Indeed, if we denote the left hand side of (2.15) by $g(\varrho)$, we obtain $g(-\frac{1}{p-1}) = q(p-1) > 0$ and $g(1) = -q < 0$, and thus at least one solution of $g(\varrho) = 0$ belongs to the desired interval. To get the MLE of $\boldsymbol{\Sigma}$ it is enough to plug $\widehat{\varrho}$ into (2.14). It is worth noting that there might be several roots of (2.15) in the desired interval. Then, the one which maximizes $\ln L(\widehat{\boldsymbol{\Psi}}_{CS}, \widehat{\boldsymbol{\Sigma}}; \mathbf{S})$ is chosen.

It is easy to see that the second equation of (2.8) and (2.13) represent the same estimator of $\boldsymbol{\Sigma}$ (positive definite). Moreover, using the property (2.9) and Lemma A.3 from Appendix A for a and b defined with (2.8), and then simplifying the first equation of (2.8) we get (2.15).

It is worth to note, that there is no iterative procedure in this approach, and even if (2.15) is nonlinear, this technique is much less time consuming. Indeed, if we compare by simulations the time necessary to determine the MLEs of ϱ and $\boldsymbol{\Sigma}$ using iterative procedure (2.8) and using direct procedure solving (2.15) and (2.14), this observation is confirmed by Figure 1. We have generated 1000 samples of size $n = 100$ from a pq multivariate normal distribution $N_{pq}(\mathbf{0}, \boldsymbol{\Psi}_{CS} \otimes \boldsymbol{\Sigma})$ for $p = 3, 10, 15, 30$ with $q = 3$ and $p = 3, 10, 15$ with $q = 5$, assuming $\varrho = 0, 0.5, 0.9$ and $\boldsymbol{\Sigma} = \mathbf{I}_q$ and randomly generated

$$\boldsymbol{\Sigma} = \begin{pmatrix} 6.020 & 4.124 & 1.584 \\ 4.124 & 4.503 & 1.039 \\ 1.584 & 1.039 & 6.326 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 6.467 & 0.438 & 1.877 & -1.049 & 2.235 \\ 0.438 & 6.188 & 0.039 & 0.854 & 0.055 \\ 1.877 & 0.039 & 5.971 & 0.245 & 0.000 \\ -1.049 & 0.854 & 0.245 & 4.129 & -0.651 \\ 2.235 & 0.055 & 0.000 & -0.651 & 7.885 \end{pmatrix}.$$

The averaged time (in seconds) of calculating MLEs on the medium class computer was computed and the results are presented on Figure 1.

As expected, it can be seen from Figure 1 that the time of determining MLEs depends on the parameters p and q for both procedures (for $q = 5$ the computing time is slightly longer than for $q = 3$), but only in the case of iterative procedure it depends also on the values of ϱ and $\boldsymbol{\Sigma}$ (red lines overlap with each other). An interesting thing is that for $\varrho = 0$ the iterative procedure can be faster than the direct one.

3. ELEs of $\boldsymbol{\Psi}_{CS} \otimes \boldsymbol{\Sigma}$. In order to find ELE, first we apply the same approach as Filipiak et al. [6] for MLE of unknown separable covariance matrix with one component structured as compound symmetry matrix. Thus, we have to differentiate $f(\boldsymbol{\Psi}_{CS}, \boldsymbol{\Sigma}; \mathbf{S})$ defined in (1.6) with respect to ϱ and $\boldsymbol{\Sigma}$.

Using the differentiation rules described by Magnus and Neudecker [12] we obtain

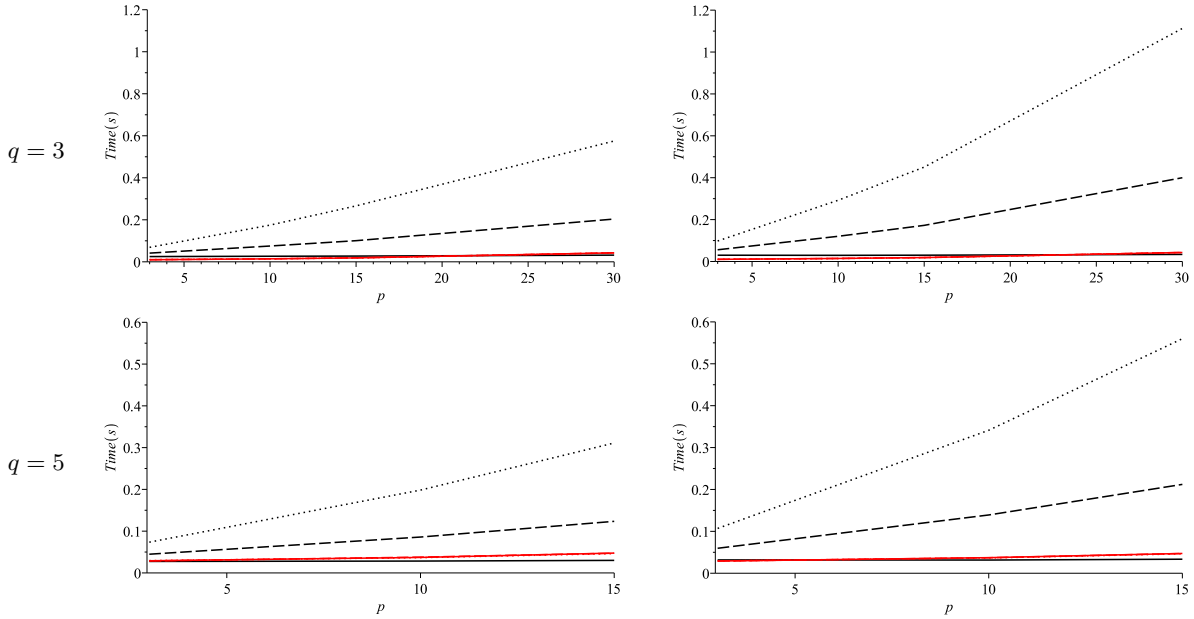
$$\begin{cases} \dfrac{\partial f}{\partial \varrho} &= \operatorname{vec}^\top \left(\mathbf{S}^{-1} - \boldsymbol{\Psi}_{CS}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) (\mathbf{I}_p \otimes \mathbf{K}_{q,p} \otimes \mathbf{I}_q) \left(\mathbf{I}_{p^2} \otimes \operatorname{vec} \boldsymbol{\Sigma}\right) \operatorname{vec}\left(\mathbf{1}_p\mathbf{1}_p^\top - \mathbf{I}_p\right) \\ \dfrac{\partial f}{\partial \boldsymbol{\Sigma}} &= \operatorname{vec}^\top \left(\mathbf{S}^{-1} - \boldsymbol{\Psi}_{CS}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) (\mathbf{I}_p \otimes \mathbf{K}_{q,p} \otimes \mathbf{I}_q) \left(\operatorname{vec} \boldsymbol{\Psi}_{CS} \otimes \mathbf{I}_{q^2}\right) \cdot \mathbf{D}_q, \end{cases}$$

