

Uniwersytet Wrocławski

Estimation of a block-structured covariance matrix

Malwina Mrowińska

Praca doktorska przygotowana
w Instytucie Matematyki
Politechniki Poznańskiej
pod kierunkiem
dr hab. inż. Katarzyny Filipiak
oraz promotora pomocniczego
dr inż. Moniki Mokrzyckiej
i przedłożona jako rozprawa doktorska
Radzie Naukowej dyscyplin
matematyka i informatyka
Uniwersytetu Wrocławskiego

Wrocław 2026

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof. Katarzyna Filipiak, and my auxiliary supervisor, Dr. Monika Mokrzycka, for their scientific guidance, invaluable support, and continuous encouragement throughout the preparation of this dissertation. I am especially grateful not only for their knowledge, experience, and countless insightful discussions, but also for their kindness, patience, and exceptionally supportive approach, which created an atmosphere in which I could develop both scientifically and personally. Their trust, openness, and willingness to share their time and ideas have had a profound influence on this work and on my development as a researcher.

I would also like to express my heartfelt gratitude to Prof. Augustyn Markiewicz, who introduced me to the field of covariance matrix estimation and invited me to participate in the seminars on Linear Algebra and Its Applications. I am deeply grateful for the time, knowledge, and experience he shared with me, as well as for many valuable discussions which significantly influenced my scientific development and shaped my interest in multivariate statistics. It was through these seminars and collaborations that my research on block-structured covariance matrices gradually began to develop.

I am also grateful to the members of the research group participating in the regular “Multivariate and Mixed Linear Models” scientific meetings in Będlewo for many inspiring discussions, valuable remarks, and for creating an atmosphere of openness and scientific curiosity. The opportunity to exchange ideas in such an environment has been an important part of my academic development.

Finally, I would like to thank my family, friends, and especially my husband for their constant support, patience, and belief in me throughout this journey. Their encouragement and understanding were an invaluable source of strength during the most demanding stages of this work, even if the subject of my research was not always easy to understand outside the scientific world.

Contents

I	Research Overview	8
1	Preliminaries	9
1.1	Model	9
1.2	Covariance matrix structures	11
1.3	Estimation methods	15
1.4	Loss functions	18
1.5	Partitioned matrices, majorization and matrix derivatives	19
2	Main results – review	23
[A]	Block matrix approximation via entropy loss function	23
[B]	Quasi shrinkage estimation of a block-structured covariance matrix	26
[C]	Block covariance matrix estimation with structured off-diagonal blocks	30
[D]	Characterization of quadratic subspaces of partitioned matrices	37
[E]	On estimation of partitioned covariance matrix with linearly structured blocks . .	41
	Bibliography	45
II	Main results – full text of articles	48
[A]	Block matrix approximation via entropy loss function	49
[B]	Quasi shrinkage estimation of a block-structured covariance matrix	66
[C]	Block covariance matrix estimation with structured off-diagonal blocks	85
[D]	Characterization of quadratic subspaces of partitioned matrices	109
[E]	On estimation of partitioned covariance matrix with linearly structured blocks . .	147

Notation

$\mathbb{R}_{n \times m}$	space of $n \times m$ real matrices
$\mathbb{R}_m$	space of real matrices of order m
$\mathbb{R}_m^s$	space of symmetric matrices of order m
$\mathbb{R}_m^>$	cone of real symmetric positive definite matrices of order m
$\mathbb{R}^n$	space of n – dimensional real column vectors
$\mathbf{a}$	column vector $\mathbf{a} \in \mathbb{R}^n$
$\mathbf{A}$	matrix $\mathbf{A}$
$\mathbf{a}', \mathbf{A}'$	transpose of $\mathbf{a}$ and $\mathbf{A}$
$\mathbf{A}^{-1}$	inverse of $\mathbf{A}$
$\mathbf{1}_n$	n – dimensional vector of ones
$\mathbf{I}_n$	identity matrix of order n
$\mathbf{D}_n$	$n^2 \times n(n+1)/2$ duplication matrix
$\mathbf{K}_{mn}$	commutation matrix of order mn
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of two matrices
$\text{vec}(\mathbf{A})$	vectorization of $\mathbf{A}$
$\text{vech}(\mathbf{A})$	vectorization of a lower triangle of symmetric matrix $\mathbf{A}$
$\text{tr}(\mathbf{A})$	trace of $\mathbf{A}$
$\kappa(\mathbf{A})$	condition number of matrix $\mathbf{A}$
$\text{Bdiag}_{i=1}^k(\mathbf{A}_i)$	block-diagonal matrix consisting of k matrices $\mathbf{A}_i$ on the diagonal
$\dim(\mathcal{A})$	dimension of space $\mathcal{A}$
$\mathcal{A} \boxplus \mathcal{B}$	direct sum of the spaces $\mathcal{A}$ and $\mathcal{B}$
$\mathbf{P}_{\mathcal{A}}(\mathbf{A})$	projection of matrix $\mathbf{A}$ onto the space $\mathcal{A}$
$\mathbf{S} = \frac{1}{n-1} \mathbf{X}(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n') \mathbf{X}$	sample covariance matrix
$\mathbf{S}_{\text{ML}} = \frac{1}{n} \mathbf{X}(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n') \mathbf{X}$	maximum likelihood estimator of covariance matrix

Abbreviations

ML	Maximum Likelihood
LS	Least Squares
QS	Quasi Shrinkage
CS	Compound Symmetry
AR(1)	Autoregression of order one

A Brief Introduction

This doctoral dissertation addresses problems in multivariate statistics, focusing on the estimation of block structures of covariance matrices describing dependencies between sub-vectors of observations, corresponding to natural groupings of observed variables. These groups may arise, for example, from different feature types, hierarchical nature of the experiment, or repeated measurements across various conditions (e.g., time points, locations, depths). In all these cases, the partitioning of the vector of variables should be reflected in the covariance matrix, providing block structures. Furthermore, specific dependence structures within each sub-vector as well as between the sub-vectors, play a crucial role in the analysis of high-dimensional data, common in various fields of science, such as, e.g., agriculture, genetics, biology, economics, or engineering. In these settings, classical statistical methods often fail or become ineffective because the number of variables (the dimension of the observation vector) is larger than, comparable to, or grows together with the sample size. As a consequence, the sample covariance matrix often becomes singular or ill-conditioned. Importantly, imposing structure on the covariance matrix reduces the number of unknown parameters, thereby making estimation feasible in large- or high-dimensional settings.

We consider block-structured covariance matrices partitioned into four sub-blocks, with square symmetric (but otherwise unstructured) diagonal blocks of possibly different sizes. This extension to rectangular off-diagonal blocks (cross-covariance matrices) provides a more flexible framework for modeling covariance structures than classic approach with equal blocks dimensions, and broadens the range of potential applications in multivariate statistics and related fields.

We begin with covariance structures whose off-diagonal blocks have one of the following two forms: all elements are equal, that corresponds to a block of a compound symmetry covariance matrix, or the off-diagonal block is proportional to the corresponding sub-block of a first-order autoregression covariance structure. Subsequently, more general classes of partitioned matrices with structured diagonal and off-diagonal blocks are studied. In particular, the construction of the off-diagonal block structure depending on diagonal blocks, such that the whole covariance matrix belongs to the quadratic subspace, is of interest.

The first important issue, studied in this thesis, is the choice of the relevant covariance structure. In the literature, various methods have been considered to solve this problem, including graphical approaches and algebraic methods based on discrepancy functions. Janiszewska et al. (2020) [A] focus on the algebraic identification of block-structured covariance matrices using the entropy loss

function, which allows selecting the structure from a given set of candidates by minimizing the discrepancy between symmetric positive definite matrix and candidate block-structured matrices. Specifically, we consider two sets of structures described above: matrices with unstructured diagonal blocks and off-diagonal blocks corresponding to off-diagonal blocks of the compound symmetry or autoregression of order one structures.

The main goal of this dissertation is to develop computationally efficient methods for estimating block-structured covariance matrices. Proposed methods offer advantages over classical approaches, also discussed in the thesis, in terms of both structural simplicity and computational cost.

In the paper by Markiewicz et al. (2024) [B] the estimators of block-structured covariance matrices with off-diagonal blocks consisting of equal elements are proposed, whereas in Mokrzycka and Mrowińska (2025) [C] block structures whose off-diagonal blocks are proportional to the corresponding sub-block of an autoregression of order one covariance matrix are considered. First, classical maximum likelihood estimators of these two specific block structures are derived. Given the limitations of maximum likelihood estimators in terms of existence and computational feasibility, in Markiewicz et al. (2024) [B] and Mokrzycka and Mrowińska (2025) [C] the alternative estimation method based on the orthogonal projection of the maximum likelihood estimator of unstructured covariance matrix onto the space of the considered block structures is developed. Proposed projections, while not always preserving positive definiteness, are improved using shrinking technique and, if necessary, algebraic correction. It is worth noting that the classic shrinkage method is adapted here to estimate structured matrices, which is a novel approach. Furthermore, because of computational complexity of proposed estimators in case of off-diagonal block corresponding to autoregression of order one, in the latter paper, four additional estimation methods are presented and compared with respect to several loss functions.

In this dissertation a special emphasis is placed on quadratic subspaces, where, among other properties, the projection preserves positive definiteness. Moreover, the inverse of any matrix in a quadratic subspace preserves the same block structure, thereby allowing the estimation of both the covariance matrix and the precision matrix, which is particularly important in graphical models. For this purpose in Filipiak et al. (2025) [D] block-diagonal representation and block-spectral decomposition have been used to characterize a quadratic space and to present the projection of a symmetric matrix onto this space. Moreover, it is shown that projection does not increase the condition number. In addition, necessary and sufficient conditions under which a linear space of partitioned matrices with a nonzero off-diagonal block, forms a quadratic space, are established.

Finally, a block-spectral representation of partitioned matrices belonging to considered space is provided.

The quadratic spaces introduced by Filipiak et al. (2025) [D] are subsequently applied in Filipiak et al. (2025) [E] to the estimation of block-structured covariance matrices with structured blocks that do not belong to any quadratic subspace. It is shown that the optimal choice of a target space in the shrinkage estimation method is a quadratic space, and that, under normality, the resulting shrinkage estimator is consistent and asymptotically unbiased. It is also demonstrated how the choice of a quadratic target space influences the statistical and algebraic properties of the resulting estimator.

To achieve the objectives of this dissertation, statistical techniques, matrix analysis, and numerical methods are employed. The properties of the introduced estimators are compared through simulation studies. The proposed methods are further applied to a real-world data example.

Organization

This thesis is divided into two parts. The first part, containing two chapters, serves as an introduction to the topic of the dissertation and presents the review of the main results obtained in the following series of papers:

- [A] **Block matrix approximation via entropy loss function**
by Malwina Janiszewska, Augustyn Markiewicz, Monika Mokrzycka,
in Applications of Mathematics, 65(6), 829–844, 2020,
- [B] **Quasi shrinkage estimation of a block-structured covariance matrix**
by Augustyn Markiewicz, Monika Mokrzycka, Malwina Mrowińska,
in Journal of Statistical Computation and Simulation, 94(14), 3093–3110, 2024,
- [C] **Block covariance matrix estimation with structured off-diagonal blocks**
by Monika Mokrzycka, Malwina Mrowińska,
in Statistical Papers, 66, 129, 2025,
- [D] **Characterization of quadratic subspaces of partitioned matrices**
by Katarzyna Filipiak, Augustyn Markiewicz, Malwina Mrowińska,
submitted,
- [E] **On estimation of partitioned covariance matrix with linearly structured blocks**
by Katarzyna Filipiak, Augustyn Markiewicz, Adam Mieldzioc, Malwina Mrowińska
in Statistical Papers, 66, 98, 2025,

along with the main references.

In the second part of the dissertation the original articles are attached.

Part I

Research Overview

Chapter 1

Preliminaries

1.1 Model

Consider an experiment with n observations, each represented as an m -dimensional random vector $\mathbf{x}_i$, $i \in \{1, \dots, n\}$. Assume that the $\mathbf{x}_i$ are mutually independent and normally distributed with mean $\mathbb{E}(\mathbf{x}_i) = \boldsymbol{\mu}$, an m -dimensional vector, and covariance matrix $\mathbb{D}(\mathbf{x}_i) = \boldsymbol{\Omega}$, a symmetric positive definite (p.d.) matrix of order m . This implies that the observation matrix $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ has a matrix normal distribution, that is

$$\mathbf{X} \sim \mathcal{N}_{m,n}(\boldsymbol{\mu}\mathbf{1}'_n, \boldsymbol{\Omega}, \mathbf{I}_n), \quad (1.1)$$

where $\mathbf{1}_n$ is the n -dimensional vector of ones and $\mathbf{I}_n$ is the identity matrix of order n .

Suppose that each observation vector $\mathbf{x}_i$, $i \in \{1, \dots, n\}$, is related to vectorizing a (multi-indices) matrix of observations, and thus consists of multiple sub-vectors, corresponding to natural groupings of observed variables. These groups may arise, for example, from different feature types, the hierarchical nature of the experiment, or repeated measurements across various conditions (time points, locations, depths, etc.). In all these cases, the partitioning of $\mathbf{x}_i$ should be reflected in the covariance matrix, providing block structure.

Assuming that each $\mathbf{x}_i$, $i \in \{1, \dots, n\}$, consists of two sub-vectors, say $\mathbf{x}_i = (\mathbf{y}'_i, \mathbf{z}'_i)'$, where $\mathbf{y}_i \in \mathbb{R}_p$, $\mathbf{z}_i \in \mathbb{R}_q$, $p + q = m$, the expectation of $\mathbf{x}_i$ is partitioned accordingly, that is $\boldsymbol{\mu} = (\boldsymbol{\mu}'_y, \boldsymbol{\mu}'_z)'$, and the covariance matrix of $\mathbf{x}_i$, $\boldsymbol{\Omega}$, is a block matrix of the form

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_1 & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}'_{12} & \boldsymbol{\Omega}_2 \end{pmatrix}, \quad (1.2)$$

where $\mathbf{\Omega}_1 \in \mathbb{R}_p^s$, $\mathbf{\Omega}_{12} \in \mathbb{R}_{p \times q}$, $\mathbf{\Omega}_2 \in \mathbb{R}_q^s$. Matrices $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ are the covariance matrices of $\mathbf{y}_i$ and $\mathbf{z}_i$, respectively, while $\mathbf{\Omega}_{12}$ is the cross-covariance matrix between $\mathbf{y}_i$ and $\mathbf{z}_i$.

The most common and well-known estimators of the unstructured covariance matrix are the sample covariance matrix, $\mathbf{S} = \frac{1}{n-1} \mathbf{X} \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n' \right) \mathbf{X}'$, which is an unbiased estimator, and the maximum likelihood (ML) estimator, given by $\mathbf{S}_{\text{ML}} = \frac{1}{n} \mathbf{X} \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n' \right) \mathbf{X}'$. For $\mathbf{\Omega}$ partitioned as in (1.2), the estimators $\mathbf{S}$ and $\mathbf{S}_{\text{ML}}$ can be partitioned accordingly, resulting in blocks $\mathbf{S}_{11} = \frac{1}{n^*} \mathbf{Y} \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n' \right) \mathbf{Y}'$, $\mathbf{S}_{22} = \frac{1}{n^*} \mathbf{Z} \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n' \right) \mathbf{Z}'$, $\mathbf{S}_{12} = \frac{1}{n^*} \mathbf{Y} \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n' \right) \mathbf{Z}'$, where $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$, $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)$, and n^* equals respectively $n - 1$ and n for an unbiased and ML estimators of $\mathbf{\Omega}_{jk}$, $j, k \in \{1, 2\}$. The estimator ($\mathbf{S}$ or $\mathbf{S}_{\text{ML}}$) employed in the subsequent considerations will be explicitly indicated in the relevant sections.

In the literature it is common to examine whether any dependence between the two groups of features is present, that is to test the hypothesis

$$H_0 : \mathbf{\Omega}_{12} = \mathbf{0} \quad \text{vs} \quad H_1 : \mathbf{\Omega}_{12} \neq \mathbf{0}.$$

A well-known approach to testing independence is the likelihood ratio test based on Wilks' Lambda statistic, originally introduced by Wilks (1932, 1935). If the null hypothesis is rejected, it is reasonable to move on to a more detailed analysis of the potential dependence structures between the sub-vectors of $\mathbf{x}_i$, for example, the same covariance between all the components of $\mathbf{y}_i$ and $\mathbf{z}_i$, resulting in $\mathbf{\Omega}_{12} = \gamma \mathbf{1}_p \mathbf{1}_q'$, where $\gamma \in \mathbb{R}$.

Similarly, one can be interested in testing specific within-group dependencies, which are revealed by the structures of $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$. The most popular covariance structures considered in the literature are, for example, sphericity, compound symmetry, Toeplitz or autoregression; cf. Anderson (1969).

Assuming that $\mathbf{x}_i$ is divided into more than two sub-vectors, the hypothesis about particular block structure of $\mathbf{\Omega}$ can be also of interest, for example, block-spherical structure (Olkin, 1973) or block-compound symmetry (Szatrowski, 1982). It is worth noting, however, that such structures are usually limited to equal dimensions of sub-vectors.

In this thesis, we consider various dependence structures, which will be presented in the next subsection. In contrast to the unstructured covariance matrix $\mathbf{\Omega}$, the matrix having a structure will be denoted by $\mathbf{\Gamma}$.

1.2 Covariance matrix structures

The unstructured covariance matrix $\mathbf{\Omega}$ has $m(m + 1)/2$ unknown parameters to estimate. This relatively large number of parameters can make estimation challenging, particularly in settings with limited sample size. In such cases, the matrix of sums of squares and products may become singular or ill-conditioned. Additional challenges arise in high-dimensional settings, where both the sample size and the dimension may be large. High-dimensional data can make the sample covariance matrix unstable, break down classical asymptotic results, and increase the cost of matrix operations.

One way to address these issues is to impose structural constraints on the covariance matrix, thereby reducing the number of unknown parameters. Depending on the nature of the experiment, specific covariance structures often arise naturally and can be motivated by prior knowledge about the data or identified using appropriate methods. Using a relevant structure leads to more stable and reasonable estimators and improves the quality of statistical inference.

One of the most extensively studied and foundational types of dependence structures is a linear structure, defined as follows.

Definition 1.2.1 (Anderson, 1973). *The covariance matrix $\mathbf{\Omega}$ of order m has a linear structure if it can be expressed as a linear combination of known, symmetric matrices $\mathbf{G}_0, \dots, \mathbf{G}_\ell$, which are assumed to be independent, that is*

$$\mathbf{\Omega} = \sum_{i=0}^{\ell} \sigma_i \mathbf{G}_i,$$

where $\sigma_i \in \mathbb{R}$, $i \in \{1, \dots, \ell\}$, are unknown parameters. Furthermore, it is assumed that there exists at least one combination $\sigma_0, \dots, \sigma_\ell$ such that $\mathbf{\Omega}$ is p.d.

The vector space with basis $\mathbf{G}_0, \dots, \mathbf{G}_\ell$ will be called a linear structure space of dimension l .

In view of the partitioned structure (1.2) of $\mathbf{\Omega}$, let us consider the following sets of matrices:

$$\begin{aligned}\mathcal{T}_1 &= \left\{ \begin{pmatrix} \mathbf{T}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \in \mathbb{R}_{p+q}^s : \mathbf{T}_{11} \in \mathcal{T}_{11} \subseteq \mathbb{R}_p^s \right\}, \\ \mathcal{T}_2 &= \left\{ \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_{22} \end{pmatrix} \in \mathbb{R}_{p+q}^s : \mathbf{T}_{22} \in \mathcal{T}_{22} \subseteq \mathbb{R}_q^s \right\}, \\ \mathcal{T}_3 &= \left\{ \begin{pmatrix} \mathbf{0} & \mathbf{T}_{12} \\ \mathbf{T}_{12}' & \mathbf{0} \end{pmatrix} \in \mathbb{R}_{p+q}^s : \mathbf{T}_{12} \in \mathcal{T}_{12} \subseteq \mathbb{R}_{p \times q} \right\},\end{aligned}$$

and a set

$$\mathcal{T} = \{\mathbf{\Gamma} = \mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_3 : \mathbf{T}_1 \in \mathcal{T}_1, \mathbf{T}_2 \in \mathcal{T}_2, \mathbf{T}_3 \in \mathcal{T}_3\}. \quad (1.3)$$

If all $\mathcal{T}_{ij}$, $i, j \in \{1, 2\}$, $i \leq j$, are linear subspaces, then $\mathcal{T}$ is a linear subspace that can be presented as

$$\mathcal{T} = \mathcal{T}_1 \boxplus \mathcal{T}_2 \boxplus \mathcal{T}_3, \quad (1.4)$$

where $\boxplus$ denotes the orthogonal direct sum. Notably, $\mathcal{T}$ contains an identity matrix. Obviously, if $\mathcal{T}_{11} = \mathbb{R}_p^s$, $\mathcal{T}_{22} = \mathbb{R}_q^s$, and $\mathcal{T}_{12} = \mathbb{R}_{p \times q}$, then $\mathcal{T}$ spans the linear space of all symmetric matrices. Such a linear space, of dimension $m(m+1)/2$, will be denoted by $\mathcal{S}$. If any of $\mathcal{T}_{ij}$, $i, j \in \{1, 2\}$, $i \leq j$ is a proper linear subspace of $\mathbb{R}_p^s$, $\mathbb{R}_q^s$, and $\mathbb{R}_{p \times q}$, respectively, then $\mathcal{T}$ is a proper linear subspace of $\mathcal{S}$.

One of the trivial examples of a linear subspace $\mathcal{T}$ is $\mathcal{T}_1 \boxplus \mathcal{T}_2$, where $\mathcal{T}_{11} = \mathbb{R}_p^s$ and $\mathcal{T}_{22} = \mathbb{R}_q^s$. In other words, $\mathcal{T}$ is a space of block-diagonal matrices of the form

$$\mathbf{\Gamma} = \begin{pmatrix} \mathbf{\Gamma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}_2 \end{pmatrix}, \quad (1.5)$$

where $\mathbf{\Gamma}_1, \mathbf{\Gamma}_2$ are arbitrary unstructured matrices of relevant order. Moreover, if $\mathbf{\Gamma}_1, \mathbf{\Gamma}_2$ are p.d., then $\mathbf{\Gamma}$ is a covariance matrix of a partitioned random vector with cross-covariance matrix equal to zero (two sub-vectors are uncorrelated).

Assume now that both, $\mathcal{T}_{11}$ and $\mathcal{T}_{22}$, are the linear spaces of matrices having compound symmetry (CS) structure, i.e.,

$$\begin{aligned}\mathcal{T}_{11} &= \{\alpha_1 \mathbf{I}_p + \beta_1 \mathbf{1}_p \mathbf{1}_p' : \alpha_1, \beta_1 \in \mathbb{R}\}, \\ \mathcal{T}_{22} &= \{\alpha_2 \mathbf{I}_q + \beta_2 \mathbf{1}_q \mathbf{1}_q' : \alpha_2, \beta_2 \in \mathbb{R}\}.\end{aligned} \quad (1.6)$$

If α_i, β_i , $i \in \{1, 2\}$, are limited, such that $\mathcal{T}_{11}$ and $\mathcal{T}_{22}$ are the cones of p.d. matrices, we can say that, in a single sub-vector, the variances of random components are equal to each other and the covariances between random components are also equal to each other.

Let us consider now a subspace $\mathcal{T} = \mathcal{T}_1 \boxplus \mathcal{T}_2 \boxplus \mathcal{T}_3$, where $\mathcal{T}_{11}$ and $\mathcal{T}_{22}$ are the linear spaces of CS matrices defined in (1.6), and $\mathcal{T}_{12}$ is a linear space of matrices proportional to the matrix of ones, i.e.

$$\mathcal{T}_{12} = \{\gamma \mathbf{1}_p \mathbf{1}_q' : \gamma \in \mathbb{R}\}. \quad (1.7)$$

If $\mathcal{T}$ is a cone of p.d. matrices, $\mathbf{\Gamma} \in \mathcal{T}$ is a covariance matrix of partitioned random vector with equal dependencies between the two sub-vectors.

Another structure space commonly studied in the literature is the space of CS structures of the form

$$\mathcal{T} = \{\alpha \mathbf{I}_m + \beta \mathbf{1}_m \mathbf{1}_m' : \alpha, \beta \in \mathbb{R}\}. \quad (1.8)$$

In view of the above considerations, it can be also represented as a special case of $\mathcal{T}_1 \boxplus \mathcal{T}_2 \boxplus \mathcal{T}_3$ with $\mathcal{T}_{11}$, $\mathcal{T}_{22}$, and $\mathcal{T}_{12}$ defined in (1.6) and (1.7), where $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2 = \gamma$.

Finally, combining (1.5) and (1.7), we obtain the space of matrices of the form

$$\mathbf{\Gamma} = \begin{pmatrix} \mathbf{\Gamma}_1 & \gamma \mathbf{1}_p \mathbf{1}_q' \\ \gamma \mathbf{1}_q \mathbf{1}_p' & \mathbf{\Gamma}_2 \end{pmatrix}. \quad (1.9)$$

It is worth noting that all the above linear spaces, except the space containing matrices defined in (1.9), are the quadratic subspaces of $\mathcal{S}$, since they are closed under squaring, that is if $\mathbf{\Gamma} \in \mathcal{T}$ then also $\mathbf{\Gamma}^2 \in \mathcal{T}$; cf. Seely (1971). Furthermore, a space of CS matrices defined in (1.8) is the commutative quadratic subspace of $\mathcal{S}$, as the product of two matrices from the space belongs to the same space and the matrices commute.

Another example of commutative quadratic subspace is the space of circular Toeplitz structures, of the form

$$\mathcal{T} = \{\alpha_0 \mathbf{I}_m + \sum_{i=1}^{m^*} \alpha_i (\mathbf{R}^i + \mathbf{R}^{i'}) : \alpha_0, \alpha_1, \dots, \alpha_{m^*} \in \mathbb{R}\},$$

where $m^* = \lfloor m/2 \rfloor$ if m is odd and $m^* = m/2 - 1$ when m is even, and $\mathbf{R}$ is a zero-one permutation matrix of order m with ones on the first sub-diagonal and the upper right corner and zeros elsewhere. Note that if $\alpha_i, i \in \{0, 1, \dots, m^*\}$, are limited such that $\mathcal{T}$ is a cone of p.d. matrices, then every matrix from this cone is a covariance matrix of a random vector, for which variances are equal to each other and the covariances of neighboring variables (arranged circularly) are equal to each other.

Similarly as in the case of a space of block-diagonal structures with CS blocks, (1.6), the space of block-diagonal matrices with $\mathcal{T}_{11}$ or $\mathcal{T}_{22}$ having circular Toeplitz structure will be of interest in this thesis; see articles [D] and [E].

Another covariance structure often considered in the literature is the first order autoregression, AR(1), where the correlations between observations decrease exponentially with time, that is

$$\mathbf{\Psi} = \gamma \left(\mathbf{I}_m + \sum_{i=1}^{m-1} \varrho^i \mathbf{H}_i \right), \quad (1.10)$$

where $\gamma > 0$, $\varrho \in (-1, 1)$, and $\mathbf{H}_i, i \in \{1, \dots, m-1\}$, are zero-one matrices of order m with ones on the i -th sub-diagonal and zeros elsewhere. Obviously, this structure is not linear.

In this thesis, similarly as in (1.9), we consider a set $\mathcal{T}$ given in (1.3) with $\mathcal{T}_{11} = \mathbb{R}_p^s, \mathcal{T}_{22} = \mathbb{R}_q^s$, and

$$\mathcal{T}_{12} = \{\gamma \mathbf{\Psi}_{12} : \gamma \in \mathbb{R}\},$$

where $\mathbf{\Psi}_{12}$ is a relevant part of the AR(1) correlation structure, that is,

$$\mathbf{\Psi}_{12} = \begin{pmatrix} \varrho^p & \varrho^{p+1} & \ddots & \ddots & \ddots & \varrho^{p+q-3} & \varrho^{p+q-2} & \varrho^{p+q-1} \\ \varrho^{p-1} & \varrho^p & \varrho^{p+1} & \ddots & \ddots & \ddots & \varrho^{p+q-3} & \varrho^{p+q-2} \\ \ddots & \varrho^{p-1} & \varrho^p & \varrho^{p+1} & \ddots & \ddots & \ddots & \varrho^{p+q-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \varrho^3 & \ddots & \ddots & \ddots & \varrho^{q-1} & \varrho^q & \varrho^{q-1} & \varrho^{q-2} \\ \varrho^2 & \varrho^3 & \ddots & \ddots & \ddots & \varrho^{q-1} & \varrho^q & \varrho^{q-1} \\ \varrho & \varrho^2 & \varrho^3 & \ddots & \ddots & \ddots & \varrho^{q-1} & \varrho^q \end{pmatrix}.$$

In other words, the considered set of structures contains matrices of the form

$$\mathbf{\Gamma} = \begin{pmatrix} \mathbf{\Gamma}_1 & \gamma \mathbf{\Psi}_{12} \\ \gamma \mathbf{\Psi}_{12}' & \mathbf{\Gamma}_2 \end{pmatrix}, \quad (1.11)$$

where $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ are arbitrary unstructured matrices of relevant order. The set consisting of matrices of the form (1.11) will be denoted by $\mathcal{T}_{\text{AR}}$. The set of p.d. matrices $\mathbf{\Gamma}$ will be denoted by $\mathcal{T}_{\text{AR}}^>$. Finally, the set of well-conditioned p.d. matrices $\mathbf{\Gamma}$ will be denoted by $\mathcal{T}_{\text{AR}}^{\gg}$.

1.3 Estimation methods

In this section, we present the estimation methods considered in the doctoral dissertation. The main objective is to determine an estimator of a block covariance matrix with a prescribed structure that satisfies the same structural constraints, is positive definite and well-conditioned, possesses desirable statistical properties, and remains computationally feasible.

Maximum likelihood estimator

It is well known that the ML estimators of the expectation $\boldsymbol{\mu}$ and the unstructured covariance matrix $\boldsymbol{\Omega}$ under the model (1.1) are, respectively, $\bar{\mathbf{X}} = \frac{1}{n}\mathbf{X}\mathbf{1}_n$ and $\mathbf{S}_{\text{ML}} = \frac{1}{n}\mathbf{X}\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\right)\mathbf{X}'$. Since the structure imposed on the covariance matrix does not affect the ML estimator of $\boldsymbol{\mu}$, substituting $\hat{\boldsymbol{\mu}}$ into the likelihood function simplifies the estimation of the structured covariance matrix $\boldsymbol{\Gamma}$ to maximizing the following log-likelihood function

$$\ln f(\boldsymbol{\Gamma}) = -\frac{nm}{2} \ln(2\pi) - \frac{n}{2} \ln |\boldsymbol{\Gamma}| - \frac{n}{2} \text{tr}(\mathbf{S}_{\text{ML}}\boldsymbol{\Gamma}^{-1}), \quad (1.12)$$

with respect to $\boldsymbol{\Gamma}$ over a set $\mathcal{T}$ of structured p.d. matrices.

In general, solving this constrained optimization problem is complex and computationally demanding, typically requiring iterative procedures. However, when the structure space is quadratic, the ML estimator is simply the projection of $\mathbf{S}_{\text{ML}}$ onto the structure space; cf. Filipiak et al. (2020).

The ML estimation over $\mathcal{T}_{\text{CS}}^{\gg}$ and $\mathcal{T}_{\text{AR}}^{\gg}$ is addressed in [B] and [C].

Least squares method

The least squares (LS) method is based on minimizing the Frobenius norm of the difference between the matrix $\mathbf{S}$ and the set of structured matrices $\boldsymbol{\Gamma}$, that is, minimizing

$$f(\mathbf{S}, \boldsymbol{\Gamma}) = \|\mathbf{S} - \boldsymbol{\Gamma}\|_{\text{F}}^2 = \text{tr}[(\mathbf{S} - \boldsymbol{\Gamma})(\mathbf{S} - \boldsymbol{\Gamma})'] \quad (1.13)$$

over a set $\mathcal{T}$ of structured p.d. matrices.

If $\boldsymbol{\Gamma}$ belongs to a linear subspace $\mathcal{T}$ of $\mathcal{S}$, spanned by an orthonormal basis $\mathbf{V}_1, \dots, \mathbf{V}_v$, then the LS estimator of $\boldsymbol{\Gamma} \in \mathcal{T}$ is the projection of the sample covariance matrix, $\mathbf{S}$, onto the linear structure space, $\mathcal{T}$, that is

$$\hat{\boldsymbol{\Gamma}}_{\text{LS}} = \boldsymbol{\Pi}_{\mathcal{T}}(\mathbf{S}) = \sum_{i=1}^v \text{tr}(\mathbf{V}_i\mathbf{S})\mathbf{V}_i. \quad (1.14)$$

Such an approach has been studied, for instance, in Ohlson and von Rosen (2010).

When the structure space is quadratic, projecting $\mathbf{S}$ onto this space provides the best quadratic unbiased estimator; see, e.g., Koziol et al. (2017). However, if the structure space is not quadratic, the resulting estimator does not always possess desirable properties, such as positive definiteness or well conditioning, and it may even become indefinite (Fuglede and Jensen, 2013). In such cases, additional techniques, such as shrinkage method, may be applied to ensure that the estimator satisfies the necessary conditions for positive definiteness and well-conditioning.

The LS estimation over $\mathcal{T}_{\text{CS}}^>$ and $\mathcal{T}_{\text{AR}}^>$ is addressed in [B] and [C].

Shrinkage method

The shrinkage method proposed by Ledoit and Wolf (2004, 2021) was developed for high-dimensional settings, where the dimension p is comparable to the sample size n . In such situations, the sample covariance matrix is known to perform poorly. When $p > n$, it is not invertible, while when the ratio p/n is less than one but not negligible, the sample covariance matrix is invertible but numerically ill-conditioned, so that its inversion substantially amplifies estimation error. Since in high-dimensional problems it is often difficult to obtain sufficiently large samples relative to p , it becomes essential to construct covariance matrix estimators that are well-conditioned. The shrinkage approach addresses this issue by constructing an estimator that reduces the mean squared error while yielding an estimator that is both invertible and well-conditioned.

Following Ledoit and Wolf (2004), let us define

$$\mathbf{\Omega}_S = (1 - \lambda)\mathbf{S} + \lambda\mathbf{T}, \quad 0 \leq \lambda \leq 1,$$

where $\mathbf{S}$ is the sample covariance matrix, $\mathbf{T}$ is a structured target matrix (e.g., a multiple of the identity matrix), and the shrinkage intensity λ is chosen to minimize the quadratic risk

$$\mathbb{E} \|\mathbf{\Omega}_S - \mathbf{\Omega}\|_F^2. \quad (1.15)$$

The optimal shrinkage intensity is given by

$$\lambda^* = \frac{\mathbb{E} \|\mathbf{S} - \mathbf{\Omega}\|_F^2}{\mathbb{E} \|\mathbf{S} - \mathbf{T}\|_F^2},$$

and to compute the above expectations the unbiasedness of $\mathbf{S}$ is used.

Since λ^* is unknown in practice, it is replaced by a consistent estimator constructed from the data, $\widehat{\lambda}$. Thus, the shrinkage estimator of unstructured $\mathbf{\Omega}$ can be presented as

$$\widehat{\mathbf{\Omega}}_S = (1 - \widehat{\lambda})\mathbf{S} + \widehat{\lambda}\mathbf{T}.$$

This yields a feasible estimator that asymptotically minimizes the mean squared error within the class of convex combinations of $\mathbf{S}$ and $\mathbf{T}$.

In this thesis, the shrinkage method is adapted to improve the properties of the LS estimator of a structured matrix $\mathbf{\Gamma}$, while preserving the structural constraints imposed on the covariance matrix. Therefore, let us define

$$\mathbf{\Gamma}_S = (1 - \lambda)\widehat{\mathbf{\Gamma}}_{LS} + \lambda\mathbf{T}, \quad (1.16)$$

where the target matrix $\mathbf{T}$ is chosen to keep the structure. Common choices for the target matrix $\mathbf{T}$ in the literature include diagonal structures or multiples of the identity, with the latter employed in Ledoit and Wolf (2004); in contrast, the works of Hannart and Naveau (2014) and Ikeda et al. (2016) focus on diagonal targets, although the shrinkage is applied to unstructured estimators, while, in our approach, the target is selected from a quadratic subspace and is not limited to diagonal or spherical targets. Since it is known that the projection of $\mathbf{S}$ onto an arbitrary quadratic subspace $\mathcal{Q}$ preserves positive definiteness, we also allow $\mathbf{\Pi}_Q(\mathbf{S})$ to serve as a target matrix.

Finally, the optimal shrinkage intensity λ for a structured covariance matrix is obtained by minimizing the quadratic risk

$$\begin{aligned} \mathbb{E} \left(\|\widehat{\mathbf{\Gamma}}_S - \mathbf{\Gamma}\|_F^2 \right) &= \text{tr} \left[\mathbb{E} (\mathbf{T}^2) \right] + \lambda \text{tr} \left[\mathbb{E} (\mathbf{T} \widehat{\mathbf{\Gamma}}_{LS}) \right] + \lambda^2 \text{tr} \left[\mathbb{E} (\widehat{\mathbf{\Gamma}}_{LS}^2) \right] \\ &\quad - 2 \text{tr} \left[\mathbb{E} (\mathbf{T} \mathbf{\Gamma}) \right] - 2\lambda \text{tr} \left[\mathbb{E} (\widehat{\mathbf{\Gamma}}_{LS}) \mathbf{\Gamma} \right] + \text{tr} [\mathbf{\Gamma}^2] \\ &= g(\lambda). \end{aligned} \quad (1.17)$$

To derive the expected values of the relevant matrices we use selected properties of the Wishart distribution. These properties summarized in Theorem 1.3.1, follow from standard results on Wishart matrices presented in ,e.g., Gupta and Nagar (2000), Holgersson and Pielaszkiewicz (2020), and Pielaszkiewicz and von Rosen (2021).

Theorem 1.3.1. *Let $\mathbf{W} \sim W_p(n, \Sigma)$, $p < n$. Then,*

- (i) $\mathbb{E}(\mathbf{W}) = n\Sigma$
- (ii) $\mathbb{E}(\mathbf{W}^2) = (n^2 + n)\Sigma^2 + n \operatorname{tr}(\Sigma)\Sigma$
- (iii) $\mathbb{E}[\operatorname{tr}(\mathbf{A}\mathbf{W})] = n \operatorname{tr}(\mathbf{A}\Sigma)$
- (iv) $\mathbb{E}[\operatorname{tr}(\mathbf{A}\mathbf{W}) \operatorname{tr}(\mathbf{B}\mathbf{W})] = n \operatorname{tr}(\mathbf{A}\Sigma\mathbf{B}\Sigma) + n \operatorname{tr}(\mathbf{A}'\Sigma\mathbf{B}\Sigma) + n^2 \operatorname{tr}(\mathbf{A}\Sigma) \operatorname{tr}(\mathbf{B}\Sigma).$

Since the optimal λ is unknown in practice, it is replaced by a consistent estimator constructed from the data, $\hat{\lambda}$. Consequently, the shrinkage estimator of the structured $\mathbf{\Gamma}$ can be presented as

$$\hat{\mathbf{\Gamma}}_S = (1 - \hat{\lambda})\hat{\mathbf{\Gamma}}_{LS} + \hat{\lambda}\mathbf{T}.$$

The explicit formulas for the optimal shrinkage intensity for the cone $\mathcal{T}_{CS}^>$ and the set $\mathcal{T}_{AR}^>$ are provided in [B] and [C]. For arbitrary block structures, the explicit form of the shrinkage intensity is given in [E].

1.4 Loss functions

In this dissertation, loss functions are used for two related goals. First, to identify an appropriate covariance structure from a set of candidate structures, the entropy loss function is applied. Second, to compare the statistical properties of the considered covariance matrix estimators, the Stein loss, inverse Stein loss, quadratic loss, and Frobenius norm are used.

If it is reasonable to believe that the dependency structure belongs to a known class of potential structures, it becomes possible to identify this structure. In the literature, this process is sometimes associated with the term regularization (cf. Lin et al., 2014, Cui et al., 2016), which traditionally refers to improving the properties of an estimator rather than identifying the structure itself. To avoid confusion, throughout this thesis we refer to this process specifically as covariance structure identification. Structure identification involves selecting, from among a set of candidate structures, the one that minimizes the discrepancy from the true covariance matrix, as measured by a chosen discrepancy function.

In this thesis, as a discrepancy function we consider the Kullback–Leibler divergence between two multivariate normal distributions that differ in their covariance matrices, say $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ (cf. Pan and Fang, 2002), also called the entropy loss function (cf. Stein, 1956, Dey and Srinivasan, 1985, James and Stein, 1961, Lin et al., 2014), having the form

$$f_E(\mathbf{\Omega}_1, \mathbf{\Omega}_2) = \text{tr} \left(\mathbf{\Omega}_1^{-1} \mathbf{\Omega}_2 \right) - \ln |\mathbf{\Omega}_1^{-1} \mathbf{\Omega}_2| - m. \quad (1.18)$$

Without loss of generality, assume that unstructured $\mathbf{\Omega}_1$ is estimated by an ML method, that means by $\mathbf{S}_{\text{ML}}$. The main idea of structure identification is to minimize $f_E(\mathbf{S}_{\text{ML}}, \mathbf{\Gamma})$ with respect to a structured $\mathbf{\Gamma} \in \mathcal{T} \subset \mathcal{S}$. In the paper [A] we consider $\mathcal{T}$ to be $\mathcal{T}_{\text{CS}}$ and $\mathcal{T}_{\text{AR}}$.

The other commonly used loss functions include:

- the quadratic loss function (cf. Anderson, 2003)

$$f_Q(\mathbf{\Omega}_1, \mathbf{\Omega}_2) = \text{tr} \left[\left(\mathbf{\Omega}_1^{-1} \mathbf{\Omega}_2 - \mathbf{I}_m \right)^2 \right],$$

- the inverse Stein loss (cf. Dey and Srinivasan, 1985)

$$f_{\text{IS}}(\mathbf{\Omega}_1, \mathbf{\Omega}_2) = \frac{1}{m} \left[\text{tr} \left(\mathbf{\Omega}_1 \mathbf{\Omega}_2^{-1} \right) - \log |\mathbf{\Omega}_1 \mathbf{\Omega}_2^{-1}| \right] - 1,$$

- the squared Frobenius norm of the form $f(\mathbf{\Omega}_1, \mathbf{\Omega}_2)$ defined in (1.13).

Similarly as in Ledoit and Wolf (2021), in this thesis all the above loss functions are used to compare the properties of various estimators of covariance structures from $\mathcal{T}_{\text{CS}}^{\gg}$ (article [B]) and $\mathcal{T}_{\text{AR}}^{\gg}$ (article [C]). For this purpose, in simulation studies, in the above formulas $\mathbf{\Omega}_1$ is replaced by the assumed covariance structure $\mathbf{\Gamma}$, while $\mathbf{\Omega}_2$ is replaced by various estimators of $\mathbf{\Gamma}$.

1.5 Partitioned matrices, majorization and matrix derivatives

To facilitate the ML estimation, we use specific properties of partitioned matrices, which are particularly useful for estimating structured covariance matrices. These properties, as formulated in Kollo and von Rosen (2005), are stated in the following propositions and allow for analytical simplification of the log-likelihood function.

Proposition 1.5.1. Let $\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$ be a partitioned non-singular matrix. If $\mathbf{A}_{22}$ is non-singular, then

$$|\mathbf{A}| = |\mathbf{A}_{22}| \cdot |\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}|,$$

and if $\mathbf{A}_{11}$ is non-singular, then

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

Proposition 1.5.2. Let $\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$ be a partitioned non-singular matrix. Then, assuming that all inverses below exist,

$$\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{C}_{22}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{C}_{11}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{C}_{22}^{-1} & \mathbf{C}_{11}^{-1} \end{pmatrix},$$

where

$$\mathbf{C}_{11} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12},$$

$$\mathbf{C}_{22} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}.$$

Matrices $\mathbf{C}_{11}$ and $\mathbf{C}_{22}$ are called Schur complements of $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ in $\mathbf{A}$, respectively.

In the thesis, we focus on the well-conditioning of the estimates. Therefore, we employ the concept of majorization (cf. Marshall et al., 2011) together with lemmas describing the relationships between the eigenvalues of various matrices.

Let $\mathbf{a}$ and $\mathbf{b}$ be two given p -dimensional vectors. Let $a_{(1)} \geq a_{(2)} \geq \dots \geq a_{(p)}$ and $b_{(1)} \geq b_{(2)} \geq \dots \geq b_{(p)}$ denote the components of $\mathbf{a}$ and $\mathbf{b}$ rearranged in decreasing order. We say that $\mathbf{a}$ is majorized by $\mathbf{b}$, and write

$$\mathbf{a} \prec \mathbf{b},$$

if for $k \in \{1, \dots, p-1\}$

$$\sum_{i=1}^k a_{(i)} \leq \sum_{i=1}^k b_{(i)} \quad \text{and} \quad \sum_{i=1}^p a_{(i)} = \sum_{i=1}^p b_{(i)}.$$

Now, let $\lambda(\mathbf{A}) = (\lambda_1(\mathbf{A}), \dots, \lambda_p(\mathbf{A}))'$, $\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \dots \geq \lambda_p(\mathbf{A})$, be the vector of the eigenvalues of a symmetric matrix $\mathbf{A}$ of order p .

Lemma 1.5.3. If $\mathbf{A}$ and $\mathbf{B}$ are symmetric matrices of order p , then

$$(\lambda_1(\mathbf{A} + \mathbf{B}), \dots, \lambda_p(\mathbf{A} + \mathbf{B})) \prec (\lambda_1(\mathbf{A}) + \lambda_1(\mathbf{B}), \dots, \lambda_p(\mathbf{A}) + \lambda_p(\mathbf{B})).$$

Lemma 1.5.4. *If $\mathbf{A}$ and $\bar{\mathbf{A}}$ are symmetric matrices of order m of the form*

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}, \quad \bar{\mathbf{A}} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{pmatrix}$$

with $\mathbf{A}_{11} \in \mathbb{R}_p^s$, $\mathbf{A}_{22} \in \mathbb{R}_q^s$, $p + q = m$, then

$$\begin{pmatrix} \lambda(\mathbf{A}_{11}) \\ \lambda(\mathbf{A}_{22}) \end{pmatrix} = \lambda(\bar{\mathbf{A}}) \prec \lambda(\mathbf{A}).$$

Throughout this dissertation, we use the vec-operator, as well as the duplication and commutation matrices, which are required for computing matrix derivatives. Their definitions can be found, for example, in Magnus and Neudecker (1986) and Kollo and von Rosen (2005).

Let $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_q)$ be a $p \times q$ matrix, where $\mathbf{a}_i, i \in \{1, \dots, q\}$, is the i -th column vector of $\mathbf{A}$. The vectorization operator, $\text{vec}(\mathbf{A})$, is an operator from $\mathbb{R}_{p \times q}$ to $\mathbb{R}_{pq}$, such that

$$\text{vec}(\mathbf{A}) = \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_q \end{pmatrix}.$$

To vectorize the transpose of $\mathbf{A}$, one can use a commutation matrix, that is a permutation matrix $\mathbf{K}_{p,q}$ of order pq satisfying

$$\mathbf{K}_{p,q} \cdot \text{vec}(\mathbf{A}) = \text{vec}(\mathbf{A}').$$

Note that $\mathbf{K}_{q,p} = \mathbf{K}_{p,q} = \mathbf{K}_{p,q}^{-1}$. Moreover, for symmetric matrices of order p , $\mathbf{K}_{p,p} = \mathbf{I}_{p^2}$.