where $\mathbf{D}_q$ is a $q \times q(q+1)/2$ zero-one duplication matrix such that for arbitrary $q \times q$ matrix $\mathbf{X}$, $\mathbf{D}_q \operatorname{vech} \mathbf{X} = \operatorname{vec} \mathbf{X}$. Since the duplication matrix does not have any effect on the solution of the above derivatives equated to zero, it can be omitted, and hence

$$\begin{cases} \operatorname{vec}^\top \left(\mathbf{1}_p\mathbf{1}_p^\top - \mathbf{I}_p\right) \left(\mathbf{I}_{p^2} \otimes \operatorname{vec}^\top \boldsymbol{\Sigma}\right) (\mathbf{I}_p \otimes \mathbf{K}_{p,q} \otimes \mathbf{I}_q) \operatorname{vec}\left(\mathbf{S}^{-1} - \boldsymbol{\Psi}_{CS}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) &= 0 \\ \left(\operatorname{vec}^\top \boldsymbol{\Psi}_{CS} \otimes \mathbf{I}_{q^2}\right) (\mathbf{I}_p \otimes \mathbf{K}_{p,q} \otimes \mathbf{I}_q) \operatorname{vec}\left(\mathbf{S}^{-1} - \boldsymbol{\Psi}_{CS}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) &= 0. \end{cases}$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 83-98, November 2018.

FIGURE 1. *The averaged time of calculating MLEs using two approaches (black color for iterative procedure and red color for direct) with respect to p, ϱ and $\mathbf{\Sigma}$: $\varrho = 0$ – solid line, $\varrho = 0.5$ – dashed line, $\varrho = 0.9$ – dotted line; $\mathbf{\Sigma} = \mathbf{I}_q$ – left column, $\mathbf{\Sigma}$ randomly generated – right column.*



Using the properties of partial trace and block trace operators described in Lemmas A.1 and A.3 given in Appendix A and formulas (8) and (15) from Magnus and Neudecker [12] we obtain

$$\begin{cases} -(p-1)\alpha\varrho^2 + [(p-2)\alpha + pq(p-1)]\,\varrho + \alpha = 0 \\ p\mathbf{\Sigma}^{-1} = \mathrm{BTr}_q\left[(\mathbf{\Psi}_{CS} \otimes \mathbf{I}_q)\,\mathbf{S}^{-1}\right] \end{cases}$$

with $\alpha = \mathrm{tr}\left\{\left[\left(\mathbf{1}_p\mathbf{1}_p^\top - \mathbf{I}_p\right) \otimes \mathbf{\Sigma}\right]\mathbf{S}^{-1}\right\}$.

Let us denote $-(p-1)\alpha\varrho^2 + [(p-2)\alpha + pq(p-1)]\,\varrho + \alpha$ by $h(\varrho)$. This function has two roots, as for $p \geq 2$ its discriminant is always positive. If $\alpha > 0$ ($\alpha < 0$) then the polynomial $h(\varrho)$ is concave (convex) and the solution $\widetilde{\varrho}$ is determined by a smaller (bigger) root of $h(\varrho)$, since this root minimizes $f(\mathbf{\Psi}_{CS}, \mathbf{\Sigma}; \mathbf{S})$. In both cases $\widetilde{\varrho}$ has the same form and the best approximation of $\mathbf{S}$ can be found by solving

$$\begin{cases} \varrho = \dfrac{-(p-2)\alpha - pq(p-1) + \sqrt{((p-2)\alpha + pq(p-1))^2 + 4(p-1)\alpha^2}}{-2(p-1)\alpha} \\ p\mathbf{\Sigma}^{-1} = \mathrm{BTr}_q\left[(\mathbf{\Psi}_{CS} \otimes \mathbf{I}_q)\,\mathbf{S}^{-1}\right]. \end{cases} \tag{3.16}$$

Since the first equation in (3.16) depends on $\mathbf{\Sigma}$ through α, and since the second one depends on ϱ through $\mathbf{\Psi}_{CS}$, one possibility of finding the solution could be the iterative flip-flop algorithm. In the next proposition we show that $\widetilde{\varrho} \in (-1/(p-1), 1)$, and thus the solution of (3.16) gives the ELEs of ϱ and $\mathbf{\Sigma}$ for which $\widetilde{\mathbf{\Psi}}_{CS}$ and $\widetilde{\mathbf{\Sigma}}$ are positive definite.

PROPOSITION 3.1. *If $\mathbf{S}$ and $\mathbf{\Sigma}$ are symmetric positive definite matrices, then $\widetilde{\mathbf{\Psi}}_{CS}$ is also positive definite. Similarly, if $\mathbf{S}$ and $\mathbf{\Psi}_{CS}$ are symmetric positive definite matrices, then $\widetilde{\mathbf{\Sigma}}$ is also positive definite.*

Proof. The necessary condition for $\mathbf{\Psi}_{CS}$ being positive definite is that $\varrho \in \left(-\frac{1}{p-1}; 1\right)$. Therefore it is enough to show that the square polynomial $h(\varrho)$ has different signs at $\varrho = -1/(p-1)$ and $\varrho = 1$. We obtain

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 83-98, November 2018.

$h\left(-\frac{1}{p-1}\right) = -pq < 0$ and $h(1) = pq(p-1) > 0$, which complete the proof of the first part of the theorem. The second part follows from positive definiteness of the block trace operator of positive definite matrix; cf. Filipiak et al. [7, Lemma 2.4]. □

We now use similar transformation of the entropy loss function (1.6) as the log-likelihood function (1.3) by applying the property $\mathrm{tr}[\mathrm{BTr}_q \mathbf{X}] = \mathrm{tr}\,\mathbf{X}$, to obtain the direct solution of ELE, analogous to MLE. For $\mathbf{\Psi} = \mathbf{\Psi}_{CS}$ we have

$$f(\mathbf{\Psi}_{CS}, \mathbf{\Sigma}; \mathbf{S}) \quad = \quad \mathrm{tr}\left[\mathbf{\Sigma} \cdot \mathrm{BTr}_q\left[\mathbf{S}^{-1}(\mathbf{\Psi}_{CS} \otimes \mathbf{I}_q)\right]\right] - \ln|\mathbf{S}^{-1}(\mathbf{\Psi}_{CS} \otimes \mathbf{\Sigma})| - pq,$$

and using spectral decomposition of $\mathbf{\Psi}_{CS}$ given by (2.11) with notation

$$\mathbf{C} = \mathrm{BTr}_q\left[(\mathbf{P}_p \otimes \mathbf{I}_q)\mathbf{S}^{-1}\right], \qquad \mathbf{D} = \mathrm{BTr}_q\left[(\mathbf{Q}_p \otimes \mathbf{I}_q)\mathbf{S}^{-1}\right]$$

we can write

$$\begin{aligned} &f(\mathbf{\Psi}_{CS}, \mathbf{\Sigma}; \mathbf{S}) \\ &\quad = c_1 \,\mathrm{tr}\,(\mathbf{\Sigma}\mathbf{C}) + c_2 \,\mathrm{tr}\,(\mathbf{\Sigma}\mathbf{D}) + \ln|\mathbf{S}| - q\ln(c_1) - q(p-1)\ln(c_2) - p\ln|\mathbf{\Sigma}| - pq. \end{aligned} \tag{3.17}$$