When considering symmetric matrices, it is suitable to vectorize only a lower triangle of a matrix. Thus, a $\text{vech}(\cdot)$ operator stands for a vector obtained from $\text{vec}(\cdot)$ by eliminating all supradiagonal entries. A unique $p^2 \times p(p+1)/2$ matrix $\mathbf{D}_p$ transforming $\text{vech}(\mathbf{A})$ into $\text{vec}(\mathbf{A})$, that is

$$\mathbf{D}_p \cdot \text{vech}(\mathbf{A}) = \text{vec}(\mathbf{A}), \tag{1.19}$$

is called a duplication matrix.

The following properties of the vec operator are used throughout this dissertation.

Lemma 1.5.5. *Let $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ be matrices such that the product $\mathbf{ABC}$ exists. Then:*

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A}) \text{vec}(\mathbf{B}).$$

Lemma 1.5.6. *Let $\mathbf{A}$ and $\mathbf{B}$ be square matrices of the same order. Then, the following relationship between the vec -operator and the trace of a matrix holds:*

$$\text{vec}'(\mathbf{A}) \cdot \text{vec}(\mathbf{B}) = \text{tr}(\mathbf{A}'\mathbf{B}).$$

Following Magnus and Neudecker (1988), we use the following convention for matrix derivatives. For a matrix-valued function $\mathbf{F} : \mathbb{R}_{p \times q} \rightarrow \mathbb{R}_{r \times s}$, the derivative with respect to a $p \times q$ matrix $\mathbf{X}$ is an $rs \times pq$ matrix defined as

$$\frac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} = \frac{\partial \text{vec } \mathbf{F}(\mathbf{X})}{\partial \text{vec}'(\mathbf{X})}.$$

In addition, if $\mathbf{G} = \mathbf{G}(\mathbf{F}) : \mathbb{R}_{r \times s} \rightarrow \mathbb{R}_{t \times u}$, then

$$\frac{\partial \mathbf{G}}{\partial \mathbf{X}} = \frac{\partial \mathbf{G}}{\partial \mathbf{F}} \cdot \frac{\partial \mathbf{F}}{\partial \mathbf{X}}. \quad (1.20)$$

Moreover, if $\mathbf{H} = \mathbf{H}(\mathbf{X}) : \mathbb{R}_{p \times q} \rightarrow \mathbb{R}_{s \times u}$, then

$$\frac{\partial (\mathbf{FH})}{\partial \mathbf{X}} = (\mathbf{H}' \otimes \mathbf{I}_r) \frac{\partial \mathbf{F}}{\partial \mathbf{X}} + (\mathbf{I}_u \otimes \mathbf{F}) \frac{\partial \mathbf{H}}{\partial \mathbf{X}}.$$

The formulas used for matrix differentiation are summarized in the following lemma.

Lemma 1.5.7. *Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{X}$ be matrices of appropriate dimensions without any pattern. Then:*

$$\begin{aligned} \frac{\partial \text{tr}(\mathbf{AX})}{\partial \mathbf{X}} &= \text{vec}'(\mathbf{A}'), & \frac{\partial \ln |\mathbf{A}|}{\partial \mathbf{A}} &= \text{vec}'[(\mathbf{A}^{-1})'], \\ \frac{\partial \mathbf{A}^{-1}}{\partial \mathbf{A}} &= -\left[(\mathbf{A}^{-1})' \otimes \mathbf{A}^{-1}\right], & \frac{\partial \mathbf{AA}'}{\partial \mathbf{A}} &= (\mathbf{I}_{p^2} + \mathbf{K}_{p,p})(\mathbf{I}_p \otimes \mathbf{A}'), \\ \frac{\partial \mathbf{A}^i}{\partial \mathbf{A}} &= \sum_{j=1}^i (\mathbf{A}')^{i-j} \otimes \mathbf{A}^{j-1}, & \frac{\partial \mathbf{AXB}}{\partial \mathbf{X}} &= \mathbf{B}' \otimes \mathbf{A}. \end{aligned}$$

For symmetric matrices, derivatives involve post-multiplying the corresponding formulas from Lemma 1.5.7 by the duplication matrix; cf. Magnus and Neudecker (1986), Filipiak et al. (2016). This follows from the property (1.19) and the chain rule (1.20), since for a symmetric matrix $\mathbf{A}$ of order p the relation $\partial \text{vec}(\mathbf{A}) / \partial \text{vech}'(\mathbf{A}) = \mathbf{D}_p$ holds.

Chapter 2

Main results – review

[A] Block matrix approximation via entropy loss function

The aim of the paper [A] is to develop a method for approximating a symmetric p.d. matrix by a symmetric block-partitioned matrix with particularly structured off-diagonal blocks. Moreover, the proposed procedure is applied in a simulation study addressing the statistical problem of covariance structure identification.

Assume that $\mathbf{\Gamma}$ is a structured block covariance matrix. We consider two covariance structures defined in (1.9) and (1.11), corresponding to the cone $\mathcal{T}_{\text{CS}}^>$ and the set $\mathcal{T}_{\text{AR}}^>$, respectively. For a given unstructured partitioned covariance matrix $\mathbf{\Omega}$, defined in (1.2), the objective is to determine a matrix from the sets $\mathcal{T}_{\text{CS}}^>$ or $\mathcal{T}_{\text{AR}}^>$ that minimizes the Kullback-Leibler discrepancy defined by the entropy loss function (1.18).

Let us partition unstructured $\mathbf{\Omega}$ into four blocks of size corresponding to the partition of $\mathbf{\Gamma}$, and denote appropriate blocks of $\mathbf{\Omega}^{-1}$ by $\mathbf{\Omega}^{jk}$, $j, k \in \{1, 2\}$, that is

$$\mathbf{\Omega}^{-1} = \begin{pmatrix} \mathbf{\Omega}^{11} & \mathbf{\Omega}^{12} \\ \mathbf{\Omega}^{21} & \mathbf{\Omega}^{22} \end{pmatrix},$$

where each block is given by Proposition 1.5.2 from Section 1.5.

Using Proposition 1.5.1 from Section 1.5, the entropy loss function (1.18) can be presented with the use of two variants of Schur complement as

$$\begin{aligned}
 f_E(\mathbf{\Omega}, \mathbf{\Gamma}) &= \text{tr}(\mathbf{\Omega}^{11}\mathbf{\Gamma}_1) + 2\text{tr}(\mathbf{\Omega}^{12}\mathbf{\Gamma}_{21}) + \text{tr}(\mathbf{\Omega}^{22}\mathbf{\Gamma}_2) + \ln|\mathbf{\Omega}| \\
 &\quad - \ln|\mathbf{\Gamma}_2| - \ln|\mathbf{\Gamma}_1 - \mathbf{\Gamma}_{12}\mathbf{\Gamma}_2^{-1}\mathbf{\Gamma}_{21}| - (p+q) \\
 &= \text{tr}(\mathbf{\Omega}^{11}\mathbf{\Gamma}_1) + 2\text{tr}(\mathbf{\Omega}^{12}\mathbf{\Gamma}_{21}) + \text{tr}(\mathbf{\Omega}^{22}\mathbf{\Gamma}_2) + \ln|\mathbf{\Omega}| \\
 &\quad - \ln|\mathbf{\Gamma}_1| - \ln|\mathbf{\Gamma}_2 - \mathbf{\Gamma}_{21}\mathbf{\Gamma}_1^{-1}\mathbf{\Gamma}_{12}| - (p+q).
 \end{aligned} \tag{A.1}$$

The following theorems are proven using the vectorization and differentiation rules described in Section 1.5.

Theorem A.1. *For a given $\mathbf{\Omega} \in \mathbb{R}_m^>$, there exists $\mathbf{\Gamma} \in \mathbb{R}_m^>$ that minimizes the entropy loss function (A.1) over $\mathcal{T}_{CS}^>$ and this minimum is attained at $\mathbf{\Gamma}_1, \mathbf{\Gamma}_2, \gamma$ satisfying the following system of equations*

$$\begin{cases}
 \mathbf{\Gamma}_1 = (\mathbf{\Omega}^{11})^{-1} + \gamma^2 \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p' \\
 \mathbf{\Gamma}_2 = (\mathbf{\Omega}^{22})^{-1} + \gamma^2 \mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}_q' \\
 \gamma = -\frac{\text{tr}(\mathbf{\Omega}^{12} \mathbf{1}_q \mathbf{1}_p')}{\text{tr}(\mathbf{\Omega}^{11} \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p')}.
 \end{cases}$$

Theorem A.2. *For a given $\mathbf{\Omega} \in \mathbb{R}_m^>$, there exists $\mathbf{\Gamma} \in \mathbb{R}_m^>$ that minimizes the entropy loss function (A.1) over $\mathcal{T}_{AR}^>$ and this minimum is attained at $\mathbf{\Gamma}_1, \mathbf{\Gamma}_2, \gamma, \varrho$ satisfying the following system of equations*

$$\begin{cases}
 \mathbf{\Gamma}_1 = (\mathbf{\Omega}^{11})^{-1} + \gamma^2 \mathbf{\Psi}_{12} \mathbf{\Gamma}_2^{-1} \mathbf{\Psi}_{12}' \\
 \mathbf{\Gamma}_2 = (\mathbf{\Omega}^{22})^{-1} + \gamma^2 \mathbf{\Psi}_{12}' \mathbf{\Gamma}_1^{-1} \mathbf{\Psi}_{12} \\
 \gamma = -\frac{\text{tr}(\mathbf{\Omega}^{21} \mathbf{\Psi}_{12})}{\text{tr}(\mathbf{V}_{11} \mathbf{\Psi}_{12} \mathbf{\Gamma}_2^{-1} \mathbf{\Psi}_{12}')} \\
 2\text{tr}(\mathbf{\Omega}^{21} \mathbf{F}) + \gamma \text{tr} \left[\mathbf{\Omega}^{11} \left(\mathbf{F} \mathbf{\Gamma}_2^{-1} \mathbf{\Psi}_{12}' + \mathbf{\Psi}_{12} \mathbf{\Gamma}_2^{-1} \mathbf{F}' \right) \right] = 0,
 \end{cases}$$

with $\mathbf{F} = \partial \mathbf{\Psi}_{12} / \partial \varrho$.

Note that to determine the above approximations, numerical procedure for solving the system of equations must be employed. For the purpose of this study, the corresponding algorithms are implemented in the R environment.

In the paper [A] it is also of interest to verify whether the entropy loss function properly identifies the underlying structure. For this purpose the simulation studies are conducted, where for various parameters settings we generate data from a matrix normal distribution with mean vector $\mathbf{0}_m$ and covariance matrix $\mathbf{\Gamma}$ belonging to $\mathcal{T}_{CS}^>$ and $\mathbf{\Gamma}$ belonging to $\mathcal{T}_{AR}^>$. Since in practice $\mathbf{\Gamma}$ is unknown, under both scenarios, it is replaced by the ML estimator of the unstructured covariance matrix, which does not have any structure with probability 1. The structure of $\mathbf{\Gamma}$ is then approximated by minimization of the entropy loss between $\mathbf{S}_{ML}$ and the set of structured covariance matrices, $f_E(\mathbf{S}_{ML}, \mathbf{\Gamma})$, with respect to $\mathbf{\Gamma} \in \mathcal{T}_{CS}^>$ and $\mathbf{\Gamma} \in \mathcal{T}_{AR}^>$, respectively, where f_E is defined in (A.1), and the one with smaller discrepancy is selected. After repeating the procedure 1000 times, the proportion of properly identified structures among 1000 simulation runs is computed.

The results of simulation studies show that the structure from $\mathcal{T}_{AR}^>$ is better identified than the structure from $\mathcal{T}_{CS}^>$ because for most of the parameter settings the proportion of properly identified structures from $\mathcal{T}_{AR}^>$ exceeds 80%. On the other hand, the proportion of properly identified structures from $\mathcal{T}_{CS}^>$ does not exceed 25% in most parameter settings, which means that even if the underlying structure is from $\mathcal{T}_{CS}^>$, the entropy loss function identified a structure from $\mathcal{T}_{AR}^>$ as the proper one. This may be due to, among other factors, the inclusion between the considered sets, $\mathcal{T}_{CS}^> \subseteq \mathcal{T}_{AR}^>$. Nevertheless, the entropy loss function can still be effectively used to approximate the covariance matrix.

[B] Quasi shrinkage estimation of a block-structured covariance matrix

The aim of the paper [B] is to propose well-conditioned estimators of the structured covariance matrix $\mathbf{\Gamma} \in \mathcal{T}_{\text{CS}}^>$, defined in (1.9), having good statistical properties. For this purpose, the ML estimation method and the LS procedure improved by shrinking are proposed, and their statistical properties are then compared.

Using vectorization properties and differentiation rules described in Section 1.5, the following theorem is proven.

Theorem B.1. *For a given matrix $\mathbf{S}_{\text{ML}} \in \mathbb{R}_m^>$, there exists a matrix $\mathbf{\Gamma} \in \mathbb{R}_m^>$ that maximizes the likelihood function (1.12) over $\mathcal{T}_{\text{CS}}^>$ and this maximum is attained at $\mathbf{\Gamma}_1, \mathbf{\Gamma}_2, \gamma$ satisfying the following system of equations*

$$\left\{ \begin{array}{l} \mathbf{\Gamma}_1 = \mathbf{S}_{11} - \gamma \left(\mathbf{C}_{22} \mathbf{\Gamma}_1^{-1} \mathbf{S}_{12} \mathbf{C}_{11}^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1^{-1} \mathbf{C}_{22} + \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{S}_{12}' \right) \\ \quad + \gamma^2 \left(\mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p' + \mathbf{C}_{22} \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{C}_{11}^{-1} \mathbf{S}_{22} \mathbf{C}_{11}^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1^{-1} \mathbf{C}_{22} \right) \\ \quad - \gamma^3 \mathbf{C}_{22} \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{C}_{11}^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1^{-1} \mathbf{S}_{12} \mathbf{C}_{11} \mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1^{-1} \mathbf{C}_{22} \\ \mathbf{\Gamma}_2 = \mathbf{S}_{22} - \gamma \left(\mathbf{C}_{11} \mathbf{\Gamma}_2^{-1} \mathbf{S}_{12}' \mathbf{C}_{22}^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{C}_{11} + \mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1^{-1} \mathbf{S}_{12} \right) \\ \quad + \gamma^2 \left(\mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}_q' + \mathbf{C}_{11} \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{C}_{22}^{-1} \mathbf{S}_{11} \mathbf{C}_{22}^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{C}_{11} \right) \\ \quad - \gamma^3 \mathbf{C}_{11} \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{C}_{22}^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{S}_{12}' \mathbf{C}_{22}^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{C}_{11} \\ \gamma = \frac{-2b - \sqrt{4b^2 - 8ac}}{-4a}, \end{array} \right.$$

where

$$\begin{aligned} a &= \text{tr} \left[\mathbf{S}_{12} \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{C}_{22}^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{C}_{22}^{-1} + \mathbf{S}_{12}' \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{C}_{11}^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{C}_{11}^{-1} \right] \\ b &= \text{tr} \left[\left(\mathbf{C}_{22}^{-1} \mathbf{S}_{11} - \mathbf{I}_p \right) \mathbf{C}_{22}^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p' + \mathbf{S}_{22} \mathbf{C}_{11}^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{C}_{11}^{-1} \right] \\ c &= \text{tr} \left[\mathbf{S}_{12} \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{C}_{22}^{-1} + \mathbf{S}_{12}' \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{C}_{11}^{-1} \right] \end{aligned}$$

and $\mathbf{C}_{11} = \mathbf{\Gamma}_2 - \gamma^2 \mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1 \mathbf{1}_p \mathbf{1}_q'$, $\mathbf{C}_{22} = \mathbf{\Gamma}_1 - \gamma^2 \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2 \mathbf{1}_q \mathbf{1}_p'$ are the Schur complements of $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$, respectively, in $\mathbf{\Gamma}$.

The solution of the above system of equations is not given in explicit form and thus it must be found numerically; note however, that the procedure can be time-consuming. In the article, we apply a gradient descent algorithm, and, in order to compute a well-conditioned estimate, we additionally impose a constraint on the condition number of the resulting estimator. Thus, the resulting estimator belongs to $\mathcal{T}_{\text{CS}}^{\gg}$.

Although ML is a standard approach to covariance matrix estimation, it becomes computationally demanding for structured models. We therefore propose a faster alternative based on the LS method, i.e., the projection of the estimator of the unstructured covariance matrix $\mathbf{\Omega}$ onto the space of structured matrices; in the considered case, projection onto the convex cone $\mathbf{T}_{\text{CS}}^>$. Observing that the projection is linear, we can split the process into two steps: projecting onto the linear space $\mathcal{T}_{\text{CS}}$, providing an explicit solution $\widehat{\mathbf{\Gamma}}_{\text{LS}}$ as in (1.14), which is not necessarily positive definite (cf. Fuglede and Jensen, 2013), and thus requires projection of $\widehat{\mathbf{\Gamma}}_{\text{LS}}$ onto the convex cone $\mathcal{T}_{\text{CS}}^>$. However, this second projection usually requires complex numerical procedures and thus can be time-consuming (cf. Filipiak et al., 2018), while our aim is to propose a faster alternative to ML technique. Therefore, instead of projecting $\widehat{\mathbf{\Gamma}}_{\text{LS}}$ onto $\mathcal{T}_{\text{CS}}^>$, the shrinkage method (with some additional algebraic correction, if necessary) is proposed as a faster alternative.

In the following theorem we give the form of $\widehat{\mathbf{\Gamma}}_{\text{LS}}$, which is obtained by minimization of the Frobenius norm defined in (1.13) over the space $\mathcal{T}_{\text{CS}}$, assuming that $\mathbf{S}_{\text{ML}}$ is a given matrix, that is

$$f(\mathbf{S}_{\text{ML}}, \widehat{\mathbf{\Gamma}}_{\text{LS}}) = \min_{\mathbf{\Gamma} \in \mathcal{T}_{\text{CS}}} \|\mathbf{S}_{\text{ML}} - \mathbf{\Gamma}\|_{\text{F}}^2. \quad (\text{B.1})$$

Note that $\mathbf{S}_{\text{ML}}$ appears naturally in the likelihood function, and thus is applied also in this method; however, a sample covariance matrix, $\mathbf{S}$, can also be used here instead of $\mathbf{S}_{\text{ML}}$. Using vectorization properties and differentiation rules described in Section 1.5, the following theorem is proven.

Theorem B.2. *For a given matrix $\mathbf{S}_{\text{ML}} \in \mathbb{R}_m^>$, there exists a symmetric matrix $\mathbf{\Gamma} \in \mathbb{R}_m^s$ that minimizes the Frobenius norm (B.1) over $\mathcal{T}_{\text{CS}}$ and this minimum is attained at*

$$\begin{cases} \mathbf{\Gamma}_1 &= \mathbf{S}_{11} \\ \mathbf{\Gamma}_2 &= \mathbf{S}_{22} \\ \gamma &= \frac{1}{pq} \mathbf{1}_p' \mathbf{S}_{12} \mathbf{1}_q. \end{cases}$$

To choose the target matrix required in the shrinkage method, we apply the following lemma, where the proof is based on majorization theory presented in Section 1.5.

Lemma B.3. *Let*

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{pmatrix}$$

be symmetric ,p.d., block-partitioned matrices of order $(p + q)$ with $\mathbf{A}_{11} \in \mathbb{R}_p^s$, $\mathbf{A}_{22} \in \mathbb{R}_q^s$. Then, the relation between their condition numbers is

$$\kappa(\mathbf{A}) \geq \kappa(\mathbf{B}).$$

From the above lemma, it follows that the condition number of a block-diagonal matrix consisting of diagonal blocks of $\mathbf{S}_{\text{ML}}$ is not greater than that of $\mathbf{S}_{\text{ML}}$. Therefore, as the target matrix in (1.16), we choose the projection of $\mathbf{S}_{\text{ML}}$ onto the space of block-diagonal matrices. Moreover, since this projection depends on the data, we denote it by $\widehat{\mathbf{T}}$ instead of $\mathbf{T}$, i.e.,

$$\widehat{\mathbf{T}} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22} \end{pmatrix}, \quad (\text{B.2})$$

obtaining

$$\mathbf{\Gamma}_S = (1 - \lambda)\widehat{\mathbf{\Gamma}}_{\text{LS}} + \lambda\widehat{\mathbf{T}}.$$

To determine the shrinkage intensity λ , we use the quadratic risk $g(\lambda)$ defined in (1.17) with $\mathbf{T}$ replaced by $\widehat{\mathbf{T}}$. Minimizing $g(\lambda)$ and using the properties of the Wishart distribution presented in Section 1.3, the following theorem is proven.

Theorem B.4. *For a given $\mathbf{\Gamma} \in \mathcal{T}_{\text{CS}}$ the minimum of the quadratic risk $g(\lambda)$ given in (1.17) is attained at*

$$\lambda = \frac{np^2q^2\gamma^2}{\mathbf{1}'_p \mathbf{\Gamma}_1 \mathbf{1}_p \mathbf{1}'_q \mathbf{\Gamma}_2 \mathbf{1}_q + np^2q^2\gamma^2}.$$

Since the shrinkage intensity given in the above theorem depends on the unknown matrix $\mathbf{\Gamma} \in \mathcal{T}_{\text{CS}}$, it is replaced by its structured estimator $\widehat{\mathbf{\Gamma}}_{\text{LS}}$, leading to

$$\widehat{\lambda} = \frac{n \left(\mathbf{1}'_p \mathbf{S}_{12} \mathbf{1}_q \right)^2}{\mathbf{1}'_p \mathbf{S}_{11} \mathbf{1}_p \mathbf{1}'_q \mathbf{S}_{22} \mathbf{1}_q + n \left(\mathbf{1}'_p \mathbf{S}_{12} \mathbf{1}_q \right)^2}.$$

The shrinkage estimator is not necessarily p.d. or well-conditioned; therefore, we additionally propose several algebraic improvements based on rescaling the off-diagonal blocks or diagonal blocks of $\widehat{\mathbf{\Gamma}}_S$. The obtained estimator with the required condition number is called the quasi shrinkage (QS) estimator. Using simulation studies it is shown that rescaling off-diagonal blocks of $\widehat{\mathbf{\Gamma}}_S$ provides the best QS estimator with respect to its normalized distance to $\mathbf{S}_{ML}$.

Statistical properties of the QS estimator, such as the Stein loss, inverse Stein loss, and quadratic loss (cf. Section 1.4), are compared with those of the ML estimator in a simulation study, where for various parameters p , q , and γ , we generate 1000 datasets of size $n = 100$ from a matrix normal distribution with mean vector $\mathbf{0}_m$ and covariance matrix $\mathbf{\Gamma} \in \mathcal{T}_{CS}^>$. The results show that the ML estimators and QS estimators are comparable with respect to the mentioned loss functions. In addition, for the true matrix $\mathbf{\Gamma}$ being close to singularity, the QS estimator has lower empirical risk than the corresponding ML estimator.

Summing up, determination of QS estimator is computationally simpler and less time-consuming than the ML estimator, although a three-steps procedure: projection onto $\mathcal{T}_{CS}$, shrinkage and algebraic correction, is involved. Since the computational cost of the ML estimation increases with the dimension, the QS method offers a reasonable alternative.

[C] Block covariance matrix estimation with structured off-diagonal blocks

The aim of the paper [C] is to propose well-conditioned estimators of the structured covariance matrix $\mathbf{\Gamma} \in \mathcal{T}_{\text{AR}}^>$, defined in (1.11), having good statistical properties. For this purpose, similarly to the previous section, the ML estimation method and the LS procedure improved by shrinking are proposed, and the statistical properties of the obtained estimates are then compared.

Let $\mathbf{F} = \partial \Psi_{12} / \partial \varrho$, where Ψ_{12} is respective off-diagonal block of AR(1) matrix defined in (1.11). Using vectorization properties and differentiation rules described in Section 1.5, the following theorem is proven.

Theorem C.1. *If for a given matrix $\mathbf{S}_{\text{ML}} \in \mathbb{R}_m^>$ there exists a unique solution to the following system of equations*

$$\left\{ \begin{array}{l} \mathbf{\Gamma}_1 = \mathbf{S}_{11} + (\gamma)^2 \Psi_{12} \mathbf{\Gamma}_2^{-1} \Psi'_{12} - \gamma \Psi_{12} \mathbf{\Gamma}_2^{-1} \mathbf{S}'_{12} - \gamma \mathbf{C}_{11} \mathbf{\Gamma}_1^{-1} \mathbf{S}_{12} \mathbf{C}_{22}^{-1} \Psi'_{12} \mathbf{\Gamma}_1^{-1} \mathbf{C}_{11} \\ \quad - (\gamma)^3 \mathbf{C}_{11} \mathbf{\Gamma}_1^{-1} \Psi_{12} \mathbf{C}_{22}^{-1} \Psi'_{12} \mathbf{\Gamma}_1^{-1} \mathbf{S}_{12} \mathbf{C}_{22}^{-1} \Psi'_{12} \mathbf{\Gamma}_1^{-1} \mathbf{C}_{11} \\ \quad + (\gamma)^2 \mathbf{C}_{11} \mathbf{\Gamma}_1^{-1} \Psi_{12} \mathbf{C}_{22}^{-1} \mathbf{S}_{22} \mathbf{C}_{22}^{-1} \Psi'_{12} \mathbf{\Gamma}_1^{-1} \mathbf{C}_{11} \\ \mathbf{\Gamma}_2 = \mathbf{S}_{22} + (\gamma)^2 \Psi'_{12} \mathbf{\Gamma}_1^{-1} \Psi_{12} - \gamma \Psi'_{12} \mathbf{\Gamma}_1^{-1} \mathbf{S}_{12} - \gamma \mathbf{C}_{22} \mathbf{\Gamma}_2^{-1} \mathbf{S}'_{12} \mathbf{C}_{11}^{-1} \Psi_{12} \mathbf{\Gamma}_2^{-1} \mathbf{C}_{22} \\ \quad - (\gamma)^3 \mathbf{C}_{22} \mathbf{\Gamma}_2^{-1} \Psi'_{12} \mathbf{C}_{11}^{-1} \Psi_{12} \mathbf{\Gamma}_2^{-1} \mathbf{S}'_{12} \mathbf{C}_{11}^{-1} \Psi_{12} \mathbf{\Gamma}_2^{-1} \mathbf{C}_{22} \\ \quad + (\gamma)^2 \mathbf{C}_{22} \mathbf{\Gamma}_2^{-1} \Psi'_{12} \mathbf{C}_{11}^{-1} \mathbf{S}_{11} \mathbf{C}_{11}^{-1} \Psi_{12} \mathbf{\Gamma}_2^{-1} \mathbf{C}_{22} \\ 0 = \text{tr} [-2\gamma \mathbf{C}_{11}^{-1} \Psi_{12} \mathbf{\Gamma}_2^{-1} \Psi'_{12} + 2\gamma \mathbf{C}_{11}^{-1} \mathbf{S}_{11} \mathbf{C}_{11}^{-1} \Psi_{12} \mathbf{\Gamma}_2^{-1} \Psi'_{12} \\ \quad - \mathbf{S}_{12} \mathbf{\Gamma}_2^{-1} \Psi'_{12} \mathbf{C}_{11}^{-1} - 2\gamma \mathbf{S}_{12} \mathbf{\Gamma}_2^{-1} \Psi'_{12} \mathbf{C}_{11}^{-1} \Psi_{12} \mathbf{\Gamma}_2^{-1} \Psi'_{12} \mathbf{C}_{11}^{-1} \\ \quad - \mathbf{C}_{22}^{-1} \mathbf{S}'_{12} \mathbf{\Gamma}_1^{-1} \Psi_{12} - 2\gamma \mathbf{S}'_{12} \mathbf{\Gamma}_1^{-1} \Psi_{12} \mathbf{C}_{22}^{-1} \Psi'_{12} \mathbf{\Gamma}_1^{-1} \Psi_{12} \mathbf{C}_{22}^{-1} \\ \quad + 2\gamma \mathbf{C}_{22}^{-1} \mathbf{S}_{22} \mathbf{C}_{22}^{-1} \Psi'_{12} \mathbf{\Gamma}_1^{-1} \Psi_{12}] \\ \gamma = \frac{-b - \sqrt{b^2 - 4ac}}{2a}, \end{array} \right.$$

where

$$\begin{aligned}
 a &= \text{tr} \left[-\Gamma_2^{-1} (\Psi'_{12} + \mathbf{F}' \mathbf{C}_{11}^{-1} \Psi_{12}) + \mathbf{C}_{11}^{-1} \mathbf{S}_{11} \mathbf{C}_{11}^{-1} (\mathbf{F} \Gamma_2^{-1} \Psi'_{12} + \Psi_{12} \Gamma_2^{-1} \mathbf{F}') \right. \\
 &\quad - \mathbf{S}_{12} \Gamma_2^{-1} \Psi'_{12} \mathbf{C}_{11}^{-1} (\mathbf{F} \Gamma_2^{-1} \Psi'_{12} + \Psi_{12} \Gamma_2^{-1} \mathbf{F}') \mathbf{C}_{11}^{-1} \\
 &\quad + \mathbf{C}_{22}^{-1} \mathbf{S}'_{12} \Gamma_1^{-1} \Psi_{12} \mathbf{C}_{22}^{-1} (\Psi'_{12} \Gamma_1^{-1} \mathbf{F} - \mathbf{F}' \Gamma_1^{-1} \Psi_{12}) \\
 &\quad \left. + \mathbf{C}_{22}^{-1} \mathbf{S}_{22} \mathbf{C}_{22}^{-1} (\Psi'_{12} \Gamma_1^{-1} \mathbf{F} + \mathbf{F}' \Gamma_1^{-1} \Psi_{12}) \right], \\
 b &= -\text{tr} \left[\mathbf{C}_{11}^{-1} \mathbf{S}_{12} \Gamma_2^{-1} \mathbf{F}' + \mathbf{C}_{22}^{-1} \mathbf{S}'_{12} \Gamma_1^{-1} \mathbf{F} \right], \\
 c &= \text{tr} [\mathbf{C}_{11} \mathbf{F}],
 \end{aligned}$$

$\mathbf{C}_{11} = \Gamma_2 - \gamma^2 \Psi_{12} \Gamma_1 \Psi'_{12}$, $\mathbf{C}_{22} = \Gamma_1 - \gamma^2 \Psi'_{12} \Gamma_2 \Psi_{12}$ are the Schur complements of Γ_1 and Γ_2 , respectively, in Γ , then this solution is the ML estimator of Γ over $\mathcal{T}_{\text{AR}}^>$.

Similarly as in Theorem B.1, the solution of the above system of equations is not given in explicit form and thus it must be found numerically. For this purpose, in the article [C], we again apply a gradient descent algorithm, and, in order to compute a well-conditioned estimate, we additionally impose a constraint on the condition number of the resulting estimator. Thus, the resulting estimator belongs to $\mathcal{T}_{\text{AR}}^{\gg}$.

As discussed in our previous paper, ML estimation becomes computationally demanding for structured covariance matrices. Since this difficulty also occurs in this case, it is natural to consider a faster alternative based on the LS and shrinkage methods.

The LS estimator of Γ , $\widehat{\Gamma}_{\text{LS}}$, can be obtained by minimizing the Frobenius norm defined in (1.13) over the space $\mathcal{T}_{\text{AR}}$, assuming that $\mathbf{S}_{\text{ML}}$ is a given matrix, that is

$$f(\mathbf{S}_{\text{ML}}, \widehat{\Gamma}_{\text{LS}}) = \min_{\Gamma \in \mathcal{T}_{\text{AR}}} \|\mathbf{S}_{\text{ML}} - \Gamma\|_F^2.$$

Note that, since $\mathcal{T}_{\text{AR}}$ is not a linear space, the minimizer of the Frobenius norm is not given by the projection formula (1.14).

Using vectorization properties and differentiation rules described in Section 1.5, the following theorem is proven.

Theorem C.2. *If for a given matrix $\mathbf{S}_{\text{ML}} \in \mathbb{R}_m^{\geq}$ there exists a unique solution to the following system of equations*

$$\left\{ \begin{array}{lcl} \mathbf{\Gamma}_1 & = & \mathbf{S}_{11} \\ \mathbf{\Gamma}_2 & = & \mathbf{S}_{22} \\ \text{tr}(\mathbf{F}\mathbf{\Psi}'_{12}) & = & \text{tr}(\mathbf{S}'_{12}\mathbf{F}) \\ \gamma & = & \frac{\text{tr}(\mathbf{S}'_{12}\mathbf{\Psi}_{12})}{\text{tr}(\mathbf{\Psi}_{12}\mathbf{\Psi}'_{12})} \end{array} \right.$$

over $\mathcal{T}_{\text{AR}}$, then this solution provides the LS estimator of $\mathbf{\Gamma} \in \mathcal{T}_{\text{AR}}$.

Observe that the LS estimator of the off-diagonal block is not given in explicit form, as the third equation is a polynomial of order $2(p + q) - 3$ in ϱ , which may admit multiple solutions in $\mathcal{T}_{\text{AR}}$, and ϱ is hidden in the last equation. The parameter $\widehat{\varrho}$ should therefore be selected from the interval $(-1, 1)$ so that $\widehat{\mathbf{\Gamma}} \in \mathcal{T}_{\text{AR}}$. Among all admissible solutions, in the article [C], for simulation purposes, we choose the one minimizing the Frobenius norm between $\mathbf{S}_{\text{ML}}$ and $\widehat{\mathbf{\Gamma}}$. Finally, to ensure positive definiteness and to improve conditioning of the Frobenius norm estimator, similarly as in [B], we employ a shrinkage method with an additional algebraic correction.

Motivated by the result of Lemma B.3, as the target matrix we again use the projection of $\mathbf{S}_{\text{ML}}$ onto the space of block-diagonal matrices, i.e., matrix defined in (B.2), and to determine the shrinkage intensity, we use the quadratic risk $g(\lambda)$ given in (1.17). Minimizing $g(\lambda)$ and using the properties of the Wishart distribution presented in Section 1.3, the following theorem is proven.

Theorem C.3. *For a given $\mathbf{\Gamma} \in \mathcal{T}_{\text{AR}}$, the minimum of the quadratic risk $g(\lambda)$ given in (1.17) is attained at*

$$\lambda = \frac{N}{D},$$

where

$$\begin{aligned} N &= n \left[\text{tr}(\mathbf{\Gamma}_1^2) + 4\gamma^2 \text{tr}(\mathbf{\Psi}_{12}\mathbf{\Psi}'_{12}) + \text{tr}(\mathbf{\Gamma}_2^2) \right] - \text{tr}^2(\mathbf{\Gamma}_1) - \text{tr}^2(\mathbf{\Gamma}_2) \\ D &= 2 \left\{ n \left[\text{tr}(\mathbf{\Gamma}_1^2) + 2\gamma^2 \text{tr}(\mathbf{\Psi}_{12}\mathbf{\Psi}'_{12}) + \text{tr}(\mathbf{\Gamma}_2^2) \right] + \text{tr}^2(\mathbf{\Gamma}_1) + \text{tr}^2(\mathbf{\Gamma}_2) \right. \\ &\quad \left. + 2\gamma^4 \text{tr}^2(\mathbf{\Psi}'_{12}\mathbf{\Gamma}_1^2\mathbf{\Psi}_{12}) + 2\gamma^4 \text{tr}^2(\mathbf{\Psi}_{12}\mathbf{\Gamma}_2^2\mathbf{\Psi}'_{12}) + 2\gamma^4 \text{tr}^2(\mathbf{\Psi}_{12}\mathbf{\Psi}'_{12}) \right\}. \end{aligned}$$

Replacing in the above theorem the unknown matrix $\mathbf{\Gamma} \in \mathcal{T}_{\text{AR}}$ by its structured estimator $\widehat{\mathbf{\Gamma}}_{\text{LS}}$

yields

$$\hat{\lambda} = \frac{\hat{N}}{\hat{D}},$$

where

$$\begin{aligned}\hat{N} &= n \left[\text{tr}(\mathbf{S}_{11}^2) + 4 \text{tr}(\mathbf{S}_{12} \mathbf{S}_{12}') + \text{tr}(\mathbf{S}_{22}^2) \right] - \text{tr}^2(\mathbf{S}_{11}) - \text{tr}^2(\mathbf{S}_{22}) \\ \hat{D} &= 2 \left\{ n \left[\text{tr}(\mathbf{S}_{11}^2) + 2 \text{tr}(\mathbf{S}_{12} \mathbf{S}_{12}') + \text{tr}(\mathbf{S}_{22}^2) \right] + \text{tr}^2(\mathbf{S}_{11}) + \text{tr}^2(\mathbf{S}_{22}) \right. \\ &\quad \left. + 2 \text{tr}^2(\mathbf{S}_{12}' \mathbf{S}_{11}^2 \mathbf{S}_{12}) + 2 \text{tr}^2(\mathbf{S}_{12} \mathbf{S}_{22}^2 \mathbf{S}_{12}') + 2 \text{tr}^2(\mathbf{S}_{12} \mathbf{S}_{12}') \right\}.\end{aligned}$$

If the shrinkage estimator is not a p.d. or well-conditioned matrix, the off-diagonal blocks are rescaled by decreasing an appropriate constant from the interval $(0, 1)$ until the estimator gets required properties. Similarly as in [B], obtained estimator with the required condition number is called the QS estimator.

Both estimators, ML and QS, are not given in explicit forms and must be computed numerically. It can be seen that the ML estimator is more complex, however, QS estimator requires a multi-step procedure, with verification of definiteness and conditioning in each step. Therefore, in [C] we also propose alternative methods of estimation of the considered structure, based on separate estimation of each component in the representation of $\mathbf{\Gamma} \in \mathcal{T}_{\text{AR}}^>$ as

$$\mathbf{\Gamma} = \mathbf{\Psi} + \begin{pmatrix} \mathbf{\Gamma}_1 - \mathbf{\Psi}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}_2 - \mathbf{\Psi}_{22} \end{pmatrix} = \mathbf{\Psi} + \mathbf{W}, \quad (\text{C.1})$$

where $\mathbf{\Psi}$ represents an AR(1) covariance structure defined in (1.10) with $\mathbf{\Psi}_{11}$ and $\mathbf{\Psi}_{22}$ being its diagonal blocks of proper size, and where $\mathbf{W}$ represents the noise.

We propose four methods related to the following, well-established, estimation techniques for AR(1) covariance structure.

The most intuitive method of estimation of AR(1) structure is given in Roy and Leiva (2008), where the estimators of ϱ and γ are given by

$$\begin{cases} \hat{\varrho} = \left(\frac{\hat{\varrho}_1^{p-1} + \hat{\varrho}_2^{\frac{p-1}{2}} + \hat{\varrho}_3^{\frac{p-1}{3}} + \cdots + \hat{\varrho}_{p-1}}{p-1} \right)^{\frac{1}{p-1}} \\ \hat{\gamma} = \frac{\text{tr}(\mathbf{S}_{\text{ML}})}{m}, \end{cases}$$

where $\hat{\varrho}_i$ is the average of the i th super-diagonal elements of $\mathbf{S}_{\text{ML}}$, $i \in \{1, \dots, p-1\}$. Because of its simplicity, it is called here *naive* estimator.

The ML estimator of AR(1) structure, given in the following theorem, can be found, for example, in Filipiak et al. (2017).

Theorem C.4 (Filipiak et al., 2017). *If for a given matrix $\mathbf{S}_{ML} \in \mathbb{R}_m^>$ there exists unique solution to the following system of equations*

$$\begin{cases} \varrho^3 [(m-1) \text{tr}(\mathbf{S}\mathbf{C}_1)] + \varrho^2 \left[\left(\frac{2-m}{2}\right) \text{tr}(\mathbf{S}_{ML}\mathbf{C}_2) \right] \\ - \varrho [\text{tr}(\mathbf{S}_{ML}) + m \text{tr}(\mathbf{S}_{ML}\mathbf{C}_1)] + \frac{m}{2} \text{tr}(\mathbf{S}_{ML}\mathbf{C}_2) = 0 \\ \gamma = \frac{1}{m(1-\varrho^2)} [\text{tr}(\mathbf{S}_{ML}) + \varrho^2 \text{tr}(\mathbf{S}_{ML}\mathbf{C}_1) - \varrho \text{tr}(\mathbf{S}_{ML}\mathbf{C}_2)] , \end{cases}$$

where $\mathbf{C}_1 = \text{diag}(0, \mathbf{1}_{m-2}, 0)$ and $\mathbf{C}_2$ is a tridiagonal matrix of order m with 0 on the main diagonal and 1 on the first super- and sub-diagonal, then it provides the ML estimator of Ψ .

The LS estimator of AR(1) structure, given in the following theorem, can be found, for example, in Cui et al. (2016).

Theorem C.5 (Cui et al., 2016). *If for a given matrix $\mathbf{S}_{ML} \in \mathbb{R}_m^>$ there exists unique solution to the following system of equations*

$$\begin{cases} - \left[m + 2 \sum_{i=1}^{m-1} (m-i) \varrho^{2i} \right] \sum_{i=1}^{m-1} i \varrho^{i-1} \text{tr}(\mathbf{S}_{ML}\mathbf{H}_i) \\ + 2 \left[\text{tr}(\mathbf{S}_{ML}) + \sum_{i=1}^{m-1} \varrho^i \text{tr}(\mathbf{S}_{ML}\mathbf{H}_i) \right] \sum_{i=1}^{m-1} (m-i) i \varrho^{2i-1} = 0 \\ \gamma = \frac{\text{tr}(\mathbf{S}_{ML}) + \sum_{i=1}^{m-1} \varrho^i \text{tr}(\mathbf{S}_{ML}\mathbf{H}_i)}{m + 2 \sum_{i=1}^{m-1} (m-i) \varrho^{2i}} , \end{cases}$$

where $\mathbf{H}_i$ is a symmetric matrix of order m with i -th super- and sub-diagonal elements equal to 1 and all other elements equal to 0, $\mathbf{H}_0 = \mathbf{I}_m$, then it provides an LS estimator of Ψ .

Finally, since for the comparison of variances in the univariate case the ratio of variances is used rather than their difference, in the fourth estimation method we propose minimization of the distance between the product $\mathbf{S}_{ML}\Psi^{-1}$ and the identity matrix in the sense of the Frobenius norm, i.e.,

$$\min_{\Psi \in \mathbb{R}_m^>} \|\mathbf{S}_{ML}\Psi^{-1} - \mathbf{I}_m\|_F^2. \quad (\text{C.2})$$

Note that in this method the tridiagonal form of the inverse of AR(1) structure plays an important role.

Using vectorization properties and differentiation rules described in Section 1.5, the following theorem is proven.

Theorem C.6. *If for a given matrix $\mathbf{S}_{\text{ML}} \in \mathbb{R}_m^>$ there exists unique solution to the following system of equations*

$$\begin{cases} a_6 \varrho^6 + a_5 \varrho^5 + a_4 \varrho^4 + a_3 \varrho^3 + a_2 \varrho^2 + a_1 \varrho + a_0 = 0 \\ \gamma = \frac{N}{D} \end{cases}$$

with

$$\begin{aligned} a_6 &= -\text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11} \mathbf{S}_{\text{ML}} \mathbf{C}_{11}) + \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22}) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11} \mathbf{S}_{\text{ML}} \mathbf{C}_{11}) \\ a_5 &= 2 \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11} \mathbf{S}_{\text{ML}}) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11}) + \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22} \mathbf{S}_{\text{ML}} \mathbf{C}_{22}) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11}) \\ &\quad - \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11} \mathbf{S}_{\text{ML}} \mathbf{C}_{22}) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22}) - 2 \text{tr}(\mathbf{S}_{\text{ML}}) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11} \mathbf{S}_{\text{ML}} \mathbf{C}_{11}) \\ a_4 &= -3 \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22} \mathbf{S}_{\text{ML}}) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11}) + \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11}) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11} \mathbf{S}_{\text{ML}} \mathbf{C}_{22}) \\ &\quad + 3 \text{tr}(\mathbf{S}_{\text{ML}}) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11} \mathbf{S}_{\text{ML}} \mathbf{C}_{22}) - \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11} \mathbf{S}_{\text{ML}} \mathbf{C}_{11}) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22}) \\ a_3 &= 2 \text{tr}(\mathbf{S}_{\text{ML}}^2) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11}) - 2 \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11}) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11} \mathbf{S}_{\text{ML}}) - 2 \text{tr}(\mathbf{S}_{\text{ML}}) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11} \mathbf{S}_{\text{ML}}) \\ &\quad + \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22}) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22} \mathbf{S}_{\text{ML}}) + \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11} \mathbf{S}_{\text{ML}} \mathbf{C}_{22}) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22}) \\ &\quad - \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11}) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22} \mathbf{S}_{\text{ML}} \mathbf{C}_{22}) - \text{tr}(\mathbf{S}_{\text{ML}}) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22} \mathbf{S}_{\text{ML}} \mathbf{C}_{22}) \\ &\quad + 2 \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11} \mathbf{S}_{\text{ML}} \mathbf{C}_{11}) \text{tr}(\mathbf{S}_{\text{ML}}) , \\ a_2 &= -\text{tr}(\mathbf{S}_{\text{ML}}^2) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22}) + 3 \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11}) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22} \mathbf{S}_{\text{ML}}) + \text{tr}(\mathbf{S}_{\text{ML}}) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22} \mathbf{S}_{\text{ML}}) \\ &\quad - 3 \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11} \mathbf{S}_{\text{ML}} \mathbf{C}_{22}) \text{tr}(\mathbf{S}_{\text{ML}}) \\ a_1 &= 2 \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11} \mathbf{S}_{\text{ML}}) \text{tr}(\mathbf{S}_{\text{ML}}) - \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22}) \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22} \mathbf{S}_{\text{ML}}) \\ &\quad + \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22} \mathbf{S}_{\text{ML}} \mathbf{C}_{22}) \text{tr}(\mathbf{S}_{\text{ML}}) - 2 \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11}) \text{tr}(\mathbf{S}_{\text{ML}}^2) \\ a_0 &= \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22}) \text{tr}(\mathbf{S}_{\text{ML}}^2) - \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22} \mathbf{S}_{\text{ML}}) \text{tr}(\mathbf{S}_{\text{ML}}) \\ N &= \text{tr}(\mathbf{S}_{\text{ML}}^2) - 2\varrho \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22} \mathbf{S}_{\text{ML}}) + \varrho^2 [2 \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11} \mathbf{S}_{\text{ML}}) - \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22} \mathbf{S}_{\text{ML}} \mathbf{C}_{22})] \\ &\quad - 2\varrho^3 \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11} \mathbf{S}_{\text{ML}} \mathbf{C}_{22}) + \varrho^4 \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11} \mathbf{S}_{\text{ML}} \mathbf{C}_{11}) \\ D &= (1 - \varrho^2) [\text{tr}(\mathbf{S}_{\text{ML}}) + \varrho^2 \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{11}) - \varrho \text{tr}(\mathbf{S}_{\text{ML}} \mathbf{C}_{22})] \end{aligned}$$

where $\mathbf{C}_{11} = \mathbf{\Psi}_{22} - \gamma^2 \mathbf{1}_q \mathbf{1}_p' \mathbf{\Psi}_{11} \mathbf{1}_p \mathbf{1}_q'$, $\mathbf{C}_{22} = \mathbf{\Psi}_{11} - \gamma^2 \mathbf{1}_p \mathbf{1}_q' \mathbf{\Psi}_{22} \mathbf{1}_q \mathbf{1}_p'$ are the Schur complements of $\mathbf{\Psi}_{11}$ and $\mathbf{\Psi}_{22}$, respectively, in $\mathbf{\Psi}$, then it provides an estimator of $\mathbf{\Psi}$ minimizing (C.2).