Differentiating the above with respect to $\mathbf{\Sigma}$ and equating the result to zero gives

$$p\widetilde{\mathbf{\Sigma}}^{-1} = c_1\mathbf{C} + c_2\mathbf{D} \tag{3.18}$$

and finally

$$\widetilde{\mathbf{\Sigma}} = p\,\mathrm{BTr}_q^{-1}\left[(\mathbf{\Psi}_{CS} \otimes \mathbf{I}_q)\,\mathbf{S}^{-1}\right] \tag{3.19}$$

where $\mathrm{BTr}_q^{-1}(\bullet)$ denotes the inverse of $\mathrm{BTr}_q(\bullet)$, and it exists as the block trace operator of positive definite matrix preserves positive definiteness (cf. Filipiak et al. [7, Lemma 2.4]).

Let us differentiate (3.17) with respect to ϱ. Equating the result to zero gives

$$(p-1)\,\mathrm{tr}(\mathbf{\Sigma}\mathbf{C}) - \mathrm{tr}(\mathbf{\Sigma}\mathbf{D}) - \frac{q(p-1)}{c_1} + \frac{q(p-1)}{c_2} = 0.$$

Plugging (3.19) into the above equality we finally get

$$q(p-1)\varrho + c_1c_2(p-1)\,\mathrm{tr}\left[(c_1\mathbf{C} + c_2\mathbf{D})^{-1}\mathbf{C}\right] - c_1c_2\,\mathrm{tr}\left[(c_1\mathbf{C} + c_2\mathbf{D})^{-1}\mathbf{D}\right] = 0. \tag{3.20}$$

Since (3.20) does not depend on $\mathbf{\Sigma}$, the solution $\widetilde{\varrho}$ of (3.20) gives us the ELE of ϱ directly. It can be observed, that $\widetilde{\varrho} \in (-1/(p-1), 1)$. Indeed, if we denote the left hand side of (3.20) by $\ell(\varrho)$, we obtain $\ell(-\frac{1}{p-1}) = -q < 0$ and $\ell(1) = q(p-1) > 0$, and thus at least one solution of $\ell(\varrho) = 0$ belongs to the desired interval. To get the ELE of $\mathbf{\Sigma}$ it is enough to plug $\widetilde{\varrho}$ into (3.19). It is worth noting that there might be several roots of (3.20) in the desired interval. Then, the one which minimizes $f(\widetilde{\mathbf{\Psi}}_{CS}, \widetilde{\mathbf{\Sigma}}; \mathbf{S})$ is chosen.

It is easy to see that the second equation of (3.16) and (3.18) represent the same estimator of $\mathbf{\Sigma}$ (positive definite). Moreover, using the property (2.9) and Lemma A.3 from Appendix A for α defined with (3.16), and then simplifying the first equation of (3.16) we get (3.20).

It is worth to note, that similarly as in MLE case, there is no iterative procedure in this approach, and even if (3.20) is nonlinear, this technique is much less time consuming. Indeed, if we compare by simulations

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 83-98, November 2018.

the time necessary to determine the ELEs of ϱ and $\mathbf{\Sigma}$ using iterative procedure (2.8) and using the direct procedure solving (3.20) and (3.19), this observation is confirmed by Figure 2 (the black and red lines, respectively, for iterative and direct approach).

Finally, we present another iterative procedure, which is surprisingly for small p better than the direct approach described above.

Let the spectral decomposition of a matrix $\mathbf{\Psi}_{CS}$ be $\mathbf{UGU}^\top$, where $\mathbf{U}$ is a $p \times p$ orthogonal matrix with columns corresponding to the orthogonalized eigenvectors of $\mathbf{\Psi}_{CS}$. It is known that $\mathbf{U}$ does not depend on ϱ and $\mathbf{G} = \operatorname{diag}(1+(p-1)\varrho, 1-\varrho, \ldots, 1-\varrho)$. Denoting $\mathbf{\Lambda} = (\mathbf{U}^\top \otimes \mathbf{I}_q)\mathbf{S}^{-1}(\mathbf{U} \otimes \mathbf{I}_q)$ we can write $\mathbf{\Lambda}$ as a block matrix $(\mathbf{\Lambda}_{ij})_{1\le i,j\le p}$. Then, the entropy loss function (1.6) can be rewritten as

$$\begin{aligned} f(\mathbf{\Psi}_{CS}, \mathbf{\Sigma}; \mathbf{\Lambda}) &= \operatorname{tr}[\mathbf{\Lambda}(\mathbf{G} \otimes \mathbf{\Sigma})] - \ln|\mathbf{G} \otimes \mathbf{\Sigma}| - \ln|\mathbf{\Lambda}| - pq \\ &= \operatorname{tr}\left[(1+(p-1)\varrho)\mathbf{\Lambda}_{11}\mathbf{\Sigma} + \sum_{i=2}^{p}(1-\varrho)\mathbf{\Lambda}_{ii}\mathbf{\Sigma}\right] \\ &\quad -q\ln(1+(p-1)\varrho) - (p-1)q\ln(1-\varrho) - p\ln|\mathbf{\Sigma}| - \ln|\mathbf{\Lambda}| - pq \\ &= \operatorname{tr}\left[(p\varrho\mathbf{\Lambda}_{11} + (1-\varrho)\operatorname{BTr}_q\mathbf{\Lambda})\mathbf{\Sigma}\right] \\ &\quad -q\ln(1+(p-1)\varrho) - (p-1)q\ln(1-\varrho) - p\ln|\mathbf{\Sigma}| - \ln|\mathbf{\Lambda}| - pq. \end{aligned}$$

To get respective derivatives of the above function with respect to ϱ and $\mathbf{\Sigma}$ we use again the matrix differentiation formulas described in Magnus and Neudecker [12] or Fackler [3]. We obtain

$$\begin{cases} \dfrac{\partial f}{\partial \varrho} = \operatorname{tr}[(p\mathbf{\Lambda}_{11} - \operatorname{BTr}_q\mathbf{\Lambda})\mathbf{\Sigma}] - \dfrac{q(p-1)}{1+(p-1)\varrho} + \dfrac{q(p-1)}{1-\varrho} \\ \dfrac{\partial f}{\partial \Sigma} = \operatorname{vec}^\top(p\varrho\mathbf{\Lambda}_{11} + (1-\varrho)\operatorname{BTr}_q\mathbf{\Lambda}) - p\cdot\operatorname{vec}^\top\mathbf{\Sigma}^{-1} \end{cases}$$

and, denoting $\operatorname{tr}[(p\mathbf{\Lambda}_{11} - \operatorname{BTr}_q\mathbf{\Lambda})\mathbf{\Sigma}]$ by β,

$$\begin{cases} -(p-1)\beta\cdot\varrho^2 + [(p-2)\beta - pq(p-1)]\cdot\varrho + \beta = 0 \\ \mathbf{\Sigma}^{-1} = \frac{1}{p}\left[p\varrho\mathbf{\Lambda}_{11} + (1-\varrho)\operatorname{BTr}_q\mathbf{\Lambda}\right]. \end{cases}$$

Let us denote $-(p-1)\beta\cdot\varrho^2 + [(p-2)\beta - pq(p-1)]\cdot\varrho + \beta$ by $u(\varrho)$. This function has two roots, as for $p \ge 2$ its discriminant is always positive. If $\beta < 0$ then the polynomial $u(\varrho)$ is convex and the solution $\widetilde{\varrho}$ is determined by a bigger root of $u(\varrho)$, while for $\beta > 0$, by a smaller root of $u(\varrho)$. Moreover, $\widetilde{\varrho} \in (-1/(p-1), 1)$, since $u(-\frac{1}{p-1}) = pq > 0$ and $u(1) = -pq(p-1) < 0$. In both cases $\widetilde{\varrho}$ has the same form and the best approximation of $\mathbf{S}$ can be found by solving

$$\begin{cases} \varrho = \dfrac{-(p-2)\beta + pq(p-1) + \sqrt{((p-2)\beta + pq(p-1))^2 + 4(p-1)\beta^2}}{-2(p-1)\beta} \\ \mathbf{\Sigma} = p\left[p\varrho\mathbf{\Lambda}_{11} + (1-\varrho)\operatorname{BTr}_q\mathbf{\Lambda}\right]^{-1}. \end{cases} \tag{3.21}$$

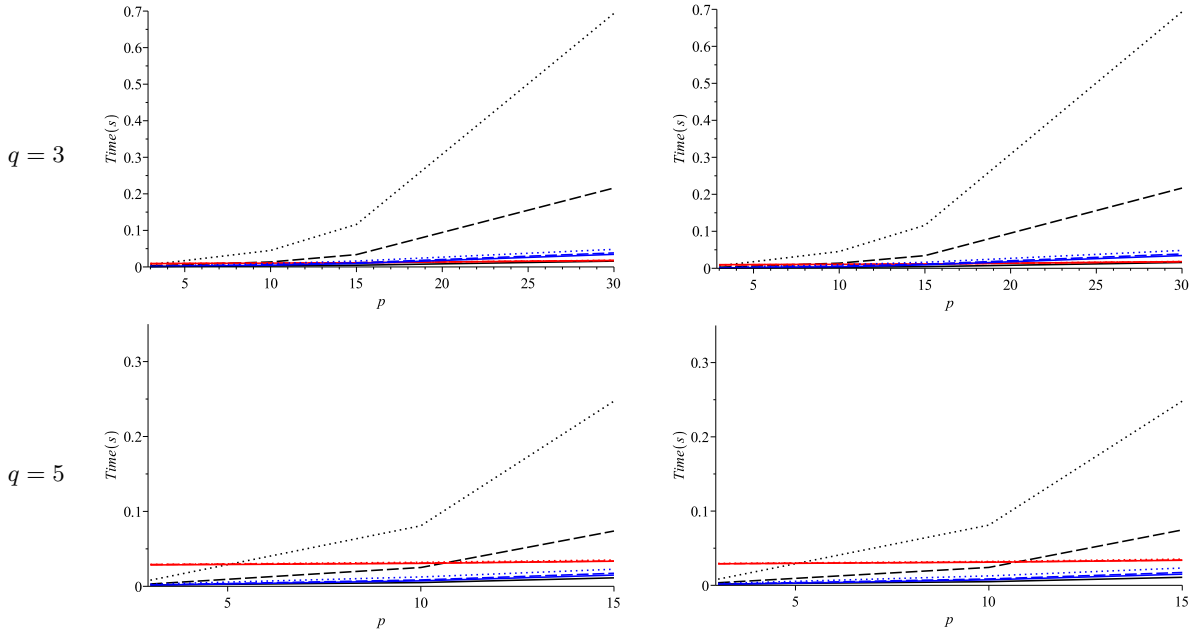
Observe, that (3.21) can be solved iteratively using flip-flop algorithm. This numerical solution gives us ELEs of ϱ and $\mathbf{\Sigma}$.

To compare the time necessary to determine the ELEs of ϱ and $\mathbf{\Sigma}$ using standard iterative ("iterative") procedure (3.16), direct ("direct") procedure solving (3.20) and (3.19), and spectral decomposition ("spectral") procedure (3.21), simulation studies were carried out. We have generated 1000 samples of size $n = 100$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 83-98, November 2018.

from a pq multivariate normal distribution $N_{pq}(\mathbf{0}, \mathbf{\Psi}_{CS} \otimes \mathbf{\Sigma})$ for $p = 3, 10, 15, 30$ with $q = 3$ and $p = 3, 10, 15$ with $q = 5$, assuming $\varrho = 0, 0.5, 0.9$ and $\mathbf{\Sigma} = \mathbf{I}_q$ and $\mathbf{\Sigma}$ randomly generated as in previous section. The averaged time (in seconds) of calculating ELEs on the medium class computer was computed and the results are presented on Figure 2.

FIGURE 2. *The averaged time of calculating ELEs using three approaches (black color for iterative procedure, red color for direct, and blue for spectral) with respect to p, ϱ and $\mathbf{\Sigma}$: $\varrho = 0$ – solid line, $\varrho = 0.5$ – dashed line, $\varrho = 0.9$ – dotted line; $\mathbf{\Sigma} = \mathbf{I}_q$ – left column, $\mathbf{\Sigma}$ randomly generated – right column.*



As expected, it can be seen from Figure 2 that the time of determining ELEs depends on the parameters p and q for all the procedures (for $q = 5$ the computing time is slightly longer than for $q = 3$), and in the case of the direct approach it does not depend on ϱ (red lines overlap with each other), whilst for spectral procedure the time changes only slightly with the change of ϱ. In all cases the time does not depend on $\mathbf{\Sigma}$. An interesting thing is that for $\varrho = 0$ the iterative procedure can be faster than the others. Moreover, for relatively small parameter p and $q = 3$, the spectral procedure can be shorter than the direct one, which also holds for every p if $q = 5$.

4. MLE and ELE comparison. In the previous sections we showed, that the algorithms for determination of MLE and ELE are quite similar, which can be well seen especially in direct algorithms. In this section we compare the properties of the solutions of the algorithms, that is the properties of MLEs and ELEs of ϱ and $\mathbf{\Sigma}$. The first conclusion which can be seen from Figures 1 and 2 is that the time necessary to obtain ELEs using direct approach is shorter than respective time necessary to obtain MLEs. To compare statistical properties of the estimators, we perform simulation studies for the parameters (p, q) as $(3, 3)$, $(10, 3)$, $(15, 3)$, $(30, 3)$ and $(3, 5)$, $(10, 5)$, $(15, 5)$. Samples of size $n = 100$ are generated 1000 times from a pq-variate normal population $N_{pq}(0; \mathbf{\Psi}_{CS} \otimes \mathbf{I}_q)$. We assume various values of ϱ in CS structure.