The diagonal matrix $\mathbf{W}$ in representation (C.1) is then estimated using the projection of $\mathbf{S}_{\text{ML}}$ onto the space of diagonal matrices, as it is given analytically via the diagonal blocks of $\mathbf{S}_{\text{ML}}$, denoted respectively $\mathbf{S}_{11}$ and $\mathbf{S}_{22}$, and the diagonal block of corresponding estimator of AR(1) structure, providing

$$\widehat{\mathbf{W}} = \begin{pmatrix} \mathbf{S}_{11} - \widehat{\Psi}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22} - \widehat{\Psi}_{22} \end{pmatrix}.$$

Obviously, the naive method is the fastest one, as it gives the estimator explicitly. The other three estimators require solving polynomial equations, but they are still much faster than ML and QS techniques.

In the article [C] we also compare all the estimators with respect to their statistical properties, such as Stein loss, inverse Stein loss and quadratic loss, defined in Section 1.4. For this purpose, we perform simulation studies, where for various parameters settings we generate 1000 datasets of size $n = 100$ from a matrix normal distribution with mean vector $\mathbf{0}_m$ and covariance matrix $\mathbf{\Gamma} \in \mathcal{T}_{\text{AR}}^>$. The results show that the naive estimators and those based on ML estimation of a whole $\mathbf{\Gamma}$ structure (see Theorem C.1) are the best with respect to all mentioned loss functions, however, the ML estimation is the most time-consuming technique.

Finally, the proposed estimation methods are illustrated using an agricultural dataset. We compare the six proposed estimators with respect to the estimates $\widehat{\gamma}$ and $\widehat{\varrho}$, the Frobenius norm of $\widehat{\mathbf{\Gamma}}_1$ and $\widehat{\mathbf{\Gamma}}_2$, and the condition numbers of the corresponding $\widehat{\mathbf{\Gamma}}$. All estimates yield similar results with respect to the considered criteria. Due to its computational efficiency, the naive estimator is recommended.

[D] Characterization of quadratic subspaces of partitioned matrices

The topic of the paper [D] is inspired by the statistical analysis of multivariate data, particularly the estimation of specific linear dependence structures within a quadratic space containing the identity matrix. For this purpose, we provide a characterization of a matrix quadratic space, and determine the form of the projection of a symmetric matrix onto this space using block-spectral decomposition. We also show that projecting $\mathbf{S}$ or $\mathbf{S}_{\text{ML}}$ onto the quadratic subspace of partitioned matrices yields a positive definite, well-conditioned estimator with desirable statistical properties.

Given the significance of partitioned covariance matrices in statistics, we establish the necessary and sufficient conditions under which the linear space of partitioned matrices, defined in (1.4), constitutes a quadratic space. Finally, we demonstrate, how to block-diagonalize partitioned matrices belonging to this space.

The concept of a matrix quadratic space in the sense of a Jordan algebra appears in Jensen (1988), where it is defined for Hermitian matrices over the real numbers, complex numbers, or quaternions. Since statistical applications typically involve real matrices, in the following proposition we provide a more suitable characterization of the relevant Jordan algebra (quadratic space) using block-diagonal representation of this space.

Proposition D.1. *For a given quadratic space $\mathcal{Q} \subseteq \mathbb{R}_m^s$, there exist integers k and $r_i, p_i, i \in \{1, \dots, k\}$, and $\sum_{i=1}^k r_i p_i = m$, such that $\mathcal{Q}$ is isomorphic to the space*

$$\mathcal{D} = \left\{ \text{Bdiag}_{i=1}^k (\mathbf{I}_{r_i} \otimes \boldsymbol{\Psi}_i) : \boldsymbol{\Psi}_i \in \mathbb{R}_{p_i}^s \right\}.$$

Moreover, the multiset $(r_i, p_i), i \in \{1, \dots, k\}$, is unique up to permutation.

Observe that, since $\mathcal{Q}$ is isomorphic to $\mathcal{D}$, every matrix $\mathbf{A} \in \mathcal{Q}$ can be written as

$$\mathbf{A} = \mathbf{P} \text{Bdiag}_{i=1}^k (\mathbf{I}_{r_i} \otimes \boldsymbol{\Psi}_i) \mathbf{P}', \quad (\text{D.1})$$

where $\mathbf{P}$ is an orthogonal matrix of order m . Moreover, if $\mathbf{P}$ is partitioned into semi-orthogonal sub-blocks, we have

$$\mathbf{P} = (\mathbf{P}_1, \dots, \mathbf{P}_k), \quad (\text{D.2})$$

where $\mathbf{P}_i \in \mathbb{R}_{m \times r_i p_i}$, $(\mathbf{P}_i \mathbf{P}_i')^2 = \mathbf{P}_i \mathbf{P}_i'$, $\mathbf{P}_i' \mathbf{P}_i = \mathbf{I}_{r_i p_i}$ and for $i \neq j$, $\mathbf{P}_i' \mathbf{P}_j = \mathbf{0}$, and we can represent $\mathbf{A}$ in block-spectral decomposition form as

$$\mathbf{A} = \sum_{i=1}^k \mathbf{P}_i (\mathbf{I}_{r_i} \otimes \Psi_i) \mathbf{P}_i'.$$

Using representation (D.1), an explicit form of the projection of a symmetric matrix onto quadratic subspace, together with its properties related to definiteness and conditioning, is given in the following proposition. To prove the relationship between condition numbers, the properties of majorization, presented in Section 1.5, are applied.

Proposition D.2. *Let $\mathcal{Q}$ be a quadratic space containing the identity matrix, and let $\mathcal{D}$ be its isomorphic representation given in Proposition D.1. The projection of a p.d. matrix $\mathbf{X}$ onto the space $\mathcal{Q}$ is given by*

$$\Pi_{\mathcal{Q}}(\mathbf{X}) = \mathbf{P} \text{Bdiag}_{i=1}^k (\mathbf{I}_{r_i} \otimes \frac{1}{r_i} \sum_{j=1}^{r_i} (\mathbf{P}_i' \mathbf{X} \mathbf{P}_i)_{jj}) \mathbf{P}',$$

where $(\mathbf{P}' \mathbf{X} \mathbf{P})_{jj}$ are the diagonal blocks of $\mathbf{P}' \mathbf{X} \mathbf{P}$ of order corresponding to the block-diagonal representation of $\mathcal{D}$, and $\mathbf{P}$ is partitioned as in (D.2). Moreover, the following properties hold:

- (a) $\Pi_{\mathcal{Q}}(\mathbf{X})$ is positive definite;
- (b) $\kappa(\Pi_{\mathcal{Q}}(\mathbf{X})) \leq \kappa(\mathbf{X})$.

In the following, we focus on the linear space $\mathcal{T}$ of partitioned matrices defined in (1.4) and we characterize their diagonal block elements, for which $\mathcal{T}$ forms a quadratic subspace.

Lemma D.3. *If $\mathcal{T}$ is a quadratic space, then $\mathcal{T}_{11}$ and $\mathcal{T}_{22}$ are quadratic subspaces. Moreover, if $\mathcal{T}$ is a commutative quadratic space, then $\mathcal{T}_{11}$ and $\mathcal{T}_{22}$ are also commutative quadratic subspaces.*

Before we formulate a necessary and sufficient condition for a linear space $\mathcal{T}_{12}$ to constitute a quadratic space $\mathcal{T}$, we provide some notation.

Let $\mathcal{T}_{11}$ and $\mathcal{T}_{22}$ be quadratic subspaces. Then matrices $\mathbf{T}_{11} \in \mathcal{T}_{11}$ and $\mathbf{T}_{22} \in \mathcal{T}_{22}$ can be represented as

$$\begin{aligned} \mathbf{T}_{11} &= \sum_{i=1}^{\ell} \lambda_i \mathbf{V}_i \mathbf{V}_i' + \sum_{i=\ell+1}^k \mathbf{V}_i (\mathbf{I}_{r_i} \otimes \Lambda_i) \mathbf{V}_i', \\ \mathbf{T}_{22} &= \sum_{i'=1}^{\ell'} \delta_{i'} \mathbf{W}_{i'} \mathbf{W}_{i'}' + \sum_{i'=\ell'+1}^{k'} \mathbf{W}_{i'} (\mathbf{I}_{r_{i'}} \otimes \Lambda_{i'}) \mathbf{W}_{i'}', \end{aligned} \tag{D.3}$$

where $\lambda_i, \mathbf{\Lambda}_i \in \mathbb{R}_{p_i \times p_i}, \delta_{i'}, \mathbf{\Lambda}_{i'} \in \mathbb{R}_{p_{i'} \times p_{i'}}$ are the eigenvalues and block-eigenvalues of $\mathbf{T}_{11}$ and $\mathbf{T}_{22}$, respectively. The matrices $\mathbf{V}_i$ and $\mathbf{W}_{i'}$ are the corresponding eigenvectors and block-eigenvectors of $\mathbf{T}_{11}$ and $\mathbf{T}_{22}$.

We define the following index sets:

$$\mathcal{J}_{11} = \{(i, i') : i \in \{1, \dots, \ell\}, i' \in \{1, \dots, \ell'\}\},$$

$$\mathcal{J}_{21} = \{(i, i') : i \in \{\ell + 1, \dots, k\}, i' \in \{1, \dots, \ell'\}\},$$

$$\mathcal{J}_{12} = \{(i, i') : i \in \{1, \dots, \ell\}, i' \in \{\ell' + 1, \dots, k'\}\},$$

$$\mathcal{J}_{22} = \{(i, i') : i \in \{\ell + 1, \dots, k\}, i' \in \{\ell' + 1, \dots, k'\}\}.$$

For $v, w \in \{1, 2\}$, we define $\mathcal{J}_{vw}^*$ to be the subset of $\mathcal{J}_{vw}$ containing only those elements in which no index i or i' appear more than once.

Let $\mathcal{T} = \mathcal{T}_1 \boxplus \mathcal{T}_2 \boxplus \mathcal{T}_3$ be a quadratic space, and let $\mathbf{T} \in \mathcal{T}$. Then,

$$\mathbf{T}^2 = \begin{pmatrix} \mathbf{T}_{11}^2 + \mathbf{T}_{12}\mathbf{T}'_{12} & \mathbf{T}_{11}\mathbf{T}_{12} + \mathbf{T}_{12}\mathbf{T}_{22} \\ \mathbf{T}'_{12}\mathbf{T}_{11} + \mathbf{T}_{22}\mathbf{T}'_{12} & \mathbf{T}'_{12}\mathbf{T}_{12} + \mathbf{T}_{22}^2 \end{pmatrix} \in \mathcal{T}.$$

By Lemma D.3 it follows that $\mathbf{T}_{11}^2 \in \mathcal{T}_{11}$ and $\mathbf{T}_{22}^2 \in \mathcal{T}_{22}$. Moreover,

$$\mathbf{T}_{12}\mathbf{T}'_{12} \in \mathcal{T}_{11}, \quad \mathbf{T}'_{12}\mathbf{T}_{12} \in \mathcal{T}_{22}.$$

This requires that $r_i = r_{i'}$ in (D.3) and imposes a specific structure on $\mathbf{T}_{12} \in \mathcal{T}_{12}$, given by $\mathbf{T}_{12} = \sum_{(i,i') \in \mathcal{J}^*} \mathbf{V}_i(\mathbf{\Gamma}_{ii'} \otimes \mathbf{G}_{ii'})\mathbf{W}'_{i'}$, where $\mathcal{J}^* \subseteq \mathcal{J}_{11}^* \cup \mathcal{J}_{21}^* \cup \mathcal{J}_{12}^* \cup \mathcal{J}_{22}^*$. Here $\mathbf{\Gamma}_{ii'}$ is an orthogonal matrix of order r_i , and $\mathbf{G}_{ii'} \in \mathbb{R}_{p_i \times p_{i'}}$ which becomes a scalar $\gamma_{ii'}$ if $(i, i') \in \mathcal{J}_{11}^*$, a column vector $\mathbf{g}_{ii'}$ if $(i, i') \in \mathcal{J}_{21}^*$, a row vector $\mathbf{g}'_{ii'}$ if $(i, i') \in \mathcal{J}_{12}^*$, and an arbitrary matrix $\mathbf{G}_{ii'}$ of appropriate size if $(i, i') \in \mathcal{J}_{22}^*$.

In other words, $\mathcal{T}_{12}$ is a subspace of $\mathcal{T}_{12}^{11} \boxplus \mathcal{T}_{12}^{21} \boxplus \mathcal{T}_{12}^{12} \boxplus \mathcal{T}_{12}^{22}$, where

$$\mathcal{T}_{12}^{vw} = \left\{ \sum_{(i,i') \in \mathcal{J}_{vw}^*} \mathbf{V}_i(\mathbf{\Gamma}_{ii'} \otimes \mathbf{G}_{ii'})\mathbf{W}'_{i'} : \mathbf{\Gamma}_{ii'}\mathbf{\Gamma}'_{ii'} = \mathbf{\Gamma}'_{ii'}\mathbf{\Gamma}_{ii'} = \mathbf{I}_{r_i}, \mathbf{G}_{ii'} \in \mathbb{R}_{p_i \times p_{i'}} \right\},$$

$v, w \in \{1, 2\}$.

The main result of the paper [D] is the following theorem.

Theorem D.4. *The linear space $\mathcal{T}$ defined in (1.4), is a quadratic subspace if and only if $\mathcal{T}_{11}$ and $\mathcal{T}_{22}$ are quadratic subspaces and $\mathcal{T}_{12}$ is a linear subspace of $\mathcal{T}_{12}^{11} \boxplus \mathcal{T}_{12}^{21} \boxplus \mathcal{T}_{12}^{12} \boxplus \mathcal{T}_{12}^{22}$. Moreover, $\mathcal{T}$ is commutative if and only if $\mathcal{T}_1$ and $\mathcal{T}_2$ are commutative, and $\dim(\mathcal{T}_3) = 0$.*

Using representation (D.3) and Theorem D.4, we can give a general formula for the block-eigenvalue characterization of $\mathbf{T} \in \mathcal{T}$ given in Proposition D.1, namely

$$\text{Bdiag} \left(\lambda_i \mathbf{I}_{r_i}, \delta_{i'} \mathbf{I}_{r_{i'}}, \mathbf{I}_{r_i} \otimes \mathbf{\Lambda}_i, \mathbf{I}_{r_{i'}} \otimes \mathbf{\Delta}_{i'}, \mathbf{I}_{r_i} \otimes \begin{pmatrix} \lambda_i & \boldsymbol{\gamma}_{ii'} \\ \boldsymbol{\gamma}_{ii'} & \delta_{i'} \end{pmatrix}, \right. \\ \left. \mathbf{I}_{r_i} \otimes \begin{pmatrix} \mathbf{\Lambda}_i & \mathbf{g}_{ii'} \\ \mathbf{g}_{ii'}' & \delta_{i'} \end{pmatrix}, \mathbf{I}_{r_i} \otimes \begin{pmatrix} \lambda_i & \mathbf{g}_{ii'}' \\ \mathbf{g}_{ii'} & \mathbf{\Delta}_{i'} \end{pmatrix}, \mathbf{I}_{r_i} \otimes \begin{pmatrix} \mathbf{\Lambda}_i & \mathbf{G}_{ii'}' \\ \mathbf{G}_{ii'} & \mathbf{\Delta}_{i'} \end{pmatrix} \right)$$

with $i \in \{1, \dots, k\}$ and $i' \in \{1, \dots, k'\}$.

Coming back to the applications of the above algebraic results in statistics, the following proposition is proven.

Proposition D.5. *Let $\mathcal{T}$ defined in (1.4) be a quadratic space. The projection $\mathbf{\Pi}_{\mathcal{T}}(\mathbf{S})$, is the best unbiased quadratic estimator of the covariance structure $\mathbf{\Gamma} \in \mathcal{T}$. Moreover, the projection $\mathbf{\Pi}_{\mathcal{T}}(\mathbf{S}_{\text{ML}})$ is the ML estimator of $\mathbf{\Gamma} \in \mathcal{T}$.*

Finally, since the inverse of any positive definite matrix in the quadratic space retains the same block structure, the presented results can also be applied in graphical models. Thus, potential applications of these results in graph theory are discussed in [D].

[E] On estimation of partitioned covariance matrix with linearly structured blocks

The aim of the paper [E] is to propose a fast estimation method of structured matrices from the linear space $\mathcal{T}$ defined in (1.4), which is not necessarily a quadratic subspace. Obviously, if $\mathcal{S}$ is a linear space of symmetric matrices defined in Section 1.2, in our consideration $\mathcal{T}$ must be a proper subspace of $\mathcal{S}$. Similarly to [B] and [C], to provide p.d. estimators, we improve the estimator obtained by projection of $\mathbf{S}$ or $\mathbf{S}_{\text{ML}}$ onto the space $\mathcal{T}$ by shrinking. However, in comparison to the results presented in [B] and [C], the paper [E] offers the following two extensions: the off-diagonal block of $\mathbf{\Gamma}$ does not have any specific structure, and the space of target matrices is an arbitrary quadratic subspace, with characterization given in [D].

Let $\mathcal{T} \subset \mathcal{S}$ be a linear, nonquadratic subspace of $\mathcal{S}$, and let $\mathcal{Q}$ be a quadratic subspace of $\mathcal{T}$. By $\mathbf{S}$ we denote a sample covariance matrix defined in Section 1.1, and by $\mathbf{\Pi}_{\mathcal{A}}(\mathbf{S})$ we denote the projection of $\mathbf{S}$ onto $\mathcal{A}$ defined in (1.14).

To determine the general formula for the optimal target matrix and shrinkage intensity, we use the quadratic risk defined in (1.15). By minimizing $\mathbb{E} \left[\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|^2 \right]$ and using the unbiasedness of the sample covariance matrix $\mathbf{S}$ together with Jensen's inequality, the following theorem is proven.

Theorem E.1. *For a given $\mathbf{\Gamma} \in \mathcal{T}$ and some $\mathbf{T} \in \mathcal{Q} \subseteq \mathcal{T}$ the minimum of the quadratic risk, i.e.,*

$$\min_{\mathbf{T} \in \mathcal{Q} \subseteq \mathcal{T}, \lambda \in \mathbb{R}} \mathbb{E} \left[\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|^2 \right] \quad \text{subject to} \quad \widehat{\mathbf{\Gamma}} = \lambda \mathbf{T} + (1 - \lambda) \mathbf{\Pi}_{\mathcal{T}}(\mathbf{S}) \in \mathcal{T},$$

is attained at

$$\mathbf{T} = \mathbf{\Pi}_{\mathcal{Q}}(\mathbf{\Gamma}), \quad \lambda = \frac{\mathbb{E}[\text{tr}(\mathbf{\Pi}_{\mathcal{T}}^2(\mathbf{S}))] - \|\mathbf{\Pi}_{\mathcal{T}}(\mathbf{\Gamma})\|^2}{\mathbb{E}[\text{tr}(\mathbf{\Pi}_{\mathcal{T}}^2(\mathbf{S}))] - \|\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{\Gamma})\|^2}.$$

Moreover, $\lambda \in [0, 1]$.

In view of the above result, assume now that the target matrix, $\mathbf{T}$, is a projection of $\mathbf{\Gamma}$ onto the quadratic space $\mathcal{Q}$. Then, under the assumption of normality and using the properties of the Wishart distribution presented in Section 1.3, the following proposition is proven.

Proposition E.2. *For a given $\mathbf{\Gamma} \in \mathcal{T}$ and $\mathcal{Q} \subseteq \mathcal{T}$, the minimum of the quadratic risk with respect to shrinkage intensity, i.e.,*

$$\min_{\lambda \in \mathbb{R}} \mathbb{E} \left[\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|^2 \right] \quad \text{subject to} \quad \widehat{\mathbf{\Gamma}} = \lambda \mathbf{\Pi}_{\mathcal{Q}}(\mathbf{\Gamma}) + (1 - \lambda) \mathbf{\Pi}_{\mathcal{T}}(\mathbf{S}) \in \mathcal{T},$$

is attained at

$$\lambda = \frac{2 \sum_{i=1}^v \text{tr}((\mathbf{V}_i \mathbf{\Gamma})^2)}{2 \sum_{i=1}^v \text{tr}((\mathbf{V}_i \mathbf{\Gamma})^2) + (n-1)(\|\mathbf{\Pi}_{\mathcal{T}}(\mathbf{\Gamma})\|^2 - \|\mathbf{\Pi}_Q(\mathbf{\Gamma})\|^2)},$$

where $\mathbf{V}_1, \dots, \mathbf{V}_v$ is an orthonormal basis of $\mathcal{T}$.

Since the true $\mathbf{\Gamma}$ remains unknown, to determine the shrinkage estimator of $\mathbf{\Gamma}$, in the above proposition $\mathbf{\Gamma}$ must be replaced by $\mathbf{S}$, which yields

$$\widehat{\mathbf{\Gamma}}_S = \widehat{\lambda} \mathbf{\Pi}_Q(\mathbf{S}) + (1 - \widehat{\lambda}) \mathbf{\Pi}_{\mathcal{T}}(\mathbf{S}) \quad (\text{E.1})$$

with

$$\widehat{\lambda} = \frac{2 \sum_{i=1}^v \text{tr}[(\mathbf{V}_i \mathbf{S})^2]}{2 \sum_{i=1}^v \text{tr}[(\mathbf{V}_i \mathbf{S})^2] + (n-1)(\|\mathbf{\Pi}_{\mathcal{T}}(\mathbf{S})\|^2 - \|\mathbf{\Pi}_Q(\mathbf{S})\|^2)}. \quad (\text{E.2})$$

Moreover, using the unbiasedness and consistency of the sample covariance matrix $\mathbf{S}$ as well as the properties of the projection, the following theorem is proven.

Theorem E.3. *Let $\widehat{\mathbf{\Gamma}}_S$ given in (E.1) be a shrinkage estimator of $\mathbf{\Gamma} \in \mathcal{T}$ under normality. Then, $\widehat{\mathbf{\Gamma}}_S$ is consistent and asymptotically unbiased. Moreover, if $Q = \mathcal{T}$, then $\widehat{\mathbf{\Gamma}}_S$ is an unbiased estimator of $\mathbf{\Gamma}$.*

In addition, we prove in [E] the following lemma, showing the importance of Q being a quadratic subspace of $\mathcal{T}$. The proof is based on the property of preserving definiteness of projection onto the quadratic subspace (cf. Fuglede and Jensen, 2013, Theorem 1) and Proposition D.2.

Lemma E.4. *Let $\widehat{\mathbf{\Gamma}}_S$ given in (E.1) be a shrinkage estimator of $\mathbf{\Gamma} \in \mathcal{T}$. A necessary condition to achieve positive definiteness or improve the conditioning of $\widehat{\mathbf{\Gamma}}_S$ is for Q to be a quadratic subspace.*

Similarly as in [B] and [C], the shrinkage estimator is not guaranteed to be p.d. and well-conditioned. Therefore, we propose to rescale $\widehat{\lambda}$ by $(1 + \eta)$, where η is chosen as small as possible to guarantee the positive definiteness and well-conditioning of the structured estimate of $\mathbf{\Gamma}$. Similarly as in previous sections, the obtained estimator with the required condition number is called the QS estimator.

To illustrate proposed estimation technique based on various target spaces, we end this section with the following example.

Example.

Let us consider a vector of observations partitioned into two sub-vectors of dimensions p and q , which may be dependent, allowing for a nonzero off-diagonal block of covariance matrix Γ . Assume that within each sub-vector the dependence between components is identical, and that the off-diagonal block is proportional to $\mathbf{1}_p \mathbf{1}_q'$. This structure, illustrated in Figure 2.1 (first column), defines the space $\mathcal{T}_{(a)}$. A stronger dependence between “nearest neighbors” leads to the structure shown in the second column, corresponding to $\mathcal{T}_{(b)}$, whereas an additional dependence between the last variable of the first sub-vector and the first variable of the second sub-vector (third column) defines the space $\mathcal{T}_{(c)}$, with $\mathcal{T}_{(a)} \subset \mathcal{T}_{(b)} \subset \mathcal{T}_{(c)}$.

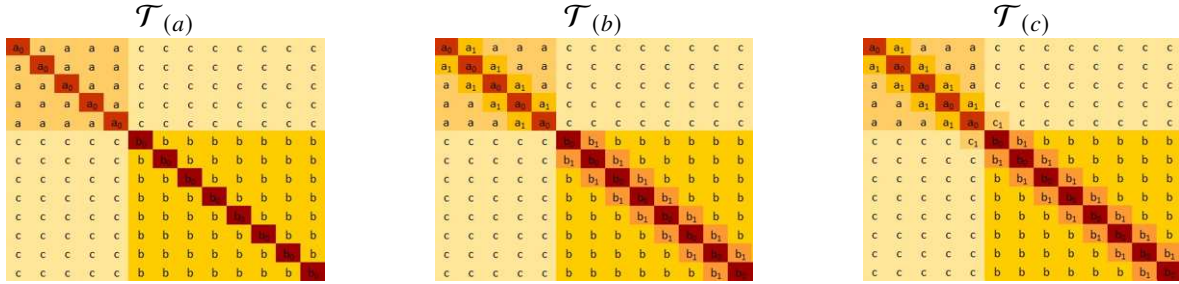


Figure 2.1: Visualization of covariance structures: $\mathcal{T}_{(a)}$ – the same dependence between the pairs of components of sub-vectors and the same dependence between components of different sub-vectors; $\mathcal{T}_{(b)}$ – additional noise on the first sub- and super-diagonals within each block; $\mathcal{T}_{(c)}$ – additional noise on the first sub- and super-diagonals of the whole covariance matrix.

In this section, we focus on the estimation of structures from $\mathcal{T}_{(c)}$, while $\mathcal{T}_{(b)}$ is presented as an example of a space that is not quadratic. For illustrative purposes the following four quadratic subspaces are chosen as the quadratic target spaces in the structured shrinkage estimation method:

$$\begin{aligned}
 Q_I &= \{\alpha \mathbf{I}_m : \alpha \in \mathbb{R}\}, \\
 Q_{II} &= \left\{ \begin{pmatrix} \alpha \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \beta \mathbf{I}_q \end{pmatrix} : \alpha, \beta \in \mathbb{R} \right\}, \\
 Q_{III} &= \left\{ \begin{pmatrix} \alpha_1 \mathbf{I}_p + \beta_1 \mathbf{1}_p \mathbf{1}_p' & \mathbf{0} \\ \mathbf{0} & \alpha_2 \mathbf{I}_q + \beta_2 \mathbf{1}_q \mathbf{1}_q' \end{pmatrix} : \alpha_1, \beta_1, \alpha_2, \beta_2 \in \mathbb{R} \right\}, \\
 Q_{IV} &= \left\{ \begin{pmatrix} \alpha_1 \mathbf{I}_p + \beta_1 \mathbf{1}_p \mathbf{1}_p' & \gamma \mathbf{1}_p \mathbf{1}_q' \\ \gamma \mathbf{1}_q \mathbf{1}_p' & \alpha_2 \mathbf{I}_q + \beta_2 \mathbf{1}_q \mathbf{1}_q' \end{pmatrix} : \alpha_1, \beta_1, \alpha_2, \beta_2, \gamma \in \mathbb{R} \right\}.
 \end{aligned}$$

Note that $Q_I \subset Q_{II} \subset Q_{III} \subset Q_{IV}$ and Q_{IV} is simply $\mathcal{T}_{(a)}$.

Projecting the sample covariance matrix $\mathbf{S}$ onto $\mathcal{T}_{(c)}$ and $\mathbf{Q}_i, i \in \{\text{I, II, III, IV}\}$, and computing optimal intensities $\widehat{\lambda}_i$ following (E.2), we obtain four shrinkage estimators,

$$\widehat{\mathbf{\Gamma}}_{S_i} = \widehat{\lambda}_i \mathbf{\Pi}_{\mathbf{Q}_i}(\mathbf{S}) + (1 - \widehat{\lambda}_i) \mathbf{\Pi}_{\mathcal{T}_{(c)}}(\mathbf{S}), \quad i \in \{\text{I, II, III, IV}\}.$$

If for a given data set the shrinkage estimator is indefinite or ill-conditioned, the final estimate should be calculated by numerical adjustment of the shrinkage intensity to obtain QS estimate.

As expected, the choice of a quadratic space of target matrices plays a crucial role in the final form of the estimator. Thus, in [E] we perform simulation studies to examine the algebraic (conditioning) and statistical (empirical risk based on loss functions defined in Section 1.4 and the Frobenius norm) properties of QS estimates of covariance structure with respect to the choice of target space and dimension of the covariance matrix. It is concluded that if the researcher is interested in an estimator which is well-conditioned, then the quadratic space the closest to the considered structure should be selected as the target space. On the other hand, if statistical properties are of main importance, a target space being a commutative quadratic subspace of the relevant structure space is more suitable.

Bibliography

- T. W. Anderson. Statistical inference for covariance matrices with linear structure. In: P. R. Krishnaiah (Ed.), *Multivariate Analysis II*, pp. 55–66, Academic Press, New York, 1969.
- T. W. Anderson. Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.*, 1(1), 135–141, 1973.
- T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, 2003.
- X. Cui, C. Lin, J. Zhao, L. Zeng, D. Zhang, J. Pan. Covariance structure regularization via Frobenius norm discrepancy. *Linear Algebra Appl.*, 510, 124–145, 2016.
- D. K. Dey, C. Srinivasan. Estimation of a covariance matrix under Stein’s loss. *Ann. Statist.*, 13(4), 1581–1591, 1985.
- B. Fuglede, S. T. Jensen. Positive projections of symmetric matrices and Jordan algebras. *Expo. Math.*, 31, 295–303, 2013.
- K. Filipiak, M. John, A. Markiewicz. Comments on maximum likelihood estimation and projections under multivariate statistical models. In: M. Singull, T. Holgersson (Eds.), *Recent Developments in Multivariate and Random Matrix Analysis*, pp. 51–66, Springer, 2020.
- K. Filipiak, D. Klein, 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.
- K. Filipiak, D. Klein, A. Roy. A comparison of likelihood ratio tests and Rao’s score test for three separable covariance matrix structures. *Biom. J.*, 59(1), 192–215, 2017.
- 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, 74–82, 2018.

- A. K. Gupta, D. K. Nagar. *Matrix Variate Distributions*. Chapman & Hall/CRC, Boca Raton, 2000.
- A. Hannart, P. Naveau. Estimating high dimensional covariance matrices: A new look at the Gaussian conjugate framework. *J. Multivariate Anal.*, 131, 149–162, 2014.
- T. Holgersson, J. Pielaszkiewicz. A Collection of moments of the Wishart distribution. In: M. Singull, T. Holgersson (Eds.), *Recent Developments in Multivariate and Random Matrix Analysis*, pp. 147–162, Springer, 2020.
- Y. Ikeda, T. Kubokawa, M. S. Srivastava. Comparison of linear shrinkage estimators of a large covariance matrix in normal and non-normal distributions. *Comput. Statist. Data Anal.*, 95, 95–108, 2016.
- W. James, C. Stein. Estimation with quadratic loss. In: J. Neyman (Ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 361–379, The Statistical Laboratory, University of California, University of California Press, 1961.
- S. T. Jensen. Covariance hypotheses which are linear in both the covariance and the inverse of covariance. *Ann. Statist.*, 16, 302–322, 1988.
- T. Kollo, D. von Rosen. *Advanced Multivariate Statistics with Matrices*. Springer, Dordrecht, 2005.
- A. Koziół, A. Roy, R. Zmyślony, R. Leiva, M. Fonseca. Best unbiased estimates for parameters of three-level multivariate data with doubly exchangeable covariance structure. *Linear Algebra Appl.*, 535, 87–104, 2017.
- O. Ledoit, M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Anal.*, 88(2), 365–411, 2004.
- O. Ledoit, M. Wolf. Shrinkage estimation of large covariance matrices: Keep it simple, statistician? Working Paper No. 327, 2021.
- L. Lin, N. J. Higham, J. Pan. Covariance structure regularization via entropy loss function. *Comput. Statist. Data Anal.*, 72, 315–327, 2014.
- J. R. Magnus, H. Neudecker. Symmetry, 0-1 matrices and Jacobians: a review. *Econometric Theory*, 2, 157–190, 1986.

- J. R. Magnus, H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, Chichester, 1988.
- A. W. Marshall, I. Olkin, B. C. Arnold. *Inequalities: Theory of Majorization and Its Applications*. 2nd ed., Springer, 2011.
- M. Ohlson, D. von Rosen. Explicit estimators of parameters in the Growth Curve model with linearly structured covariance matrices. *J. Multivariate Anal.*, 101, 1284–1295, 2010.
- I. Olkin. Testing and estimation for structures which are circular symmetric in blocks. In: D. G. Kabe and R. P. Gupta (Eds.), *Multivariate Statistical Inference*, pp. 183–195, North-Holland, 1973.
- J.-X. Pan, K.-T. Fang. *Growth Curve Models and Statistical Diagnostics*. Springer, New York, 2002.
- J. Pielaszkiewicz, D. von Rosen. Multivariate moments in multivariate analysis. In: K. Filipiak, A. Markiewicz, D. von Rosen (Eds.), *Multivariate, Multilinear and Mixed Linear Models*, pp. 41–92, Springer, 2021.
- A. Roy, R. Leiva. Likelihood ratio tests for triply multivariate data with structured correlation on spatial repeated measurements. *Statist. Probab. Lett.*, 78(13), 1971–1980, 2008.
- J. Seely. Quadratic Subspaces and Completeness. *Ann. Math. Statist.*, 42(2), 710–721, 1971.
- C. Stein. *On Some Problems in Multivariate Analysis. Part I*. Technical Report, Department of Statistics, Stanford University, 1956.
- T. H. Szatrowski. Testing and estimation in the block compound symmetry problem. *J. Educ. Stat.*, 7(1), 3–18, 1982.
- S. S. Wilks. Certain generalizations in the analysis of variance. *Biometrika*, 24(3–4), 471–494, 1932.
- S. S. Wilks. On the independence of k sets of normally distributed statistical variables. *Econometrica*, 3(3), 309–326, 1935.

Part II

Main results – full text of articles

[A] **Block matrix approximation via entropy loss function**

by Malwina Janiszewska, Augustyn Markiewicz, and Monika Mokrzycka

Applications of Mathematics, **65(6)**, 829–844, 2020.

Contribution:

- Malwina Janiszewska (40%) – literature review, formulation and proofs of theoretical results (Theorems 3.1, 3.2), implementation of numerical algorithms in R, simulation studies, interpretation of the results, writing – original draft preparation and editing
- Augustyn Markiewicz (20%) – conceptualization and methodological supervision
- Monika Mokrzycka (40%) – conceptualization, methodology, formulation and proofs of theoretical results (Theorems 3.1, 3.2), implementation of numerical algorithms in R, simulation studies, interpretation of the results, writing – editing and general supervision

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.

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

$$(2.1) \quad \mathbf{\Psi} = \delta \mathbf{1}_p \mathbf{1}_q',$$

▷ the matrix of the form

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

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

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

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

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

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 $\mathbf{B}$

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

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

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

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

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

with two variants of Schur complement (2.6) used.

3. APPROXIMATION

In this section we present the best approximation of a given $\mathbf{\Omega} \in \mathbb{R}_m^{\geq}$ 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

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

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:

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

where $\mathbf{D}_p$ is a $(p+q)^2 \times (p+q)(p+q+1)/2$ duplication matrix and $\text{vec}(\cdot)$ 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 $\mathbf{\Gamma}_1, \mathbf{\Gamma}_2, \delta$ satisfying the following system of equations:*

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

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

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

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 $\mathbf{\Gamma}_1, \mathbf{\Gamma}_2, \delta$

and ϱ . Derivatives have the following forms:

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

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

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

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 $\mathbf{\Gamma}$ from Theorem 3.2 can be only a local minimum (cf. Lin et al. (2014) [12]).

Theorem 3.2. *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 $\mathcal{S}_2$ and this minimum is attained at $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$, δ , ϱ satisfying the following system of equations:*

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

Observe that the equations received by differentiating the entropy loss function with respect to Γ_1 , Γ_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, $i = 1, \dots, n$. 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 $\boldsymbol{\Omega}$, i.e. $\mathbf{x}_i \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Omega})$ with $m = p + q$. Using vectors of observations for each object, we construct the observation matrix as $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ having matrix normal distribution $N_{m,n}(\boldsymbol{\mu}\mathbf{1}'_n, \boldsymbol{\Omega}, \mathbf{I}_n)$. We are interested in relations between characteristics from the first and second group described in the covariance matrix $\boldsymbol{\Omega}_{12}$, 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 $\boldsymbol{\Omega}$ 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 $\boldsymbol{\Omega}$, i.e. $\mathbf{X} \sim N_{p+q,n}(\mathbf{0}, \boldsymbol{\Omega}, \mathbf{I}_n)$.

The choice of the most appropriate structure is based on the estimation of matrix $\boldsymbol{\Omega}$ and determination of the nearest structure in the sense of some discrepancy function. In real life applications, the covariance matrix $\boldsymbol{\Omega}$ is usually unknown; thus we replace it with a maximum likelihood estimator of $\boldsymbol{\Omega}$, i.e. $\mathbf{S} = \frac{1}{n}\mathbf{X}(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n)\mathbf{X}'$. By minimizing 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

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

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 $\Gamma_2 = \mathbf{I}_q$, while in Theorem 3.2: $\Gamma_2 = \mathbf{I}_q$ and $\varrho = 0.5$. The other variables, δ and Γ_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 Ω we generate 1000 matrices $\mathbf{X}$ and for each of them we determine the maximum likelihood estimator of Ω , i.e. the matrix $\mathbf{S}$ and discrepancies ζ_1 and ζ_2 . Then the averaged results of discrepancies, i.e. $\bar{\zeta}_1$ and $\bar{\zeta}_2$, are computed.

We expect that when the true matrix $\Omega \in \mathcal{S}_1$ 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 $\Omega \in \mathcal{S}_2$ we expect the opposite rule, that is

$$\begin{aligned}\Omega \in \mathcal{S}_1 &\Rightarrow \bar{\zeta}_1 \leq \bar{\zeta}_2, \\ \Omega \in \mathcal{S}_2 &\Rightarrow \bar{\zeta}_2 \leq \bar{\zeta}_1.\end{aligned}$$

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 $\Omega \in \mathcal{S}_1$ 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 $\Omega \in \mathcal{S}_2$ are presented. We can see that the averaged discrepancies satisfy $\bar{\zeta}_2 \leq \bar{\zeta}_1$ for each value of parameters p , q and δ (except one case). Moreover, the true $\Omega \in \mathcal{S}_2$ 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

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

and

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

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

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

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

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

and

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

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 $\hat{\mathbf{\Gamma}}_{AR}$, $\zeta_1 = 0.07429 > 0.0447 = \zeta_2$.

On the other hand, the discrepancy values obtained in simulation studies are rather small. This means that the considered approximations ($\hat{\mathbf{\Gamma}}_{CS}$ and $\hat{\mathbf{\Gamma}}_{AR}$) 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.

q	δ	$p = 2$		π	$p = 3$		π
		$\bar{\zeta}_1$	$\bar{\zeta}_2$		$\bar{\zeta}_1$	$\bar{\zeta}_2$	
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 $\bar{\zeta}_1$, $\bar{\zeta}_2$ and proportion π (in %) of proper structure identification among 1000 simulation runs in the case of $\Omega \in \mathcal{S}_1$ with $n = 100$ for various values of δ .

p	q	ϱ	$\delta = 1$			$\delta = 2$		
			$\bar{\zeta}_1$	$\bar{\zeta}_2$	π	$\bar{\zeta}_1$	$\bar{\zeta}_2$	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

p	q	ϱ	$\delta = 1$			$\delta = 2$		
			$\bar{\zeta}_1$	$\bar{\zeta}_2$	π	$\bar{\zeta}_1$	$\bar{\zeta}_2$	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 $\bar{\zeta}_1$, $\bar{\zeta}_2$ 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 classification 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.



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.

[B] Quasi shrinkage estimation of a block-structured covariance matrix

by Augustyn Markiewicz, Monika Mokrzycka, and Malwina Mrowińska

Journal of Statistical Computation and Simulation, **94(14)**, 3093–3110, 2024.

Contribution:

- Augustyn Markiewicz (20%) – conceptualization, methodological supervision, proposing adaptation of the shrinkage estimation method to block-structured matrices to determine positive definite estimators of covariance structure
- Monika Mokrzycka (40%) – methodology, formulation and proofs of theoretical results (Theorems 3.1, 4.1, 4.3), proposing algebraic corrections of shrinkage estimators, proposing loss functions for estimators comparison, implementation of numerical algorithms in R, interpretation of the results of simulation studies, writing – editing and general supervision
- Malwina Mrowińska (40%) – literature review, methodology, formulation and proofs of theoretical results (Theorems 3.1, 4.1, 4.3, and Lemma 4.2), implementation of numerical algorithms in R, simulation studies, empirical risk analysis, algebraic correction analysis, writing – original draft preparation and editing



Quasi shrinkage estimation of a block-structured covariance matrix

A. Markiewicz ^a, M. Mokrzycka ^b and M. Mrowińska ^c

^aDepartment of Mathematical and Statistical Methods, Poznań University of Life Sciences, Poznań, Poland;

^bInstitute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland; ^cInstitute of Mathematics, Poznań University of Technology, Poznań, Poland

ABSTRACT

In this paper, we study the estimation of a block covariance matrix with linearly structured off-diagonal blocks. We consider estimation based on the least squares method, which has some drawbacks. These estimates are not always well conditioned and may not even be definite. We propose a new estimation procedure providing a structured positive definite and well-conditioned estimator with good statistical properties. The least squares estimator is improved with the use of a shrinkage method and an additional algebraic approach. The resulting so-called quasi shrinkage estimator is compared with the structured maximum likelihood estimator.

ARTICLE HISTORY

Received 8 November 2022

Accepted 3 June 2024

KEYWORDS

Covariance matrix; block linear structure; maximum likelihood estimation; orthogonal projection; shrinkage method

1. Introduction

Estimation of covariance matrices is widely used in multivariate data analysis and is applied in various fields including agronomy, biology, medicine, economics, and geography. Recently, more and more characteristics have come to be measured on sample units. Often the number of measurements is higher than the number of units, which leads to a problem of overparameterization. Having some knowledge about the structure, we can reduce the number of parameters. In the literature, commonly considered structures of covariance matrices include, for example, compound symmetry (cf. [1]), first-order autoregression (cf. [2]), and banded Toeplitz matrices (cf. [3,4]). Other structures considered include a separability (Kronecker product) structure (cf. [5–9]) and the block compound symmetry structure (cf. [10–12]). We consider the specific partitioning of the matrix into four sub-blocks, with squared diagonal blocks, not necessarily of the same order; cf. [13,14]. The diagonal unstructured blocks are arbitrary symmetric square matrices and the off-diagonal blocks have some structure. Our main object of interest is the cross-covariance matrix, and more precisely, inference about the structured off-diagonal blocks. The structure we focus on is linear, with the off-diagonal blocks consisting of all elements equal (they can be considered as off-diagonal blocks of the compound symmetry structure), and arbitrary diagonal blocks. In the case of a structured covariance matrix, the common maximum likelihood estimation procedure can lead to a complex numerical problem. In this paper,

CONTACT M. Mrowińska  malwina.mrowinska@put.poznan.pl

© 2024 Informa UK Limited, trading as Taylor & Francis Group

we consider the computationally simpler least squares method; cf. [15]. This method has good statistical properties; however, the estimator obtained may be indefinite. We propose an improvement to this method, based on shrinkage and an additional algebraic approach, giving an estimator with good statistical properties that is also positive definite and well conditioned. We call the new approach the quasi shrinkage method, and the new estimator the quasi shrinkage estimator.

The paper is organized as follows. Section 2 presents the statistical background. The results of maximum likelihood estimation and the least squares estimation and its improvements are presented in Sections 3 and 4, respectively. A comparison of the estimators based on a simulation study is given in Section 5. Conclusions and recommendations are discussed in the final section.

2. Statistical background

Assume an experiment in which characteristics are measured on n experimental units. An observation has the form of a vector $\mathbf{x}_i$, $i = 1, \dots, n$, that can be divided into two subvectors, $\mathbf{x}_i = (\mathbf{x}'_{ip}, \mathbf{x}'_{iq})'$. These correspond to two groups of characteristics $\mathbf{x}_{ip}$ and $\mathbf{x}_{iq}$, with numbers of features p and q , respectively. The observations from different experimental units are independent and have the normal distribution

$$\mathbf{x}_i \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Omega}),$$

where $m = p + q$, $\boldsymbol{\mu} = (\boldsymbol{\mu}'_p, \boldsymbol{\mu}'_q)'$ is an unknown mean vector, and $\boldsymbol{\Omega}$ is an unknown positive definite (p.d.) covariance matrix. The matrix of all observations $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ is matrix normally distributed

$$\mathbf{X} \sim N_{m,n}(\boldsymbol{\mu} \mathbf{1}'_n, \boldsymbol{\Omega}, \mathbf{I}_n),$$

where $\mathbf{1}_n$ is an n -dimensional vector of ones and $\mathbf{I}_n$ is an identity matrix of order n .

We focus on the dependency between two groups of characteristics, which is given by the off-diagonal block of the covariance matrix

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}'_{12} & \boldsymbol{\Omega}_{22} \end{pmatrix}.$$

The covariance matrix $\boldsymbol{\Omega}$ has $m(m+1)/2$ unknown parameters to estimate, while in the case of a structured covariance matrix, this number decreases to the dimension of the structure.

Following [16], we define a linear covariance structure as

$$\Theta = \{\boldsymbol{\Gamma} \in \mathcal{S} | \boldsymbol{\Gamma} \text{ is positive definite}\},$$

where $\mathcal{S}$ is a linear subspace of the space of symmetric matrices defined by [17].

We consider the following structured matrices:

$$\boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{\Gamma}_1 & \gamma \mathbf{1}_p \mathbf{1}'_q \\ \gamma \mathbf{1}_q \mathbf{1}'_p & \boldsymbol{\Gamma}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Gamma}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_2 \end{pmatrix} + \gamma \begin{pmatrix} \mathbf{0} & \mathbf{1}_p \mathbf{1}'_q \\ \mathbf{1}_q \mathbf{1}'_p & \mathbf{0} \end{pmatrix}, \quad (1)$$

where $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ are symmetric p.d. matrices of order p and q , respectively, and $\gamma \in \mathbb{R}$. In [18], a similar cross-covariance structure was considered for metabolomics data.

We assume that $\mathbf{\Gamma}$ is a p.d. covariance matrix, thus $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ are p.d. matrices and γ is such that $\mathbf{\Gamma}$ is p.d.

Based on Theorem 1.12. of [19,20] we give Lemma 2.1.

Lemma 2.1: *Let*

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

be a $(p+q) \times (p+q)$ partitioned matrix. If $\mathbf{A}_{11} : p \times p$ and $\mathbf{C}_{11 \bullet 2}$ are p.d. or $\mathbf{A}_{22} : q \times q$ and $\mathbf{C}_{22 \bullet 1}$ are p.d., then the matrix $\mathbf{A}$ is p.d., where the respective Schur complements are given in the following forms:

$$\begin{aligned} \mathbf{C}_{11 \bullet 2} &= \mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}'_{12}, \\ \mathbf{C}_{22 \bullet 1} &= \mathbf{A}_{22} - \mathbf{A}'_{12} \mathbf{A}_{11}^{-1} \mathbf{A}_{12}. \end{aligned}$$

Observe that, if $\mathbf{\Gamma}$ is positive definite, then the Schur complements of $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ in $\mathbf{\Gamma}$:

$$\begin{aligned} \mathbf{C}_{11 \bullet 2} &= \mathbf{\Gamma}_1 - \gamma^2 \mathbf{1}_p \mathbf{1}'_q \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}'_p \\ \mathbf{C}_{22 \bullet 1} &= \mathbf{\Gamma}_2 - \gamma^2 \mathbf{1}_q \mathbf{1}'_p \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}'_q \end{aligned} \quad (2)$$

are positive definite matrices. Thus, for any q -dimensional vector $\mathbf{v}$ it should hold that

$$\mathbf{v}' \mathbf{\Gamma}_2 \mathbf{v} - \gamma^2 \mathbf{v}' \mathbf{1}_q \mathbf{1}'_p \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}'_q \mathbf{v} > 0,$$

and after transformation, we obtain

$$\gamma^{-2} < \mathbf{1}'_p \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \frac{\mathbf{v}' \mathbf{1}_q \mathbf{1}'_q \mathbf{v}}{\mathbf{v}' \mathbf{\Gamma}_2 \mathbf{v}}$$

Now, following [21, Theorem 2.5], we obtain that

$$\max_{\mathbf{v}} \frac{\mathbf{v}' \mathbf{1}_q \mathbf{1}'_q \mathbf{v}}{\mathbf{v}' \mathbf{\Gamma}_2 \mathbf{v}} = \mathbf{1}'_q \mathbf{\Gamma}_2^{-1} \mathbf{1}_q$$

and consequently

$$\gamma^{-2} < \frac{1}{\mathbf{1}'_p \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}'_q \mathbf{\Gamma}_2^{-1} \mathbf{1}_q}.$$

$$\text{Therefore, } \gamma \in \left(-\frac{1}{\sqrt{\mathbf{1}'_p \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}'_q \mathbf{\Gamma}_2^{-1} \mathbf{1}_q}}; \frac{1}{\sqrt{\mathbf{1}'_p \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}'_q \mathbf{\Gamma}_2^{-1} \mathbf{1}_q}} \right).$$

Our purpose is to determine a structured p.d. and well-conditioned estimator of the covariance matrix with good statistical properties. Therefore, we are looking for an estimator

where the condition number $\kappa(\mathbf{A})$ of the form

$$\kappa(\mathbf{A}) = \frac{\lambda_{\max}(\mathbf{A})}{\lambda_{\min}(\mathbf{A})}, \quad (3)$$

with maximal ($\lambda_{\max}(\mathbf{A})$) and minimal ($\lambda_{\min}(\mathbf{A})$) eigenvalues of the matrix $\mathbf{A}$, is smaller than a fixed κ_0 . Thus, we consider the structure $\Theta_0 \subset \Theta$:

$$\Theta_0 = \{\mathbf{\Gamma} \in \mathcal{S} | \mathbf{\Gamma} \text{ is p.d. and } \kappa(\mathbf{\Gamma}) < \kappa_0\}.$$

The unknown true $\mathbf{\Gamma}$ will be estimated in four steps. In the first step, we determine the maximum likelihood estimator (MLE) of the unstructured covariance matrix $\mathbf{S} = \frac{1}{n}\mathbf{X}(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n')\mathbf{X}'$. In the second step, we transform $\mathbf{S}$ into a structured estimator $\widehat{\mathbf{\Gamma}}_{LS}$ using least squares estimation (cf. [15]). The resulting estimator belongs to the linear space $\mathcal{B}$, which is a sum of two subspaces: the linear subspace of block-diagonal matrices, $\mathcal{B}_1$, and the subspace of off-diagonal matrices with all elements equal, $\mathcal{B}_2$:

$$\mathcal{B} = \mathcal{B}_1 \boxplus \mathcal{B}_2,$$

where

$$\begin{aligned} \mathcal{B}_1 &= \left\{ \mathbf{A} \in \mathbb{R}_{p+q}^s : \mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{pmatrix}, \mathbf{A}_1 \in \mathbb{R}_p^s, \mathbf{A}_2 \in \mathbb{R}_q^s \right\} \\ \mathcal{B}_2 &= \left\{ \mathbf{B} \in \mathbb{R}_{p+q}^s : \mathbf{B} = b \begin{pmatrix} \mathbf{0} & \mathbf{1}_p\mathbf{1}_q' \\ \mathbf{1}_q\mathbf{1}_p' & \mathbf{0} \end{pmatrix}, b \in \mathbb{R} \right\} \end{aligned}$$

with $\mathbb{R}_d^s$ being the set of symmetric matrices of order d and

$$\dim \mathcal{B} = p(p+1)/2 + q(q+1)/2 + 1.$$

Observe that this estimation method does not ensure good properties such as positive definiteness or well-conditioning. Therefore, in the third step, to improve the least square estimator, the shrinkage method is used. The shrinkage estimator $\widehat{\mathbf{\Gamma}}_S \in \mathcal{B}$ has better properties than $\widehat{\mathbf{\Gamma}}_{LS}$; however, the problem of indefiniteness may still exist. Then, in the last step, the required properties are obtained using an algebraic approach. The final estimator $\widehat{\mathbf{\Gamma}}_{QS} \in \Theta_0$ is called the quasi shrinkage estimator (QSE).

The quasi shrinkage estimator is compared with the MLE, which is obtained by maximizing the likelihood function with respect to p.d. $\mathbf{\Gamma} \in \Theta_0$:

$$\mathcal{L}(\mathbf{S}, \mathbf{\Gamma}) = (2\pi)^{-\frac{1}{2}nm} |\mathbf{\Gamma}|^{-\frac{1}{2}n} \exp \left[-\frac{1}{2} \text{tr}(\mathbf{S}\mathbf{\Gamma}^{-1}) \right]. \quad (4)$$

In general, solving the constrained problem is complex and time-consuming, and requires iterative procedures (described in Section 3). However, if Θ_0 were a quadratic space, then the MLE of the covariance matrix would be a projection of $\mathbf{S}$ onto Θ_0 , or equivalently, the MLE would be the minimum of the Frobenius norm over $\mathbf{\Gamma} \in \Theta_0$; cf. [22].

The projection onto non-quadratic space as an estimation procedure was studied in [15]. This method, called least squares, is based on minimizing the difference between the matrix $\mathbf{S}$ and the block matrix $\mathbf{\Gamma} \in \mathcal{B}$ in the sense of the Frobenius norm:

$$f(\mathbf{S}, \mathbf{\Gamma}) = \|\mathbf{S} - \mathbf{\Gamma}\|_F^2 = \text{tr}[(\mathbf{S} - \mathbf{\Gamma})(\mathbf{S} - \mathbf{\Gamma})']. \quad (5)$$

The resulting estimator $\hat{\mathbf{\Gamma}}_{LS}$ does not always have good properties such as positive definiteness and well-conditioning, and it may even be indefinite; cf. [23].

For example, for the matrix

$$\mathbf{S} = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 9 & 2 \\ 0 & 2 & 1 \end{pmatrix},$$

which is p.d., the estimate obtained by minimizing the Frobenius norm (5):

$$\hat{\mathbf{\Gamma}}_{LS} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 9 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

is indefinite.

Therefore, we propose an improvement of the estimation procedure to provide the estimator with good statistical properties, which is p.d. and well-conditioned. The next sections describe the methods used to obtain both the MLE and the QSE.

3. Maximum likelihood estimation

The MLE is obtained via maximization of the likelihood function (4) with given $\mathbf{S}$ over $\mathbf{\Gamma} \in \Theta_0$, and is denoted as $\hat{\mathbf{\Gamma}}_{ML}$. The problem of maximizing $\mathcal{L}(\mathbf{S}, \mathbf{\Gamma})$ over the considered set has the following form:

$$\mathcal{L}(\mathbf{S}, \hat{\mathbf{\Gamma}}_{ML}) = \max_{\mathbf{\Gamma} \in \Theta_0} \mathcal{L}(\mathbf{S}, \mathbf{\Gamma}).$$

It is equivalent to maximizing the log-likelihood function in the following form:

$$\ln \mathcal{L}(\mathbf{S}, \mathbf{\Gamma}) = \frac{nm}{2} \ln(2\pi) - \frac{n}{2} \ln |\mathbf{\Gamma}| - \frac{n}{2} \text{tr}(\mathbf{S}\mathbf{\Gamma}^{-1}).$$

Using respective Schur complements (2), we obtain

$$\mathbf{\Gamma}^{-1} = \begin{pmatrix} \mathbf{C}_{11\bullet 2}^{-1} & -\mathbf{\Gamma}_1^{-1} \gamma \mathbf{1}_p \mathbf{1}_q' \mathbf{C}_{22\bullet 1}^{-1} \\ -\mathbf{\Gamma}_2^{-1} \gamma \mathbf{1}_q \mathbf{1}_p' \mathbf{C}_{11\bullet 2}^{-1} & \mathbf{C}_{22\bullet 1}^{-1} \end{pmatrix}$$

and

$$\begin{aligned} |\mathbf{\Gamma}| &= |\mathbf{\Gamma}_2| \cdot |\mathbf{C}_{11\bullet 2}| \\ &= |\mathbf{\Gamma}_1| \cdot |\mathbf{C}_{22\bullet 1}|. \end{aligned}$$

The matrix $\mathbf{S}$ can be presented in partitioned form:

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

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

The maximum of the likelihood function over Θ_0 is given by Theorem 3.1.

Theorem 3.1: For a given matrix $\mathbf{S} \in \mathbb{R}^>$, there exists a matrix $\mathbf{\Gamma} \in \mathbb{R}^>$ that maximizes the likelihood function (4) over Θ_0 , and this maximum is attained at $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$, γ satisfying the following system of equations:

$$\left\{ \begin{array}{l} \mathbf{\Gamma}_1 = \mathbf{S}_{11} - \gamma \left(\mathbf{C}_{11 \bullet 2} \mathbf{\Gamma}_1^{-1} \mathbf{S}_{12} \mathbf{C}_{22 \bullet 1}^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1^{-1} \mathbf{C}_{11 \bullet 2} + \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{S}'_{12} \right) \\ \quad + \gamma^2 \left(\mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p' + \mathbf{C}_{11 \bullet 2} \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{C}_{22 \bullet 1}^{-1} \mathbf{S}_{22} \mathbf{C}_{22 \bullet 1}^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1^{-1} \mathbf{C}_{11 \bullet 2} \right) \\ \quad - \gamma^3 \mathbf{C}_{11 \bullet 2} \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{C}_{22 \bullet 1}^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1^{-1} \mathbf{S}_{12} \mathbf{C}_{22 \bullet 1} \mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1^{-1} \mathbf{C}_{11 \bullet 2} \\ \mathbf{\Gamma}_2 = \mathbf{S}_{22} - \gamma \left(\mathbf{C}_{22 \bullet 1} \mathbf{\Gamma}_2^{-1} \mathbf{S}'_{12} \mathbf{C}_{11 \bullet 2}^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{C}_{22 \bullet 1} + \mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1^{-1} \mathbf{S}_{12} \right) \\ \quad + \gamma^2 \left(\mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}_q' + \mathbf{C}_{22 \bullet 1} \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{C}_{11 \bullet 2}^{-1} \mathbf{S}_{11} \mathbf{C}_{11 \bullet 2}^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{C}_{22 \bullet 1} \right) \\ \quad - \gamma^3 \mathbf{C}_{22 \bullet 1} \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{C}_{11 \bullet 2}^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{S}'_{12} \mathbf{C}_{11 \bullet 2}^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{C}_{22 \bullet 1} \\ \gamma = \frac{-2b - \sqrt{4b^2 - 8ac}}{-4a}, \end{array} \right.$$

where

$$\begin{aligned} a &= \text{tr} \left[\mathbf{S}_{12} \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{C}_{11 \bullet 2}^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{C}_{11 \bullet 2}^{-1} + \mathbf{S}'_{12} \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{C}_{22 \bullet 1}^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{C}_{22 \bullet 1}^{-1} \right] \\ b &= \text{tr} \left[(\mathbf{C}_{11 \bullet 2}^{-1} \mathbf{S}_{11} - \mathbf{I}_p) \mathbf{C}_{11 \bullet 2}^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p' + \mathbf{S}_{22} \mathbf{C}_{22 \bullet 1}^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{C}_{22 \bullet 1}^{-1} \right] \\ c &= \text{tr} \left[\mathbf{S}_{12} \mathbf{\Gamma}_2^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{C}_{11 \bullet 2}^{-1} + \mathbf{S}'_{12} \mathbf{\Gamma}_1^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{C}_{22 \bullet 1}^{-1} \right]. \end{aligned}$$

Proof: To obtain the maximum of the log-likelihood function over Θ_0 , we differentiate $\ln \mathcal{L}(\mathbf{S}, \mathbf{\Gamma})$ with respect to $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$ and γ . The derivatives have the following forms:

$$\begin{aligned} \frac{\partial \ln \mathcal{L}}{\partial \mathbf{\Gamma}_1} &= -\frac{n}{2} \left\{ \text{vec}'(\mathbf{C}_{11 \bullet 2}^{-1}) - \text{vec}'(\mathbf{S}_{11})(\mathbf{C}_{11 \bullet 2}^{-1} \otimes \mathbf{C}_{11 \bullet 2}^{-1}) + \text{vec}'(\mathbf{S}_{12} \mathbf{\Gamma}_2^{-1} \gamma \mathbf{1}_q \mathbf{1}_p')' \right. \\ &\quad \times (\mathbf{C}_{11 \bullet 2}^{-1} \otimes \mathbf{C}_{11 \bullet 2}^{-1}) + \text{vec}'(\mathbf{S}_{12}) \cdot [(\mathbf{C}_{22 \bullet 1}^{-1} \otimes \mathbf{I}_p)(\gamma \mathbf{1}_q \mathbf{1}_p' \otimes \mathbf{I}_p)(\mathbf{\Gamma}_1^{-1} \otimes \mathbf{\Gamma}_1^{-1}) \\ &\quad + (\mathbf{I}_q \otimes \mathbf{\Gamma}_1^{-1})(\mathbf{I}_q \otimes \gamma \mathbf{1}_p \mathbf{1}_q')(\mathbf{C}_{11 \bullet 2}^{-1} \otimes \mathbf{C}_{22 \bullet 1}^{-1})(\mathbf{1}_q \mathbf{1}_p' \otimes \gamma^2 \mathbf{1}_q \mathbf{1}_p')(\mathbf{\Gamma}_1^{-1} \otimes \mathbf{\Gamma}_1^{-1})] \\ &\quad \left. - \text{vec}'(\mathbf{S}_{22})(\mathbf{C}_{22 \bullet 1}^{-1} \otimes \mathbf{C}_{22 \bullet 1}^{-1})(\mathbf{1}_q \mathbf{1}_p' \otimes \gamma^2 \mathbf{1}_q \mathbf{1}_p')(\mathbf{\Gamma}_1^{-1} \otimes \mathbf{\Gamma}_1^{-1}) \right\} \cdot \mathbf{D}_p \\ \frac{\partial \ln \mathcal{L}}{\partial \mathbf{\Gamma}_2} &= -\frac{n}{2} \left\{ \text{vec}'(\mathbf{C}_{22 \bullet 1}^{-1}) - \text{vec}'(\mathbf{S}_{11})(\mathbf{C}_{11 \bullet 2}^{-1} \otimes \mathbf{C}_{11 \bullet 2}^{-1})(\mathbf{1}_p \mathbf{1}_q' \otimes \gamma^2 \mathbf{1}_p \mathbf{1}_q')(\mathbf{\Gamma}_2^{-1} \otimes \mathbf{\Gamma}_2^{-1}) \right. \\ &\quad + \text{vec}'(\mathbf{S}'_{12}) \cdot [(\mathbf{C}_{11 \bullet 2}^{-1} \otimes \mathbf{I}_q)(\gamma \mathbf{1}_p \mathbf{1}_q' \otimes \mathbf{I}_q)(\mathbf{\Gamma}_2^{-1} \otimes \mathbf{\Gamma}_2^{-1}) \\ &\quad + (\mathbf{I}_p \otimes \mathbf{\Gamma}_2^{-1} \gamma \mathbf{1}_q \mathbf{1}_p')(\mathbf{C}_{11 \bullet 2}^{-1} \otimes \mathbf{C}_{11 \bullet 2}^{-1})(\mathbf{1}_p \mathbf{1}_q' \otimes \gamma^2 \mathbf{1}_p \mathbf{1}_q')(\mathbf{\Gamma}_2^{-1} \otimes \mathbf{\Gamma}_2^{-1})] \\ &\quad \left. + \text{vec}'(\mathbf{1}_q \mathbf{1}_p' \gamma \mathbf{\Gamma}_1^{-1} \mathbf{S}_{12})(\mathbf{C}_{22 \bullet 1}^{-1} \otimes \mathbf{C}_{22 \bullet 1}^{-1}) - \text{vec}'(\mathbf{S}_{22})(\mathbf{C}_{22 \bullet 1}^{-1} \otimes \mathbf{C}_{22 \bullet 1}^{-1}) \right\} \cdot \mathbf{D}_q \end{aligned}$$

$$\begin{aligned}
\frac{\partial \ln \mathcal{L}}{\partial \gamma} = & -\frac{n}{2} \left\{ -2\gamma \operatorname{vec}'(\mathbf{C}_{11\bullet 2}^{-1}) \operatorname{vec}(\mathbf{1}_p \mathbf{1}_q' \Gamma_2^{-1} \mathbf{1}_q \mathbf{1}_p') \right. \\
& + 2\gamma \operatorname{vec}'(\mathbf{S}_{11}) (\mathbf{C}_{11\bullet 2}^{-1} \otimes \mathbf{C}_{11\bullet 2}^{-1}) \operatorname{vec}(\mathbf{1}_p \mathbf{1}_q' \Gamma_2^{-1} \mathbf{1}_q \mathbf{1}_p') \\
& - \operatorname{vec}'(\mathbf{S}_{12}') \cdot [(\mathbf{C}_{11\bullet 2}^{-1} \otimes \mathbf{I}_q) \operatorname{vec}(\Gamma_2^{-1} \mathbf{1}_q \mathbf{1}_p') \\
& + 2\gamma (\mathbf{I}_p \otimes \Gamma_2^{-1} \gamma \mathbf{1}_q \mathbf{1}_p') (\mathbf{C}_{11\bullet 2}^{-1} \otimes \mathbf{C}_{11\bullet 2}^{-1}) \operatorname{vec}(\mathbf{1}_p \mathbf{1}_q' \Gamma_2^{-1} \mathbf{1}_q \mathbf{1}_p')] \\
& - \operatorname{vec}'(\Gamma_1^{-1} \mathbf{S}_{12}) [(\mathbf{C}_{22\bullet 1}^{-1} \otimes \mathbf{I}_p) \operatorname{vec}(\mathbf{1}_p \mathbf{1}_q') \\
& + 2\gamma (\mathbf{I}_q \otimes \gamma \mathbf{1}_p \mathbf{1}_q') (\mathbf{C}_{22\bullet 1}^{-1} \otimes \mathbf{C}_{22\bullet 1}^{-1}) \operatorname{vec}(\mathbf{1}_q \mathbf{1}_p' \Gamma_1^{-1} \mathbf{1}_p \mathbf{1}_q')] \\
& \left. + 2\gamma \operatorname{vec}'(\mathbf{S}_{22}) (\mathbf{C}_{22\bullet 1}^{-1} \otimes \mathbf{C}_{22\bullet 1}^{-1}) \operatorname{vec}(\mathbf{1}_q \mathbf{1}_p' \Gamma_1^{-1} \mathbf{1}_p \mathbf{1}_q') \right\},
\end{aligned}$$

where $\mathbf{D}_k$ is a $k^2 \times k(k+1)/2$ duplication matrix and $\operatorname{vec}(\cdot)$ is an operator stacking the columns of a given matrix one below another; cf. [24,25]. The formulas for matrix differentiation can be found in [26].

Equating the above derivatives to zero, we can omit the duplication matrices since they do not have any effect on the result. Then using some formulas, properties of the $\operatorname{vec}(\cdot)$ operator, and some transformations (cf. [20,24]), we obtain the maximum likelihood equations:

$$\begin{aligned}
& (\mathbf{C}_{11\bullet 2}^{-1}) - (\mathbf{C}_{11\bullet 2}^{-1} \mathbf{S}_{11} \mathbf{C}_{11\bullet 2}^{-1}) + \gamma (\mathbf{C}_{11\bullet 2}^{-1} \mathbf{1}_p \mathbf{1}_q' \Gamma_2^{-1} \mathbf{S}_{12} \mathbf{C}_{11\bullet 2}^{-1}) \\
& + \gamma (\Gamma_1^{-1} \mathbf{S}_{12} \mathbf{C}_{22\bullet 1}^{-1} \mathbf{1}_q \mathbf{1}_p' \Gamma_1^{-1}) + \gamma^3 (\Gamma_1^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{C}_{22\bullet 1}^{-1} \mathbf{1}_q \mathbf{1}_p' \Gamma_1^{-1} \mathbf{S}_{12} \mathbf{C}_{22\bullet 1}^{-1} \mathbf{1}_q \mathbf{1}_p' \Gamma_1^{-1}) \\
& - \gamma^2 (\Gamma_1^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{C}_{22\bullet 1}^{-1} \mathbf{S}_{22} \mathbf{C}_{22\bullet 1}^{-1} \mathbf{1}_q \mathbf{1}_p' \Gamma_1^{-1}) = 0
\end{aligned} \tag{6}$$

$$\begin{aligned}
& (\mathbf{C}_{22\bullet 1}^{-1}) - \gamma^2 (\Gamma_2^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{C}_{11\bullet 2}^{-1} \mathbf{S}_{11} \mathbf{C}_{11\bullet 2}^{-1} \mathbf{1}_p \mathbf{1}_q' \Gamma_2^{-1}) \\
& + \gamma (\Gamma_2^{-1} \mathbf{S}_{12}' \mathbf{C}_{11\bullet 2}^{-1} \mathbf{1}_p \mathbf{1}_q' \Gamma_2^{-1}) + \gamma^3 (\Gamma_2^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{C}_{11\bullet 2}^{-1} \mathbf{1}_p \mathbf{1}_q' \Gamma_2^{-1} \mathbf{S}_{12}' \mathbf{C}_{11\bullet 2}^{-1} \mathbf{1}_p \mathbf{1}_q' \Gamma_2^{-1}) \\
& + \gamma (\mathbf{C}_{22\bullet 1}^{-1} \mathbf{1}_q \mathbf{1}_p' \Gamma_1^{-1} \mathbf{S}_{12} \mathbf{C}_{22\bullet 1}^{-1}) - (\mathbf{C}_{22\bullet 1}^{-1} \mathbf{S}_{22} \mathbf{C}_{22\bullet 1}^{-1}) = 0
\end{aligned} \tag{7}$$

$$\begin{aligned}
& -2\gamma \operatorname{tr}(\mathbf{C}_{11\bullet 2}^{-1} \mathbf{1}_p \mathbf{1}_q' \Gamma_2^{-1} \mathbf{1}_q \mathbf{1}_p') + 2\gamma \operatorname{tr}(\mathbf{C}_{11\bullet 2}^{-1} \mathbf{S}_{11} \mathbf{C}_{11\bullet 2}^{-1} \mathbf{1}_p \mathbf{1}_q' \Gamma_2^{-1} \mathbf{1}_q \mathbf{1}_p') \\
& - \operatorname{tr}(\mathbf{C}_{11\bullet 2}^{-1} \mathbf{S}_{12} \Gamma_2^{-1} \mathbf{1}_q \mathbf{1}_p') - 2\gamma^2 \operatorname{tr}(\mathbf{C}_{11\bullet 2}^{-1} \mathbf{S}_{12} \Gamma_2^{-1} \mathbf{1}_q \mathbf{1}_p' \mathbf{C}_{11\bullet 2}^{-1} \mathbf{1}_p \mathbf{1}_q' \Gamma_2^{-1} \mathbf{1}_q \mathbf{1}_p') \\
& - \operatorname{tr}(\mathbf{C}_{22\bullet 1}^{-1} \mathbf{S}_{12}' \Gamma_1^{-1} \mathbf{1}_p \mathbf{1}_q') - 2\gamma^2 \operatorname{tr}(\mathbf{C}_{22\bullet 1}^{-1} \mathbf{S}_{12}' \Gamma_1^{-1} \mathbf{1}_p \mathbf{1}_q' \mathbf{C}_{22\bullet 1}^{-1} \mathbf{1}_q \mathbf{1}_p' \Gamma_1^{-1} \mathbf{1}_p \mathbf{1}_q') \\
& + 2\gamma \operatorname{tr}(\mathbf{C}_{22\bullet 1}^{-1} \mathbf{S}_{22} \mathbf{C}_{22\bullet 1}^{-1} \mathbf{1}_q \mathbf{1}_p' \Gamma_1^{-1} \mathbf{1}_p \mathbf{1}_q') = 0
\end{aligned} \tag{8}$$

We transform the Equations (6) and (7) by multiplying them by the appropriate Schur complement ($\mathbf{C}_{11\bullet 2}$ or $\mathbf{C}_{22\bullet 1}$, given in (2)), and then we obtain implicit formulas for Γ_1 and Γ_2 . In equation (8), the implicit formula for γ is obtained using properties of the trace operator. Thus, the system of equations has the form given in Theorem 3.1. ■

Note that the above system of equations is not given in an explicit form, and determining the solution is challenging and time-consuming, requiring the use of numerical methods. Solving the system of equations given in Theorem 3.1 requires the use of an iterative method: the so-called flip-flop algorithm. In general, the flip-flop algorithm needs a starting assumption. In our case, two parameters have to be assumed (for example, $\mathbf{\Gamma}_1 = \mathbf{S}_{11}$, $\mathbf{\Gamma}_2 = \mathbf{S}_{22}$) and the third parameter is computed. Then, new parameters are computed iteratively based on the previous values. The calculations are repeated until the stopping rule (requiring the difference between respective parameters from two steps to be small enough) is fulfilled.

Alternatively, the MLE can be determined by a method using derivatives – the gradient descent algorithm. This approach also requires a starting assumption, but this time for all parameters (for example, $\mathbf{\Gamma}_1 = \mathbf{S}_{11}$, $\mathbf{\Gamma}_2 = \mathbf{S}_{22}$ and $\gamma = 0$). Then, all parameters are improved using appropriate derivatives. The calculations are repeated until the stopping rule is fulfilled. Both methods provide estimators of $\mathbf{\Gamma}$.

4. Improved least squares estimation

Since maximum likelihood estimation is numerically complex and has no explicit form when a non-quadratic linear structure is considered, we use the least squares method. This approach provides an explicit estimator, as it is a projection of the matrix $\mathbf{S}$ onto a linear space. It should be noted that for a non-quadratic space such a projection does not preserve definiteness. To overcome this problem the projection onto a convex cone of p.d. structured matrices can be used; cf. [3]. However, this is also time-consuming; thus, we propose to use shrinkage estimation and an additional algebraic improvement if necessary.

The shrinkage method was considered in [27] as a way of improving the estimator of the covariance matrix. This method is used when the estimator is ill-conditioned and may be non-invertible, which is a common problem in covariance matrix estimation in high-dimensional settings. The method is used in [27] for an unstructured sample covariance matrix $\mathbf{S}$, and in [4] for a structured covariance matrix (banded Toeplitz structure) obtained by the least squares method. In this paper, we also improve the least squares estimator, which is presented in the next subsection.

4.1. Least squares estimator

The least squares method was applied in [15] to the estimation of a structured covariance matrix. This method is based on minimizing the distance between a sample covariance matrix and structured $\mathbf{\Gamma}$ in the sense of the Frobenius norm. The least squares estimator of $\mathbf{\Gamma}$ given by (1) is denoted as $\widehat{\mathbf{\Gamma}}_{LS}$, and it can be obtained by minimization of the Frobenius norm (5) with given $\mathbf{S}$ over the space $\mathcal{B}$:

$$f(\mathbf{S}, \widehat{\mathbf{\Gamma}}_{LS}) = \min_{\mathbf{\Gamma} \in \mathcal{B}} f(\mathbf{S}, \mathbf{\Gamma}) = \min_{\mathbf{\Gamma} \in \mathcal{B}} \|\mathbf{S} - \mathbf{\Gamma}\|_F^2,$$

where

$$\begin{aligned} f(\mathbf{S}, \mathbf{\Gamma}) = & \text{tr}(\mathbf{S}^2) - 2 \text{tr}(\mathbf{S}_{11} \mathbf{\Gamma}_1) - 4\gamma \text{tr}(\mathbf{S}_{12} \mathbf{1}_q \mathbf{1}_p') - 2 \text{tr}(\mathbf{S}_{22} \mathbf{\Gamma}_2) + \text{tr}(\mathbf{\Gamma}_1^2) \\ & + 2\gamma^2 pq + \text{tr}(\mathbf{\Gamma}_2^2). \end{aligned}$$

The minimum over the set $\mathcal{B}$ can be obtained by differentiation $f(\mathbf{S}, \hat{\mathbf{\Gamma}}_{LS})$ with respect to $\mathbf{\Gamma}_1, \mathbf{\Gamma}_2$ and γ . The derivatives have the following forms:

$$\begin{aligned}\frac{\partial f}{\partial \mathbf{\Gamma}_1} &= \left\{ \text{vec}'(\mathbf{I}_p) [(\mathbf{\Gamma}_1 \otimes \mathbf{I}_p) + (\mathbf{I}_p \otimes \mathbf{\Gamma}_1)] - 2 \text{vec}'(\mathbf{S}_{11}) \right\} \mathbf{D}_p \\ \frac{\partial f}{\partial \mathbf{\Gamma}_2} &= \left\{ -2 \text{vec}'(\mathbf{S}_{22}) + \text{vec}'(\mathbf{I}_q) [(\mathbf{\Gamma}_2 \otimes \mathbf{I}_q) + (\mathbf{I}_q \otimes \mathbf{\Gamma}_2)] \right\} \mathbf{D}_q \\ \frac{\partial f}{\partial \gamma} &= -4 \mathbf{1}'_p \mathbf{S}_{12} \mathbf{1}_q + 4\gamma pq.\end{aligned}$$

On equating the derivatives to zero and applying some transformations, we obtain the following theorem.

Theorem 4.1: For a given matrix $\mathbf{S} \in \mathbb{R}^>$, there exists a symmetric matrix $\mathbf{\Gamma} \in \mathbb{R}^s$ that minimizes the Frobenius norm (5) over $\mathcal{B}$, and this minimum is attained at

$$\begin{cases} \mathbf{\Gamma}_1 = \mathbf{S}_{11} \\ \mathbf{\Gamma}_2 = \mathbf{S}_{22} \\ \gamma = \frac{1}{pq} \mathbf{1}'_p \mathbf{S}_{12} \mathbf{1}_q. \end{cases}$$

The minimum attained in Theorem 4.1 is a global minimum of the convex Frobenius norm function (cf. [1]) and is given in an explicit form. Moreover, the estimator, $\hat{\mathbf{\Gamma}}_{LS}$, is consistent; cf. [15]. The solution given in Theorem 4.1 is the projection of the matrix $\mathbf{S}$ onto the space $\mathcal{B} = \mathcal{B}_1 \boxplus \mathcal{B}_2$.

The projection onto $\mathcal{B}$ does not preserve positive definiteness, since the space $\mathcal{B}$ is not quadratic. Recall that a quadratic space $\mathcal{P}$ is a linear space with the property that $\mathbf{A} \in \mathcal{P}$ implies $\mathbf{A}^2 \in \mathcal{P}$ (cf. [28]). However, the projections of some matrices onto this space are p.d. The subspace $\mathcal{B}_1$ is a quadratic space; consequently, the projection onto $\mathcal{B}_1$ preserves matrix definiteness (cf. [28]). However, $\mathcal{B}_2$ is not a quadratic subspace, and in addition, the projection of any matrix onto $\mathcal{B}_2$ is not definite.

The least squares estimator has good statistical properties such as consistency, while it is not always p.d. and well-conditioned and may even be indefinite (see the example in Section 2). Therefore, in the next sections, we propose two methods to improve it.

4.2. Shrinkage method

In this section, we use the shrinkage method developed in [27,29–31] to restore positive definiteness and well-conditioning to the unstructured sample covariance matrix. These properties are required in multivariate analysis when the inverse of the estimator of the covariance matrix is needed.

We apply this method to improve the estimator $\hat{\mathbf{\Gamma}}_{LS}$. A shrinkage estimator (SE), denoted as $\hat{\mathbf{\Gamma}}_S$, is a convex combination of a given target matrix $\mathbf{T}$ and the least squares estimator of the covariance matrix:

$$\hat{\mathbf{\Gamma}}_S = (1 - \eta)\mathbf{T} + \eta\hat{\mathbf{\Gamma}}_{LS}, \quad (9)$$

where the target matrix is p.d. and the shrinkage intensity $\eta \in [0; 1]$. Note that $\hat{\mathbf{\Gamma}}_{LS}$ has the considered structure, and thus $\hat{\mathbf{\Gamma}}_S$ also has the structure, when $\mathbf{T}$ is appropriately chosen.

In the literature, common sets of target matrices are the linear space of matrices proportional to the identity matrix or the linear space of diagonal matrices. These spaces are quadratic, and the empirical target matrix $\widehat{\mathbf{T}}$ is in fact the projection of $\mathbf{S}$ onto an appropriate space; cf. [32,33]. We propose the space $\mathcal{B}_1$ as the target set, that is, the largest quadratic subspace of the considered structure. Thus, the empirical target matrix has the following form:

$$\widehat{\mathbf{T}} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22} \end{pmatrix}.$$

Moreover, this is the MLE of the projection of $\mathbf{\Gamma}$ onto $\mathcal{B}_1$; cf. [34]. Note that $\widehat{\mathbf{T}}$ is not less well-conditioned than the matrix $\mathbf{S}$, which follows from Lemma 4.2 based on Theorem A.1. of [35].

Lemma 4.2: *Let*

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{pmatrix}$$

be $(p+q) \times (p+q)$ p.d. partitioned matrices with $\mathbf{A}_{11} : p \times p$ and $\mathbf{A}_{22} : q \times q$. Thus, the relation between their condition numbers $\kappa(*)$, defined in (3), is

$$\kappa(\mathbf{A}) \geq \kappa(\mathbf{B}).$$

Applying the empirical target matrix $\widehat{\mathbf{T}}$, the shrinkage estimator can be rewritten as:

$$\widehat{\mathbf{\Gamma}}_S = \widehat{\mathbf{T}} + \eta(\widehat{\mathbf{\Gamma}}_{LS} - \widehat{\mathbf{T}}).$$

Note that $\widehat{\mathbf{\Gamma}}_{LS} - \widehat{\mathbf{T}}$ is the projection of $\mathbf{S}$ onto $\mathcal{B}_2$ – the orthogonal complement of the quadratic subspace $\mathcal{B}_1 \in \mathcal{B}$.

Now, we determine an optimal η using quadratic risk as the optimization criterion, as proposed in [27].

Let denote the quadratic risk by

$$\begin{aligned} E(\|\widehat{\mathbf{\Gamma}}_S - \mathbf{\Gamma}\|_F^2) &= \text{tr} \left[E(\widehat{\mathbf{T}}^2) \right] + \eta \text{tr} \left[E(\widehat{\mathbf{T}}\widehat{\mathbf{\Gamma}}_{LS}) \right] + \eta^2 \text{tr} \left[E(\widehat{\mathbf{\Gamma}}_{LS}^2) \right] \\ &\quad - 2 \text{tr} \left[E(\widehat{\mathbf{T}}\mathbf{\Gamma}) \right] - 2\eta \text{tr} \left[E(\widehat{\mathbf{\Gamma}}_{LS} \mathbf{\Gamma}) \right] + \text{tr} \left[\mathbf{\Gamma}^2 \right] \\ &= g(\eta). \end{aligned} \tag{10}$$

Since the quadratic risk is convex, the minimum is attained at the stationary point where the derivative of $g(\eta)$ with respect to η is equal to 0, which can be written as

$$\text{tr} \left[E(\widehat{\mathbf{T}}\widehat{\mathbf{\Gamma}}_{LS}) \right] + 2\eta \text{tr} \left[E(\widehat{\mathbf{\Gamma}}_{LS}^2) \right] - 2 \text{tr} \left[E(\widehat{\mathbf{\Gamma}}_{LS} \mathbf{\Gamma}) \right] = 0.$$

This equation can be solved using the properties of the Wishart distribution given in [36,37]. Note that the matrix $\mathbf{S}$ and its diagonal blocks have Wishart distributions with $n-1$ degrees of freedom and appropriate covariance matrices:

$$\begin{aligned} n\mathbf{S} &\sim W_m(n-1, \mathbf{\Omega}), \\ n\mathbf{S}_{11} &\sim W_p(n-1, \mathbf{\Omega}_{11}), \\ n\mathbf{S}_{22} &\sim W_q(n-1, \mathbf{\Omega}_{22}). \end{aligned}$$

Thus, applying the properties of the Wishart distribution, we obtain the following equation:

$$\begin{aligned} & \frac{n-1}{n^2} \left\{ n\text{tr}(\mathbf{\Gamma}_{11}^2) + \text{tr}^2(\mathbf{\Gamma}_{11}) + n\text{tr}(\mathbf{\Gamma}_{22}^2) + \text{tr}^2(\mathbf{\Gamma}_{22}) + 2\eta \left[n\text{tr}(\mathbf{\Gamma}_{11}^2) + \text{tr}^2(\mathbf{\Gamma}_{11}) + n\text{tr}(\mathbf{\Gamma}_{22}^2) + \text{tr}^2(\mathbf{\Gamma}_{22}) \right] \right\} \\ & + \frac{4\eta}{pq} \cdot \frac{n-1}{n^2} \left\{ \text{tr} \left(\mathbf{\Gamma}_{12} \mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_{12} \mathbf{1}_q \mathbf{1}_p' \right) + \text{tr} \left(\mathbf{1}_q \mathbf{1}_p' \mathbf{\Gamma}_{11} \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_{22} \right) + (n-1) \left[\text{tr} \left(\mathbf{\Gamma}_{12} \mathbf{1}_q \mathbf{1}_p' \right) \right]^2 \right\} \\ & - 2 \cdot \frac{n-1}{n} \left[\text{tr}(\mathbf{\Gamma}_{11}^2) + \frac{2}{pq} \text{tr}^2(\mathbf{\Gamma}_{12} \mathbf{1}_q \mathbf{1}_p') + \text{tr}(\mathbf{\Gamma}_{22}^2) \right] = 0 \end{aligned}$$

The solution is given in Theorem 4.3.

Theorem 4.3: *The minimum of the quadratic risk $g(\eta)$ given in (10) is attained at*

$$\eta = \frac{n \left(\mathbf{1}_p' \mathbf{\Gamma}_{12} \mathbf{1}_q \right)^2}{\mathbf{1}_p' \mathbf{\Gamma}_{11} \mathbf{1}_p \mathbf{1}_q' \mathbf{\Gamma}_{22} \mathbf{1}_q + n \left(\mathbf{1}_p' \mathbf{\Gamma}_{12} \mathbf{1}_q \right)^2}.$$

The shrinkage intensity η , given in Theorem 4.3, is a function of the unknown $\mathbf{\Gamma}$, and its estimate can be obtained by replacing $\mathbf{\Gamma}$ by $\mathbf{S}$:

$$\hat{\eta} = \frac{n \left(\mathbf{1}_p' \mathbf{S}_{12} \mathbf{1}_q \right)^2}{\mathbf{1}_p' \mathbf{S}_{11} \mathbf{1}_p \mathbf{1}_q' \mathbf{S}_{22} \mathbf{1}_q + n \left(\mathbf{1}_p' \mathbf{S}_{12} \mathbf{1}_q \right)^2}.$$

Note that $\hat{\eta} \in [0; 1]$ and $\mathbf{S}$ is consistent when $n \rightarrow \infty$ and the dimension of the data-generating model is fixed, then $\hat{\eta} \rightarrow 1$ with probability one.

The estimator $\hat{\mathbf{\Gamma}}_S$ has the form

$$\hat{\mathbf{\Gamma}}_S = \hat{\mathbf{T}} + \hat{\eta}(\hat{\mathbf{\Gamma}}_{LS} - \hat{\mathbf{T}}).$$

Observe that $\hat{\mathbf{\Gamma}}_{LS}$ is a consistent estimator of $\mathbf{\Gamma}$, thus $\hat{\mathbf{\Gamma}}_S$ is consistent, because $\hat{\mathbf{\Gamma}}_S \rightarrow \hat{\mathbf{\Gamma}}_{LS}$; cf. [15].

Finally, the shrinkage estimator takes the following form:

$$\hat{\mathbf{\Gamma}}_S = \begin{pmatrix} \mathbf{S}_{11} & \hat{\gamma}_S \mathbf{1}_p \mathbf{1}_q' \\ \hat{\gamma}_S \mathbf{1}_q \mathbf{1}_p' & \mathbf{S}_{22} \end{pmatrix}$$

with $\hat{\gamma}_S = \hat{\eta} \cdot \hat{\gamma}_{LS}$.

In our problem, unlike in the classical approach, the use of the shrinkage method does not always lead to a p.d. or well-conditioned estimator. Therefore, we propose an estimation procedure that provides an estimator with the required properties, which we call the quasi shrinkage estimator (QSE). This approach consists of two steps. Firstly, we derive the shrinkage estimator $\hat{\mathbf{\Gamma}}_S$. Secondly, if it is p.d. and well-conditioned then $\hat{\mathbf{\Gamma}}_{QS} = \hat{\mathbf{\Gamma}}_S$; otherwise we improve it using an algebraic approach, presented in the next section.

4.3. Algebraic approach

In the previous section, we studied the shrinkage method for estimator construction. The estimates obtained are not always p.d. or well-conditioned, and therefore an improvement is needed to receive $\hat{\mathbf{\Gamma}}_{QS}$ having these properties. In this section, we propose an algebraic method based on the properties of the Schur complement (Lemma 2.1).

Recall that the diagonal blocks $\mathbf{S}_{11}$ and $\mathbf{S}_{22}$ are p.d., while off-diagonal blocks can affect the definiteness of the whole matrix $\widehat{\mathbf{\Gamma}}_S$. To obtain a p.d. and well-conditioned estimator, we propose – for use only when $\widehat{\mathbf{\Gamma}}_S$ does not satisfy these properties – the following three corrections:

$$(1) \quad \widehat{\mathbf{\Gamma}}_{S1} = \begin{pmatrix} \mathbf{S}_{11} & \beta \cdot \widehat{\gamma}_S \mathbf{1}_p \mathbf{1}_q' \\ \beta \cdot \widehat{\gamma}_S \mathbf{1}_q \mathbf{1}_p' & \mathbf{S}_{22} \end{pmatrix},$$

with the highest $\beta \in [0, 1]$ such that $\mathbf{S}_{11} - \beta^2 \cdot \widehat{\gamma}_S \mathbf{1}_p \mathbf{1}_q' \mathbf{S}_{22}^{-1} \widehat{\gamma}_S \mathbf{1}_q \mathbf{1}_p'$ is p.d. and $\widehat{\mathbf{\Gamma}}_{S1}$ is well-conditioned;

$$(2) \quad \widehat{\mathbf{\Gamma}}_{S2} = \begin{pmatrix} \alpha \cdot \mathbf{S}_{11} & \widehat{\gamma}_S \mathbf{1}_p \mathbf{1}_q' \\ \widehat{\gamma}_S \mathbf{1}_q \mathbf{1}_p' & \alpha \cdot \mathbf{S}_{22} \end{pmatrix},$$

with minimal $\alpha \geq 1$ such that $\alpha \mathbf{S}_{11} - \frac{1}{\alpha} \cdot \widehat{\gamma}_S \mathbf{1}_p \mathbf{1}_q' \mathbf{S}_{22}^{-1} \widehat{\gamma}_S \mathbf{1}_q \mathbf{1}_p'$ is p.d. and $\widehat{\mathbf{\Gamma}}_{S2}$ is well-conditioned;

$$(3) \quad \widehat{\mathbf{\Gamma}}_{S3} = \begin{pmatrix} \alpha \cdot \mathbf{S}_{11} & \beta \cdot \widehat{\gamma}_S \mathbf{1}_p \mathbf{1}_q' \\ \beta \cdot \widehat{\gamma}_S \mathbf{1}_q \mathbf{1}_p' & \alpha \cdot \mathbf{S}_{22} \end{pmatrix},$$

where $\beta \in [0, 1]$ and $\alpha \geq 1$ such that $\alpha \mathbf{S}_{11} - \frac{\beta^2}{\alpha} \cdot \widehat{\gamma}_S \mathbf{1}_p \mathbf{1}_q' \mathbf{S}_{22}^{-1} \widehat{\gamma}_S \mathbf{1}_q \mathbf{1}_p'$ is p.d. and $\widehat{\mathbf{\Gamma}}_{S3}$ is well-conditioned.

Note that the third correction can be rewritten as

$$\widehat{\mathbf{\Gamma}}_{S3} = \begin{pmatrix} \alpha \cdot \mathbf{S}_{11} & \beta \cdot \widehat{\gamma}_S \mathbf{1}_p \mathbf{1}_q' \\ \beta \cdot \widehat{\gamma}_S \mathbf{1}_q \mathbf{1}_p' & \alpha \cdot \mathbf{S}_{22} \end{pmatrix} = \alpha \begin{pmatrix} \mathbf{S}_{11} & \frac{\beta}{\alpha} \cdot \widehat{\gamma}_S \mathbf{1}_p \mathbf{1}_q' \\ \frac{\beta}{\alpha} \cdot \widehat{\gamma}_S \mathbf{1}_q \mathbf{1}_p' & \mathbf{S}_{22} \end{pmatrix}$$

and after excluding the α outside the matrix, we obtain the estimator corresponding to $\widehat{\mathbf{\Gamma}}_{S1}$. Similarly,

$$\widehat{\mathbf{\Gamma}}_{S3} = \begin{pmatrix} \alpha \cdot \mathbf{S}_{11} & \beta \cdot \widehat{\gamma}_S \mathbf{1}_p \mathbf{1}_q' \\ \beta \cdot \widehat{\gamma}_S \mathbf{1}_q \mathbf{1}_p' & \alpha \cdot \mathbf{S}_{22} \end{pmatrix} = \beta \begin{pmatrix} \frac{\alpha}{\beta} \cdot \mathbf{S}_{11} & \widehat{\gamma}_S \mathbf{1}_p \mathbf{1}_q' \\ \widehat{\gamma}_S \mathbf{1}_q \mathbf{1}_p' & \frac{\alpha}{\beta} \cdot \mathbf{S}_{22} \end{pmatrix}$$

after excluding β outside the matrix, we obtain the estimator corresponding to $\widehat{\mathbf{\Gamma}}_{S2}$. Thus, in the simulation study in Section 5, we compare only $\widehat{\mathbf{\Gamma}}_{S1}$ and $\widehat{\mathbf{\Gamma}}_{S2}$. Our aim is to identify and recommend the improvement providing a p.d. and well-conditioned estimator which is closer to $\widehat{\mathbf{\Gamma}}_S$ in the sense of the Frobenius norm.

5. Simulations

In this section, we perform simulation study

- (i) to detect how many shrinkage estimators are not positive definite and what affects this;
- (ii) to compare the proposed improvements with respect to a given criterion and indicate the best one;
- (iii) to compare MLE and QSE with respect to certain criteria and recommend one of them.

The data are generated from the matrix normal distribution $\mathbf{X} \sim N_{p+q,n}(\mathbf{0}, \mathbf{\Gamma}, \mathbf{I}_n)$, for parameters (p, q) from the set $\{(3, 2), (3, 10)\}$, sample size $n = 100$ and a p.d. matrix $\mathbf{\Gamma}$ with the structure under consideration as under consideration as given in (1). For each (p, q) , the simulation procedure starts by generating a p.d. $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ for which some integer γ 's are chosen such that $\mathbf{\Gamma}$ is p.d. The simulation is repeated 1000 times with fixed $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ for chosen (p, q) . In this paper, we present results for $(p, q) \in \{(3, 2), (3, 10)\}$.