Table 1 contains the bias (B) of $\widehat{\varrho}$ and $\widetilde{\varrho}$, as well as their mean square errors (MSEs). It can be seen

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 83-98, November 2018.

that the MLEs are a little less biased than ELEs and similarly the MSE of $\widehat{\varrho}$ is usually smaller than MSE of $\widetilde{\varrho}$. The superiority of MLE over ELE can be seen especially for bigger values of p. From Table S.1 in Supplementary Material it can be noticed, that the variability of both estimators is comparable.

TABLE 1
Bias and MSE of the estimators of ϱ for different values of p, q

		$q=3$				$q=5$			
		MLE		ELE		MLE		ELE	
p	ϱ	B($\widehat{\varrho}$)	MSE($\widehat{\varrho}$)	B($\widetilde{\varrho}$)	MSE($\widetilde{\varrho}$)	B($\widehat{\varrho}$)	MSE($\widehat{\varrho}$)	B($\widetilde{\varrho}$)	MSE($\widetilde{\varrho}$)
3	0	0.0010	0.0011	0.0022	0.0012	-0.0004	0.0007	0.0004	0.0008
3	0.5	-0.0001	0.0011	0.0010	0.0012	-0.0011	0.0007	-0.0004	0.0008
3	0.9	-0.0003	0.0001	0.0000	0.0001	-0.0005	0.0001	-0.0003	0.0001
10	0	-0.0001	0.0001	0.0006	0.0001	0.0004	0.0000	0.0009	0.0001
10	0.5	-0.0012	0.0006	0.0004	0.0008	0.0007	0.0003	0.0014	0.0007
10	0.9	-0.0006	0.0001	-0.0002	0.0001	0.0001	0.0000	0.0002	0.0001
15	0	-0.0002	0.0000	0.0004	0.0001	0.0003	0.0000	0.0014	0.0001
15	0.5	-0.0015	0.0005	0.0000	0.0010	0.0006	0.0003	0.0030	0.0013
15	0.9	-0.0008	0.0001	-0.0005	0.0001	0.0001	0.0000	0.0004	0.0002
30	0	0.0000	0.0000	0.0022	0.0001	—	—	—	—
30	0.5	-0.0011	0.0004	0.0074	0.0048	—	—	—	—
30	0.9	-0.0006	0.0001	-0.0001	0.0007	—	—	—	—

In the case of estimation of $\mathbf{\Sigma}$, the bias of $\widehat{\mathbf{\Sigma}}$ and $\widetilde{\mathbf{\Sigma}}$ as well as their variances and MSEs, and the standard deviations of the estimators (for every element of the estimator of $\mathbf{\Sigma}$ separately) were computed. The results for expectation and standard deviation are presented in Tables S.2 – S.5 in Supplementary Material. Recall that if we denote by $\mathbf{\Theta}$ the estimator of the unknown parameter $\mathbf{\Sigma}$ (where $\mathbf{\Theta}$ is the MLE or ELE of $\mathbf{\Sigma}$), then $\mathrm{B}(\mathbf{\Theta}) = \mathrm{E}(\mathbf{\Theta}) - \mathbf{\Sigma}$, $\mathrm{Var}(\mathbf{\Theta}) = \mathrm{E}\{\mathrm{vec}[\mathbf{\Theta} - \mathrm{E}(\mathbf{\Theta})]\,\mathrm{vec}^T[\mathbf{\Theta} - \mathrm{E}(\mathbf{\Theta})]\}$, and $\mathrm{MSE}(\mathbf{\Theta}) = E[\mathrm{vec}(\mathbf{\Theta} - \mathbf{\Sigma})\,\mathrm{vec}^T(\mathbf{\Theta} - \mathbf{\Sigma})]$, whilst $\mathrm{SD}(\mathbf{\Theta})$ is calculated separately for every element of $\mathbf{\Theta}$. We do not present variance and MSE of the estimators, as they are $q^2 \times q^2$ matrices, however, these values are available from the authors on the request. Observe, that there exists a lot of possible methods of comparison of two covariance matrices, e.g. total variability measure or generalized variance, nevertheless, since we want to compare variance and MSE matrices for two estimators instead of the estimators itself, we use Frobenius norm. On Figures 3, 4 and 5 the values of, respectively, Frobenius norm of $B(\widehat{\mathbf{\Sigma}})$ and $B(\widetilde{\mathbf{\Sigma}})$, $\mathrm{Var}(\widehat{\mathbf{\Sigma}})$ and $\mathrm{Var}(\widetilde{\mathbf{\Sigma}})$ as well as Frobenius norm of $\mathrm{MSE}(\widehat{\mathbf{\Sigma}})$ and $\mathrm{MSE}(\widetilde{\mathbf{\Sigma}})$, are presented.

From Figure 3 it can be noticed, that the solid, dashed and dotted lines overlap in both of the cases, which means that the bias of the estimators does not depend on the true value of ϱ. Moreover, the bias of $\widetilde{\mathbf{\Sigma}}$ is greater than the bias of $\widehat{\mathbf{\Sigma}}$, and increases quickly with the increase of p and q, whilst in the case of MLE it remains on the same level. It can be observed from Figure 4 that the variability of MLEs are greater than ELEs, and this difference increases with the increase of p and q. From Figure 5 it can be noticed, that the solid, dashed and dotted lines overlap in both of the cases, which means that the MSEs of the estimators do not depend on the true value of ϱ. Moreover, the MSE of $\widetilde{\mathbf{\Sigma}}$ is greater than MSE of $\widehat{\mathbf{\Sigma}}$, and increases quickly with the increase of p and q, whilst in the case of MLE it remains on the same level.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 83-98, November 2018.

FIGURE 3. *Frobenius norm of bias of MLE (black line) and ELE (red line) with respect to p, q and ϱ: q = 3 - left column, q = 5 - right column; ϱ = 0 - solid line, ϱ = 0.5 - dashed line, ϱ = 0.9 - dotted line.*

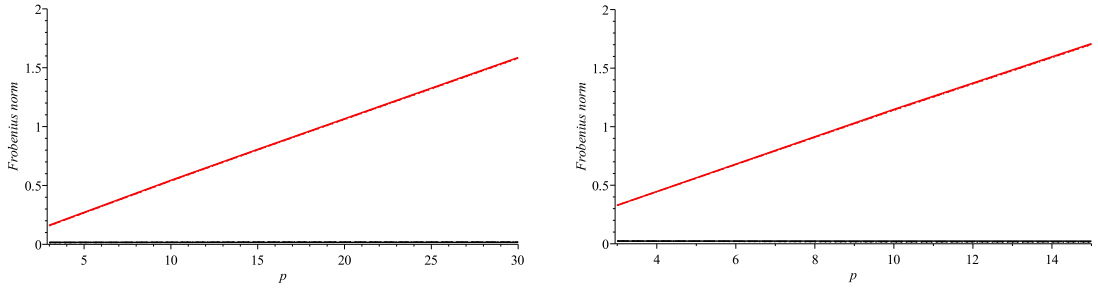


FIGURE 4. *Frobenius norm of* $\mathrm{Var}(\widehat{\boldsymbol{\Sigma}})$ *(black line) and of* $\mathrm{Var}(\widetilde{\boldsymbol{\Sigma}})$ *(red line) with respect to p, q and ϱ: q = 3 – left column, q = 5 – right column; ϱ = 0 – solid line, ϱ = 0.5 – dashed line, ϱ = 0.9 – dotted line.*