5.1. Properties of shrinkage estimators

The main goal of this paper is to compare MLEs and SEs. The time needed to compute MLEs is much longer than for SEs. This results from the MLE determination procedure, in which the inverse of the estimator is required for each step. Consequently, the resulting MLEs are p.d. On the other hand, we obtain SEs much faster than MLEs, but they are not always p.d. The assumed covariance matrix $\mathbf{\Gamma}$ has an impact on the definiteness of the estimator. With the increasing absolute value of γ , the condition number of the matrix $\mathbf{\Gamma}$ increases and the matrix $\mathbf{\Gamma}$ approaches singularity; consequently, the number of non-p.d. shrinkage estimates among 1000 simulation runs is greater than for a $\mathbf{\Gamma}$ with better conditioning (see Table 1).

Since the positive definiteness of the estimator is a necessary property, we need to use an additional algebraic improvement, as described in Section 4.3, to obtain a p.d. estimator. Our aim is to compare the two proposed improvements with respect to the distance between the improved $\hat{\mathbf{\Gamma}}_S$ and the matrix $\mathbf{S}$, and to indicate the better one. In this study, only non-p.d. estimates are considered.

We present relative distances between chosen single non-p.d. $\hat{\mathbf{\Gamma}}_S$ and the matrix $\mathbf{S}$ given by

$$\zeta = \frac{\|\mathbf{S} - \hat{\mathbf{\Gamma}}_S\|_F^2}{\|\mathbf{S}\|_F^2}$$

for $(p, q) \in \{(3, 2), (3, 10)\}$.

Corrections are based on scaling off-diagonal blocks of $\hat{\mathbf{\Gamma}}_S$ by such β or by scaling diagonal blocks of $\hat{\mathbf{\Gamma}}_S$ by such α that the estimate obtained is p.d. and well-conditioned. Based on the simulation study, we choose the improvement that gives better results in the sense of the distances:

$$\zeta_{S1} = \frac{\|\mathbf{S} - \hat{\mathbf{\Gamma}}_{S1}\|_F^2}{\|\mathbf{S}\|_F^2}, \quad \zeta_{S2} = \frac{\|\mathbf{S} - \hat{\mathbf{\Gamma}}_{S2}\|_F^2}{\|\mathbf{S}\|_F^2}.$$

Table 1. Condition numbers of $\mathbf{\Gamma}$, denoted $\kappa(\mathbf{\Gamma})$, and numbers of non-p.d. shrinkage estimates ν among 1000 simulation runs for $p = 3$ and given values of q and γ .

q	γ	$\kappa(\mathbf{\Gamma})$	ν	q	γ	$\kappa(\mathbf{\Gamma})$	ν
2	-4	46.09	38	10	-5	43.13	10
2	-3	20.67	0	10	-3	31.53	0
2	-1	11.57	0	10	-1	29.78	0
2	0	10.85	0	10	0	29.64	0
2	1	11.57	0	10	1	29.78	0
2	3	20.67	2	10	3	31.53	0
2	4	46.09	41	10	5	43.13	7

Table 2. Examples of single distances ζ for non-p.d. $\widehat{\Gamma}_S$, condition number of $\widehat{\Gamma}_{S1}$, β calculated in the first step, $\alpha = \frac{1}{\beta}$, and ratios of ζ_{S1} , ζ_{S2} to ζ ; condition number of $\widehat{\Gamma}_{S2}$, α calculated in the first step, $\beta = \frac{1}{\alpha}$, and ratios of ζ_{S2} , ζ_{S1} to ζ , for $p = 3$ and appropriate q and γ .

q	γ	ζ	$\kappa(\widehat{\Gamma}_{S1})$	β	$\frac{\zeta_{S1}}{\zeta}$	α	$\frac{\zeta_{S2}}{\zeta}$	$\kappa(\widehat{\Gamma}_{S2})$	α	$\frac{\zeta_{S2}}{\zeta}$	β	$\frac{\zeta_{S1}}{\zeta}$
2	-4	0.01	36.93	0.94	1.04	1.06	1.35	48.65	1.16	3.42	0.86	1.19
2	-4	0.01	44.29	0.91	1.07	1.10	1.85	49.75	1.26	6.90	0.79	1.32
2	-4	0.01	48.05	0.92	1.05	1.09	2.25	49.02	1.38	23.49	0.72	1.41
2	4	0.02	36.73	0.91	1.03	1.09	11.48	49.39	1.26	3.49	0.80	1.13
2	4	0.03	33.55	0.92	1.02	1.08	9.34	48.34	1.21	2.31	0.83	1.08
2	4	0.01	34.99	0.90	1.05	1.11	15.90	48.35	1.30	7.21	0.77	1.24
10	-5	0.01	47.62	0.88	1.03	1.14	2.73	49.84	1.99	80.59	0.50	1.33
10	-5	0.02	49.80	0.86	1.03	1.16	2.83	49.23	1.54	17.91	0.65	1.16
10	-5	0.02	49.03	0.90	1.01	1.11	1.66	49.74	1.72	27.55	0.58	1.12
10	5	0.01	49.93	0.91	1.02	1.09	1.76	49.72	1.56	28.57	0.64	1.22
10	5	0.01	46.24	0.90	1.03	1.11	2.61	49.47	1.56	47.36	0.64	1.28
10	5	0.01	46.21	0.90	1.03	1.11	2.18	49.41	1.49	22.39	0.67	1.28

In the case when $\widehat{\Gamma}_S$ is not p.d. or not well-conditioned, simulations are performed as follows:

- (1) ζ is calculated
- (2) β is determined and ζ_{S1} is calculated
- (3) For β as given in 2, $\alpha = \frac{1}{\beta}$ and ζ_{S2} is calculated
- (4) α is determined and ζ_{S2} is calculated
- (5) For α as given in 4, $\beta = \frac{1}{\alpha}$ and ζ_{S1} is calculated

The estimators obtained based on β and $\alpha = \frac{1}{\beta}$ have the same conditioning (as do the estimators obtained based on α and $\beta = \frac{1}{\alpha}$). We calculate the ratios ζ_{S1}/ζ and ζ_{S2}/ζ , and the estimator for which the ratio is closer to one is determined to give better results.

From Table 2, it can be seen that in both cases (when β is calculated in the first step and when β is calculated in the second step) $\widehat{\Gamma}_{S1}$ is less distant from $\mathbf{S}$ than $\widehat{\Gamma}_{S2}$ for all considered cases ($\zeta_{S1}/\zeta < \zeta_{S2}/\zeta$).

Therefore, the first improvement gives better results.

Note that, when diagonal blocks of $\widehat{\Gamma}_S$ are scaled, $p(p+1)/2 + q(q+1)/2$ parameters are modified, while when off-diagonal blocks of $\widehat{\Gamma}_S$ are scaled, only one parameter is changed. Moreover, $\widehat{\Gamma}_{S1}$ preserves diagonal blocks which are projections of $\mathbf{S}$ onto quadratic subspaces. Thus, we recommend the first improvement.

5.2. Comparison of properties of MLEs and QSEs

In this section we compare the statistical properties of MLE and QSE with respect to selected loss functions; cf. [30].

We consider the following functions:

- Stein loss

$$L_{\text{Stein}}(\mathbf{\Gamma}, \widehat{\mathbf{\Gamma}}) = \frac{1}{m} [\text{tr}(\mathbf{\Gamma}^{-1}\widehat{\mathbf{\Gamma}}) - \log |\mathbf{\Gamma}^{-1}\widehat{\mathbf{\Gamma}}|] - 1$$

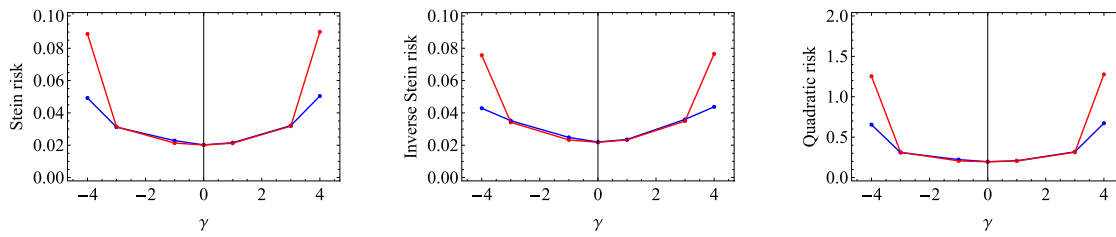


Figure 1. Empirical risks for MLE (red line) and QSE (blue line) for $(p, q) = (3, 2)$.

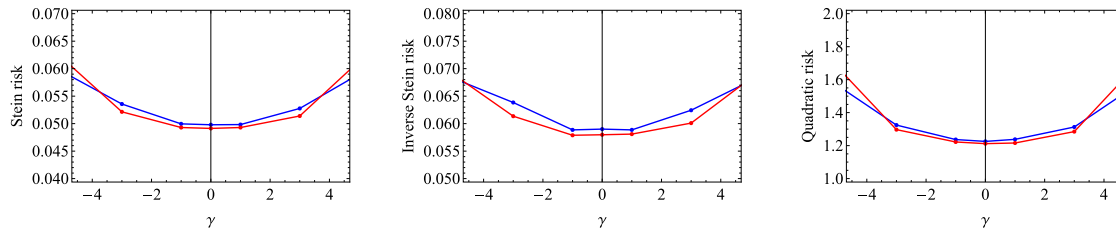


Figure 2. Empirical risks for MLE (red line) and QSE (blue line) for $(p, q) = (3, 10)$.

- inverse Stein loss

$$L_{IStein}(\mathbf{\Gamma}, \hat{\mathbf{\Gamma}}) = \frac{1}{m} \left[\text{tr}(\mathbf{\Gamma} \hat{\mathbf{\Gamma}}^{-1}) - \log |\mathbf{\Gamma} \hat{\mathbf{\Gamma}}^{-1}| \right] - 1$$

- quadratic loss

$$L_Q(\mathbf{\Gamma}, \hat{\mathbf{\Gamma}}) = \text{tr}(\mathbf{\Gamma}^{-1} \hat{\mathbf{\Gamma}})^2 - 2 \text{tr}(\mathbf{\Gamma}^{-1} \hat{\mathbf{\Gamma}}) + m.$$

Based on the simulation study, we generate data for chosen parameters (p, q) , $n = 100$ and p.d. $\mathbf{\Gamma}$, where $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ are fixed p.d. matrices, and chosen γ s such that $\mathbf{\Gamma}$ is p.d. Figures 1 and 2 present examples of graphs for averaged losses of γ functions for $(p, q) \in \{(3, 2), (3, 10)\}$. Since the simulations are repeated 1000 times, the results presented are empirical risks.

Recall that increasing the absolute value of γ causes $\mathbf{\Gamma}$ to be close to a singular matrix. This affects both of the considered estimators, $\hat{\mathbf{\Gamma}}_{ML}$ and $\hat{\mathbf{\Gamma}}_{QS}$; as a result, their condition numbers increase. For $(p, q) = (3, 2)$, the estimators are close to singularity for the absolute value of γ equal to 4. From Figure 1, we can observe that for small $|\gamma|$ the QSEs and the MLEs are comparable, and for greater $|\gamma|$, the empirical risks of MLEs are larger.

For $(p, q) = (3, 10)$, when the absolute value of γ is equal to 5, the matrix $\hat{\mathbf{\Gamma}}$ is close to the singular matrix (see Figure 2). It can be seen that for small values of $|\gamma|$, the empirical risks of QSEs are slightly larger than for MLEs, but when estimators are close to singularity, in most cases the empirical risks for MLEs are larger than for QSEs.

6. Discussion

Maximum likelihood is a common method for covariance matrix estimation. However, when the covariance matrix is structured, maximum likelihood estimation becomes challenging and time-consuming. Thus, we propose an alternative, much faster approach based

on the least squares method. This method allows one to determine estimators in a simple form, with good statistical properties, such as consistency. However, the problem of definiteness or conditioning may arise. This follows from the fact that the space of the considered structure is not quadratic. In response to this difficulty, a shrinkage method and an additional algebraic approach leading to QSE were proposed. The improvement methods are based on scaling off-diagonal blocks of the shrinkage estimators without modifying diagonal blocks. Finally, the QSE is p.d. and well-conditioned, and it is obtained much faster than the corresponding MLE, even though the correction procedure is needed.

Our simulation study presents that MLEs and QSEs are comparable in the sense of the considered loss functions, and QSEs can even have better properties than MLEs. This occurs when the true covariance matrix is close to singularity. Moreover, the determination of QSEs is computationally simple (much less time-consuming) than the determination of MLEs, even though an improvement is needed. The time needed for MLE determination increases with the increasing value of $p + q$. Thus, we recommend estimation using the quasi shrinkage method.

This method can also be used for high-dimensional datasets, where problems with definiteness and ill-conditioning of the sample covariance matrix $\mathbf{S}$ can appear. In this paper, we assumed that $\mathbf{S}$ is p.d. and well-conditioned. However, if this assumption were not satisfied, a double shrinkage method could be used; cf. [33]. The double shrinkage method is based on the use of a convex combination of two target matrices and an estimator of the covariance matrix. Candidates for target sets may be, for instance, spherical or diagonal sets of matrices; cf. [33]. Following the use of double shrinkage, we recommend applying the quasi shrinkage method proposed in this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

A. Markiewicz  <http://orcid.org/0000-0001-5473-3419>

M. Mokrzycka  <http://orcid.org/0000-0002-4512-845X>

M. Mrowińska  <http://orcid.org/0000-0002-7183-5174>

References

- [1] Cui X, Lin C, Zhao J, et al. Covariance structure regularization via Frobenius norm discrepancy. *Linear Algebra Appl.* 2016;510:124–145. doi: [10.1016/j.laa.2016.08.013](https://doi.org/10.1016/j.laa.2016.08.013)
- [2] Lin L, Higham NJ, Pan J. Covariance structure regularization via entropy loss function. *Comput Statist Data Anal.* 2014;72:315–327. doi: [10.1016/j.csda.2013.10.004](https://doi.org/10.1016/j.csda.2013.10.004)
- [3] Filipiak K, Markiewicz A, Mieldzioc A, et al. On projection of a positive definite matrix on a cone of nonnegative definite Toeplitz matrices. *Electronic J Linear Algebra.* 2018;33:74–82. doi: [10.13001/1081-3810.3750](https://doi.org/10.13001/1081-3810.3750)
- [4] John M, Mieldzioc A. The comparison of the estimators of banded Toeplitz covariance structure under the high-dimensional multivariate model. *Comm Statist Simulation Comput.* 2020;49(3):734–752. doi: [10.1080/03610918.2019.1614622](https://doi.org/10.1080/03610918.2019.1614622)
- [5] Lu N, Zimmerman DL. The likelihood ratio test for a separable covariance matrix. *Stat Probab Lett.* 2005;73(4):449–457. doi: [10.1016/j.spl.2005.04.020](https://doi.org/10.1016/j.spl.2005.04.020)

- [6] Srivastava M, von Rosen T, von Rosen D. Models with a Kronecker product covariance structure: estimation and testing. *Math Methods Statist.* 2008;17(4):357–370. doi: [10.3103/S1066530708040066](https://doi.org/10.3103/S1066530708040066)
- [7] Filipiak K, Klein D. Approximation with a Kronecker product structure with one component as compound symmetry or autoregression. *Electronic J Linear Algebra.* 2018;559:11–33. doi: [10.1016/j.laa.2018.08.031](https://doi.org/10.1016/j.laa.2018.08.031)
- [8] Filipiak K, Klein D, Mokrzycka M. Estimators comparison of separable covariance structure with one component as compound symmetry matrix. *Electronic J Linear Algebra.* 2018;33:83–98. doi: [10.13001/1081-3810.3740](https://doi.org/10.13001/1081-3810.3740)
- [9] Filipiak K, Klein D, Markiewicz A, et al. Approximation with a Kronecker product structure with one component as compound symmetry or autoregression via entropy loss function. *Linear Algebra Appl.* 2021;610:625–646. doi: [10.1016/j.laa.2020.10.013](https://doi.org/10.1016/j.laa.2020.10.013)
- [10] Szatrowski TH. Estimation and testing for block compound symmetry and other patterned covariance matrices with linear and non-linear structure. Technical report No. 107, Dept. of Statistics, Stanford University; 1976.
- [11] Szatrowski TH. Testing and estimation in the block compound symmetry problem. *J Educ Stat.* 1982;7(1):3–18. doi: [10.3102/10769986007001003](https://doi.org/10.3102/10769986007001003)
- [12] Filipiak K, Klein D. Estimation and testing the covariance structure of doubly multivariate data. In: Filipiak K, Markiewicz A, von Rosen D. editors. *Multivariate, Multilinear and Mixed Linear Models*, Springer; 2021. p. 131–156.
- [13] Szczepańska-Álvarez A, Hao C, Liang Y, et al. Estimation equations for multivariate linear models with Kronecker structured covariance matrices. *Commun Stat Theory Methods.* 2017;46(16):7902–7915. doi: [10.1080/03610926.2016.1165852](https://doi.org/10.1080/03610926.2016.1165852)
- [14] Janiszewska M, Markiewicz A, Mokrzycka M. Block matrix approximation via entropy loss function. *Appl Math (Prague).* 2020;65(6):829–844. doi: [10.21136/AM](https://doi.org/10.21136/AM)
- [15] Ohlson M, Von Rosen D. Explicit estimators of parameters in the growth curve model with linearly structured covariance matrices. *J Multivar Anal.* 2010;101(5):1284–1295. doi: [10.1016/j.jmva.2009.12.023](https://doi.org/10.1016/j.jmva.2009.12.023)
- [16] Jensen ST. Covariance hypotheses which are linear in both the covariance and the inverse covariance. *Ann Stat.* 1988;16(1):302–322. doi: [10.1214/aos/1176350707](https://doi.org/10.1214/aos/1176350707)
- [17] Anderson TW. Asymptotically efficient estimation of covariance matrices with linear structure. *Ann Stat.* 1973;1(1):135–141. doi: [10.1214/aos/1193342389](https://doi.org/10.1214/aos/1193342389)
- [18] Mieldzioc A, Mokrzycka M, Sawikowska A. Identification of block-structured covariance matrix on an example of metabolomic data. *Separations.* 2021;8(11):205–752. doi: [10.3390/separations8110205](https://doi.org/10.3390/separations8110205)
- [19] Zhang F. *The schur complement and its applications*. New York: Springer; 2005.
- [20] Kollo T, Von Rosen D. *Advanced multivariate statistics with matrices*. Dordrecht: Springer; 2005.
- [21] Härdle WK, Simar L. *Applied multivariate statistical analysis*. 4th ed. Switzerland: Springer; 2015.
- [22] Filipiak K, John M, Markiewicz A. Comments on Maximum Likelihood Estimation and Projections Under Multivariate Statistical Models. In: Holgersson T, Singull M editors. *Recent Developments in Multivariate and Random Matrix Analysis*, Springer; 2020. p. 51–66.
- [23] Fuglede B, Jensen ST. Positive projections of symmetric matrices and Jordan algebras. *Expo Math.* 2013;31(3):295–303. doi: [10.1016/j.exmath.2013.01.005](https://doi.org/10.1016/j.exmath.2013.01.005)
- [24] Magnus J, Neudecker H. Symmetry, 0-1 matrices and Jacobians, a review. *Econom Theory.* 1986;2(2):157–190. doi: [10.1017/S0266466600011476](https://doi.org/10.1017/S0266466600011476)
- [25] Magnus J, Neudecker H. *Matrix differential calculus with applications in statistics and econometrics*. revised edition. Chichester: Wiley; 1999.
- [26] Fackler PL. Notes on matrix calculus. 2005. Available from: <http://www4.ncsu.edu/~pfackler/MatCalc.pdf>.
- [27] Ledoit O, Wolf M. A well-conditioned estimator for large-dimensional covariance matrices. *J Multivar Anal.* 2004;88(2):365–411. doi: [10.1016/S0047-259X\(03\)00096-4](https://doi.org/10.1016/S0047-259X(03)00096-4)

- [28] Seely J. Quadratic subspaces and completeness. *Ann Stat.* 1971;42(2):710–721. doi: [10.1214/aoms/1177693420](https://doi.org/10.1214/aoms/1177693420)
- [29] Bodnar T, Gupta AK, Parolya N. Direct shrinkage estimation of large dimensional precision matrix. *J Multivar Anal.* 2016;146:223–236. doi: [10.1016/j.jmva.2015.09.010](https://doi.org/10.1016/j.jmva.2015.09.010)
- [30] Ledoit O, Wolf M. Shrinkage estimation of large covariance matrices: keep it simple, statistician? Working Paper No. 327; 2021.
- [31] Bodnar O, Bodnar T, Parolya N. Recent advances in shrinkage-based high-dimensional inference. *J Multivariate Anal.* 2022;188:104826. doi: [10.1016/j.jmva.2021.104826](https://doi.org/10.1016/j.jmva.2021.104826)
- [32] Hannart A, Naveau P. Estimating high dimensional covariance matrices: a new look at the Gaussian conjugate framework. *J Multivar Anal.* 2014;131:149–162. doi: [10.1016/j.jmva.2014.06.001](https://doi.org/10.1016/j.jmva.2014.06.001)
- [33] Ikeda Y, Kubokawa T, Srivastava MS. Comparison of linear shrinkage estimators of a large covariance matrix in normal and non-normal distributions. *Comput Stat Data Anal.* 2016;95:95–108. doi: [10.1016/j.csda.2015.09.011](https://doi.org/10.1016/j.csda.2015.09.011)
- [34] Zehna WP. Invariance of maximum likelihood estimators. *Ann Math Statist.* 1966;37(3): 744–744. doi: [10.1214/aoms/1177699475](https://doi.org/10.1214/aoms/1177699475)
- [35] Marshall AW, Olkin I, Arnold BC. Inequalities: theory of majorization and its applications. 2nd ed. New York: Springer; 2011. 784.
- [36] Holgersson T, Pielaszkiewicz J. A collection of moments of the Wishart distribution. In: Holgersson T, Singull M, editors. Recent developments in multivariate and random matrix analysis. Springer; 2020. p. 147–162.
- [37] Pielaszkiewicz J, Von Rosen D. Multivariate moments in multivariate analysis. In: Filipiak K, Markiewicz A, von Rosen D. editors. Multivariate, multilinear and mixed linear models, Springer; 2021. p. 41–92.

[C] **Block covariance matrix estimation with structured off-diagonal blocks**

by Monika Mokrzycka and Malwina Mrowińska

Statistical Papers, **66**, 129, 2025.

Contribution:

- Monika Mokrzycka (50%) – conceptualization and methodological supervision, determination of the optimal shrinkage parameter, proposal of alternative estimation approaches, implementation of numerical algorithms in R, providing and interpretation of the results using real-world data example, writing – editing and general supervision
- Malwina Mrowińska (50%) – conceptualization, methodology, literature review, characterization of the covariance structure estimators (Theorem 1, Theorem 2, shrinkage estimator, Theorem 5), implementation of numerical algorithms in R, simulation studies, interpretations, writing – original draft preparation and editing



Block covariance matrix estimation with structured off-diagonal blocks

Monika Mokrzycka¹ · Malwina Mrowińska²

Received: 24 October 2024 / Revised: 21 June 2025
© The Author(s) 2025

Abstract

Under the multivariate model with partitioned vector of observations various estimators of a block covariance matrix with structured cross-covariance matrices are proposed. It is assumed that the structure of the off-diagonal block of the covariance matrix corresponds to the appropriate part of autoregression of the order one structure, AR(1). The maximum likelihood and least squares estimators are determined and four new estimators are proposed. Comparison of estimates using simulation studies and real data example is demonstrated, respectively. The simulation studied suggested that maximum likelihood and intuitive estimates have the best statistical properties and the latter are computationally simple.

Keywords Partitioned covariance matrix · Autoregression of order one · Maximum likelihood estimation · Least squares estimation · Shrinking

1 Introduction

Nowadays, in almost every field of science, in particular in agriculture, medicine, economics, biology, or geography, more and more complex experiments are performed resulting in huge multivariate data sets. One of the issues in such cases is the estimation of the relationships between observations. However, in many situations, experts have some knowledge about the possible structure of the covariance matrix, which has to be taken into account.

✉ Malwina Mrowińska
malwina.mrowinska@put.poznan.pl
Monika Mokrzycka
mmok@igr.poznan.pl

¹ Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland

² Institute of Mathematics, Poznan University of Technology, Poznań, Poland

Let us assume an experiment where several features are measured repeatedly in time. One of the most appropriate structures of the covariance matrix is the autoregression of order one (AR(1)), where the correlations between observations decay as the time interval between them increases, cf. Lin et al (2014), Cui et al (2016). However, in some experiments, additional factors, such as, e.g., weather conditions (drought or heavy rains), stages of development, or the nature of the phenomenon under study, may affect the results. Moreover, measurement frequencies may also be related to this additional factor. Therefore, in this paper, we consider the block covariance structure partitioned into four subblocks of AR(1), not necessarily of the same dimensions, with unstructured diagonal blocks. It is worth mentioning that similar structures were considered in the literature; Markiewicz et al (2024) proposed estimation methods for the covariance structure with an off-diagonal block proportional to the matrix of ones, while Janiszewska et al (2020) proposed approximation techniques for both, AR(1) or equal correlations, structures.

It is well-known that the estimator of an unstructured covariance matrix is proportional to the matrix of sums and squares. It is also known that the choice of the proper covariance structure is crucial for data analysis. Therefore, the aim of this paper is to propose and compare several estimation methods based on maximization of the likelihood function or the least squares estimation. Since in considered setup both methods are complex or do not provide a positive definite estimator, further improvements, for instance, based on the shrinking method (Ledoit and Wolf 2004) or rescaling (Markiewicz et al 2024), are presented. We also propose alternative estimation methods, where the covariance matrix is presented as the sum of the matrix having AR(1) structure and the block diagonal matrix and each component of the sum is estimated separately.

All estimates considered are compared by simulation study with respect to four loss functions and computing time. Moreover, they are illustrated and compared by real data example.

The paper is organized as follows. Section 2 presents the model and the form of the covariance structure. The maximum likelihood and least squares estimation methods are presented in Sect. 3. The alternative methods based on separate estimation of AR(1) structure and block diagonal matrix are given in Sect. 4. In Sects. 5 and 6 the comparison of the estimates using simulation studies and real data examples is demonstrated, respectively. The paper ends with conclusions.

2 Model

Assume the experiment where two types of variables are measured on n experimental units. Observations from each unit can be stored in an m -dimensional vector $\mathbf{z}_i = (\mathbf{x}_i, \mathbf{y}_i)$, $i = 1, \dots, n$, with p -dimensional vector $\mathbf{x}_i$ and q -dimensional vector $\mathbf{y}_i$, $m = p + q$. Assume that $\mathbf{z}_i$, $i = 1, \dots, n$, are independent, normally distributed random vectors, i.e.,

$$\mathbf{z}_i \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Gamma}),$$

where $\boldsymbol{\mu} = (\boldsymbol{\mu}'_x, \boldsymbol{\mu}'_y)'$ is an unknown mean vector and $\boldsymbol{\Gamma}$ is an unknown positive definite (p.d.) covariance matrix of the form

$$\boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{\Gamma}_{11} & \boldsymbol{\Gamma}_{12} \\ \boldsymbol{\Gamma}'_{12} & \boldsymbol{\Gamma}_{22} \end{pmatrix},$$

with $\boldsymbol{\Gamma}_{11}$ ($\boldsymbol{\Gamma}_{22}$) being a covariance matrix of the first (second) type of variables, and $\boldsymbol{\Gamma}_{12}$ being a covariance matrix between two types of variables (cross-covariance matrix). The matrix of all observations, $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)$, then is a normally distributed matrix

$$\mathbf{Z} \sim N_{m,n}(\boldsymbol{\mu}\mathbf{1}'_n, \boldsymbol{\Gamma}, \mathbf{I}_n),$$

where $\mathbf{1}_n$ is n -dimensional vector of ones and $\mathbf{I}_n$ is an identity matrix of order n .

Assume that correlations between observations decrease exponentially with the distance between variables and this correlation structure can be represented by autoregression of order one process, AR(1), of the form

$$\sigma^2 \begin{pmatrix} 1 & \varrho & \varrho^2 & \dots & \varrho^{m-1} \\ \varrho & 1 & \varrho & \dots & \varrho^{m-2} \\ \varrho^2 & \varrho & 1 & \dots & \varrho^{m-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \varrho^{m-1} & \varrho^{m-2} & \varrho^{m-3} & \dots & 1 \end{pmatrix}, \quad (1)$$

which will be denoted by $\boldsymbol{\Psi}$ to distinguish from unstructured covariance matrix $\boldsymbol{\Gamma}$. This matrix is p.d. for $\sigma^2 > 0$ and $-1 \leq \varrho \leq 1$. Such a structure is the most common when a characteristic is measured in time. Since in our model two groups of variables related to two different conditions of the experiment are observed, additional noise in diagonal blocks of covariance matrix is considered. Partitioning matrix (1) into four blocks, where the first diagonal block is of order p and the second is of order q , the covariance matrix of $\mathbf{Z}$ can be represented as

$$\boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{\Gamma}_1 & \sigma^2 \begin{pmatrix} \varrho^p & \varrho^{p+1} & \dots & \varrho^{p+q-1} \\ \varrho^{p-1} & \dots & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots \\ \varrho^2 & \dots & \dots & \dots \\ \varrho & \varrho^2 & \dots & \varrho^q \end{pmatrix} \\ \sigma^2 \begin{pmatrix} \varrho^p & \varrho^{p+1} & \dots & \varrho^{p+q-1} \\ \varrho^{p-1} & \dots & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots \\ \varrho^2 & \dots & \dots & \dots \\ \varrho & \varrho^2 & \dots & \varrho^q \end{pmatrix}' & \boldsymbol{\Gamma}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Gamma}_1 & \sigma^2 \mathbf{A} \\ \sigma^2 \mathbf{A}' & \boldsymbol{\Gamma}_2 \end{pmatrix}, \quad (2)$$

where $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ are symmetric p.d. matrices of order p and q respectively ($\boldsymbol{\Gamma}_1 \in \mathbb{R}_p^{>}$, $\boldsymbol{\Gamma}_2 \in \mathbb{R}_q^{>}$) such that $\boldsymbol{\Gamma}$ is p.d. matrix of order m having the structure defined in (2), $\boldsymbol{\Gamma} \in \mathcal{S}_m^{>}$. Note that $\boldsymbol{\Gamma} \in \mathcal{S}_m$ is a matrix with the structure (2) with $\sigma^2 > 0$ and $-1 \leq \varrho \leq 1$ not necessarily p.d. To check positive definiteness of $\boldsymbol{\Gamma}$, it is enough to check whether Schur complement $\boldsymbol{\Gamma}_2 - \sigma^4 \mathbf{A}' \boldsymbol{\Gamma}_1^{-1} \mathbf{A}$ is p.d. ($\boldsymbol{\Gamma}_1$ is p.d. by the assumption), or equivalently $\boldsymbol{\Gamma}_1 - \sigma^4 \mathbf{A} \boldsymbol{\Gamma}_2^{-1} \mathbf{A}'$ is p.d. ($\boldsymbol{\Gamma}_2$ is p.d. by the assumption).

3 Estimation of block covariance matrix

In this section we determine the maximum likelihood and least squares estimators of $\Gamma \in \mathcal{S}_m$. Since the second estimate is not necessarily p.d. the improvement of its definiteness via shrinking is proposed.

3.1 Maximum likelihood estimation

It is well known that maximum likelihood estimators MLEs of μ and the covariance matrix without any structure are, respectively, $\bar{\mathbf{Z}} = \frac{1}{n} \mathbf{Z} \mathbf{1}_n$ and $\mathbf{S} = \frac{1}{n} \mathbf{Z} (\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n') \mathbf{Z}'$. Since the possible structure of covariance matrix does not influence the MLE of μ , to determine MLE of $\Gamma \in \mathcal{S}_m^>$ one has to maximize the log-likelihood function of the form

$$\ln \mathcal{L}(\Gamma) = \frac{nm}{2} \ln(2\pi) - \frac{n}{2} \ln |\Gamma| - \frac{n}{2} \text{tr}(\mathbf{S}\Gamma^{-1}) \quad (3)$$

over $\Gamma \in \mathcal{S}_m^>$.

It is known that for partitioned matrices (cf. Kollo and von Rosen 2005)

$$\Gamma^{-1} = \begin{pmatrix} \mathbf{C}_{11}^{-1} & -\sigma^2 \Gamma_1^{-1} \mathbf{A} \mathbf{C}_{22}^{-1} \\ -\sigma^2 \Gamma_2^{-1} \mathbf{A}' \mathbf{C}_{11}^{-1} & \mathbf{C}_{22}^{-1} \end{pmatrix}$$

and $|\Gamma| = |\Gamma_2| \cdot |\mathbf{C}_{11}| = |\Gamma_1| \cdot |\mathbf{C}_{22}|$, where

$$\mathbf{C}_{11} = \Gamma_1 - (\sigma^2)^2 \mathbf{A} \Gamma_2^{-1} \mathbf{A}', \quad \mathbf{C}_{22} = \Gamma_2 - (\sigma^2)^2 \mathbf{A}' \Gamma_1^{-1} \mathbf{A}.$$

Therefore, the log-likelihood function (3) can be rewritten as

$$\ln \mathcal{L}(\Gamma) = \frac{nm}{2} \ln(2\pi) - \frac{n}{2} \ln |\Gamma_2| - \frac{n}{2} \ln |\mathbf{C}_{11}| - \frac{n}{2} [\text{tr}(\mathbf{S}_{11} \mathbf{C}_{11}^{-1}) - \sigma^2 \text{tr}(\mathbf{S}_{12} \Gamma_2^{-1} \mathbf{A}' \mathbf{C}_{11}^{-1}) - \sigma^2 \text{tr}(\mathbf{S}_{12}' \Gamma_1^{-1} \mathbf{A} \mathbf{C}_{22}^{-1}) + \text{tr}(\mathbf{S}_{22} \mathbf{C}_{22}^{-1})] \quad (4)$$

or

$$\ln \mathcal{L}(\Gamma) = \frac{nm}{2} \ln(2\pi) - \frac{n}{2} \ln |\Gamma_1| - \frac{n}{2} \ln |\mathbf{C}_{22}| - \frac{n}{2} [\text{tr}(\mathbf{S}_{11} \mathbf{C}_{11}^{-1}) - \sigma^2 \text{tr}(\mathbf{S}_{12} \Gamma_2^{-1} \mathbf{A}' \mathbf{C}_{11}^{-1}) - \sigma^2 \text{tr}(\mathbf{S}_{12}' \Gamma_1^{-1} \mathbf{A} \mathbf{C}_{22}^{-1}) + \text{tr}(\mathbf{S}_{22} \mathbf{C}_{22}^{-1})]. \quad (5)$$

Differentiating (4) with respect to Γ_1 , (5) with respect to Γ_2 and any of the above with respect to σ^2 and ϱ and using the following properties of the $\text{vec}(\cdot)$ operator, which is an operator stacking the columns of a given matrix one below another:

$$\text{vec} \mathbf{ABC} = (\mathbf{C}' \otimes \mathbf{A}) \text{vec} \mathbf{B},$$

$$\text{vec}' \mathbf{A} \cdot \text{vec} \mathbf{B} = \text{tr}(\mathbf{A}' \mathbf{B}),$$

$$\mathbf{K}_{s,t} \cdot \text{vec } \mathbf{A} = \text{vec } \mathbf{A}',$$

where $\mathbf{K}_{s,t}$ is an $st \times st$ commutation matrix (cf. Magnus and Neudecker 1986; Kollo and von Rosen 2005), we obtain

$$\begin{aligned} \frac{\partial \ln \mathcal{L}}{\partial \Gamma_1} &= -\frac{n}{2} \left\{ \text{vec}' \left[\mathbf{C}_{11}^{-1} + \sigma^2 \mathbf{C}_{11}^{-1} \mathbf{A} \Gamma_2^{-1} \mathbf{S}'_{12} \mathbf{C}_{11}^{-1} + \sigma^2 \Gamma_1^{-1} \mathbf{S}_{12} \mathbf{C}_{22}^{-1} \mathbf{A}' \Gamma_1^{-1} \right. \right. \\ &\quad \left. \left. + (\sigma^2)^3 \Gamma_1^{-1} \mathbf{A} \mathbf{C}_{22}^{-1} \mathbf{A}' \Gamma_1^{-1} \mathbf{S}_{12} \mathbf{C}_{22}^{-1} \mathbf{A}' \Gamma_1^{-1} - (\sigma^2)^2 \Gamma_1^{-1} \mathbf{A} \mathbf{C}_{22}^{-1} \mathbf{S}_{22} \mathbf{C}_{22}^{-1} \mathbf{A}' \Gamma_1^{-1} \right. \right. \\ &\quad \left. \left. - \mathbf{C}_{11}^{-1} \mathbf{S}_{11} \mathbf{C}_{11}^{-1} \right] \right\} \cdot \mathbf{D}_p \\ \frac{\partial \ln \mathcal{L}}{\partial \Gamma_2} &= -\frac{n}{2} \left\{ \text{vec}' \left[\mathbf{C}_{22}^{-1} + \sigma^2 \mathbf{C}_{22}^{-1} \mathbf{A}' \Gamma_1^{-1} \mathbf{S}_{12} \mathbf{C}_{22}^{-1} + \sigma^2 \Gamma_2^{-1} \mathbf{S}'_{12} \mathbf{C}_{11}^{-1} \mathbf{A} \Gamma_2^{-1} \right. \right. \\ &\quad \left. \left. + (\sigma^2)^3 \Gamma_2^{-1} \mathbf{A}' \mathbf{C}_{11}^{-1} \mathbf{A} \Gamma_2^{-1} \mathbf{S}'_{12} \mathbf{C}_{11}^{-1} \mathbf{A} \Gamma_2^{-1} - (\sigma^2)^2 \Gamma_2^{-1} \mathbf{A}' \mathbf{C}_{11}^{-1} \mathbf{S}_{11} \mathbf{C}_{11}^{-1} \mathbf{A} \Gamma_2^{-1} \right. \right. \\ &\quad \left. \left. - \mathbf{C}_{22}^{-1} \mathbf{S}_{22} \mathbf{C}_{22}^{-1} \right] \right\} \cdot \mathbf{D}_q \\ \frac{\partial \ln \mathcal{L}}{\partial \sigma^2} &= -\frac{n}{2} \left\{ \text{tr} \left[-2\sigma^2 \mathbf{C}_{11}^{-1} \mathbf{A} \Gamma_2^{-1} \mathbf{A}' + 2\sigma^2 \mathbf{C}_{11}^{-1} \mathbf{S}_{11} \mathbf{C}_{11}^{-1} \mathbf{A} \Gamma_2^{-1} \mathbf{A}' \right. \right. \\ &\quad \left. \left. - \mathbf{S}_{12} \Gamma_2^{-1} \mathbf{A}' \mathbf{C}_{11}^{-1} - 2\sigma^2 \mathbf{S}_{12} \Gamma_2^{-1} \mathbf{A}' \mathbf{C}_{11}^{-1} \mathbf{A} \Gamma_2^{-1} \mathbf{A}' \mathbf{C}_{11}^{-1} \right. \right. \\ &\quad \left. \left. - \mathbf{C}_{22}^{-1} \mathbf{S}'_{12} \Gamma_1^{-1} \mathbf{A} - 2\sigma^2 \mathbf{S}'_{12} \Gamma_1^{-1} \mathbf{A} \mathbf{C}_{22}^{-1} \mathbf{A}' \Gamma_1^{-1} \mathbf{A} \mathbf{C}_{22}^{-1} \right. \right. \\ &\quad \left. \left. + 2\sigma^2 \mathbf{C}_{22}^{-1} \mathbf{S}_{22} \mathbf{C}_{22}^{-1} \mathbf{A}' \Gamma_1^{-1} \mathbf{A} \right] \right\} \\ \frac{\partial \ln \mathcal{L}}{\partial \varrho} &= -\frac{n}{2} \left\{ \text{tr} \left[-(\sigma^2)^2 \Gamma_2^{-1} \mathbf{A}' + \mathbf{C}_{11}^{-1} \mathbf{F} - (\sigma^2)^2 \mathbf{C}_{11}^{-1} \mathbf{A} \Gamma_2^{-1} \mathbf{F}' \right. \right. \\ &\quad \left. \left. + (\sigma^2)^2 \Gamma_2^{-1} \mathbf{A}' \mathbf{C}_{11}^{-1} \mathbf{S}_{11} \mathbf{C}_{11}^{-1} \mathbf{F} + (\sigma^2)^2 \mathbf{C}_{11}^{-1} \mathbf{S}_{11} \mathbf{C}_{11}^{-1} \mathbf{A} \Gamma_2^{-1} \mathbf{F}' \right. \right. \\ &\quad \left. \left. - \sigma^2 \mathbf{C}_{11}^{-1} \mathbf{S}_{12} \Gamma_2^{-1} \mathbf{F}' - (\sigma^2)^2 \mathbf{S}_{12} \Gamma_2^{-1} \mathbf{A}' \mathbf{C}_{11}^{-1} \mathbf{F} \Gamma_2^{-1} \mathbf{A}' \mathbf{C}_{11}^{-1} \right. \right. \\ &\quad \left. \left. - \sigma^2 \mathbf{C}_{22}^{-1} \mathbf{S}'_{12} \Gamma_1^{-1} \mathbf{F} + (\sigma^2)^2 \mathbf{C}_{22}^{-1} \mathbf{S}'_{12} \Gamma_1^{-1} \mathbf{A} \mathbf{C}_{22}^{-1} \mathbf{A}' \Gamma_1^{-1} \mathbf{F}' \right. \right. \\ &\quad \left. \left. - (\sigma^2)^2 \mathbf{S}_{12} \Gamma_2^{-1} \mathbf{A}' \mathbf{C}_{11}^{-1} \mathbf{A} \Gamma_2^{-1} \mathbf{F}' \mathbf{C}_{11}^{-1} - (\sigma^2)^2 \mathbf{C}_{22}^{-1} \mathbf{S}'_{12} \Gamma_1^{-1} \mathbf{A} \mathbf{C}_{22}^{-1} \mathbf{F}' \Gamma_1^{-1} \mathbf{A} \right. \right. \\ &\quad \left. \left. + (\sigma^2)^2 \mathbf{C}_{22}^{-1} \mathbf{S}_{22} \mathbf{C}_{22}^{-1} \mathbf{A}' \Gamma_1^{-1} \mathbf{F} + (\sigma^2)^2 \mathbf{S}_{22} \mathbf{C}_{22}^{-1} \mathbf{F}' \Gamma_1^{-1} \mathbf{A} \mathbf{C}_{22}^{-1} \right] \right\} \end{aligned}$$

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

$$\mathbf{F} = \begin{pmatrix} p\varrho^{p-1} & (p+1)\varrho^p & \cdots & (p+q-1)\varrho^{p+q-2} \\ (p-1)\varrho^{p-2} & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ 2\varrho & \cdots & \cdots & \cdots \\ 1 & 2\varrho & \cdots & q\varrho^{q-1} \end{pmatrix}, \quad (6)$$

and $\mathbf{D}_k$ is a $k^2 \times k(k+1)/2$ duplication matrix; cf. Magnus and Neudecker (1986) or Magnus and Neudecker (1999). The formulas for matrix differentiation can be found in Fackler (2005).

Equating the above derivatives to zero, we can omit the duplication matrices, since they do not have any effect on the result. Then transforming the equations obtained, we get the following theorem.

Theorem 1 If there exists unique solution for a given matrix $\mathbf{S} \in \mathbb{R}^>$ to the following system of equations

$$\begin{cases} \Gamma_1 = \mathbf{S}_{11} + (\sigma^2)^2 \mathbf{A} \Gamma_2^{-1} \mathbf{A}' - \sigma^2 \mathbf{A} \Gamma_2^{-1} \mathbf{S}_{12}' - \sigma^2 \mathbf{C}_{11} \Gamma_1^{-1} \mathbf{S}_{12} \mathbf{C}_{22}^{-1} \mathbf{A}' \Gamma_1^{-1} \mathbf{C}_{11} \\ \quad - (\sigma^2)^3 \mathbf{C}_{11} \Gamma_1^{-1} \mathbf{A} \mathbf{C}_{22}^{-1} \mathbf{A}' \Gamma_1^{-1} \mathbf{S}_{12} \mathbf{C}_{22}^{-1} \mathbf{A}' \Gamma_1^{-1} \mathbf{C}_{11} \\ \quad + (\sigma^2)^2 \mathbf{C}_{11} \Gamma_1^{-1} \mathbf{A} \mathbf{C}_{22}^{-1} \mathbf{S}_{22} \mathbf{C}_{22}^{-1} \mathbf{A}' \Gamma_1^{-1} \mathbf{C}_{11} \\ \Gamma_2 = \mathbf{S}_{22} + (\sigma^2)^2 \mathbf{A}' \Gamma_1^{-1} \mathbf{A} - \sigma^2 \mathbf{A}' \Gamma_1^{-1} \mathbf{S}_{12} - \sigma^2 \mathbf{C}_{22} \Gamma_2^{-1} \mathbf{S}_{12}' \mathbf{C}_{11}^{-1} \mathbf{A} \Gamma_2^{-1} \mathbf{C}_{22} \\ \quad - (\sigma^2)^3 \mathbf{C}_{22} \Gamma_2^{-1} \mathbf{A}' \mathbf{C}_{11}^{-1} \mathbf{A} \Gamma_2^{-1} \mathbf{S}_{12}' \mathbf{C}_{11}^{-1} \mathbf{A} \Gamma_2^{-1} \mathbf{C}_{22} \\ \quad + (\sigma^2)^2 \mathbf{C}_{22} \Gamma_2^{-1} \mathbf{A}' \mathbf{C}_{11}^{-1} \mathbf{S}_{11} \mathbf{C}_{11}^{-1} \mathbf{A} \Gamma_2^{-1} \mathbf{C}_{22} \\ 0 = \text{tr} [-2\sigma^2 \mathbf{C}_{11}^{-1} \mathbf{A} \Gamma_2^{-1} \mathbf{A}' + 2\sigma^2 \mathbf{C}_{11}^{-1} \mathbf{S}_{11} \mathbf{C}_{11}^{-1} \mathbf{A} \Gamma_2^{-1} \mathbf{A}' \\ \quad - \mathbf{S}_{12} \Gamma_2^{-1} \mathbf{A}' \mathbf{C}_{11}^{-1} - 2\sigma^2 \mathbf{S}_{12} \Gamma_2^{-1} \mathbf{A}' \mathbf{C}_{11}^{-1} \mathbf{A} \Gamma_2^{-1} \mathbf{A}' \mathbf{C}_{11}^{-1} \\ \quad - \mathbf{C}_{22}^{-1} \mathbf{S}_{12}' \Gamma_1^{-1} \mathbf{A} - 2\sigma^2 \mathbf{S}_{12}' \Gamma_1^{-1} \mathbf{A} \mathbf{C}_{22}^{-1} \mathbf{A}' \Gamma_1^{-1} \mathbf{A} \mathbf{C}_{22}^{-1} \\ \quad + 2\sigma^2 \mathbf{C}_{22}^{-1} \mathbf{S}_{22} \mathbf{C}_{22}^{-1} \mathbf{A}' \Gamma_1^{-1} \mathbf{A}] \\ \sigma^2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}, \end{cases}$$

where

$$\begin{aligned} a &= \text{tr} [-\Gamma_2^{-1} (\mathbf{A}' + \mathbf{F}' \mathbf{C}_{11}^{-1} \mathbf{A}) + \mathbf{C}_{11}^{-1} \mathbf{S}_{11} \mathbf{C}_{11}^{-1} (\mathbf{F} \Gamma_2^{-1} \mathbf{A}' + \mathbf{A} \Gamma_2^{-1} \mathbf{F}') \\ &\quad - \mathbf{S}_{12} \Gamma_2^{-1} \mathbf{A}' \mathbf{C}_{11}^{-1} (\mathbf{F} \Gamma_2^{-1} \mathbf{A}' + \mathbf{A} \Gamma_2^{-1} \mathbf{F}') \mathbf{C}_{11}^{-1} \\ &\quad + \mathbf{C}_{22}^{-1} \mathbf{S}_{12}' \Gamma_1^{-1} \mathbf{A} \mathbf{C}_{22}^{-1} (\mathbf{A}' \Gamma_1^{-1} \mathbf{F} - \mathbf{F}' \Gamma_1^{-1} \mathbf{A}) \\ &\quad + \mathbf{C}_{22}^{-1} \mathbf{S}_{22} \mathbf{C}_{22}^{-1} (\mathbf{A}' \Gamma_1^{-1} \mathbf{F} + \mathbf{F}' \Gamma_1^{-1} \mathbf{A})], \\ b &= -\text{tr} [\mathbf{C}_{11}^{-1} \mathbf{S}_{12} \Gamma_2^{-1} \mathbf{F}' + \mathbf{C}_{22}^{-1} \mathbf{S}_{12}' \Gamma_1^{-1} \mathbf{F}], \\ c &= \text{tr} [\mathbf{C}_{11} \mathbf{F}] \end{aligned}$$

over $\mathcal{S}_m^>$ then it is maximum likelihood estimator, denoted by $\hat{\Gamma}_{BML}$.

Observe that all equations depend on all unknown parameters; consequently, the procedure to determine MLE is challenging and time-consuming. When no solution exists within the cone of positive definite matrices $\mathcal{S}_m^>$, this type of estimator is not available and an alternative approach must be adopted. One of the possible methods to solve the above system of equations is the gradient decent method (cf. Boyd and Vandenberghe 2004), which will be used in Section 5. Naturally, this method yields a single solution.

The log-likelihood function is a concave function with respect to inverse of unstructured covariance matrix; cf. Barndorff-Nielsen (1978), Anderson and Olkin (1985). However, for a considered structure, it is challenging to show this property. Nevertheless, the gradient decent method always converges to the local optimum, and it is crucial to choose the starting point to find the global optimum. In this paper, our starting point is $\Gamma_1 = \mathbf{S}_{11}$, $\Gamma_2 = \mathbf{S}_{22}$, $\sigma^2 = 1$ and $\varrho = 0.5$, where diagonal blocks are diagonal blocks of unstructured MLE.

3.2 Least squares estimation

The second considered estimation method, least squares, is based on minimizing the distance between a sample covariance matrix and structured Γ in the sense of the Frobenius norm, i.e., minimization of

$$f(\Gamma) = \|\mathbf{S} - \Gamma\|_F^2 = \text{tr} [(\mathbf{S} - \Gamma)(\mathbf{S} - \Gamma)']$$

(cf. Ohlson and von Rosen 2010) over $\mathcal{S}_m$. However, since this minimizer is not necessarily positive definite or well-conditioned, we improve it using the shrinkage method and, if necessary, another improvement based on rescaling off-diagonal block through decreasing σ^2 .

3.2.1 Frobenius norm minimization

For a given $\mathbf{S}$, we define

$$\hat{\mathbf{\Gamma}}_F = \arg \min_{\mathbf{\Gamma} \in \mathcal{S}_m} f(\mathbf{\Gamma}).$$

The above minimum can be obtained by differentiation of

$$f(\mathbf{\Gamma}) = \text{tr}(\mathbf{S}^2) - 2 \text{tr}(\mathbf{S}_{11}\mathbf{\Gamma}_1) - 4\sigma^2 \text{tr}(\mathbf{S}_{12}\mathbf{A}') - 2 \text{tr}(\mathbf{S}_{22}\mathbf{\Gamma}_2) + \text{tr}(\mathbf{\Gamma}_1^2) + 2\sigma^4 \text{tr}(\mathbf{A}\mathbf{A}') + \text{tr}(\mathbf{\Gamma}_2^2) \quad (7)$$

with respect to $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$, σ^2 and ϱ . Respective derivatives have the following forms:

$$\frac{\partial f}{\partial \mathbf{\Gamma}_1} = \left\{ -2 \text{vec}'(\mathbf{S}_{11}) + \text{vec}'(\mathbf{I}_p) [(\mathbf{\Gamma}_1 \otimes \mathbf{I}_p) + (\mathbf{I}_p \otimes \mathbf{\Gamma}_1)] \right\} \mathbf{D}_p$$

$$\frac{\partial f}{\partial \mathbf{\Gamma}_2} = \left\{ -2 \text{vec}'(\mathbf{S}_{22}) + \text{vec}'(\mathbf{I}_q) [(\mathbf{\Gamma}_2 \otimes \mathbf{I}_q) + (\mathbf{I}_q \otimes \mathbf{\Gamma}_2)] \right\} \mathbf{D}_q$$

$$\frac{\partial f}{\partial \sigma^2} = -4 \text{vec}'(\mathbf{S}_{12}) \text{vec}(\mathbf{A}) + 4\sigma^2 \text{vec}'(\mathbf{I}_p) \text{vec}(\mathbf{A}\mathbf{A}')$$

$$\frac{\partial f}{\partial \varrho} = -4\sigma^2 \text{vec}'(\mathbf{S}_{12}) \text{vec}(\mathbf{F}) + 2\sigma^2 \text{vec}'(\mathbf{I}_p) (\mathbf{I}_p^2 + \mathbf{K}_{p,p})(\mathbf{A} \otimes \mathbf{I}_p) \text{vec}(\mathbf{F})$$

with $\mathbf{F}$ being a derivative of $\mathbf{A}$ with respect to ϱ given in (6).

Equating the derivatives to zero, we obtain the following theorem.

Theorem 2 If there exists unique solution for a given matrix $\mathbf{S} \in \mathbb{R}^>$ to the following system of equations

$$\begin{cases} \mathbf{\Gamma}_1 &= \mathbf{S}_{11} \\ \mathbf{\Gamma}_2 &= \mathbf{S}_{22} \\ \text{tr}(\mathbf{F}\mathbf{A}') &= \text{tr}(\mathbf{S}_{12}'\mathbf{F}) \\ \sigma^2 &= \frac{\text{tr}(\mathbf{S}_{12}'\mathbf{A})}{\text{tr}(\mathbf{A}\mathbf{A}')} \end{cases}$$

over $\mathcal{S}_m$, then it provides the Frobenius norm estimator, denoted by $\hat{\mathbf{\Gamma}}_F$.

Observe that in the above system of equations the third equation is a polynomial of order $2(p+q)-3$, which may have no solution within the space of symmetric matrices $\mathcal{S}_m$. In such cases, an alternative estimator must be chosen. Moreover, the equation may have multiple solutions, or the solution may be any real number. However, the estimator $\hat{\varrho}$ should be chosen from the interval $(-1; 1)$ in such a way that

$\Gamma \in \mathcal{S}_m$. Among all admissible roots, one that has the smallest value of the Frobenius norm between the sample covariance matrix $\mathbf{S}$ and the estimator $\hat{\Gamma}$ should be chosen. Moreover, since we consider covariance matrices, the final least squares estimator should be p.d. that is $\hat{\Gamma}_{BLS} \in \mathcal{S}_m^>$. To satisfy this property, if it is not present, we improve $\hat{\Gamma}_F$ using shrinkage.

3.2.2 Improvement by shrinkage method

Shrinking is a well-known method of estimator improvement broadly applied by many authors, e.g., Ledoit and Wolf (2004), Bodnar et al (2016), Ledoit and Wolf (2021), Bodnar et al (2022). Following Ledoit and Wolf (2004), shrinkage estimator, $\hat{\Gamma}_S$, is a convex combination of a given p.d. target matrix $\mathbf{T}$ and the estimator of covariance matrix, $\hat{\Gamma}$, that is

$$\hat{\Gamma}_S = (1 - \eta)\mathbf{T} + \eta\hat{\Gamma}, \quad (8)$$

where $\eta \in [0; 1]$ is called shrinkage intensity. Since in our paper we consider structured matrices, as a target matrix, we use a block diagonal matrix of the form

$$\hat{\mathbf{T}} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22} \end{pmatrix},$$

where $\mathbf{S}_{ii}$, $i = 1, 2$, are the diagonal blocks of a sample covariance matrix $\mathbf{S}$. Note that it depends on the data and thus we denote it by $\hat{\mathbf{T}}$ instead of $\mathbf{T}$. Moreover, $\hat{\mathbf{T}}$ is simply the MLE of the block diagonal covariance matrix. Naturally, as the estimator of Γ we use $\hat{\Gamma}_F$; however, we keep the notation $\hat{\Gamma}$ as this method will also be used in future sections.

Observing that the diagonal blocks of $\hat{\mathbf{T}}$ and $\hat{\Gamma}$ are equal, the shrinkage estimator can be rewritten as:

$$\hat{\Gamma}_S = \hat{\mathbf{T}} + \eta(\hat{\Gamma} - \hat{\mathbf{T}}).$$

Following Ledoit and Wolf (2004) the shrinkage intensity should be determined by minimization of quadratic risk

$$g(\eta) = \mathbb{E} \left(\|\hat{\Gamma}_S - \Gamma\|_F^2 \right) = \text{tr} \left[\mathbb{E} \left(\hat{\mathbf{T}}^2 \right) \right] + \eta \text{tr} \left[\mathbb{E} \left(\hat{\mathbf{T}}\hat{\Gamma} \right) \right] + \eta^2 \text{tr} \left[\mathbb{E} \left(\hat{\Gamma}^2 \right) \right] - 2 \text{tr} \left[\mathbb{E} \left(\hat{\mathbf{T}}\Gamma \right) \right] - 2\eta \text{tr} \left[\mathbb{E} \left(\hat{\Gamma}\Gamma \right) \right] + \text{tr} \left[\Gamma^2 \right]. \quad (9)$$

Since the quadratic risk is convex, the stationary point gives the minimum. Differentiating $g(\eta)$ and equating it to zero gives the following equality

$$\text{tr} \left[\mathbb{E} \left(\hat{\mathbf{T}}\hat{\Gamma} \right) \right] + 2\eta \text{tr} \left[\mathbb{E} \left(\hat{\Gamma}^2 \right) \right] - 2 \text{tr} \left[\mathbb{E} \left(\hat{\Gamma}\Gamma \right) \right] = 0. \quad (10)$$

An important role in this equation plays the matrix $\mathbf{S}$. It is known that $\mathbf{S}$ has the Wishart distribution, that is, $n\mathbf{S} \sim W_m(n-1, \mathbf{\Gamma})$. Moreover,

$$\begin{aligned} n\mathbf{S}_{11} &\sim W_p(n-1, \mathbf{\Gamma}_{11}), \\ n\mathbf{S}_{22} &\sim W_q(n-1, \mathbf{\Gamma}_{22}). \end{aligned}$$

Therefore, equation (10) can be solved using the properties of Wishart distribution given by Holgersson and Pielaszkiewicz (2020) and Pielaszkiewicz and von Rosen (2021).

Lemma 1 (Holgersson and Pielaszkiewicz 2020)

Let $\mathbf{W} \sim W_p(n, \mathbf{\Sigma})$, $p < n$. Then,

- (i) $E(\mathbf{W}^2) = (n^2 + n)\mathbf{\Sigma}^2 + n \operatorname{tr}(\mathbf{\Sigma})\mathbf{\Sigma}$
- (ii) $E[\operatorname{tr}(\mathbf{A}\mathbf{W})] = n \operatorname{tr}(\mathbf{A}\mathbf{\Sigma})$.

Applying this lemma to (10) we obtain

$$\begin{aligned} \operatorname{tr} [E(\widehat{\mathbf{T}}\widehat{\mathbf{T}})] &= \frac{n-1}{n^2} [n \operatorname{tr}(\mathbf{\Gamma}_{11}^2) + \operatorname{tr}^2(\mathbf{\Gamma}_{11}) + n \operatorname{tr}(\mathbf{\Gamma}_{22}^2) + \operatorname{tr}^2(\mathbf{\Gamma}_{22})] \\ \operatorname{tr} [E(\widehat{\mathbf{T}})\mathbf{\Gamma}] &= \frac{n-1}{n} [\operatorname{tr}(\mathbf{\Gamma}_{11}^2) + 2 \operatorname{tr}(\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{12}') + \operatorname{tr}(\mathbf{\Gamma}_{22}^2)] \\ \operatorname{tr} [E(\widehat{\mathbf{T}}^2)] &= \frac{n-1}{n^2} \left\{ n [\operatorname{tr}(\mathbf{\Gamma}_{11}^2) + 2 \operatorname{tr}(\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{12}') + \operatorname{tr}(\mathbf{\Gamma}_{22}^2)] + \operatorname{tr}^2(\mathbf{\Gamma}_{11}) + \operatorname{tr}^2(\mathbf{\Gamma}_{22}) \right. \\ &\quad \left. + 2 \operatorname{tr}^2(\mathbf{\Gamma}_{11}^2\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{12}') + 2 \operatorname{tr}^2(\mathbf{\Gamma}_{22}^2\mathbf{\Gamma}_{12}'\mathbf{\Gamma}_{12}) + 2 \operatorname{tr}^2(\mathbf{\Gamma}_{12}'\mathbf{\Gamma}_{12}) \right\} \end{aligned}$$

and, in consequence,

$$\begin{aligned} &\frac{n-1}{n^2} \left\{ n \operatorname{tr}(\mathbf{\Gamma}_{11}^2) + \operatorname{tr}^2(\mathbf{\Gamma}_{11}) + n \operatorname{tr}(\mathbf{\Gamma}_{22}^2) + \operatorname{tr}^2(\mathbf{\Gamma}_{22}) + 2\eta \left[n \left(\operatorname{tr}(\mathbf{\Gamma}_{11}^2) + 2 \operatorname{tr}(\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{12}') + \operatorname{tr}(\mathbf{\Gamma}_{22}^2) \right) \right. \right. \\ &\quad \left. \left. + \operatorname{tr}^2(\mathbf{\Gamma}_{11}) + \operatorname{tr}^2(\mathbf{\Gamma}_{22}) + 2 \operatorname{tr}^2(\mathbf{\Gamma}_{11}^2\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{12}') + 2 \operatorname{tr}^2(\mathbf{\Gamma}_{22}^2\mathbf{\Gamma}_{12}'\mathbf{\Gamma}_{12}) + 2 \operatorname{tr}^2(\mathbf{\Gamma}_{12}'\mathbf{\Gamma}_{12}) \right] \right\} \\ &\quad - \frac{2(n-1)}{n} [\operatorname{tr}(\mathbf{\Gamma}_{11}^2) + 2 \operatorname{tr}(\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{12}') + \operatorname{tr}(\mathbf{\Gamma}_{22}^2)] = 0, \end{aligned}$$

which is simply the linear equation in η .

Therefore, the minimum is attained at

$$\eta = \frac{N}{D},$$

where

$$\begin{aligned}
N &= n \left[\text{tr}(\mathbf{\Gamma}_{11}^2) + 4 \text{tr}(\mathbf{\Gamma}_{12} \mathbf{\Gamma}_{12}') + \text{tr}(\mathbf{\Gamma}_{22}^2) \right] - \text{tr}^2(\mathbf{\Gamma}_{11}) - \text{tr}^2(\mathbf{\Gamma}_{22}) \\
D &= 2 \left\{ n \left[\text{tr}(\mathbf{\Gamma}_{11}^2) + 2 \text{tr}(\mathbf{\Gamma}_{12} \mathbf{\Gamma}_{12}') + \text{tr}(\mathbf{\Gamma}_{22}^2) \right] + \text{tr}^2(\mathbf{\Gamma}_{11}) + \text{tr}^2(\mathbf{\Gamma}_{22}) \right. \\
&\quad \left. + 2 \text{tr}^2(\mathbf{\Gamma}_{11} \mathbf{\Gamma}_{12} \mathbf{\Gamma}_{12}') + 2 \text{tr}^2(\mathbf{\Gamma}_{22} \mathbf{\Gamma}_{12}' \mathbf{\Gamma}_{12}) + 2 \text{tr}^2(\mathbf{\Gamma}_{12}' \mathbf{\Gamma}_{12}) \right\}
\end{aligned} \quad (11)$$

It can be checked that

$$\hat{\mathbf{\Gamma}}_S = \hat{\mathbf{T}} + \frac{N}{D} (\hat{\mathbf{\Gamma}} - \hat{\mathbf{T}})$$

with N and D defined in (11) still does not need to be positive definite or well-conditioned. In such cases to get a least squares estimator $\hat{\mathbf{\Gamma}}_{BLS}$, we further improve $\hat{\mathbf{\Gamma}}_S$ by rescaling off-diagonal blocks via decreasing σ^2 . This is a numerical procedure which ends up when the condition number of p.d. $\hat{\mathbf{\Gamma}}_{BLS}$ gets below a given threshold; for more details, see Markiewicz et al (2024).

Summing up, the algorithm for determining $\hat{\mathbf{\Gamma}}_{BLS}$ is as follows.

1. Determine $\hat{\mathbf{\Gamma}}_F$ from Theorem 2.
2. If $\hat{\mathbf{\Gamma}}_F$ is p.d. and well-conditioned $\hat{\mathbf{\Gamma}}_{BLS} := \hat{\mathbf{\Gamma}}_F$ and stop the algorithm, else determine $\hat{\mathbf{\Gamma}}_S$.
3. If $\hat{\mathbf{\Gamma}}_S$ is p.d. and well-conditioned $\hat{\mathbf{\Gamma}}_{BLS} := \hat{\mathbf{\Gamma}}_S$, else improve $\hat{\mathbf{\Gamma}}_S$ by rescaling off-diagonal blocks via decreasing σ^2 (repeatedly multiplying by some number smaller than 1 (e.g., 0.99) until the estimator has required properties), which ends up with the p.d. and well-condition $\hat{\mathbf{\Gamma}}_{BLS}$.

Algorithm 1 Determination of the block least squares estimate

4 Alternative methods of estimation

Note that estimators considered above are not given in explicit forms and solving systems of equations is time-consuming. The advantage of MLE is positive definiteness, while, the disadvantage of LS is multi-step procedure where there is a need to check positive definiteness of estimator at each step. Therefore, in this section, we propose alternative methods of estimation of the considered structure.

Recall that the considered covariance structure is relevant to experiment in which correlations decrease with the distance between variables (AR(1)) with additional noise in diagonal blocks. Thus, matrix $\mathbf{\Gamma}$ given in (2) can be rewritten as

$$\mathbf{\Gamma} = \mathbf{\Psi} + \mathbf{W} = \mathbf{\Psi} + \begin{pmatrix} \mathbf{\Gamma}_1 - \mathbf{\Psi}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}_2 - \mathbf{\Psi}_{22} \end{pmatrix}, \quad (12)$$

where $\mathbf{\Psi} = (\Psi_{ij})$, $i, j = 1, 2$, represents AR(1) structure given in (1) with relevant dimensions of blocks and $\mathbf{W}$ represents the noise.

In this section, we propose alternative methods of estimation based on the estimation of $\mathbf{\Psi}$ and $\mathbf{W}$ separately, that is,

$$\widehat{\Gamma} = \widehat{\Psi} + \widehat{\mathbf{W}}.$$

Maximum likelihood estimation of the AR(1) structure Ψ was considered in times series models by Azzalini (1981). Nowadays, the AR(1) structure of covariance matrix is widely studied and can be estimated by several methods, e.g., ML (cf. Filipiak et al 2017), LS (cf. Cui et al 2016), intuitive method based on the properties of the entries of AR(1) structure (cf. Roy and Leiva 2008; Žežula 2006), as well as the method based on the inverse of AR(1) structure. The latter estimator is obtained by minimizing the difference between $\mathbf{S}\Psi^{-1}$ and the identity matrix in the sense of the Frobenius norm. Note also that the estimator obtained by intuitive method will be called the naive estimator. The proposed estimators will be denoted respectively by $\widehat{\Psi}_{ML}$, $\widehat{\Psi}_{LS}$, $\widehat{\Psi}_N$ and $\widehat{\Psi}_{INV}$.

Due to (12) estimator of $\mathbf{W}$ is chosen as

$$\widehat{\mathbf{W}} = \begin{pmatrix} \mathbf{S}_{11} - \widehat{\Psi}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22} - \widehat{\Psi}_{22} \end{pmatrix}$$

as the MLE of the block diagonal covariance matrix is simply the projection of $\mathbf{S}$ onto the space of diagonal matrices, i.e.,

$$\begin{pmatrix} \mathbf{S}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22} \end{pmatrix}$$

(Filipiak and Klein 2021; Markiewicz et al 2024) and $\widehat{\Psi}_{ii}$, $i = 1, 2$, are the estimators obtained by one of the methods proposed in the previous paragraph. Therefore, the respective estimator of $\mathbf{W}$ will be denoted by $\widehat{\mathbf{W}}_{ML}$, $\widehat{\mathbf{W}}_{LS}$, $\widehat{\mathbf{W}}_N$, $\widehat{\mathbf{W}}_{INV}$.

The estimators are the sum of $\widehat{\Psi}_k + \widehat{\mathbf{W}}_k$, $k \in \{ML, LS, N, INV\}$. However, all of them are not always positive definite or well conditioned, therefore, similarly as in the Sect. 3.2, we use the shrinkage method given in Sect. 3.2.2 to improve them and an additional approach based on rescaling off-diagonal blocks via decreasing σ^2 , if needed. Summing up, we propose four estimators of Γ : $\widehat{\Gamma}_{ML}$, $\widehat{\Gamma}_{LS}$, $\widehat{\Gamma}_N$, $\widehat{\Gamma}_{INV}$, which are p.d. and well-conditioned.

In the following sections, we present four mentioned methods of estimation of AR(1) structure.

4.1 Maximum likelihood estimation

Filipiak et al (2017) considered the doubly multivariate model with covariance matrix as a Kronecker product of two matrices, where one of the factors has an AR(1) structure. Therefore, to determine $\widehat{\Psi}_{ML}$ it is enough to assume the dimension of the second factor equal to one. Therefore, we can formulate the following theorem.

Theorem 3 If there exists unique solution for a given matrix $\mathbf{S} \in \mathbb{R}^{>}$ to the following system of equations

$$\begin{cases} \varrho^3 [(m-1) \operatorname{tr}(\mathbf{S}\mathbf{C}_1)] + \varrho^2 \left[\left(\frac{2-m}{2} \right) \operatorname{tr}(\mathbf{S}\mathbf{C}_2) \right] - \varrho [\operatorname{tr}(\mathbf{S}) + m \operatorname{tr}(\mathbf{S}\mathbf{C}_1)] + \frac{m}{2} \operatorname{tr}(\mathbf{S}\mathbf{C}_2) = 0 \\ \sigma^2 = \frac{1}{m(1-\varrho^2)} [\operatorname{tr}(\mathbf{S}) + \varrho^2 \operatorname{tr}(\mathbf{S}\mathbf{C}_1) - \varrho \operatorname{tr}(\mathbf{S}\mathbf{C}_2)], \end{cases}$$

where $\mathbf{C}_1 = \operatorname{diag}(0, \mathbf{1}'_{m-2}, 0)$ and $\mathbf{C}_2$ is a tridiagonal $m \times m$ matrix with 0 on the main diagonal, 1 on the first superdiagonal and on the first subdiagonal, then it is maximum likelihood estimator, denoted by $\hat{\Psi}_{ML}$.

Similarly to Theorem 2, the estimator is obtained by finding the root of a polynomial; however, in this case, the polynomial is of order 3. It may have no root in the interval $(-1; 1)$ - then an alternative estimator must be selected, or may have multiple roots, then the one with the maximum value of the likelihood function should be chosen.

4.2 Least squares estimation

Following Cui et al (2016), we can formulate Theorem 4.

Theorem 4 If there exists unique solution for a given matrix $\mathbf{S} \in \mathbb{R}^>$ to the following system of equation

$$\begin{cases} - \left[m + 2 \sum_{i=1}^{m-1} (m-i) \varrho^{2i} \right] \sum_{i=1}^{m-1} i \varrho^{i-1} \operatorname{tr}(\mathbf{S}\mathbf{H}_i) \\ + 2 \left[\operatorname{tr}(\mathbf{S}) + \sum_{i=1}^{m-1} \varrho^i \operatorname{tr}(\mathbf{S}\mathbf{H}_i) \right] \sum_{i=1}^{m-1} (m-i) i \varrho^{2i-1} = 0 \\ \sigma^2 = \frac{\operatorname{tr}(\mathbf{S}) + \sum_{i=1}^{m-1} \varrho^i \operatorname{tr}(\mathbf{S}\mathbf{H}_i)}{m + 2 \sum_{i=1}^{m-1} (m-i) \varrho^{2i}}, \end{cases}$$

where $\mathbf{H}_i$ is a symmetric matrix with i -th superdiagonal and subdiagonal elements equal to 1 and all other elements equal to 0, $\mathbf{H}_0 = \mathbf{I}_m$, then it is Frobenius norm estimator denoted as $\hat{\Psi}_{LS} \in \mathbb{R}^>$.

In contrast to Ψ_{ML} in the least squares estimation method $\hat{\varrho}$ is obtained as the solution to a polynomial of order $3m - 4$. Similarly to the block least squares approach, this polynomial may have no root within the interval $(-1; 1)$ - an alternative estimator must be selected, or may have multiple roots. In such a case, we choose the root from the interval $(-1; 1)$ for which the Frobenius norm between $\mathbf{S}$ and the estimator of $\mathbf{\Gamma}$ has the minimum value.

We now demonstrate that the estimator of σ^2 is positive. It is evident that the denominator in the second equation remains positive. To examine the numerator, observe that it can be rewritten as follows:

$$\operatorname{tr}(\mathbf{S}) + \sum_{i=1}^{m-1} \varrho^i \operatorname{tr}(\mathbf{S}\mathbf{H}_i) = \operatorname{tr} \left[\mathbf{S}(\mathbf{I}_m + \sum_{i=1}^{m-1} \varrho^i \mathbf{H}_i) \right].$$

By assumption, both $\mathbf{S}$ and the matrix $\mathbf{I}_m + \sum_{i=1}^{m-1} \varrho^i \mathbf{H}_i$ are positive definite. Let us apply the Cholesky decomposition $\mathbf{S} = \mathbf{A}'\mathbf{A}$, where $\mathbf{A}$ is a full-rank matrix. Substituting, we obtain $\text{tr} \left[\mathbf{S} \left(\mathbf{I}_m + \sum_{i=1}^{m-1} \varrho^i \mathbf{H}_i \right) \right] = \text{tr} \left[\mathbf{A} \left(\mathbf{I}_m + \sum_{i=1}^{m-1} \varrho^i \mathbf{H}_i \right) \mathbf{A}' \right]$. According to Kollo and von Rosen (2005) (Proposition 1.1.5.), $\mathbf{A} \left(\mathbf{I}_m + \sum_{i=1}^{m-1} \varrho^i \mathbf{H}_i \right) \mathbf{A}'$ is positive definite. Hence, its trace is positive, which implies that the estimator of σ^2 is indeed positive.

4.3 Naive estimation

Following Roy and Leiva (2008) estimator $\hat{\Psi}_N \in \mathbb{R}^>$ of the AR(1) structure is given by

$$\begin{cases} \hat{\varrho} = \left(\frac{\hat{\varrho}_1^{p-1} + \hat{\varrho}_2^{\frac{p-1}{2}} + \hat{\varrho}_3^{\frac{p-1}{3}} + \cdots + \hat{\varrho}_{p-1}}{p-1} \right)^{\frac{1}{p-1}} \\ \hat{\sigma}^2 = \frac{\text{tr}(\mathbf{S})}{m}, \end{cases}$$

where $\hat{\varrho}_i$ is the average of the i th superdiagonal elements of $\mathbf{S}$, $i = 1, \dots, p-1$.

4.4 Estimation via AR(1) inverse

In this section, we propose the method based on minimizing the distance between $\mathbf{S}\Psi^{-1}$ and identity matrix in the sense of the Frobenius norm, i.e.,

$$h(\hat{\Psi}) = \min_{\Psi \in \mathbb{R}^>} h(\Psi^{-1}) = \min_{\Psi \in \mathbb{R}^>} \|\mathbf{S}\Psi^{-1} - \mathbf{I}_m\|_F^2.$$

The method uses three-diagonal form of the inverse of AR(1) structure as

$$\Psi^{-1} = \frac{1}{\sigma^2} \cdot \frac{1}{1-\varrho^2} (\mathbf{I}_m + \varrho^2 \mathbf{C}_1 - \varrho \mathbf{C}_2), \quad (13)$$

where $\mathbf{C}_1$ and $\mathbf{C}_2$ are defined in Theorem 3. Note that, $\mathbf{C}_2 = \mathbf{H}_1$, where $\mathbf{H}_1$ is defined in Theorem 4.

Using (13)

$$\begin{aligned} h(\Psi^{-1}) = & -\frac{1}{\sigma^2} \cdot \frac{1}{(1-\varrho^2)^2} [\text{tr}(\mathbf{S}^2) + 2\varrho^2 \text{tr}(\mathbf{S}\mathbf{C}_1\mathbf{S}) - 2\varrho \text{tr}(\mathbf{S}\mathbf{C}_2\mathbf{S}) \\ & + \varrho^4 \text{tr}(\mathbf{S}\mathbf{C}_1\mathbf{S}\mathbf{C}_1) - 2\varrho^3 \text{tr}(\mathbf{S}\mathbf{C}_1\mathbf{S}\mathbf{C}_2) + \varrho^2 \text{tr}(\mathbf{S}\mathbf{C}_2\mathbf{S}\mathbf{C}_2)] \\ & - \frac{2}{\sigma^2} \cdot \frac{1}{(1-\varrho^2)^2} [\text{tr} \mathbf{S} + \varrho^2 \text{tr}(\mathbf{S}\mathbf{C}_1) - \varrho \text{tr} \mathbf{S}\mathbf{C}_2]. \end{aligned} \quad (14)$$

Differentiating (14) with respect to σ^2 and ϱ we obtain

$$\begin{aligned}\frac{\partial h}{\partial \sigma^2} &= -\frac{2}{(\sigma^2)^3} \cdot \frac{1}{(1-\varrho^2)^2} [\text{tr}(\mathbf{S}^2) + 2\varrho^2 \text{tr}(\mathbf{SC}_1\mathbf{S}) - 2\varrho \text{tr}(\mathbf{SC}_2\mathbf{S}) + \varrho^4 \text{tr}(\mathbf{SC}_1\mathbf{SC}_1) \\ &\quad - 2\varrho^3 \text{tr}(\mathbf{SC}_1\mathbf{SC}_2) + \varrho^2 \text{tr}(\mathbf{SC}_2\mathbf{SC}_2)] \\ &\quad + \frac{2}{(\sigma^2)^2} \cdot \frac{1}{(1-\varrho^2)} [\text{tr}(\mathbf{S}) + \varrho^2 \text{tr}(\mathbf{SC}_1) - \varrho \text{tr}(\mathbf{SC}_2)], \\ \frac{\partial h}{\partial \varrho} &= \frac{1}{(\sigma^2)^2} \cdot \frac{1}{(1-\varrho^2)^3} [4 \text{tr}(\mathbf{S}^2) + 2\varrho(1+\varrho^2)(2 \text{tr}(\mathbf{SC}_1\mathbf{S}) + \text{tr}(\mathbf{SC}_2\mathbf{SC}_2)) - \\ &\quad - 2(1+3\varrho^2) \text{tr}(\mathbf{SC}_2\mathbf{S}) + 4\varrho^3 \text{tr}(\mathbf{SC}_1\mathbf{SC}_1) - 2(3\varrho^2 + \varrho^4) \text{tr}(\mathbf{SC}_1\mathbf{SC}_2)] \\ &\quad + \frac{1}{(\sigma^2)} \cdot \frac{1}{(1-\varrho^2)^2} [-4\varrho(\text{tr}(\mathbf{S}) + \text{tr}(\mathbf{SC}_1)) + 2(1+\varrho^2) \text{tr}(\mathbf{SC}_2)].\end{aligned}$$

Equating it to zero and simplifying, we get the following theorem.

Theorem 5 If there exists unique solution for a given matrix $\mathbf{S} \in \mathbb{R}^>$ to the following system

$$\begin{cases} a_6\varrho^6 + a_5\varrho^5 + a_4\varrho^4 + a_3\varrho^3 + a_2\varrho^2 + a_1\varrho + a_0 = 0 \\ \sigma^2 = \frac{\text{tr}(\mathbf{S}^2) - 2\varrho \text{tr}(\mathbf{SC}_2\mathbf{S}) + \varrho^2 [2 \text{tr}(\mathbf{SC}_1\mathbf{S}) - \text{tr}(\mathbf{SC}_2\mathbf{SC}_2)] - 2\varrho^3 \text{tr}(\mathbf{SC}_1\mathbf{SC}_2) + \varrho^4 \text{tr}(\mathbf{SC}_1\mathbf{SC}_1)}{(1-\varrho^2) [\text{tr}(\mathbf{S}) + \varrho^2 \text{tr}(\mathbf{SC}_1) - \varrho \text{tr}(\mathbf{SC}_2)]} \end{cases}$$

with

$$\begin{aligned}a_6 &= -\text{tr}(\mathbf{SC}_1\mathbf{SC}_1) + \text{tr}(\mathbf{SC}_2) \text{tr}(\mathbf{SC}_1\mathbf{SC}_1) \\ a_5 &= 2 \text{tr}(\mathbf{SC}_1\mathbf{S}) \text{tr}(\mathbf{SC}_1) + \text{tr}(\mathbf{SC}_2\mathbf{SC}_2) \text{tr}(\mathbf{SC}_1) - \text{tr}(\mathbf{SC}_1\mathbf{SC}_2) \text{tr}(\mathbf{SC}_2) \\ &\quad - 2 \text{tr}(\mathbf{S}) \text{tr}(\mathbf{SC}_1\mathbf{SC}_1) \\ a_4 &= -3 \text{tr}(\mathbf{SC}_2\mathbf{S}) \text{tr}(\mathbf{SC}_1) + \text{tr}(\mathbf{SC}_1) \text{tr}(\mathbf{SC}_1\mathbf{SC}_2) + 3 \text{tr}(\mathbf{S}) \text{tr}(\mathbf{SC}_1\mathbf{SC}_2) \\ &\quad - \text{tr}(\mathbf{SC}_1\mathbf{SC}_1) \text{tr}(\mathbf{SC}_2) \\ a_3 &= 2 \text{tr}(\mathbf{S}^2) \text{tr}(\mathbf{SC}_1) - 2 \text{tr}(\mathbf{SC}_1) \text{tr}(\mathbf{SC}_1\mathbf{S}) - 2 \text{tr}(\mathbf{S}) \text{tr}(\mathbf{SC}_1\mathbf{S}) + \text{tr}(\mathbf{SC}_2) \text{tr}(\mathbf{SC}_2\mathbf{S}) \\ &\quad + \text{tr}(\mathbf{SC}_1\mathbf{SC}_2) \text{tr}(\mathbf{SC}_2) - \text{tr}(\mathbf{SC}_1) \text{tr}(\mathbf{SC}_2\mathbf{SC}_2) - \text{tr}(\mathbf{S}) \text{tr}(\mathbf{SC}_2\mathbf{SC}_2) \\ &\quad + 2 \text{tr}(\mathbf{SC}_1\mathbf{SC}_1) \text{tr}(\mathbf{S}) \\ a_2 &= -\text{tr}(\mathbf{S}^2) \text{tr}(\mathbf{SC}_2) + 3 \text{tr}(\mathbf{SC}_1) \text{tr}(\mathbf{SC}_2\mathbf{S}) + \text{tr}(\mathbf{S}) \text{tr}(\mathbf{SC}_2\mathbf{S}) - 3 \text{tr}(\mathbf{SC}_1\mathbf{SC}_2) \text{tr}(\mathbf{S}) \\ a_1 &= 2 \text{tr}(\mathbf{SC}_1\mathbf{S}) \text{tr}(\mathbf{S}) - \text{tr}(\mathbf{SC}_2) \text{tr}(\mathbf{SC}_2\mathbf{S}) + \text{tr}(\mathbf{SC}_2\mathbf{SC}_2) \text{tr}(\mathbf{S}) - 2 \text{tr}(\mathbf{SC}_1) \text{tr}(\mathbf{S}^2) \\ a_0 &= \text{tr}(\mathbf{SC}_2) \text{tr}(\mathbf{S}^2) - \text{tr}(\mathbf{SC}_2\mathbf{S}) \text{tr}(\mathbf{S})\end{aligned}$$

then it is estimator of AR(1) structure minimizing function (14) denoted as $\hat{\Psi}_{INV} \in \mathbb{R}^>$.

Similarly to the least squares approaches, the polynomial may have no solution within the interval $(-1; 1)$ and then an alternative estimator must be considered. If multiple solutions exist within this interval, the root that has the smallest value of the Frobenius norm $\|\mathbf{S}\Psi^{-1} - \mathbf{I}_m\|_F^2$ is selected. The proof that $\sigma^2 > 0$ is not straightforward. Hence, when σ^2 is found to be negative, it is appropriate to employ a different estimator.

5 Simulation study

In this section, we perform a simulation study to compare the proposed estimators of the block covariance matrix with respect to the following criteria:

- Squared Frobenius norm $f(\hat{\Gamma}_k) = \|\hat{\Gamma}_k - \Gamma\|_F^2$
- Loss functions (Ledoit and Wolf 2021)
 - Stein loss

$$L_{Stein}(\hat{\Gamma}_k) = \frac{1}{m} \left[\text{tr}(\Gamma^{-1} \hat{\Gamma}_k) - \log |\Gamma^{-1} \hat{\Gamma}_k| \right] - 1$$

- Inverse Stein loss

$$L_{IStein}(\hat{\Gamma}_k) = \frac{1}{m} \left[\text{tr}(\Gamma \hat{\Gamma}_k^{-1}) - \log |\Gamma \hat{\Gamma}_k^{-1}| \right] - 1$$

- Quadratic loss

$$L_Q(\hat{\Gamma}_k) = \text{tr}(\Gamma^{-1} \hat{\Gamma}_k)^2 - 2 \text{tr}(\Gamma^{-1} \hat{\Gamma}_k) + m$$

- Computing time,

where $k \in \{BML, BLS, ML, LS, N, INV\}$. The estimator which gives the smallest values of considered function is the best one. Obviously, $\hat{\Gamma}_F$ gives the minimum of the first criterion; however, this estimator is not under comparison, since it is not p.d. Note also, that considered $\hat{\Gamma}_{BLS}$ estimator is the improvement of $\hat{\Gamma}_F$ and thus it is not necessarily the same as $\hat{\Gamma}_F$; see Algorithm 1.

Simulations are conducted for data generated from matrix normal distribution, $\mathbf{Z} \sim N_{p+q,n}(\mathbf{0}, \Gamma, \mathbf{I}_n)$, where $(p, q) \in \{(3, 7), (5, 10)\}$, $n = 100$ and a p.d. matrix Γ with the structure considered in (2) with $\sigma^2 = 1$, $\varrho \in \{-0.7, -0.5, 0, 0.5, 0.7\}$ and Γ_1 and Γ_2 are randomly generated matrices of proper orders, such that Γ is p.d. matrix and all considered estimators exist. The exact forms of Γ_1 and Γ_2 are available from the authors on request.

Figures 1, 2, 3 present boxplots of the values of squared Frobenius norm and losses for the considered estimators of block covariance matrix for $(p, q) = (3, 7)$ and $\varrho \in \{-0.5, 0, 0.5\}$, respectively, for 100 runs. We can observe that for each ϱ all estimators are comparable in the sense of squared Frobenius norm; however, it can be noticed that $\hat{\Gamma}_{BML}$ and $\hat{\Gamma}_N$ give the smallest values of all three loss functions than remaining estimators, while $\hat{\Gamma}_{INV}$ gives the highest values and the most varied results.

Apart from the values of loss functions, the computing time can also be of interest since some of the estimators are very simple to obtain ($\hat{\Gamma}_N$), some of them need numerical solving of complex system of equations ($\hat{\Gamma}_{BML}$) or just numerical solving of polynomials ($\hat{\Gamma}_{BLS}$, $\hat{\Gamma}_{ML}$, $\hat{\Gamma}_{LS}$, $\hat{\Gamma}_{INV}$). Moreover, all estimators except $\hat{\Gamma}_{BML}$ need improvement by shrinking and rescaling. Therefore, in Table 1 the average time of computing $\hat{\Gamma}_k$, $k \in \{BML, BLS, ML, LS, N, INV\}$ over 100 runs is given. It can be observed that the computing time of $\hat{\Gamma}_{BML}$ is the longest one, which means that the complexity of the BML algorithm implies longer computing time than

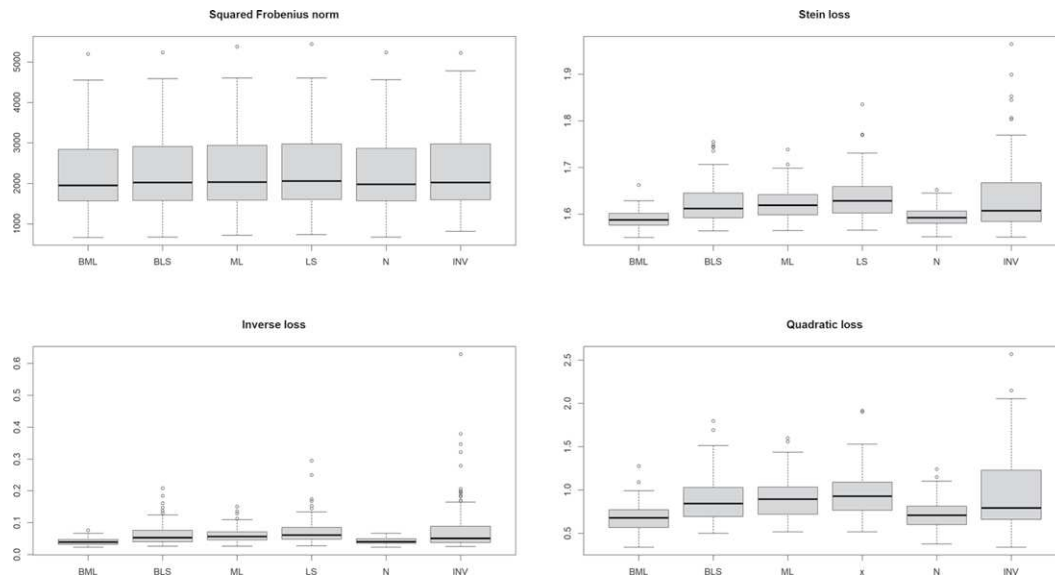


Fig. 1 The boxplots of values of squared Frobenius norm and loss functions for $\hat{\Gamma}_k$, $k \in \{BML, BLS, ML, LS, N, INV\}$, and $(p, q, \varrho) = (3, 7, -0.5)$

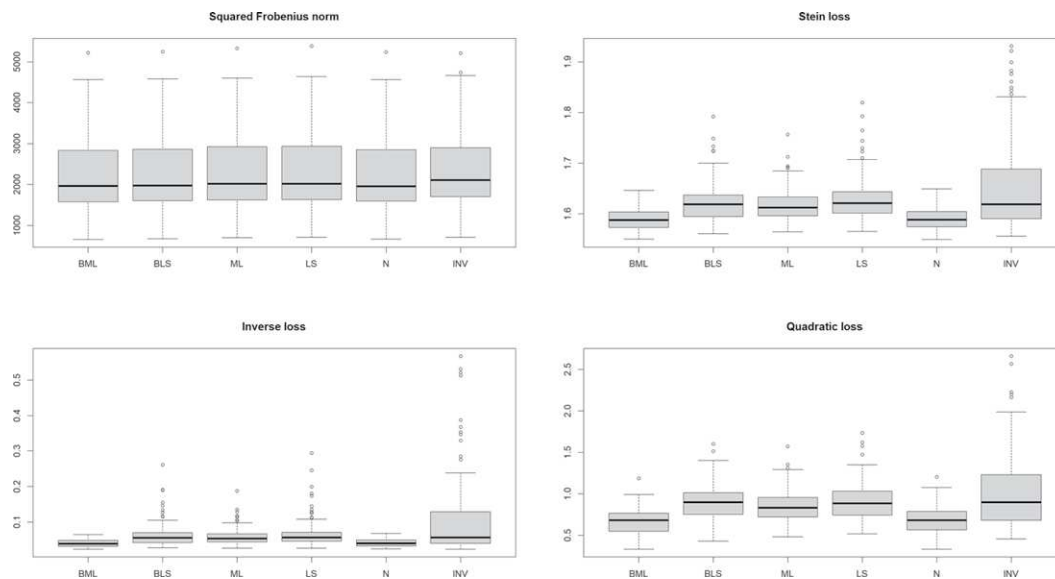


Fig. 2 The boxplots of values of squared Frobenius norm and loss functions for $\hat{\Gamma}_k$, $k \in \{BML, BLS, ML, LS, N, INV\}$, and $(p, q, \varrho) = (3, 7, 0)$

improvement of other estimators via shrinking and rescaling. On the other hand $\hat{\Gamma}_N$ is the simplest estimator; however, its improvement requires additional time, which makes general computing time comparable with the ML, LS and INV method. Note that BLS had the shortest computational time.

In summary, for $(p, q) = (3, 7)$, the naive estimator $\hat{\Gamma}_N$ seems to be the best one.