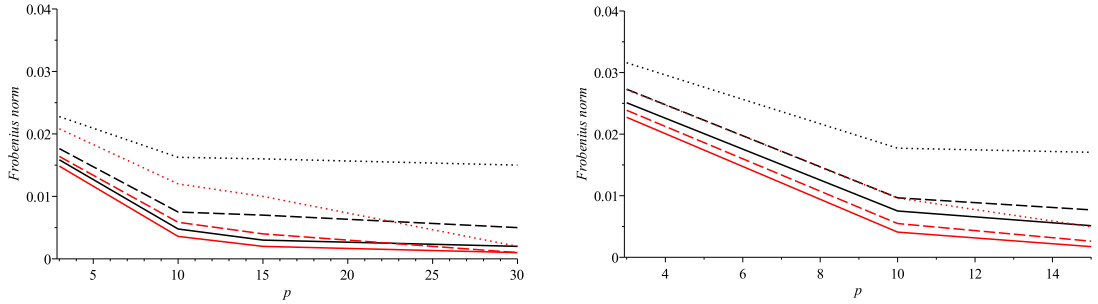
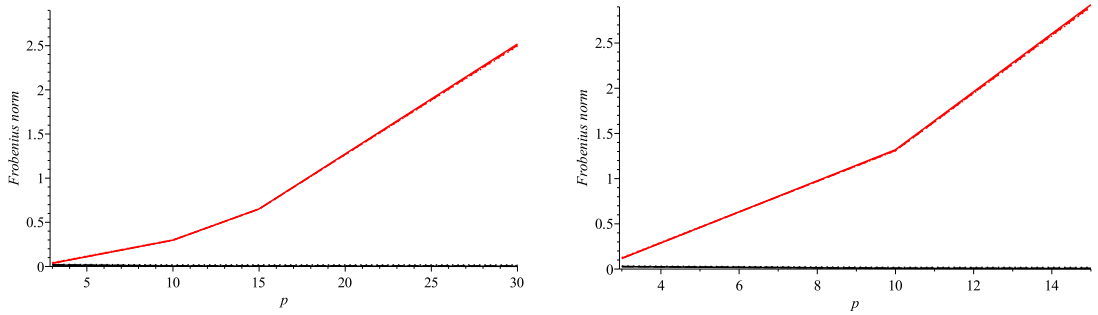


FIGURE 5. *Frobenius norm of* $\mathrm{MSE}(\widehat{\boldsymbol{\Sigma}})$ *(black line) and of* $\mathrm{MSE}(\widetilde{\boldsymbol{\Sigma}})$ *(red line) with respect to p, q and ϱ: q = 3 – left column, q = 5 – right column; ϱ = 0 – solid line, ϱ = 0.5 – dashed line, ϱ = 0.9 – dotted line.*



Finally, we compare the MLE and ELE with respect to entropy loss as the alternative to MSE. Observe, that the entropy loss function (1.5), which is a Kullback-Leibler divergence between two multivariate normal distributions which differ in covariance matrices, requires the true $\boldsymbol{\Omega}$ to be known. However, since this matrix is unknown, we replaced it by its estimator, $\mathbf{S}$, and then we determine the minimum of (1.6). Therefore, the values of loss function given in (1.6) must be smaller for ELE than for MLE, which is confirmed in Table 2 (column denoted by f_S) and on Figure 6 (second column). However, there is no such rule for the values of (1.5). In the column denoted by f_Ω in Table 2 and on Figure 6 (first column) the values of (1.5) for MLE

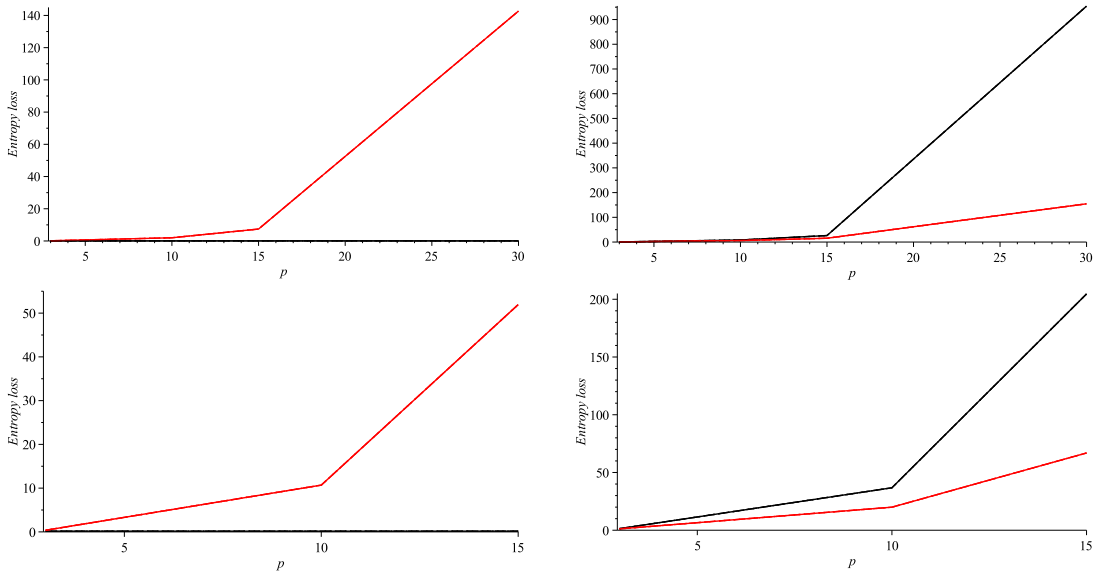
Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 83-98, November 2018.

and ELE are presented. Since the entropy loss function is invariant with respect to the linear transformation of the data (cf. James and Stein [8]), in particular it has the same value for every ϱ in $\mathbf{\Psi}_{CS}$ structure.