In Figs. 4, 5, 6 the boxplots of the values of squared Frobenius norm and loss functions for the considered estimators for $(p, q) = (5, 10)$, $\varrho \in \{-0.5, 0, 0.5\}$ and 100 runs are presented. Similarly as in the previous case the values of the Frobenius norm

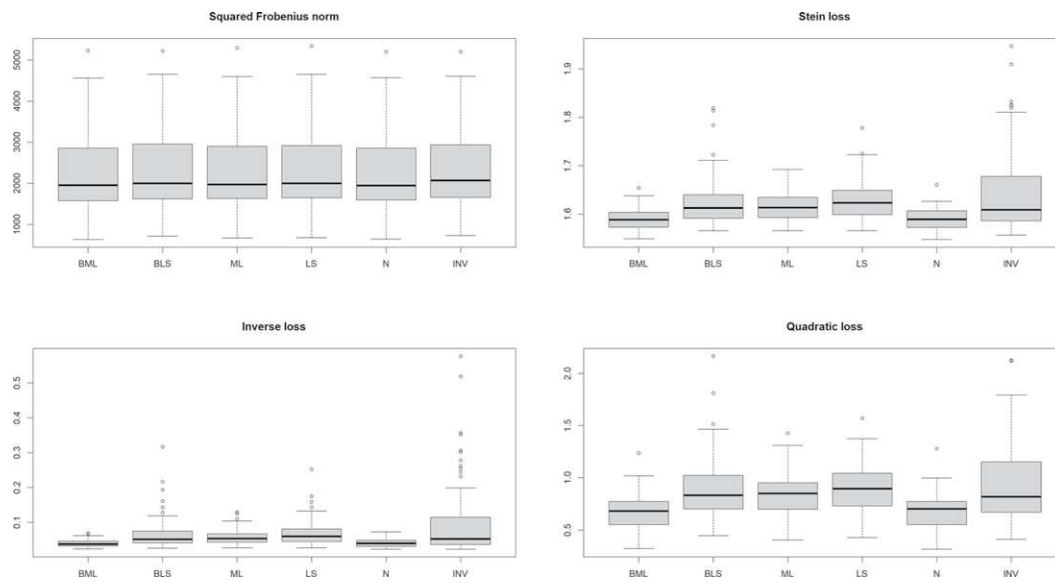


Fig. 3 The boxplots of values of squared Frobenius norm and loss functions for $\hat{\Gamma}_k$, $k \in \{BML, BLS, ML, LS, N, INV\}$, and $(p, q, \varrho) = (3, 7, 0.5)$

Table 1 Average time (in seconds) needed to determine estimates of Γ for $(p, q) = (3, 7)$, $n = 100$ over 100 simulation runs

ϱ	BML	BLS	ML	LS	N	INV
-0.7	1.866	0.441	1.054	1.045	1.056	1.033
-0.5	3.585	0.381	1.031	1.040	1.044	1.015
0	6.954	0.416	1.038	1.033	1.029	1.014
0.5	4.713	0.448	1.061	1.061	1.057	1.034
0.7	4.864	0.432	1.037	1.028	1.038	0.987

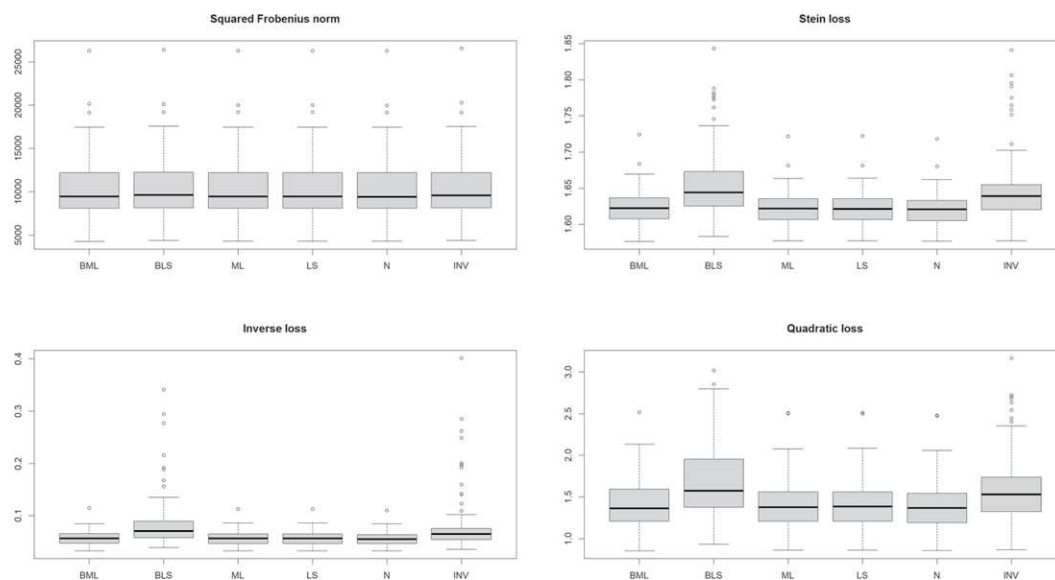


Fig. 4 The boxplots of values of squared Frobenius norm and loss functions for $\hat{\Gamma}_k$, $k \in \{BML, BLS, ML, LS, N, INV\}$, and $(p, q, \varrho) = (5, 10, -0.5)$

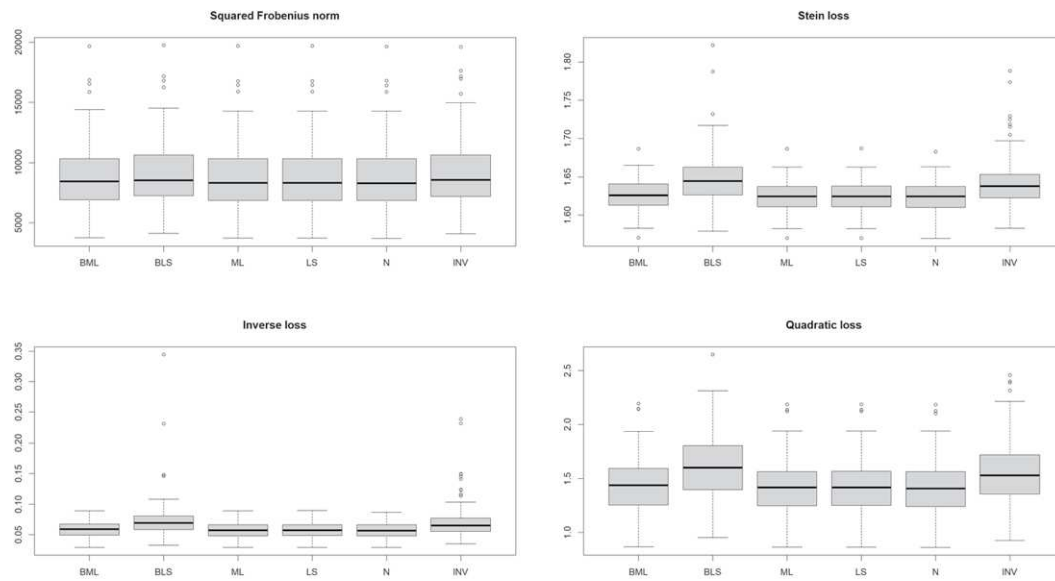


Fig. 5 The boxplots of values of squared Frobenius norm and loss functions for $\hat{\Gamma}_k$, $k \in \{BML, BLS, ML, LS, N, INV\}$, and $(p, q, \varrho) = (5, 10, 0)$

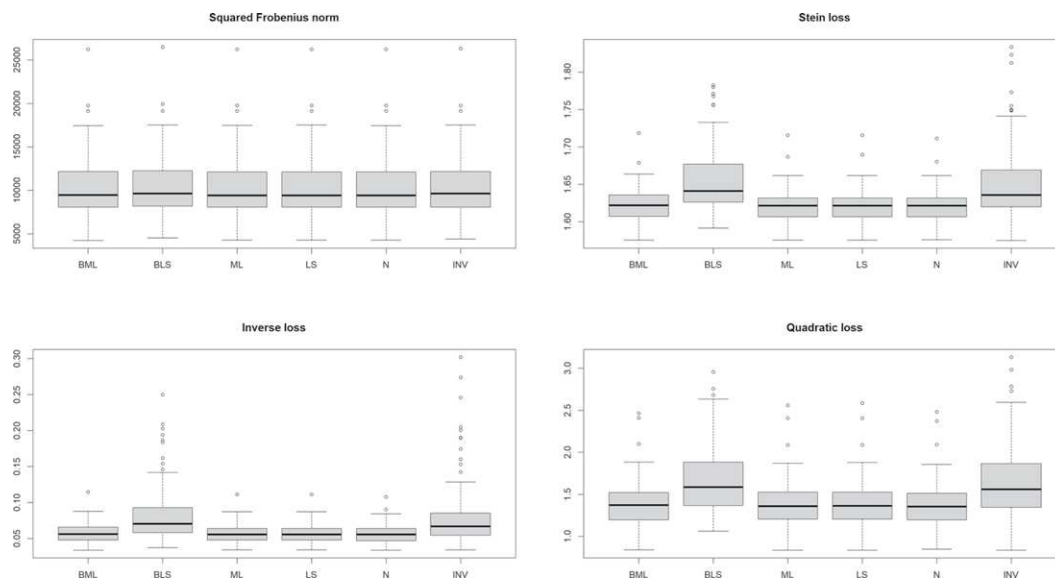


Fig. 6 The boxplots of values of squared Frobenius norm and loss functions for $\hat{\Gamma}_k$, $k \in \{BML, BLS, ML, LS, N, INV\}$, and $(p, q, \varrho) = (5, 10, 0.5)$

are very similar to each other. In contrast to smaller values of p and q the loss functions do not indicate that any estimator is the best one. On the other hand, $\hat{\Gamma}_{INV}$ and $\hat{\Gamma}_{BLS}$ produce the highest values of loss functions and are the least stable one. From the point of view of computing time (see Table 2) $\hat{\Gamma}_{BML}$ still requires the longest time, while $\hat{\Gamma}_{ML}$ and $\hat{\Gamma}_N$ are the fastest one.

Therefore, we would recommend $\hat{\Gamma}_{ML}$ and $\hat{\Gamma}_N$ to be used by practitioners. Moreover, for greater value (p, q) , the computing time for $\hat{\Gamma}_{BML}$ can be very high, even

Table 2 Average time (in seconds) needed to determine estimates for $(p, q) = (5, 10)$, $n = 100$ over 100 simulation runs

ϱ	BML	BLS	ML	LS	N	INV
-0.7	58.2951	0.0128	0.0008	0.0047	0.0011	0.0121
-0.5	60.6165	0.0223	0.0014	0.0084	0.0017	0.0681
0	86.1992	0.0017	0.0010	0.0055	0.0012	0.0378
0.5	60.2047	0.0382	0.0014	0.0078	0.0018	0.0427
0.7	57.4322	0.0017	0.0010	0.0056	0.0012	0.0269

Table 3 Examples of computing time (in seconds) needed to determine estimates for chosen (p, q) , $n = 100$ and $\varrho = 0.5$

p	q	BML	BLS	ML	LS	N	INV
30	20	2211.50	2.39	2.78	2.85	2.61	2.60
26	22	5161.74	2.38	2.77	2.99	2.81	2.55
30	25	10213.93	2.71	2.79	3.24	2.74	2.72
40	30	24087.21	3.76	3.86	5.02	3.87	3.80

more than one hour for one example, while others estimators are not as much time-consuming (see Table 3).

Observe that the values of squared Frobenius norm and loss functions do not depend on the value of ϱ , however, the case for $\varrho = 0$ is specific. Assuming normal distribution $\varrho = 0$ means that the two subvectors are independent. Therefore, we now compare the proposed estimators with the block-diagonal matrix estimator, which has explicit form:

$$\hat{\Gamma}_P := \begin{pmatrix} \mathbf{S}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22} \end{pmatrix}.$$

However, before this estimator would be applied, the test of independence of two subvectors, i.e., the test of hypothesis

$$H_0 : \Psi_{12} = \mathbf{0} \quad \text{vs} \quad H_1 : \Psi_{12} \neq \mathbf{0}$$

should be performed. In this paper, we use the likelihood ratio test of the form

$$\Lambda = \frac{|\mathbf{S}|}{|\mathbf{S}_{11}||\mathbf{S}_{22}|},$$

with the rejection region $\Lambda \leq \Lambda_\alpha$, where Λ_α is a critical value of Wilks Λ distribution with $(p, q, n - 1 - q)$ degrees of freedom and significance level α ; cf. Rencher (1995).

Using the same simulated data as previously, in Figs. 7 and 8 we see the box-plots given in Figs. 2 and 5, respectively, extended by the values of the squared Frobenius norm and all loss functions for $\hat{\Gamma}_P$. It can be seen that $\hat{\Gamma}_P$ behaves the same way as $\hat{\Gamma}_{BML}$ and $\hat{\Gamma}_N$ for $(p, q) = (3, 7)$ and $\hat{\Gamma}_{BML}$, $\hat{\Gamma}_{ML}$, $\hat{\Gamma}_{LS}$ and $\hat{\Gamma}_N$ for $(p, q) = (5, 10)$.

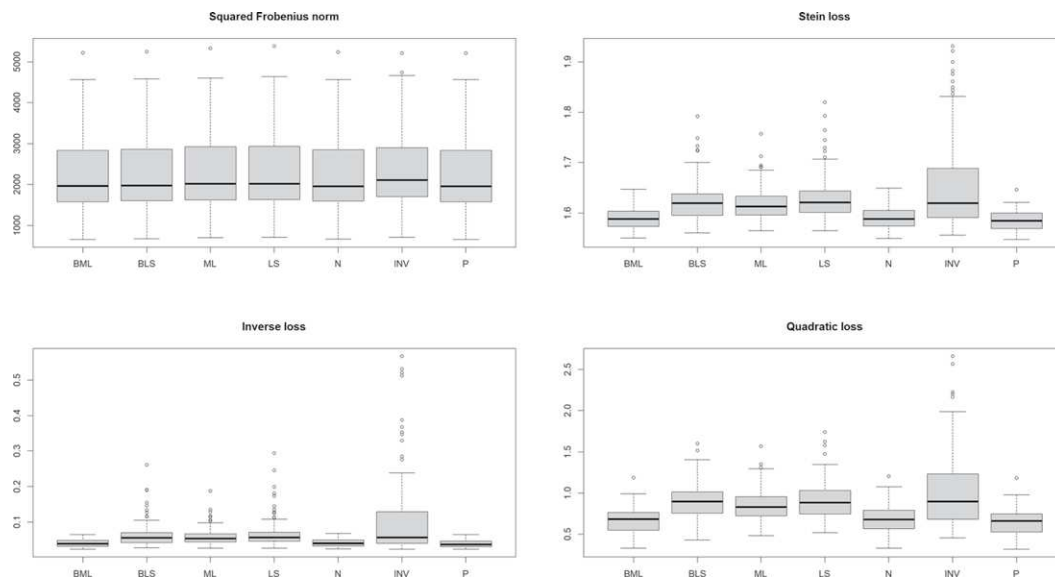


Fig. 7 The boxplots of values of squared Frobenius norm and loss functions for $\hat{\Gamma}_k$, $k \in \{BML, BLS, ML, LS, N, INV, P\}$, and $(p, q, \varrho) = (3, 7, 0)$

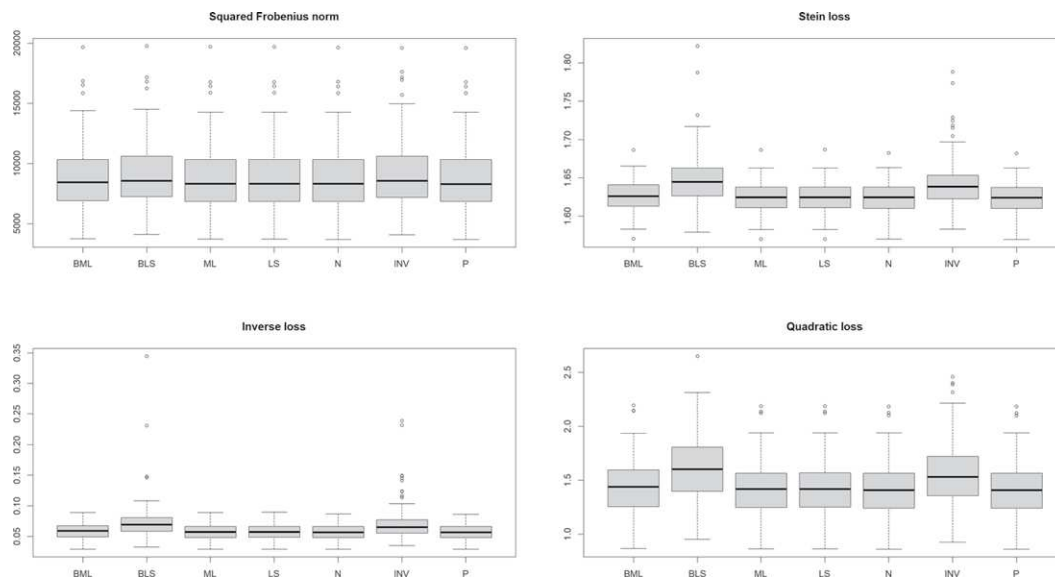


Fig. 8 The boxplots of values of squared Frobenius norm and loss functions for $\hat{\Gamma}_k$, $k \in \{BML, BLS, ML, LS, N, INV, P\}$, and $(p, q, \varrho) = (5, 10, 0)$

Table 4 Estimates of σ^2 , ϱ and Frobenius norms of the estimates of $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$, and respective condition number, κ , and computation time (in seconds) of $\hat{\mathbf{\Gamma}}_k$, $k \in \{BML, BLS, ML, LS, N, INV\}$

	BML	BLS	ML	LS	N	INV
$\hat{\sigma}^2$	23.42	12.37	26.41	25.64	26.39	39.18
$\hat{\varrho}$	0.22	0.63	0.51	0.57	0.48	0.54
$\ \hat{\mathbf{\Gamma}}_1\ _F$	66.33	66.58	66.58	66.58	66.58	66.58
$\ \hat{\mathbf{\Gamma}}_2\ _F$	93.40	94.77	94.77	94.77	94.77	94.77
$\kappa(\hat{\mathbf{\Gamma}})$	16.65	19.56	17.72	18.31	17.66	33.81
time	43.5896	0.0010	0.0040	0.0100	0.0030	0.0030

Since the independence test is available in every statistical package, if we only suspect that the subvectors may be independent, it is worth testing it and then, if H_0 is not rejected, to use the above estimation method.

6 Real data example

We refer to the experiment carried out on a high-throughput platform, where spring barley plants were grown under controlled greenhouse conditions and image-based phenotyping was conducted for 32 days, cf. Mikołajczak et al (2020). Among all characteristics taken into account in this experiment, we choose the geometry trait based on fluorescence (branchpoint count) measured on 84 plants ($n = 84$) at six time points at the beginning of the experiment ($p = 6$) and three time points at the end of the experiment ($q = 3$). Since we consider two different time intervals, we may group all the observations into two subvectors. Moreover, since the observations change over time, the autoregression structure of the covariance matrix would be a reasonable choice. Furthermore, since the observed characteristic depends on the phase of growth of the plant, one can imagine additional noise in the diagonal blocks of the covariance matrix. Therefore, we assume that the covariance matrix of the branchpoint count has the structure given in (2).

First, we verify the independence. Since $\Lambda = 0.67$ and $\Lambda_{0.05}(6, 3, 80) = 0.69$ at the significance level 0.05, the null hypothesis about the independence of subvectors is rejected. Thus, in further calculations, we compare the estimators $\hat{\mathbf{\Gamma}}_k$, $k \in \{BML, BLS, ML, LS, N, INV\}$ with respect to the estimate $\hat{\sigma}^2$, $\hat{\varrho}$, the Frobenius norm of $\hat{\mathbf{\Gamma}}_1$, $\hat{\mathbf{\Gamma}}_2$, and additionally the condition number of $\hat{\mathbf{\Gamma}}_k$ which is simply the ratio between the largest and the smallest eigenvalue. The results of the computation are given in Table 4.

As expected, estimates of diagonal blocks of $\hat{\mathbf{\Gamma}}_k$, $k \in \{ML, LS, N, INV\}$, give the same Frobenius norm; in fact, these estimates are equal, and this equality follows by the assumptions of the methods. It can be seen that for $k \in \{BML, BLS\}$ the values of the Frobenius norm of $\hat{\mathbf{\Gamma}}_1$ and $\hat{\mathbf{\Gamma}}_2$ are slightly different. Similarly, the estimates of ϱ are close to each other for $k \in \{ML, LS, N, INV\}$, while for $k \in \{BML, BLS\}$ differ significantly. When comparing $\hat{\sigma}^2$ we get similar results for $k \in \{BML, ML, LS, N\}$ and significantly different for $k \in \{BLS, INV\}$.

Finally, observe that the condition number of $\widehat{\Gamma}_{INV}$ is the highest; however, all the estimates are well-conditioned.

In summary, since all results are relatively close to each other, in view of the results of simulation studies, we would recommend using $\widehat{\Gamma}_N$ for further analysis.

7 Conclusions

In this paper, methods for estimating covariance matrices with structured cross-covariance matrices are proposed. In the considered structure, the off-diagonal block of the covariance matrix corresponds to the appropriate part of the AR(1) structure. We consider six methods: block maximum likelihood, block least squares, maximum likelihood, least squares, naive, and method based on the inverse of AR (1).

As can be observed, the best statistical properties were obtained for the block maximum likelihood and the naive method. However, block maximum likelihood is the most time-consuming approach, while the naive method is faster. Therefore, we recommend the naive method, especially for a high value of p and q .

It should also be emphasized that, if researcher has suspicion that off-diagonal blocks are zero matrices, then it is worth to test it. If the hypothesis of independence of two vectors is not rejected (cross-covariance is zero matrix), then the projection of unstructured maximum likelihood estimator onto the space of block diagonal matrices is the simplest estimator.

Acknowledgements This work was supported by the NAWA Bekker program under project no. BPN/BEK/2023/1/00142 (M. Mokrzycka) and the Poznań University of Technology under Grant no. 0213/SBAD/0119 (M. Mrowińska).

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Anderson T, Olkin I (1985) Maximum-likelihood estimation of the parameters of a multivariate normal distribution. *Linear Algebra Appl* 70:147–171
- Azzalini A (1981) Replicated observations of low order autoregressive time series. *J Time Ser Anal* 2(2):63–70
- Barndorff-Nielsen O (1978) *Information and Exponential Families in Statistical Theory*. Wiley, New York
- Bodnar O, Bodnar T, Parolya N (2022) Recent advances in shrinkage-based high-dimensional inference. *J Multivar Anal* 188:104826

- Bodnar T, Gupta AK, Parolya N (2016) Direct shrinkage estimation of large dimensional precision matrix. *J Multivar Anal* 146:223–236
- Boyd S, Vandenberghe L (2004) *Convex Optimization*. Cambridge University Press, Cambridge
- Cui X, Lin C, Zhao J et al (2016) Covariance structure regularization via frobenius norm discrepancy. *Linear Algebra Appl* 510:124–145
- Fackler P (2005) Notes on matrix calculus. Available from: <http://www4.ncsu.edu/~pfackler/MatCalc.pdf>
- Filipiak K, Klein D (2021) Estimation and testing of the covariance structure of doubly multivariate data. In: Holgersson T, Singull M (eds) *Recent Developments in Multivariate and Random Matrix Analysis*. Springer, Berlin, pp 113–130
- Filipiak K, Klein D, Roy A (2017) A comparison of likelihood ratio tests and rao's score test for three separable covariance matrix structures. *Biom J* 59(1):192–215
- Holgersson T, Pielaszkiewicz J (2020) A collection of moments of the wishart distribution. In: Holgersson T, Singull M (eds) *Recent developments in multivariate and random matrix analysis*. Springer, Berlin, pp 147–162
- Janiszewska M, Markiewicz A, Mokrzycka M (2020) Block matrix approximation via entropy loss function. *Appl Math* 65(6):829–844
- Kollo T, von Rosen D (2005) *Advanced multivariate statistics with matrices*. Springer, Dordrecht
- Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. *J Multivar Anal* 88(2):365–411
- Ledoit O, Wolf M (2021) Shrinkage estimation of large covariance matrices: keep it simple, statistician? Working Paper No. 327
- Lin L, Higham NJ, Pan J (2014) Covariance structure regularization via entropy loss function. *Comput Statist Data Anal* 72:315–327
- Magnus J, Neudecker H (1986) Symmetry, 0–1 matrices and jacobians, a review. *Econom Theory* 2:157–190
- Magnus J, Neudecker H (1999) *Matrix differential calculus with applications in statistics and econometrics*. Wiley, Hoboken
- Markiewicz A, Mokrzycka M, Mrowińska M (2024) Quasi shrinkage estimation of a block-structured covariance matrix. *J Stat Comput Simul* 94(14):1–18
- Mikołajczak K, Ogrodowicz P, Ćwiek-Kupczynska H et al (2020) Image phenotyping of spring barley (*Hordeum vulgare* L.) RIL population under drought: selection of traits and biological interpretation. *Front Plant Sci* 11:743
- Ohlson M, von Rosen D (2010) Explicit estimators of parameters in the growth curve model with linearly structured covariance matrices. *J Multivar Anal* 101:1284–1295
- Pielaszkiewicz J, von Rosen D (2021) Multivariate moments in multivariate analysis. In: Filipiak K, Markiewicz A, von Rosen D (eds) *Multivariate, multilinear and mixed linear models*. Springer, Cham, pp 41–92
- Rencher A (1995) *Methods of multivariate analysis*. John Wiley & Sons Inc, New York
- Roy A, Leiva R (2008) Likelihood ratio tests for triply multivariate data with structured correlation on spatial repeated measurements. *Statistics Probab Lett Elsevier* 78(13):1971–1980
- Žežula I (2006) Special variance structures in the growth curve model. *J Multivar Anal* 97:606–618

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[D] Characterization of quadratic subspaces of partitioned matrices

by Katarzyna Filipiak, Augustyn Markiewicz, and Malwina Mrowińska.

Submitted.

Contribution:

- Katarzyna Filipiak (25%) – conceptualization and methodological supervision, characterization of the quadratic subspaces in terms of block-spectral decomposition, characterization of the subspaces required for quadracity of the space of partitioned matrices, graph theory applications, writing – review and editing
- Augustyn Markiewicz (25%) – conceptualization and methodological supervision, characterization of a quadratic space (Proposition 2.1) and determination of algebraic properties of quadratic spaces, applications to systems of linear equations
- Malwina Mrowińska (50%) – methodology, literature review, determination and proving the form of the projection onto a quadratic space and its properties (Proposition 2.2), formulation and proof of Lemma 3.1 and Theorem 3.2, providing examples, characterization of block-spectral decomposition of arbitrary quadratic subspace, statistical applications of the results, writing – original draft preparation and editing

Characterization of quadratic subspaces of partitioned matrices

Katarzyna Filipiak^{a,*}, Augustyn Markiewicz^b, Malwina Mrowińska^a

^a*Institute of Mathematics, Poznan University of Technology, Poznań, Poland*

^b*Department of Mathematical and Statistical Methods, Poznan University of Life Sciences, Poznań, Poland*

Abstract

We characterize a matrix quadratic space, namely the Jordan algebra of real symmetric matrices, and determine the form of the projection of a symmetric matrix onto this space using block-spectral decomposition. Furthermore, we show that projecting a given positive definite matrix onto the quadratic subspace of partitioned matrices preserves definiteness and improves conditioning. In addition, we establish necessary and sufficient conditions under which a linear space of partitioned matrices, with a nonzero off-diagonal block, forms a quadratic space. We also demonstrate a procedure for block-diagonalizing partitioned matrices belonging to this space.

Finally, we present applications of the presented results in multivariate statistics and graph theory. We also point out the advantages of the presented characterization of quadratic spaces for computing matrix inverses.

Keywords:

quadratic space, Jordan algebra, partitioned matrix, block-spectral decomposition, projection, estimation, graph

MSC: 15A30, 15A18, 05C50, 62H12

*Corresponding author

Email addresses: `katarzyna.filipiak@put.poznan.pl` (Katarzyna Filipiak),
`augustyn.markiewicz@up.poznan.pl` (Augustyn Markiewicz),
`malwina.mrowinska@put.poznan.pl` (Malwina Mrowińska)

1. Introduction

In this paper, we study quadratic spaces $\mathcal{Q}$, that is, linear subspaces of the space of real symmetric matrices of order m , denoted by $\mathbb{R}_m^s$, such that $\mathbf{A} \in \mathcal{Q}$ implies $\mathbf{A}^2 \in \mathcal{Q}$; cf. [1]. Following [2, p. 305], a more appropriate term for quadratic subspace is a Jordan algebra with the product $\mathbf{A} \circ \mathbf{B} = \frac{1}{2}(\mathbf{AB} + \mathbf{BA})$, where $\mathbf{A}$ and $\mathbf{B}$ belong to the vector space of symmetric matrices over $\mathbb{R}$. Note that [2] considered a more general case of Hermitian matrices over $\mathcal{H}$, where $\mathcal{H}$ is either the set of real numbers, complex numbers or quaternions, giving the characterization of matrix Jordan algebra in his Theorem 1. Since our focus is on the space of real symmetric matrices, for simplicity, we will use the term quadratic space. In addition, we consider quadratic spaces containing identity matrix. Moreover, we express $\mathcal{Q} \subseteq \mathbb{R}_m^s$ in terms of the block-spectral decomposition (cf. [3, 4, 5]) of $\mathbf{A} \in \mathcal{Q}$, which corresponds to the characterization given in [2, Theorem 1]. Furthermore, as shown in [6], the projection of a matrix $\mathbf{X} \in \mathbb{R}_m^s$ onto the quadratic space $\mathcal{Q}$, denoted by $\Omega_{\mathcal{Q}}(\mathbf{X})$, preserves definiteness. One of the main objectives of this paper is to determine the explicit form of $\Omega_{\mathcal{Q}}(\mathbf{X})$ in terms of block-spectral decomposition. We also establish that the condition number of the projection, $\kappa(\Omega_{\mathcal{Q}}(\mathbf{X}))$, does not exceed the condition number of the original positive definite (p.d.) matrix $\mathbf{X}$.

Furthermore, we consider the space of partitioned matrices, $\mathcal{T} \subseteq \mathbb{R}_{p+q}^s$, which consists of two diagonal blocks from the linear subspaces of symmetric matrices, $\mathcal{A} \subseteq \mathbb{R}_p^s$ and $\mathcal{B} \subseteq \mathbb{R}_q^s$, while the off-diagonal block belongs to the linear subspace $\mathcal{C} \subseteq \mathbb{R}_{p \times q}$. The primary objective of this paper is to establish the conditions under which $\mathcal{T}$ forms a quadratic space. A necessary condition is clearly that $\mathcal{A}$ and $\mathcal{B}$ are quadratic spaces. Therefore, assuming that $\mathcal{A}$ and $\mathcal{B}$ are quadratic, we characterize the nontrivial subspace $\mathcal{C}$ for which $\mathcal{T}$ is also quadratic. Additionally, we show that $\mathcal{T}$ is a commutative quadratic space if and only if $\mathcal{A}$ and $\mathcal{B}$ are commutative quadratic spaces and $\mathcal{C}$ is trivial. Recall that a quadratic space $\mathcal{K}$ is said to be commutative if the commutation property holds for any two matrices, i.e., $\mathbf{A}, \mathbf{B} \in \mathcal{K} \Rightarrow \mathbf{AB} = \mathbf{BA} \in \mathcal{K}$; cf. [1]. Furthermore, we demonstrate how to block-diagonalize partitioned matrices belonging to $\mathcal{T}$ with a nontrivial subspaces $\mathcal{C}$. It is worth noting that the block-spectral decomposition can, for instance, be applied to the projection onto a quadratic space. Finally, we highlight that all presented results can be extended to partitioned matrices with more than two diagonal blocks.

The topic of this paper is motivated by statistical problems in which particular dependence structures between observations of multivariate data are of interest. Notably, the structure of a partitioned covariance matrix is particularly relevant when analyzing the relationships between two sets of variables or mitigating high-dimensionality issues. In the literature, partitioned covariance matrices belonging to quadratic subspaces have been extensively studied; see, e.g., [7, 8, 9]. It has been shown that in such cases, desirable estimators – such as unbiased or maximum likelihood estimators – can be obtained by projecting the corresponding estimator of an unstructured covariance matrix onto this space; cf. [10], [11], [12]. However, most studies on structured partitioned matrices assume equal-sized blocks. The results presented in this paper allow us to relax this assumption by considering rectangular off-diagonal blocks in the covariance matrix. It is also worth noting that a recent work by [13] proposed an estimation method for covariance matrices with non-equal-sized blocks. However, their approach assumes a specific block structure, whereas the results in this paper apply to arbitrary structures within a quadratic space. Note, finally, that recently [14] applied the structure proposed by [13] to economic data.

The identification of the most suitable covariance structure is a fundamental problem in statistics. Various approaches to addressing this issue can be found in the literature, including graphical methods ([15], [16], [17]) and algebraic techniques ([18], [19]). In graphical models, researchers typically rely on precision matrices (inverse of covariance matrices) and are often interested in sparse representation. However, a key challenge is that even if the precision matrix is sparse, the covariance matrix is not necessarily sparse unless the structure belongs to a quadratic space. Conversely, while a covariance matrix from a quadratic space does not need to be sparse, its inverse still preserves the underlying structure. Since graphical models applied to pattern recognition frequently involve graph theory, we use adjacency matrices of undirected graphs to illustrate the main results of this paper.

Finally, when discussing precision matrices within a quadratic space, we assume that the inverse exists, which motivates the assumption that the identity matrix (or, more generally, any nonsingular matrix) belongs to this space. Moreover, the block-diagonal structure of matrices in a quadratic subspace enables a more efficient computation of the inverse, as it reduces the problem to inverting matrices of smaller order. This property is particularly valuable in numerical linear algebra, where the inversion of smaller matrices improves both computational efficiency and numerical stability.

The paper is organized as follows. In Section 2 we provide a general characterization of quadratic spaces and determine the projection of a p.d. matrix onto such a space. We also show that the projection preserves positive definiteness and does not degrade the conditioning of the projected matrix. The main results of the paper are presented in Section 3, where we characterize the quadratic space of partitioned matrices and provide examples of constructing spaces of partitioned matrices consisting of four subblocks, including cases with nonzero off-diagonal blocks. Additionally, we demonstrate how a partitioned matrix from a quadratic space can be transformed into a block-diagonal form, offering a general characterization of the quadratic space structure. Finally, in Section 4, we discuss potential applications of our results in statistics and graph theory.

2. Characterization of quadratic spaces and projection

Denote a block-diagonal matrix consisting of k matrices $\mathbf{A}_i$ on the diagonal by $\text{Bdiag}(\mathbf{A}_1, \dots, \mathbf{A}_k)$, or, in a shorter version, $\text{Bdiag}_{i=1}^k(\mathbf{A}_i)$. Following Theorem 1 and a remark on p. 320 of [2], the following proposition holds.

Proposition 2.1. *For a given quadratic space $\mathcal{Q} \subseteq \mathbb{R}_m^s$, there exist integers k and $r_i, p_i, i \in \{1, \dots, k\}$, and $\sum_{i=1}^k r_i p_i = m$, such that $\mathcal{Q}$ is isomorphic to the space*

$$\mathcal{D} = \{ \mathbf{D} \in \mathbb{R}_m^s : \mathbf{D} = \text{Bdiag}_{i=1}^k(\mathbf{I}_{r_i} \otimes \mathbf{\Psi}_i), \mathbf{\Psi}_i \in \mathbb{R}_{p_i}^s \}.$$

Moreover, the multiset $(r_i, p_i), i \in \{1, \dots, k\}$, is unique up to permutation.

Obviously, since each $\mathbf{\Psi}_i$ is an arbitrary matrix from $\mathbb{R}_{p_i}^s$, in particular it can be equal to $\mathbf{I}_{p_i}$. It means that the quadratic space $\mathcal{D}$ contains identity matrix, as it was assumed in the Introduction.

Note that in the space $\mathcal{D}$, the matrices $\mathbf{\Psi}_i$ are arbitrary symmetric matrices of order p_i . Moreover, it is straightforward to verify that if all $p_i = 1$, then $\mathcal{Q}$ is commutative.

Observe that since $\mathcal{Q}$ is isomorphic to $\mathcal{D}$, there exists an orthogonal matrix of order m , say $\mathbf{P}$, such that for every matrix $\mathbf{A} \in \mathcal{Q}$ the following holds

$$\mathbf{A} = \mathbf{P} \text{Bdiag}_{i=1}^k(\mathbf{I}_{r_i} \otimes \mathbf{\Psi}_i) \mathbf{P}^\top. \quad (1)$$

Note that for a given $\mathbf{P}$, a single space $\mathcal{Q}$ is generated. If each $r_i = 1$, then different choices of $\mathbf{P}$ generate different quadratic spaces. On the other hand,

if there exists at least one $r_i \geq 2$, then two different choices of $\mathbf{P}$ can generate the same quadratic space $\mathcal{Q}$. Indeed, if $\mathbf{P}$ is partitioned into semiorthogonal subblocks, we have

$$\mathbf{P} = (\mathbf{P}_1, \dots, \mathbf{P}_k), \quad (2)$$

where $\mathbf{P}_i \in \mathbb{R}_{m \times r_i p_i}$, $(\mathbf{P}_i \mathbf{P}_i^\top)^2 = \mathbf{P}_i \mathbf{P}_i^\top$, $\mathbf{P}_i^\top \mathbf{P}_i = \mathbf{I}_{r_i p_i}$ and for $i \neq j$, $\mathbf{P}_i^\top \mathbf{P}_j = \mathbf{0}$. Then every $\mathbf{A} \in \mathcal{Q}$ can be expressed as

$$\mathbf{A} = \sum_{i=1}^k \mathbf{P}_i (\mathbf{I}_{r_i} \otimes \Psi_i) \mathbf{P}_i^\top. \quad (3)$$

For $r_i \geq 2$, the matrix $\mathbf{P}_i$ can be partitioned into r_i blocks, each consisting of p_i columns, such that each subblock is the block-eigenvector corresponding to the block-eigenvalue Ψ_i (having algebraic multiplicity r_i); cf. [3, 4, 5]. Moreover, any linear combination of these r_i subblocks is also a block-eigenvector corresponding to Ψ_i . Therefore, the linear matrix space generated by the subblocks of $\mathbf{P}_i$ is referred to as a block-eigenspace corresponding to Ψ_i , and the representation (3) is called a block-spectral decomposition of $\mathbf{A}$.

In addition, if $r_i \geq 1$ and $p_i = 1$ for every $i \in \{1, \dots, k\}$, then all Ψ_i become scalars and (3) reduces to the standard spectral decomposition. Moreover, this quadratic subspace is commutative.

If there exists i such that $p_i > 1$ then $\mathcal{Q}$ is not commutative anymore, as this space is isomorphic with the space of block-diagonal matrices, instead of diagonal matrices. As a consequence, each $\mathbf{A} \in \mathcal{Q}$ can be represented as

$$\mathbf{A} = \sum_{i=1}^{\ell} \psi_i \mathbf{P}_i \mathbf{P}_i^\top + \sum_{i=\ell+1}^k \mathbf{P}_i (\mathbf{I}_{r_i} \otimes \Psi_i) \mathbf{P}_i^\top, \quad (4)$$

where $0 \leq \ell \leq k$, and if $\ell = 0$ (or $\ell = k$) then the first (or second) component vanishes.

Each $\mathbf{P}_i (\mathbf{I}_{r_i} \otimes \Psi_i) \mathbf{P}_i^\top$, $i \in \{1, \dots, k\}$, generates a quadratic space $\mathcal{B}_i \subset \mathcal{Q}$. Thus, for $i \neq j$, it follows that

$$\dim(\mathcal{B}_i) = \frac{p_i(p_i+1)}{2} \quad \text{and} \quad \dim(\mathcal{B}_i \boxplus \mathcal{B}_j) = \frac{p_i(p_i+1)}{2} + \frac{p_j(p_j+1)}{2},$$

where $\boxplus$ denotes the orthogonal sum of two spaces.

We now establish two key properties of the projection of a p.d. matrix $\mathbf{X}$ onto the space $\mathcal{Q}$, related to its definiteness and condition number. Recall

that the condition number of a p.d. matrix $\mathbf{X}$, denoted by $\kappa(\mathbf{X})$, is given by the ratio of its largest to its smallest eigenvalue.

Proposition 2.2. *Let $\mathcal{Q}$ be a quadratic space containing the identity matrix, and let $\mathcal{D}$ be its isomorphic representation given in Proposition 2.1. The projection of a p.d. matrix $\mathbf{X}$ onto the space $\mathcal{Q}$ is given by*

$$\Omega_{\mathcal{Q}}(\mathbf{X}) = \mathbf{P} \text{Bdiag}_{i=1}^k (\mathbf{I}_{r_i} \otimes \frac{1}{r_i} \sum_{j=1}^{r_i} (\mathbf{P}_i^\top \mathbf{X} \mathbf{P}_i)_{jj}) \mathbf{P}^\top,$$

where $(\mathbf{P}^\top \mathbf{X} \mathbf{P})_{jj}$ are the diagonal blocks of $\mathbf{P}^\top \mathbf{X} \mathbf{P}$ of order corresponding to the block-diagonal representation of $\mathcal{D}$, and where $\mathbf{P}$ is partitioned as in (2). Moreover, the following properties hold:

- (a) $\Omega_{\mathcal{Q}}(\mathbf{X})$ is positive definite;
- (b) $\kappa(\Omega_{\mathcal{Q}}(\mathbf{X})) \leq \kappa(\mathbf{X})$.

Proof. Let $\mathbf{X}$ be a p.d. matrix. The projection of $\mathbf{X}$ onto the space $\mathcal{Q}$ is given by $\arg \min_{\mathbf{B} \in \mathcal{Q}} \|\mathbf{X} - \mathbf{B}\|_F^2$, where $\|\cdot\|_F$ denotes the Frobenius norm. From

(1) it follows that $\|\mathbf{X} - \mathbf{B}\|_F^2 = \|\mathbf{P}^\top (\mathbf{X} - \mathbf{B}) \mathbf{P}\|_F^2$, which implies that the projection of $\mathbf{X}$ onto $\mathcal{Q}$ is positive definite if and only if the projection of $\mathbf{P}^\top \mathbf{X} \mathbf{P}$ onto $\mathcal{D}$ is positive definite.

Define $\mathbf{Y} = \mathbf{P}^\top \mathbf{X} \mathbf{P}$.

Assume first that $\mathcal{D}$, as defined in Proposition 2.1, consists of matrices whose diagonal blocks are all distinct, i.e., $r_i = 1$, $i \in \{1, \dots, k\}$. In this case, the matrix $\mathbf{Y}$ can be partitioned as $(\mathbf{Y}_{\ell\ell'})_{\ell, \ell' \in \{1, \dots, k\}}$, where each diagonal block is of order p_i . The projection of $\mathbf{Y}$ onto $\mathcal{D}$ is given by a block-diagonal matrix whose diagonal blocks correspond to the respective diagonal blocks of $\mathbf{Y}$. A detailed proof is provided in the Appendix.

Assume now that $\mathcal{D}$ consists of matrices with identical diagonal blocks, i.e., matrices of the form $(\mathbf{I}_k \otimes \mathbf{\Psi})$, where $\mathbf{\Psi}$ is a matrix of order m/k . The matrix $\mathbf{Y}$ can then be partitioned as $(\mathbf{Y}_{\ell\ell'})_{\ell, \ell' \in \{1, \dots, k\}}$, with all blocks of order m/k . By minimizing the corresponding Frobenius norm and applying [20, Lemma 2.1], the projection of $\mathbf{Y}$ onto $\mathcal{D}$ is given by a block-diagonal matrix whose each diagonal block equals the average of the corresponding diagonal blocks of $\mathbf{Y}$. A detailed proof is provided in the Appendix.

By combining the above cases, we obtain the desired form of the projection.

We prove parts (a) and (b) separately.

- (a) Positive definiteness. Since $\mathbf{X}$ is p.d., its congruence transformation $\mathbf{Y} = \mathbf{P}^\top \mathbf{X} \mathbf{P}$ is also p.d. By the definition of the projection onto $\mathcal{Q}$, the diagonal blocks of $\mathbf{Y}$ remain p.d., which implies the positive definiteness of $\boldsymbol{\Omega}_{\mathcal{Q}}(\mathbf{X})$.
- (b) Condition number. Let $\boldsymbol{\lambda}(\cdot)$ denote the vector of eigenvalues of a given matrix, and define $\boldsymbol{\Omega} = \boldsymbol{\Omega}_{\mathcal{Q}}(\mathbf{X})$ and $\mathbf{Y} = \mathbf{P}^\top \mathbf{X} \mathbf{P}$. Using the majorization result from [21, Theorem G.1, p. 329], we obtain

$$\boldsymbol{\lambda}(\boldsymbol{\Omega}) = \boldsymbol{\lambda}(\text{Bdiag}_{i=1}^k(\mathbf{I}_{r_i} \otimes \frac{1}{r_i} \sum_{j=1}^{r_i} \mathbf{Y}_{jj})) \prec \boldsymbol{\lambda}(\text{Bdiag}_{i=1}^k(\text{Bdiag}_{j=1}^{r_i}(\mathbf{Y}_{jj}))),$$

where the symbol $\prec$ denotes majorization. Applying [21, Theorem C.1, p. 308], it follows that

$$\boldsymbol{\lambda}((\text{Bdiag}_{i=1}^k(\text{Bdiag}_{j=1}^{r_i}(\mathbf{Y}_{jj}))) \prec \boldsymbol{\lambda}(\mathbf{Y}) = \boldsymbol{\lambda}(\mathbf{X}).$$

From the definition of majorization, we immediately obtain

$$\lambda_{\max}(\boldsymbol{\Omega}) \leq \lambda_{\max}(\mathbf{X}), \quad \lambda_{\min}(\boldsymbol{\Omega}) \geq \lambda_{\min}(\mathbf{X}).$$

Thus, the condition number satisfies

$$\frac{\lambda_{\max}(\boldsymbol{\Omega})}{\lambda_{\min}(\boldsymbol{\Omega})} \leq \frac{\lambda_{\max}(\mathbf{X})}{\lambda_{\min}(\mathbf{X})},$$

which completes the proof. □

Note that the proof of (a) is a simpler version of the proof of the “if” part of Theorem 1 in [6].

3. Characterization of quadratic spaces of partitioned matrices

Consider the following linear spaces

$$\begin{aligned} \mathcal{T}_1 &= \{\mathbf{T}_1 \in \mathbb{R}_{p+q}^s : \mathbf{T}_1 = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \mathbf{A} \in \mathcal{A} \subseteq \mathbb{R}_p^s\}, \\ \mathcal{T}_2 &= \{\mathbf{T}_2 \in \mathbb{R}_{p+q}^s : \mathbf{T}_2 = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix} \in \mathbb{R}_{p+q}^s, \mathbf{B} \in \mathcal{B} \subseteq \mathbb{R}_q^s\}, \\ \mathcal{T}_3 &= \{\mathbf{T}_3 \in \mathbb{R}_{p+q}^s : \mathbf{T}_3 = \begin{pmatrix} \mathbf{0} & \mathbf{C} \\ \mathbf{C}^\top & \mathbf{0} \end{pmatrix} \in \mathbb{R}_{p+q}^s, \mathbf{C} \in \mathcal{C} \subseteq \mathbb{R}_{p \times q}\}, \end{aligned} \tag{5}$$

where $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are linear subspaces of $\mathbb{R}_p^s$, $\mathbb{R}_q^s$, and $\mathbb{R}_{p \times q}$, respectively. From the orthogonality of $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_3$ we can define a linear space

$$\mathcal{T} = \mathcal{T}_1 \boxplus \mathcal{T}_2 \boxplus \mathcal{T}_3, \quad (6)$$

for which

$$\dim \mathcal{T} = \dim(\mathcal{T}_1) + \dim(\mathcal{T}_2) + \dim(\mathcal{T}_3).$$

Notably, $\mathcal{T}$ contains identity matrix.

If $\mathcal{A} = \mathbb{R}_p^s$, $\mathcal{B} = \mathbb{R}_q^s$, and $\mathcal{C} = \mathbb{R}_{p \times q}$, then $\mathcal{T}$ spans the linear space of all symmetric matrices, say $\mathcal{S}$. If any of $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ is a proper linear subspace of $\mathbb{R}_p^s$, $\mathbb{R}_q^s$, and $\mathbb{R}_{p \times q}$, respectively, then $\mathcal{T}$ is a proper linear subspace of $\mathcal{S}$.

The aim of this section is to characterize $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ such that $\mathcal{T}$ is a quadratic space. The necessary condition for $\mathcal{T}$ to be a quadratic space is established in the following lemma.

Lemma 3.1. *If $\mathcal{T}$ is a quadratic space, then $\mathcal{A}$ and $\mathcal{B}$ are quadratic subspaces. Moreover, if $\mathcal{T}$ is a commutative quadratic space, then $\mathcal{A}$ and $\mathcal{B}$ are also commutative quadratic subspaces.*

Proof. Assume that $\mathcal{T}$ is a quadratic space. Since the matrix $\mathbf{C}$ belongs to a linear space, one possible choice is $\mathbf{C} = \mathbf{0}$. Then, $\mathcal{T}_1 \boxplus \mathcal{T}_2$ forms a quadratic space of block-diagonal matrices, implying that if $\mathbf{T} \in \mathcal{T}_1 \boxplus \mathcal{T}_2$ then $\mathbf{T}^2 \in \mathcal{T}_1 \boxplus \mathcal{T}_2$ as well. The block-diagonal structure of $\mathbf{T}^2$ ensures that $\mathcal{T}_1$ and $\mathcal{T}_2$ are quadratic subspaces.