TABLE 2
Loss of MLE and ELE with respect to $\mathbf{\Omega} = \mathbf{\Psi}_{CS} \otimes \mathbf{\Sigma}$ *(*f_Ω*) or* $\mathbf{S}$ *(*f_S*) for different values of* p *and* q

	$q = 3$				$q = 5$			
	MLE		ELE		MLE		ELE	
p	f_Ω	f_S	f_Ω	f_S	f_Ω	f_S	f_Ω	f_S
3	0.0697	0.4556	0.1194	0.4101	0.1646	1.3779	0.3731	1.1626
10	0.0707	8.2433	1.9932	5.8713	0.1605	36.7526	10.6878	19.9532
15	0.0714	26.2225	7.3970	15.3668	0.1648	204.5887	51.9409	66.8870
30	0.0728	954.1915	142.6975	154.6277	—	—	—	—

FIGURE 6. *Entropy loss function for MLEs (black line) and ELEs (red line) using either true* $\mathbf{\Omega}$ *(left column) or* $\mathbf{S}$ *(right column) for* $q = 3$ *(first row) and* $q = 5$ *(second row).*



It can be seen from Table 2 and Figure 6 that the values of f_Ω for MLEs are quite stable for a given q, whilst for ELE they increase rapidly with the increase of either p or q. This difference can be explained by the fact of replacing the true $\mathbf{\Omega}$ in (1.5) by $\mathbf{S}$, which does not need to be a good estimator of structured covariance matrix, especially if n is not much greater than pq. The problem of using better estimators of structured $\mathbf{\Omega}$ will be considered in the future work.

5. Conclusions. If we are interested in the estimation of the separable covariance structures with one component structured as CS matrix, there is no such simple relation as in the case of separable covariance structure with both components unstructured (compare (1.4) and (1.7)). We can see only some connection when we compare solutions of (2.14) and (2.15) for MLEs with (3.19) and (3.20) for ELEs.

From Figures 1 and 2 we can conclude that the algorithms obtained by preliminary transformation of

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 83-98, November 2018.

likelihood function or entropy loss function are usually much faster than algorithms obtained by differentiation of (1.3) or (1.6). The only exception can appear if true ϱ is equal to 0. Moreover, in both cases the time necessary to get the estimates using direct procedure does not depend on the true $\mathbf{\Sigma}$, but only in the case of ELE the iterative procedure does not depend on the true $\mathbf{\Sigma}$ too.

From Table 1, Figure 3 and Tables S.1, S.2 and S.4 of Supplementary Material it can be seen, that the MLEs of ϱ and $\mathbf{\Sigma}$ are less biased than respective ELEs. Moreover, from Tables 1 and 2 as well as from Figures 5 and 6 it can be noticed, that the MLEs have much smaller MSE and loss than ELEs. It means, that the MLE surpasses ELE with respect to the biasedness, mean square errors, as well as the value of loss. Moreover, MLE is less demanding from the point of view of sample size: entropy loss function (1.6) requires positive definiteness of $\mathbf{S}$, i.e. the sample size bigger than pq, whilst in the case of MLE it is enough to have $n > p/q + q/p + 1$ (cf. Soloveychik and Trushin [20]). The only advantage of ELE is its stability, which can be observed on Figure 4 and Tables S.1, S.3 and S.5 of Supplementary Material.

Summing up we can conclude, that the ELE of structured covariance matrix is not a good estimator, however, its determination is necessary from the point of view of prespecification of the covariance structure. In such a problem it is worth to use direct algorithm to determine ELE and the minimum of the entropy loss function, instead of iterative procedures based on direct differentiation of (1.6) or on the spectral decomposition of CS structure, especially for large p and q. Finally, it can be observed from Figures 1 and 2 that the time necessary to obtain ELEs using direct approach is shorter than respective time necessary to obtain MLEs.

In this paper we assumed normality of the data and it can be easily seen that the MLE of covariance structure is strongly connected with the distribution requirement. The superiority of the MLE over ELE, however, can change if the normality assumption is not satisfied. The extended comparison of the estimators for non-Gaussian observations will be considered in the future work.

Acknowledgments. The authors thank the Reviewer for all the constructive comments, which significantly improved this paper. The authors also thank the Mathematical Research and Conference Center of the Polish Academy of Sciences, Będlewo, Poland, for providing the opportunity and support for this paper.
This research is partially supported by Statutory Activities No. 04/43/DSPB/0101 (Katarzyna Filipiak) and by the Slovak Research and Development Agency under the contract no. APVV-17-0568 and grant VEGA MŠ SR 1/0311/18 (Daniel Klein).

REFERENCES

[1] X. Cui, C. Li, J. Zhao, L. Zeng, D. Zhang, J. Pan. Covariance structure regularization via Frobenius norm discrepancy. *Linear Algebra Appl.*, 510:124–145, 2016.
[2] D.K. Dey and C. Srinivasan. Estimation of a covariance matrix under Stein's loss. *Ann. Statist.*, 13(4):1581–1591, 1985.
[3] P.L. Fackler. Notes on matrix calculus. http://www4.ncsu.edu/~pfackler/MatCalc.pdf, 2005.
[4] K. Filipiak and D. Klein. Approximation with a Kronecker product structure with one component as compound symmetry or autoregression. *Linear Algebra Appl.*, 559:11–33, 2018.
[5] K. Filipiak, D. Klein, and A. Roy. Score test for a separable covariance structure with the first component as compound symmetric correlation matrix. *J. Multivariate Anal.*, 150:105–124, 2016.
[6] K. Filipiak, D. Klein, and A. Roy. A comparison of likelihood ratio tests and Rao's score test for three separable covariance matrix structures. *Biometrical J.*, 59:192–215, 2017.
[7] K. Filipiak, D. Klein, and E. Vojtková. The properties of partial trace and block trace operators of partitioned matrix. *Electron. J. Linear Algebra*, 33:3–15, 2018.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 83-98, November 2018.

[8] W. James and C. Stein. Estimation with quadratic loss. *In: Neyman, J. (ed.) Proceedings of the Fourth Berkeley Symposium. In: Mathematical Statistics and Probability, vol. 1. (pp. 361–379). The Statistical Laboratory, University of California, June 30–July 30, 1960.* University of California Press, 1961.

[9] T. Kollo and D. von Rosen. *Advanced Multivariate Statistics with Matrices.* Dordrecht, Springer, 2005.

[10] L. Lin, N. Higham, and J. Pan. Covariance structure regularization via entropy loss function. *Comput. Statist. Data Anal.*, 72:315–327, 2014.

[11] N. Lu and D. Zimmerman. The likelihood ratio test for a separable covariance matrix. *Statist. Probab. Lett.*, 73:449–457, 2005.

[12] J. Magnus and H. Neudecker. Symmetry, 0-1 matrices and Jacobians, a review. *Econom. Theory*, 2:157–190, 1986.

[13] A.M. Manceur and P. Dutilleul. Unbiased modified likelihood ratio tests for simple and double separability of a variance-covariance structure. *Statist. Probab. Lett.*, 83:631–636, 2013.

[14] M. Mitchell, M. Genton, and M. Gumpertz. A likelihood ratio test for separability of covariances. *J. Multivariate Anal.*, 97:1025–1043, 2006.

[15] A. Roy. A note on testing of Kronecker product covariance structures for doubly multivariate data. *Proceedings of American Statistics and Associations Statistical Computing Section (pp. 2157–2162).* Seattle, WA, 2007.

[16] A. Roy and R. Khattree. On implementation of a test for Kronecker product covariance structure for multivariate repeated measures data. *Stat. Methodol.*, 2:297–306, 2005.

[17] A. Roy and R. Khattree. Testing the hypothesis of a Kronecker product covariance matrix in multivariate repeated measures data. *Proceedings of the 30th Annual SAS Users Group International Conference (SUGI 30).* Philadelphia, PA, 2005.

[18] A. Roy and R. Leiva. Likelihood ratio tests for triply multivariate data with structured correlation on spatial repeated measurements. *Statist. Probab. Lett.*, 78:1971–1980, 2008.

[19] S.L. Simpson. An adjusted likelihood ratio test for separability in unbalanced multivariate repeated measures data. *Statistics and Methodology*, 7:511–519, 2010.

[20] I. Soloveychik and D. Trushin. Gaussian and robust Kronecker product covariance estimation: Existence and uniqueness. *J. Multivariate Anal.*, 149:92–113, 2016.

[21] M. Srivastava, R. von Rosen, and D. von Rosen. Models with a Kronecker product covariance structure: estimation and testing. *Mathematics and Methods Statistics*, 17:357–370, 2008.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 83-98, November 2018.

Appendix A: Properties of partial trace and block trace operators.

LEMMA A.1. *(Filipiak et al. [7, Corollary 2.10]) For arbitrary matrices* $\mathbf{A} : mp \times mp$ *and* $\mathbf{B}, \mathbf{C} : p \times p$*, the following relations hold:*

$$\begin{aligned}\operatorname{vec}\left\{\operatorname{PTr}_p\left[(\mathbf{I}_m \otimes \mathbf{B})\mathbf{A}(\mathbf{I}_m \otimes \mathbf{C}^\top)\right]\right\} &= [\mathbf{I}_{m^2} \otimes \operatorname{vec}^\top(\mathbf{B}^\top\mathbf{C})](\mathbf{I}_m \otimes \mathbf{K}_{m,p} \otimes \mathbf{I}_p)\operatorname{vec}\mathbf{A}\\ \operatorname{vec}\left\{\operatorname{BTr}_m\left[(\mathbf{B} \otimes \mathbf{I}_m)\mathbf{A}(\mathbf{C}^\top \otimes \mathbf{I}_m)\right]\right\} &= [\operatorname{vec}^\top(\mathbf{B}^\top\mathbf{C}) \otimes \mathbf{I}_{m^2}](\mathbf{I}_p \otimes \mathbf{K}_{p,m} \otimes \mathbf{I}_m)\operatorname{vec}\mathbf{A}.\end{aligned}$$

LEMMA A.2. *(Filipiak et al. [7, Lemma 2.7]) For arbitrary matrices* $\mathbf{A} : mp \times mp$ *and* $\mathbf{B} : m \times m$*, the following relations hold:*

$$\begin{aligned}\operatorname{PTr}_m\left[(\mathbf{I}_p \otimes \mathbf{B})\mathbf{A}\right] &= \operatorname{PTr}_m\left[\mathbf{A}(\mathbf{I}_p \otimes \mathbf{B})\right]\\ \operatorname{BTr}_p\left[(\mathbf{B} \otimes \mathbf{I}_p)\mathbf{A}\right] &= \operatorname{BTr}_p\left[\mathbf{A}(\mathbf{B} \otimes \mathbf{I}_p)\right]\end{aligned}$$

LEMMA A.3. *(Filipiak et al. [7, Lemma 2.11]) For arbitrary matrices* $\mathbf{A} : mp \times mp$ *and* $\mathbf{B}, \mathbf{C} : m \times n$*, the following relations hold:*

$$\begin{aligned}\operatorname{PTr}_p[(\mathbf{B}^\top \otimes \mathbf{I}_p)\mathbf{A}(\mathbf{C} \otimes \mathbf{I}_p)] &= \mathbf{B}^\top \cdot \operatorname{PTr}_p(\mathbf{A}) \cdot \mathbf{C},\\ \operatorname{BTr}_n[(\mathbf{I}_p \otimes \mathbf{B}^\top)\mathbf{A}(\mathbf{I}_p \otimes \mathbf{C})] &= \mathbf{B}^\top \cdot \operatorname{BTr}_m(\mathbf{A}) \cdot \mathbf{C}.\end{aligned}$$

Appendix B: ELE of $\boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}$. For an arbitrary given positive definite $\mathbf{S}$ the entropy loss function has the form

$$f(\boldsymbol{\Psi}, \boldsymbol{\Sigma}; \mathbf{S}) = \operatorname{tr}\left[\mathbf{S}^{-1}(\boldsymbol{\Psi} \otimes \boldsymbol{\Sigma})\right] - \ln\left|\mathbf{S}^{-1}(\boldsymbol{\Psi} \otimes \boldsymbol{\Sigma})\right| - pq.$$

In order to find the best approximation, we differentiate $f(\boldsymbol{\Psi}, \boldsymbol{\Sigma}; \mathbf{S})$ with respect to $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$. Using the differentiation rules for symmetric matrices presented by Magnus and Neudecker [12] or Fackler [3], we obtain

$$\begin{cases}\dfrac{\partial f}{\partial \boldsymbol{\Psi}} &= \operatorname{vec}^\top\left(\mathbf{S}^{-1} - \boldsymbol{\Psi}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)(\mathbf{I}_p \otimes \mathbf{K}_{q,p} \otimes \mathbf{I}_q)\left(\mathbf{I}_{p^2} \otimes \operatorname{vec}\boldsymbol{\Sigma}\right) \cdot \mathbf{D}_p\\ \dfrac{\partial f}{\partial \boldsymbol{\Sigma}} &= \operatorname{vec}^\top\left(\mathbf{S}^{-1} - \boldsymbol{\Psi}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)(\mathbf{I}_p \otimes \mathbf{K}_{q,p} \otimes \mathbf{I}_q)\left(\operatorname{vec}\boldsymbol{\Psi} \otimes \mathbf{I}_{q^2}\right) \cdot \mathbf{D}_q.\end{cases}$$

Equating the above derivatives to zero, we get the following system of equations

$$\begin{cases}\left(\mathbf{I}_{p^2} \otimes \operatorname{vec}^\top\boldsymbol{\Sigma}\right)(\mathbf{I}_p \otimes \mathbf{K}_{p,q} \otimes \mathbf{I}_q)\operatorname{vec}\left(\mathbf{S}^{-1} - \boldsymbol{\Psi}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) &= 0\\ \left(\operatorname{vec}^\top\boldsymbol{\Psi} \otimes \mathbf{I}_{q^2}\right)(\mathbf{I}_p \otimes \mathbf{K}_{p,q} \otimes \mathbf{I}_q)\operatorname{vec}\left(\mathbf{S}^{-1} - \boldsymbol{\Psi}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) &= 0.\end{cases}$$

Since the duplication matrix does not have any effect on the solution of the above system of equations, it is just omitted. Using the properties of partial trace and block trace operators described in Lemma A.1 and A.3 from Appendix A and formula (8) from Magnus and Neudecker [12] we obtain (1.7).