Similarly, if $\mathcal{T}$ is a commutative quadratic space, the product of any two matrices in $\mathcal{T}_1 \boxplus \mathcal{T}_2$ remains symmetric. Again, from the block-diagonal form of such a product, it follows that $\mathcal{T}_1$ and $\mathcal{T}_2$ are commutative quadratic subspaces. $\square$

For $\mathcal{A}$ and $\mathcal{B}$ being quadratic subspaces, we now formulate the characterization of $\mathcal{C}$ based on the quadraticity of $\mathcal{T}$. Consider the spectral representation of $\mathbf{A} \in \mathcal{A}$ and $\mathbf{B} \in \mathcal{B}$ given in (4) with, respectively, $\mathbf{V} = (\mathbf{V}_1, \dots, \mathbf{V}_k)$ and $\mathbf{W} = (\mathbf{W}_1, \dots, \mathbf{W}_{k'})$ as in (2). Matrices $\mathbf{V}_i$ and $\mathbf{W}_{i'}$ are the block-eigenvectors of any matrix from, respectively, $\mathcal{A}$ and $\mathcal{B}$, corresponding to the block-eigenvalues of a given $\mathbf{A} \in \mathcal{A}$ and $\mathbf{B} \in \mathcal{B}$. Denoting the eigenvalues and block-eigenvalues of $\mathcal{A}$ (respectively $\mathcal{B}$) by $\lambda_i \in \mathbb{R}$ and $\Lambda_i \in \mathbb{R}^{p_i \times p_i}$

(respectively $\delta_{i'} \in \mathbb{R}$, $\Delta_{i'} \in \mathbb{R}_{p_{i'} \times p_{i'}}$), we have

$$\begin{aligned} \mathbf{A} &= \sum_{i=1}^{\ell} \lambda_i \mathbf{V}_i \mathbf{V}_i^{\top} + \sum_{i=\ell+1}^k \mathbf{V}_i (\mathbf{I}_{r_i} \otimes \Lambda_i) \mathbf{V}_i^{\top}, \\ \mathbf{B} &= \sum_{i'=1}^{\ell'} \delta_{i'} \mathbf{W}_{i'} \mathbf{W}_{i'}^{\top} + \sum_{i'=\ell'+1}^{k'} \mathbf{W}_{i'} (\mathbf{I}_{r_{i'}} \otimes \Delta_{i'}) \mathbf{W}_{i'}^{\top}. \end{aligned} \quad (7)$$

Due to (3) we have $r(\mathbf{V}_i) = r_i p_i$, $r(\mathbf{W}_{i'}) = r_{i'} p_{i'}$, with $\sum_{i=1}^k r_i p_i = p$, $\sum_{i'=1}^{k'} r_{i'} p_{i'} = q$. Recall that if $\ell = k$ and $\ell' = k'$, then $\mathbf{A}$ and $\mathbf{B}$ belong to a commutative quadratic space. Furthermore, if $i \in \{1, \dots, \ell\}$ then each p_i equals 1, and similarly, if $i' \in \{1, \dots, \ell'\}$, then each $p_{i'}$ equals 1. This means that the indices $i \in \{1, \dots, \ell\}$, $i' \in \{1, \dots, \ell'\}$ correspond to the block-spectral decomposition of the commutative part of $\mathbf{A}$ and $\mathbf{B}$, while $i \in \{\ell+1, \dots, k\}$, $i' \in \{\ell'+1, \dots, k'\}$ correspond to the non-commutative parts of $\mathbf{A}$ and $\mathbf{B}$, respectively. Based on this observation, we define the following index sets:

$$\begin{aligned} \mathcal{J}_{11} &= \{(i, i') : i \in \{1, \dots, \ell\}, i' \in \{1, \dots, \ell'\}\}, \\ \mathcal{J}_{21} &= \{(i, i') : i \in \{\ell+1, \dots, k\}, i' \in \{1, \dots, \ell'\}\}, \\ \mathcal{J}_{12} &= \{(i, i') : i \in \{1, \dots, \ell\}, i' \in \{\ell'+1, \dots, k'\}\}, \\ \mathcal{J}_{22} &= \{(i, i') : i \in \{\ell+1, \dots, k\}, i' \in \{\ell'+1, \dots, k'\}\}. \end{aligned}$$

For $v, w \in \{1, 2\}$, we define $\mathcal{J}_{vw}^*$ to be a subset of $\mathcal{J}_{vw}$ containing only those elements in which no index i or i' appear more than once. Note, that $\mathcal{J}_{vw}^*$ can be a proper subspace of $\mathcal{J}_{vw}$, as well as $\emptyset$. In our consideration we assume that at least one of $\mathcal{J}_{uv}^*$ is nonempty set; otherwise we have a trivial case with $\mathbf{T}_3 = \mathbf{0}$.

Let $\mathcal{T} = \mathcal{T}_1 \boxplus \mathcal{T}_2 \boxplus \mathcal{T}_3$, as defined in (5), be a quadratic space, and let $\mathbf{T} \in \mathcal{T}$. Then,

$$\mathbf{T}^2 = \begin{pmatrix} \mathbf{A}^2 + \mathbf{C}\mathbf{C}^{\top} & \mathbf{A}\mathbf{C} + \mathbf{C}\mathbf{B} \\ \mathbf{C}^{\top}\mathbf{A} + \mathbf{B}\mathbf{C}^{\top} & \mathbf{C}^{\top}\mathbf{C} + \mathbf{B}^2 \end{pmatrix} \in \mathcal{T}.$$

By Lemma 3.1 it follows that $\mathbf{A}^2 \in \mathcal{A}$ and $\mathbf{B}^2 \in \mathcal{B}$. Moreover,

$$\mathbf{C}\mathbf{C}^{\top} \in \mathcal{A}, \quad \mathbf{C}^{\top}\mathbf{C} \in \mathcal{B}. \quad (8)$$

This requires that $r_i = r_{i'}$ and imposes a specific structure on $\mathbf{C} \in \mathcal{C}$ given by $\mathbf{C} = \sum_{(i,i') \in \mathcal{J}^*} \mathbf{V}_i(\mathbf{\Gamma}_{ii'} \otimes \mathbf{G}_{ii'})\mathbf{W}_{i'}^\top$, where $\mathcal{J}^* \subseteq \mathcal{J}_{11}^* \cup \mathcal{J}_{21}^* \cup \mathcal{J}_{12}^* \cup \mathcal{J}_{22}^*$.

Here $\mathbf{\Gamma}_{ii'}$ is an orthogonal matrix of order r_i , and $\mathbf{G}_{ii'} \in \mathbb{R}_{p_i \times p_{i'}}$ is a scalar if $(i, i') \in \mathcal{J}_{11}^*$, a column vector if $(i, i') \in \mathcal{J}_{21}^*$, a row vector if $(i, i') \in \mathcal{J}_{12}^*$, and an arbitrary matrix of appropriate size if $(i, i') \in \mathcal{J}_{22}^*$.

In other words, $\mathcal{C}$ is a subspace of $\mathcal{C}_{11} \boxplus \mathcal{C}_{21} \boxplus \mathcal{C}_{12} \boxplus \mathcal{C}_{22}$, where

$$\mathcal{C}_{vw} = \left\{ \sum_{(i,i') \in \mathcal{J}_{vw}^*} \mathbf{V}_i(\mathbf{\Gamma}_{ii'} \otimes \mathbf{G}_{ii'})\mathbf{W}_{i'}^\top : \mathbf{\Gamma}_{ii'}\mathbf{\Gamma}_{ii'}^\top = \mathbf{\Gamma}_{ii'}^\top\mathbf{\Gamma}_{ii'} = \mathbf{I}_{r_i}, \mathbf{G}_{ii'} \in \mathbb{R}_{p_i \times p_{i'}} \right\},$$

$v, w \in \{1, 2\}$. Note that the spaces $\mathcal{C}_{vw}$ are mutually orthogonal.

We are now ready to prove the main result of this section.

Theorem 3.2. *The linear space $\mathcal{T} = \mathcal{T}_1 \boxplus \mathcal{T}_2 \boxplus \mathcal{T}_3$, as defined in (5), is a quadratic subspace if and only if $\mathcal{A}$ and $\mathcal{B}$ are quadratic subspaces and $\mathcal{C}$ is a linear subspace of $\mathcal{C}_{11} \boxplus \mathcal{C}_{21} \boxplus \mathcal{C}_{12} \boxplus \mathcal{C}_{22}$. Moreover, $\mathcal{T}$ is commutative if and only if $\mathcal{T}_1$ and $\mathcal{T}_2$ are commutative, and $\dim(\mathcal{T}_3) = 0$.*

Proof. Sufficiency.

Let $\mathbf{T} \in \mathcal{T} = \mathcal{T}_1 \boxplus \mathcal{T}_2 \boxplus \mathcal{T}_3$, where the $\mathcal{T}_i$ are defined as in (5), $\mathcal{A}$ and $\mathcal{B}$ are quadratic spaces, and $\mathcal{C}$ is a linear subspace of $\mathcal{C}_{11} \boxplus \mathcal{C}_{21} \boxplus \mathcal{C}_{12} \boxplus \mathcal{C}_{22}$. It suffices to show that for $\mathbf{C} \in \mathcal{C}$, the conditions (8) and

$$\mathbf{AC} + \mathbf{CB} \in \mathcal{C} \tag{9}$$

hold.

Assume that $\mathbf{C} \in \mathcal{C}_{11}$. Since $\mathbf{G}_{ii'} = g_{ii'}$, we have

$$\begin{aligned} \mathbf{CC}^\top &= \sum_{(i,i') \in \mathcal{J}_{11}^*} g_{ii'}^2 \mathbf{V}_i \mathbf{V}_i^\top \in \mathcal{A}, \\ \mathbf{C}^\top \mathbf{C} &= \sum_{(i,i') \in \mathcal{J}_{11}^*} g_{ii'}^2 \mathbf{W}_{i'} \mathbf{W}_{i'}^\top \in \mathcal{B}, \\ \mathbf{AC} + \mathbf{CB} &= \sum_{i=1}^l \lambda_i \mathbf{V}_i \mathbf{V}_i^\top \cdot \sum_{(i,i') \in \mathcal{J}_{11}^*} \mathbf{V}_i(\mathbf{\Gamma}_{ii'} \otimes g_{ii'})\mathbf{W}_{i'}^\top \\ &\quad + \sum_{(i,i') \in \mathcal{J}_{11}^*} \mathbf{V}_i(\mathbf{\Gamma}_{ii'} \otimes g_{ii'})\mathbf{W}_{i'}^\top \cdot \sum_{i'=1}^{l'} \delta_{i'} \mathbf{W}_{i'} \mathbf{W}_{i'}^\top \\ &= \sum_{(i,i') \in \mathcal{J}_{11}^*} \mathbf{V}_i(\mathbf{\Gamma}_{ii'} \otimes (\lambda_i + \delta_{i'})g_{ii'})\mathbf{W}_{i'}^\top \in \mathcal{C}_{11}. \end{aligned}$$

Similarly, for $\mathbf{C} \in \mathcal{C}_{21}$, since $\mathbf{G}_{ii'} = \mathbf{g}_{ii'}$, we have

$$\begin{aligned}
\mathbf{C}\mathbf{C}^\top &= \sum_{(i,i') \in J_{21}^*} \mathbf{V}_i (\mathbf{I}_{r_i} \otimes \mathbf{g}_{ii'} \mathbf{g}_{ii'}^\top) \mathbf{V}_i^\top = \sum_{(i,i') \in J_{21}^*} \mathbf{V}_i (\mathbf{I}_{r_i} \otimes \tilde{\Lambda}_i) \mathbf{V}_i^\top \in \mathcal{A}, \\
\mathbf{C}^\top \mathbf{C} &= \sum_{(i,i') \in J_{21}^*} \mathbf{W}_{i'} (\mathbf{I}_{r_i} \otimes \mathbf{g}_{ii'}^\top \mathbf{g}_{ii'}) \mathbf{W}_{i'}^\top = \sum_{(i,i') \in J_{21}^*} \tilde{\delta}_{i'} \mathbf{W}_{i'} \mathbf{W}_{i'}^\top \in \mathcal{B}, \\
\mathbf{A}\mathbf{C} + \mathbf{C}\mathbf{B} &= \sum_{i=\ell+1}^k \mathbf{V}_i (\mathbf{I}_{r_i} \otimes \Lambda_i) \mathbf{V}_i^\top \cdot \sum_{(i,i') \in J_{21}^*} \mathbf{V}_i (\Gamma_{ii'} \otimes \mathbf{g}_{ii'}) \mathbf{W}_{i'}^\top \\
&\quad + \sum_{(i,i') \in J_{21}^*} \mathbf{V}_i (\Gamma_{ii'} \otimes \mathbf{g}_{ii'}) \mathbf{W}_{i'}^\top \cdot \sum_{i'=1}^{\ell'} \delta_{i'} \mathbf{W}_{i'} \mathbf{W}_{i'}^\top \\
&= \sum_{(i,i') \in J_{21}^*} \mathbf{V}_i (\Gamma_{ii'} \otimes \Lambda_i \mathbf{g}_{ii'}) \mathbf{W}_{i'}^\top + \sum_{(i,i') \in J_{21}^*} \mathbf{V}_i (\Gamma_{ii'} \otimes \delta_{i'} \mathbf{g}_{ii'}) \mathbf{W}_{i'}^\top \\
&= \sum_{(i,i') \in J_{21}^*} \mathbf{V}_i (\Gamma_{ii'} \otimes (\tilde{\mathbf{g}}_{ii'} + \tilde{\tilde{\mathbf{g}}}_{ii'})) \mathbf{W}_{i'}^\top \in \mathcal{C}_{21},
\end{aligned}$$

while for $\mathbf{C} \in \mathcal{C}_{12}$, since $\mathbf{G}_{ii'} = \mathbf{g}_{ii'}^\top$, we obtain

$$\begin{aligned}
\mathbf{C}\mathbf{C}^\top &= \sum_{(i,i') \in J_{12}^*} \mathbf{V}_{i'} (\mathbf{I}_{r_{i'}} \otimes \mathbf{g}_{ii'}^\top \mathbf{g}_{ii'}) \mathbf{V}_{i'}^\top = \sum_{(i,i') \in J_{12}^*} \tilde{\lambda}_i \mathbf{V}_i \mathbf{V}_i^\top \in \mathcal{A}, \\
\mathbf{C}^\top \mathbf{C} &= \sum_{(i,i') \in J_{12}^*} \mathbf{W}_{i'} (\mathbf{I}_{r_{i'}} \otimes \mathbf{g}_{ii'} \mathbf{g}_{ii'}^\top) \mathbf{W}_{i'}^\top \\
&= \sum_{(i,i') \in J_{12}^*} \mathbf{W}_{i'} (\mathbf{I}_{r_{i'}} \otimes \tilde{\Delta}_{i'}) \mathbf{W}_{i'}^\top \in \mathcal{B}, \\
\mathbf{A}\mathbf{C} + \mathbf{C}\mathbf{B} &= \sum_{i=1}^l \lambda_i \mathbf{V}_i \mathbf{V}_i^\top \cdot \sum_{(i,i') \in J_{12}^*} \mathbf{V}_i (\Gamma_{ii'} \otimes \mathbf{g}_{ii'}^\top) \mathbf{W}_{i'}^\top \\
&\quad + \sum_{(i,i') \in J_{12}^*} \mathbf{V}_i (\Gamma_{ii'} \otimes \mathbf{g}_{ii'}^\top) \mathbf{W}_{i'}^\top \cdot \sum_{i'=\ell'+1}^{k'} \mathbf{W}_{i'} (\mathbf{I}_{r_{i'}} \otimes \Delta_{i'}) \mathbf{W}_{i'}^\top \\
&= \sum_{(i,i') \in J_{12}^*} \mathbf{V}_i (\Gamma_{ii'} \otimes \lambda_i \mathbf{g}_{ii'}^\top) \mathbf{W}_{i'}^\top + \sum_{(i,i') \in J_{12}^*} \mathbf{V}_i (\Gamma_{ii'} \otimes \mathbf{g}_{ii'}^\top \Delta_{i'}) \mathbf{W}_{i'}^\top \\
&= \sum_{(i,i') \in J_{12}^*} \mathbf{V}_i (\Gamma_{ii'} \otimes (\tilde{\mathbf{g}}_{ii'}^\top + \tilde{\tilde{\mathbf{g}}}_{ii'}^\top)) \mathbf{W}_{i'}^\top \in \mathcal{C}_{12}.
\end{aligned}$$

Finally, for $\mathbf{C} \in \mathcal{C}_{22}$ we have

$$\begin{aligned}
\mathbf{C}\mathbf{C}^\top &= \sum_{(i,i') \in J_{22}^*} \mathbf{V}_i (\mathbf{I}_{r_i} \otimes \mathbf{G}_{ii'} \mathbf{G}_{ii'}^\top) \mathbf{V}_i^\top \\
&= \sum_{(i,i') \in J_{22}^*} \mathbf{V}_i (\mathbf{I}_{r_i} \otimes \tilde{\Lambda}_i) \mathbf{V}_i^\top \in \mathcal{A}, \\
\mathbf{C}^\top \mathbf{C} &= \sum_{(i,i') \in J_{22}^*} \mathbf{W}_{i'} (\mathbf{I}_{r_{i'}} \otimes \mathbf{G}_{ii'}^\top \mathbf{G}_{ii'}) \mathbf{W}_{i'}^\top \\
&= \sum_{(i,i') \in J_{22}^*} \mathbf{W}_{i'} (\mathbf{I}_{r_i} \otimes \tilde{\Delta}_{i'}) \mathbf{W}_{i'}^\top \in \mathcal{B}, \\
\mathbf{A}\mathbf{C} + \mathbf{C}\mathbf{B} &= \sum_{i=\ell+1}^k \mathbf{V}_i (\mathbf{I}_{r_i} \otimes \Lambda_i) \mathbf{V}_i^\top \cdot \sum_{(i,i') \in J_{22}^*} \mathbf{V}_i (\Gamma_{ii'} \otimes \mathbf{G}_{ii'}) \mathbf{W}_{i'}^\top \\
&\quad + \sum_{(i,i') \in J_{22}^*} \mathbf{V}_i (\Gamma_{ii'} \otimes \mathbf{G}_{ii'}) \mathbf{W}_{i'}^\top \cdot \sum_{i'=\ell'+1}^{k'} \mathbf{W}_{i'} (\mathbf{I}_{r_{i'}} \otimes \Delta_{i'}) \mathbf{W}_{i'}^\top \\
&= \sum_{(i,i') \in J_{22}^*} \mathbf{V}_i (\Gamma_{ii'} \otimes (\tilde{\mathbf{G}}_{ii'} + \tilde{\tilde{\mathbf{G}}}_{ii'})) \mathbf{W}_{i'}^\top \in \mathcal{C}_{22}.
\end{aligned}$$

Since, for each $\mathbf{C} \in \mathcal{C}_{vw}$, $v, w \in \{1, 2\}$, we have shown that $\mathbf{C}\mathbf{C}^\top \in \mathcal{A}$ and $\mathbf{C}^\top \mathbf{C} \in \mathcal{B}$ and $\mathbf{A}\mathbf{C} + \mathbf{C}\mathbf{B} \in \mathcal{C}_{vw}$, it follows that conditions (8) and (9) also hold for any combination of matrices from $\mathcal{C}_{vw}$, $v, w \in \{1, 2\}$. Thus, $\mathcal{T}$ is a quadratic space.

The necessity follows directly from the definition of $\mathcal{C}_{vw}$, $v, w \in \{1, 2\}$, given prior to the theorem.

Commutativity.

Assume now that $\mathcal{A}$ and $\mathcal{B}$ are commutative quadratic spaces. If $\mathcal{C} = \{\mathbf{0}\}$, the result follows trivially.

Suppose that $\mathcal{T}$, satisfying the assumptions of the theorem, is a commutative quadratic space. By Lemma 3.1 it follows that $\mathcal{A}$ and $\mathcal{B}$ are commutative quadratic subspaces. It remains to show that for any $\mathbf{T}_1, \mathbf{T}_2 \in \mathcal{T}$, the product $\mathbf{T}_1 \mathbf{T}_2$ is symmetric. Clearly,

$$\mathbf{T}_1 \mathbf{T}_2 = \begin{pmatrix} \mathbf{A}_1 \mathbf{A}_2 + \mathbf{C}_1 \mathbf{C}_2^\top & \mathbf{A}_1 \mathbf{C}_2 + \mathbf{C}_1 \mathbf{B}_2 \\ \mathbf{C}_1^\top \mathbf{A}_2 + \mathbf{B}_1 \mathbf{C}_2^\top & \mathbf{C}_1^\top \mathbf{C}_2 + \mathbf{B}_1 \mathbf{B}_2 \end{pmatrix}.$$

Assume that $\mathcal{A} \neq \{\mathbf{0}\}$ and $\mathcal{B} \neq \{\mathbf{0}\}$, and that $\mathbf{C} \in \mathcal{C}$, $\mathbf{C} \neq \mathbf{0}$. A necessary condition for the symmetry of $\mathbf{T}_1 \mathbf{T}_2$ is $\mathbf{A}_1 \mathbf{C}_2 + \mathbf{C}_1 \mathbf{B}_2 = (\mathbf{C}_1^\top \mathbf{A}_2 + \mathbf{B}_1 \mathbf{C}_2^\top)^\top$

for arbitrary $\mathbf{A}_1, \mathbf{A}_2 \in \mathcal{A}$, $\mathbf{B}_1, \mathbf{B}_2 \in \mathcal{B}$, $\mathbf{C}_1, \mathbf{C}_2 \in \mathcal{C}$. Since all spaces are linear and contain an identity matrix of the appropriate order, we set $\mathbf{A}_1 = \mathbf{I}_p$, $\mathbf{B}_1 = \mathbf{I}_q$, $\mathbf{C}_1 = \mathbf{C}$, and $\mathbf{A}_2 = \alpha \mathbf{I}_p$, $\mathbf{B}_2 = \beta \mathbf{I}_q$ and $\mathbf{C}_2 = \gamma \mathbf{C}$ with $\alpha, \beta, \gamma \in \mathbb{R}$. Substituting these into the symmetry condition yields

$$\gamma \mathbf{C} + \beta \mathbf{C} = \alpha \mathbf{C} + \gamma \mathbf{C}.$$

This equality holds for arbitrary α and β only if $\mathbf{C} = \mathbf{0}$. $\square$

Example 3.1

Let $p = 3$ and $q = 4$, and let $\mathcal{A} = \{a_0 \mathbf{I}_3 + a \mathbf{1}_3 \mathbf{1}_3^\top : a_0, a \in \mathbb{R}\}$, $\mathcal{B} = \{b_0 \mathbf{I}_4 + b_1(\mathbf{H} + \mathbf{H}^\top) + b_2 \mathbf{H}^2 : b_0, b_1, b_2 \in \mathbb{R}\}$, where $\mathbf{1}_3$ is a 3-dimensional vector of ones, $\mathbf{H}$ is a 4×4 zero-one matrix with ones on the subdiagonal and in the upper-right corner. Both $\mathcal{A}$ and $\mathcal{B}$ are quadratic subspaces. Using spectral decompositions of any $\mathbf{A} \in \mathcal{A}$ and $\mathbf{B} \in \mathcal{B}$, we obtain

$$\mathbf{A} = \lambda_1 \mathbf{V}_1 \mathbf{V}_1^\top + \lambda_2 \mathbf{V}_2 \mathbf{V}_2^\top, \quad \mathbf{B} = \delta_1 \mathbf{W}_1 \mathbf{W}_1^\top + \delta_2 \mathbf{W}_2 \mathbf{W}_2^\top + \delta_3 \mathbf{W}_3 \mathbf{W}_3^\top,$$

where $\mathbf{V}_1 = \frac{1}{\sqrt{3}} \mathbf{1}_3$, $\mathbf{V}_2$ consists of two orthonormal eigenvectors that are orthogonal to $\mathbf{1}_3$, and $\mathbf{W}_1 = \frac{1}{2} \mathbf{1}_4$, $\mathbf{W}_2 = (\mathbf{w}_2 : \mathbf{w}_3)$, $\mathbf{W}_3 = \mathbf{w}_4$, with $\mathbf{w}_i$, $i \in \{2, 3, 4\}$, being orthonormal eigenvectors corresponding to eigenvalue δ_2 (multiplicity 2), and δ_3 (multiplicity 1), respectively. All of them are orthogonal to $\mathbf{1}_4$; for the exact form of the eigenvectors, see e.g. [22] and [23]. Since the quadratic subspaces $\mathcal{A}$ and $\mathcal{B}$ are commutative, in representation (7) matrices $\mathbf{A}$ and $\mathbf{B}$ have only scalar eigenvalues. Therefore, when choosing $\mathbf{C}$, we consider only matrices belonging to $\mathcal{C}_{11}$.

The index set J_{11} related to the spectral decomposition of $\mathbf{A}$ and $\mathbf{B}$ is given by $\{(1, 1), (1, 3), (2, 2)\}$. A subset J_{11}^* may take one of the following forms: $\{(1, 1)\}$, $\{(1, 3)\}$, $\{(2, 2)\}$, $\{(1, 1), (2, 2)\}$ and $\{(1, 3), (2, 2)\}$. The set $\{(1, 1), (1, 3)\}$ is not allowed, because each index can appear at most once. Thus, the possible choices of $\mathcal{C}_{11}$ are:

$$\begin{aligned} \mathcal{C}_{11}^{(I)} &= \{\gamma_1 \mathbf{1}_3 \mathbf{1}_4^\top : \gamma_1 \in \mathbb{R}\}, \\ \mathcal{C}_{11}^{(II)} &= \{\gamma_2 \mathbf{1}_3 \mathbf{w}_4^\top : \gamma_2 \in \mathbb{R}\}, \\ \mathcal{C}_{11}^{(III)} &= \{\gamma_3 \mathbf{V}_2 \mathbf{\Gamma} \mathbf{W}_2^\top : \gamma_3 \in \mathbb{R}, \quad \mathbf{\Gamma} \mathbf{\Gamma}^\top = \mathbf{\Gamma}^\top \mathbf{\Gamma} = \mathbf{I}_2\}, \end{aligned}$$

or

$$\begin{aligned} \mathcal{C}_{11}^{(IV)} &= \mathcal{C}_{11}^{(I)} \boxplus \mathcal{C}_{11}^{(III)}, \\ \mathcal{C}_{11}^{(V)} &= \mathcal{C}_{11}^{(II)} \boxplus \mathcal{C}_{11}^{(III)}. \end{aligned}$$

As can be noticed, for $\mathbf{C} \in \mathcal{C}_{11}^{(I)}$ or $\mathbf{C} \in \mathcal{C}_{11}^{(II)}$, the matrix $\mathbf{\Gamma}$ reduces to a scalar (± 1), while for $\mathbf{C} \in \mathcal{C}_{11}^{(III)}$, matrix $\mathbf{\Gamma}$ is an arbitrarily 2×2 orthogonal matrix.

Summing up, partitioned matrices belonging to the quadratic space with nonzero off-diagonal blocks take the following forms:

$$\left(\begin{array}{ccc|cccc} a_0 & a & a & \gamma_1 & \gamma_1 & \gamma_1 & \gamma_1 \\ a & a_0 & a & \gamma_1 & \gamma_1 & \gamma_1 & \gamma_1 \\ a & a & a_0 & \gamma_1 & \gamma_1 & \gamma_1 & \gamma_1 \\ \hline \gamma_1 & \gamma_1 & \gamma_1 & b_0 & b_1 & b_2 & b_1 \\ \gamma_1 & \gamma_1 & \gamma_1 & b_1 & b_0 & b_1 & b_2 \\ \gamma_1 & \gamma_1 & \gamma_1 & b_2 & b_1 & b_0 & b_1 \\ \gamma_1 & \gamma_1 & \gamma_1 & b_1 & b_2 & b_1 & b_0 \end{array} \right) \quad \text{for } \mathbf{C} \in \mathcal{C}_{11}^{(I)},$$

$$\left(\begin{array}{ccc|cccc} a_0 & a & a & \gamma_2^* & -\gamma_2^* & \gamma_2^* & -\gamma_2^* \\ a & a_0 & a & \gamma_2^* & -\gamma_2^* & \gamma_2^* & -\gamma_2^* \\ a & a & a_0 & \gamma_2^* & -\gamma_2^* & \gamma_2^* & -\gamma_2^* \\ \hline \gamma_2^* & \gamma_2^* & \gamma_2^* & b_0 & b_1 & b_2 & b_1 \\ -\gamma_2^* & -\gamma_2^* & -\gamma_2^* & b_1 & b_0 & b_1 & b_2 \\ \gamma_2^* & \gamma_2^* & \gamma_2^* & b_2 & b_1 & b_0 & b_1 \\ -\gamma_2^* & -\gamma_2^* & -\gamma_2^* & b_1 & b_2 & b_1 & b_0 \end{array} \right) \quad \begin{array}{l} \text{for } \mathbf{C} \in \mathcal{C}_{11}^{(II)}, \quad \gamma_2^* = \frac{1}{2}\gamma_2, \\ \text{and } \mathbf{w}_4 = \frac{1}{2}(1, -1, 1, -1)^\top, \end{array}$$

$$\left(\begin{array}{ccc|cccc} a_0 & a & a & \sqrt{3}\gamma_3^* & \gamma_3^* & -\sqrt{3}\gamma_3^* & -\gamma_3^* \\ a & a_0 & a & 0 & -2\gamma_3^* & 0 & 2\gamma_3^* \\ a & a & a_0 & -\sqrt{3}\gamma_3^* & \gamma_3^* & \sqrt{3}\gamma_3^* & -\gamma_3^* \\ \hline \sqrt{3}\gamma_3^* & 0 & -\sqrt{3}\gamma_3^* & b_0 & b_1 & b_2 & b_1 \\ \gamma_3^* & -2\gamma_3^* & \gamma_3^* & b_1 & b_0 & b_1 & b_2 \\ -\sqrt{3}\gamma_3^* & 0 & \sqrt{3}\gamma_3^* & b_2 & b_1 & b_0 & b_1 \\ - & \gamma_3^* & 2\gamma_3^* & b_1 & b_2 & b_1 & b_0 \end{array} \right)$$

for $\mathbf{C} \in \mathcal{C}_{11}^{(III)}$, $\mathbf{\Gamma} = \mathbf{I}_2$, $\gamma_3^* = \frac{1}{2\sqrt{3}}\gamma_3$, and

$$\mathbf{V}_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} \sqrt{3} & 1 \\ 0 & -2 \\ -\sqrt{3} & 1 \end{pmatrix}, \quad \mathbf{W}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The resulting matrices can also be combined in cases $\mathbf{C} \in \mathcal{C}_{11}^{(IV)}$ and $\mathbf{C} \in \mathcal{C}_{11}^{(V)}$. $\square$

Recall that each matrix in a quadratic subspace is isomorphic to a block-diagonal matrix. However, since the off-diagonal block $\mathbf{C}$ of $\mathbf{T} \in \mathcal{T}$ is nonzero, the isomorphic block-diagonal matrix is not merely a combination of the block-eigenvalues of $\mathbf{A}$ and $\mathbf{B}$; the structure of $\mathbf{C}$ plays a crucial role. Before formulating a general procedure for transforming $\mathbf{T}$ into a block-diagonal form with block-eigenvalues on the diagonal, we analyze specific examples to illustrate the method.

Example 3.1 - cont.

Let $\mathbf{C} = \mathcal{C}_{11}^{(1)}$. Then

$$\begin{aligned} \mathbf{T} &= \begin{pmatrix} \lambda_1 \mathbf{V}_1 \mathbf{V}_1^\top & \gamma_1 \mathbf{V}_1 \mathbf{W}_1^\top \\ \gamma_1 \mathbf{W}_1 \mathbf{V}_1^\top & \delta_1 \mathbf{W}_1 \mathbf{W}_1^\top \end{pmatrix} + \begin{pmatrix} \lambda_2 \mathbf{V}_2 \mathbf{V}_2^\top & \mathbf{0}_{3 \times 4} \\ \mathbf{0}_{4 \times 3} & \mathbf{0}_{4 \times 4} \end{pmatrix} \\ &+ \begin{pmatrix} \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 4} \\ \mathbf{0}_{4 \times 3} & \delta_2 \mathbf{W}_2 \mathbf{W}_2^\top \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 4} \\ \mathbf{0}_{4 \times 3} & \delta_3 \mathbf{W}_3 \mathbf{W}_3^\top \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{V}_1 & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{4 \times 1} & \mathbf{W}_1 \end{pmatrix} \begin{pmatrix} \lambda_1 & \gamma_1 \\ \gamma_1 & \delta_1 \end{pmatrix} \begin{pmatrix} \mathbf{V}_1^\top & \mathbf{0}_{1 \times 4} \\ \mathbf{0}_{1 \times 3} & \mathbf{W}_1^\top \end{pmatrix} \\ &+ \lambda_2 \begin{pmatrix} \mathbf{V}_2 \\ \mathbf{0}_{4 \times 2} \end{pmatrix} (\mathbf{V}_2^\top, \mathbf{0}_{2 \times 4}) \\ &+ \delta_2 \begin{pmatrix} \mathbf{0}_{3 \times 2} \\ \mathbf{W}_2 \end{pmatrix} (\mathbf{0}_{2 \times 3}, \mathbf{W}_2^\top) + \delta_3 \begin{pmatrix} \mathbf{0}_{3 \times 1} \\ \mathbf{W}_3 \end{pmatrix} (\mathbf{0}_{1 \times 3}, \mathbf{W}_3^\top) \\ &= \mathbf{P}_I \cdot \text{Bdiag} \left(\begin{pmatrix} \lambda_1 & \gamma_1 \\ \gamma_1 & \delta_1 \end{pmatrix}, \lambda_2 \mathbf{I}_2, \delta_2 \mathbf{I}_2, \delta_3 \right) \cdot \mathbf{P}_I^\top \end{aligned}$$

with

$$\mathbf{P}_I = \left(\begin{pmatrix} \mathbf{V}_1 & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{4 \times 1} & \mathbf{W}_1 \end{pmatrix}, \begin{pmatrix} \mathbf{V}_2 \\ \mathbf{0}_{4 \times 2} \end{pmatrix}, \begin{pmatrix} \mathbf{0}_{3 \times 2} \\ \mathbf{W}_2 \end{pmatrix}, \begin{pmatrix} \mathbf{0}_{3 \times 1} \\ \mathbf{W}_3 \end{pmatrix} \right).$$

Similarly, for $\mathbf{C} = \mathcal{C}_{II}^{(2)}$, we have

$$\mathbf{T} = \mathbf{P}_{II} \cdot \text{Bdiag} \left(\begin{pmatrix} \lambda_1 & \gamma_2 \\ \gamma_2 & \delta_3 \end{pmatrix}, \lambda_2 \mathbf{I}_2, \delta_1, \delta_2 \mathbf{I}_2 \right) \cdot \mathbf{P}_{II}^\top$$

where

$$\mathbf{P}_{II} = \left(\begin{pmatrix} \mathbf{V}_1 & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{4 \times 1} & \mathbf{W}_3 \end{pmatrix}, \begin{pmatrix} \mathbf{V}_2 \\ \mathbf{0}_{4 \times 2} \end{pmatrix}, \begin{pmatrix} \mathbf{0}_{3 \times 1} \\ \mathbf{W}_1 \end{pmatrix}, \begin{pmatrix} \mathbf{0}_{3 \times 2} \\ \mathbf{W}_2 \end{pmatrix} \right).$$

Assume now $\mathbf{C} = \mathcal{C}_{11}^{(\text{III})}$. Then

$$\begin{aligned} \mathbf{T} = & \lambda_1 \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{0}_{4 \times 1} \end{pmatrix} (\mathbf{V}_1^\top, \mathbf{0}_{1 \times 4}) \\ & + \begin{pmatrix} \mathbf{V}_2 & \mathbf{0}_{3 \times 2} \\ \mathbf{0}_{4 \times 2} & \mathbf{W}_2 \end{pmatrix} \begin{pmatrix} \lambda_2 \mathbf{I}_2 & \gamma_3 \mathbf{\Gamma} \\ \gamma_3 \mathbf{\Gamma}^\top & \delta_2 \mathbf{I}_2 \end{pmatrix} \begin{pmatrix} \mathbf{V}_2^\top & \mathbf{0}_{2 \times 4} \\ \mathbf{0}_{2 \times 3} & \mathbf{W}_2^\top \end{pmatrix} \\ & + \delta_1 \begin{pmatrix} \mathbf{0}_{3 \times 1} \\ \mathbf{W}_1 \end{pmatrix} (\mathbf{0}_{1 \times 3}, \mathbf{W}_1^\top) + \delta_3 \begin{pmatrix} \mathbf{0}_{3 \times 1} \\ \mathbf{W}_3 \end{pmatrix} (\mathbf{0}_{1 \times 3}, \mathbf{W}_3^\top). \end{aligned}$$

Since $\mathbf{\Gamma}$ is orthogonal, the second component of the above sum can also be expressed as

$$\begin{aligned} & \begin{pmatrix} \mathbf{V}_2 \mathbf{\Gamma} & \mathbf{0}_{3 \times 2} \\ \mathbf{0}_{4 \times 2} & \mathbf{W}_2 \end{pmatrix} \begin{pmatrix} \lambda_2 \mathbf{I}_2 & \gamma_3 \mathbf{I}_2 \\ \gamma_3 \mathbf{I}_2 & \delta_2 \mathbf{I}_2 \end{pmatrix} \begin{pmatrix} \mathbf{\Gamma}^\top \mathbf{V}_2^\top & \mathbf{0}_{2 \times 4} \\ \mathbf{0}_{2 \times 3} & \mathbf{W}_2^\top \end{pmatrix} = \\ & \begin{pmatrix} \mathbf{V}_2 \mathbf{\Gamma} & \mathbf{0}_{3 \times 2} \\ \mathbf{0}_{4 \times 2} & \mathbf{W}_2 \end{pmatrix} \left(\begin{pmatrix} \lambda_2 & \gamma_3 \\ \gamma_3 & \delta_2 \end{pmatrix} \otimes \mathbf{I}_2 \right) \begin{pmatrix} \mathbf{\Gamma}^\top \mathbf{V}_2^\top & \mathbf{0}_{2 \times 4} \\ \mathbf{0}_{2 \times 3} & \mathbf{W}_2^\top \end{pmatrix}. \end{aligned}$$

Following [24], we use the commutation matrix $\mathbf{K}_{2,2}$ to switch the order of the Kronecker product:

$$\begin{pmatrix} \lambda_2 & \gamma_3 \\ \gamma_3 & \delta_2 \end{pmatrix} \otimes \mathbf{I}_2 = \mathbf{K}_{2,2} (\mathbf{I}_2 \otimes \begin{pmatrix} \lambda_2 & \gamma_3 \\ \gamma_3 & \delta_2 \end{pmatrix}) \mathbf{K}_{2,2}.$$

Summing up, for $\mathbf{C} \in \mathcal{C}_{11}^{(\text{III})}$ we obtain

$$\mathbf{T} = \mathbf{P}_{\text{III}} \cdot \text{Bdiag}(\lambda_1, \mathbf{I}_2 \otimes \begin{pmatrix} \lambda_2 & \gamma_3 \\ \gamma_3 & \delta_2 \end{pmatrix}, \delta_1, \delta_3) \cdot \mathbf{P}_{\text{III}}^\top$$

where

$$\mathbf{P}_{\text{III}} = \left(\begin{pmatrix} \mathbf{V}_1 \\ \mathbf{0}_{4 \times 1} \end{pmatrix}, \begin{pmatrix} \mathbf{V}_2 \mathbf{\Gamma} & \mathbf{0}_{3 \times 2} \\ \mathbf{0}_{4 \times 2} & \mathbf{W}_2 \end{pmatrix} \mathbf{K}_{2,2}, \begin{pmatrix} \mathbf{0}_{3 \times 1} \\ \mathbf{W}_1 \end{pmatrix}, \begin{pmatrix} \mathbf{0}_{3 \times 1} \\ \mathbf{W}_3 \end{pmatrix} \right).$$

Combining $\mathbf{P}_{\text{I}}$ with $\mathbf{P}_{\text{III}}$ for $\mathbf{C} = \mathcal{C}_{11}^{(\text{IV})}$ and $\mathbf{P}_{\text{II}}$ with $\mathbf{P}_{\text{III}}$ for $\mathbf{C} = \mathcal{C}_{11}^{(\text{V})}$ we obtain

$$\mathbf{T} = \mathbf{P}_{\text{IV}} \cdot \text{Bdiag}\left(\begin{pmatrix} \lambda_1 & \gamma_1 \\ \gamma_1 & \delta_1 \end{pmatrix}, \mathbf{I}_2 \otimes \begin{pmatrix} \lambda_2 & \gamma_3 \\ \gamma_3 & \delta_2 \end{pmatrix}, \delta_3\right) \cdot \mathbf{P}_{\text{IV}}^\top$$

where

$$\mathbf{P}_{\text{IV}} = \left(\begin{pmatrix} \mathbf{V}_1 & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{4 \times 1} & \mathbf{W}_1 \end{pmatrix}, \begin{pmatrix} \mathbf{V}_2 \mathbf{\Gamma} & \mathbf{0}_{3 \times 2} \\ \mathbf{0}_{4 \times 2} & \mathbf{W}_2 \end{pmatrix} \mathbf{K}_{2,2}, \begin{pmatrix} \mathbf{0}_{3 \times 1} \\ \mathbf{W}_3 \end{pmatrix} \right)$$

and

$$\mathbf{T} = \mathbf{P}_V \cdot \text{Bdiag}\left(\begin{pmatrix} \lambda_1 & \gamma_2 \\ \gamma_2 & \delta_3 \end{pmatrix}, \mathbf{I}_2 \otimes \begin{pmatrix} \lambda_2 & \gamma_3 \\ \gamma_3 & \delta_2 \end{pmatrix}, \delta_1\right) \cdot \mathbf{P}_V^\top$$

where

$$\mathbf{P}_V = \left(\begin{pmatrix} \mathbf{V}_1 & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{4 \times 1} & \mathbf{W}_3 \end{pmatrix}, \begin{pmatrix} \mathbf{V}_2 \mathbf{\Gamma} & \mathbf{0}_{3 \times 2} \\ \mathbf{0}_{4 \times 2} & \mathbf{W}_2 \end{pmatrix} \mathbf{K}_{2,2}, \begin{pmatrix} \mathbf{0}_{3 \times 1} \\ \mathbf{W}_1 \end{pmatrix} \right).$$

Finally, if $\mathcal{C} = \{\mathbf{0}\}$, then $\mathcal{T}$ is a commutative quadratic subspace, as $\mathbf{A} \in \mathcal{A}$ and $\mathbf{B} \in \mathcal{B}$ have only scalar eigenvalues, and hence $\mathcal{T}$ is isomorphic to

$$\{\text{Bdiag}(\lambda_1, \lambda_2 \mathbf{I}_2, \delta_1, \delta_2 \mathbf{I}_2, \delta_3) : \lambda_1, \lambda_2, \delta_1, \delta_2, \delta_3 \in \mathbb{R}\}.$$

However, introducing a nonzero off-diagonal block $\mathbf{C} \in \mathcal{C}_{11}^{(h)}$, $h \in \{\text{I, II, III, IV, V}\}$, results in a matrix that no longer belongs to the above space; in each case, the resulting space is isomorphic to a space of block-diagonal matrices with at least one 2×2 block-eigenvalue. $\square$

Let us now consider another example, in which one of the diagonal blocks of $\mathbf{T}$ does not belong to a commutative quadratic subspace.

Example 3.2. Let $\mathcal{A}$ be the same as in Example 3.1 and let $\mathcal{B} = \{\mathbf{I}_2 \otimes \mathbf{\Delta} : \mathbf{\Delta} \in \mathbb{R}_2^s\}$. Then, $\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2)$ is the same as in Example 3.1, and $\mathbf{W} = \mathbf{W}_1 = \mathbf{I}_4$. Then, the only possible choice of $\mathcal{C}$ is $\mathcal{C}_{12} = \{\mathbf{V}_2(\mathbf{\Gamma} \otimes \mathbf{g}^\top) : \mathbf{\Gamma} \mathbf{\Gamma}^\top = \mathbf{\Gamma}^\top \mathbf{\Gamma} = \mathbf{I}_2, \mathbf{g} \in \mathbb{R}_{2 \times 1}\}$. Thus, we can write

$$\mathbf{T} = \lambda_1 \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{0}_{4 \times 1} \end{pmatrix} (\mathbf{V}_1^\top, \mathbf{0}_{1 \times 4}) + \begin{pmatrix} \lambda_2 \mathbf{V}_2 \mathbf{V}_2^\top & \mathbf{V}_2(\mathbf{\Gamma} \otimes \mathbf{g}^\top) \\ (\mathbf{\Gamma}^\top \otimes \mathbf{g}) \mathbf{V}_2^\top & \mathbf{I}_2 \otimes \mathbf{\Delta} \end{pmatrix}.$$

Observe that the second component in the sum can also be expressed as

$$\begin{pmatrix} \mathbf{V}_2 \mathbf{\Gamma} & \mathbf{0}_{3 \times 4} \\ \mathbf{0}_{4 \times 2} & \mathbf{I}_4 \end{pmatrix} \underbrace{\begin{pmatrix} \mathbf{I}_2 \otimes \lambda_2 & \mathbf{I}_2 \otimes \mathbf{g}^\top \\ \mathbf{I}_2 \otimes \mathbf{g} & \mathbf{I}_2 \otimes \mathbf{\Delta} \end{pmatrix}}_{\mathbf{U}} \begin{pmatrix} \mathbf{\Gamma}^\top \mathbf{V}_2^\top & \mathbf{0}_{2 \times 4} \\ \mathbf{0}_{4 \times 3} & \mathbf{I}_4 \end{pmatrix}.$$

By permuting the Kronecker products in $\mathbf{U}$ twice (see [24, formula (24)]), we can rewrite

$$\begin{aligned} \mathbf{U} &= \begin{pmatrix} \mathbf{K}_{2,1}(\lambda_2 \otimes \mathbf{I}_2) \mathbf{K}_{1,2} & \mathbf{K}_{2,1}(\mathbf{g}^\top \otimes \mathbf{I}_2) \mathbf{K}_{2,2} \\ \mathbf{K}_{2,2}(\mathbf{g} \otimes \mathbf{I}_2) \mathbf{K}_{1,2} & \mathbf{K}_{2,2}(\mathbf{\Delta} \otimes \mathbf{I}_2) \mathbf{K}_{2,2} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{K}_{1,2} & \mathbf{0}_{2 \times 4} \\ \mathbf{0}_{4 \times 2} & \mathbf{K}_{2,2} \end{pmatrix} \left(\begin{pmatrix} \lambda_2 & \mathbf{g}^\top \\ \mathbf{g} & \mathbf{\Delta} \end{pmatrix} \otimes \mathbf{I}_2 \right) \begin{pmatrix} \mathbf{K}_{2,1} & \mathbf{0}_{2 \times 4} \\ \mathbf{0}_{4 \times 2} & \mathbf{K}_{2,2} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{I}_2 & \mathbf{0}_{2 \times 4} \\ \mathbf{0}_{4 \times 2} & \mathbf{K}_{2,2} \end{pmatrix} \mathbf{K}_{3,2} (\mathbf{I}_2 \otimes \begin{pmatrix} \lambda_2 & \mathbf{g}^\top \\ \mathbf{g} & \mathbf{\Delta} \end{pmatrix}) \mathbf{K}_{2,3} \begin{pmatrix} \mathbf{I}_2 & \mathbf{0}_{2 \times 4} \\ \mathbf{0}_{4 \times 2} & \mathbf{K}_{2,2} \end{pmatrix}. \end{aligned}$$

Since $\mathbf{K}_{1,2} = \mathbf{K}_{2,1} = \mathbf{I}_2$, we obtain

$$\mathbf{T} = \mathbf{P} \text{Bdiag}(\lambda_1, (\mathbf{I}_2 \otimes \begin{pmatrix} \lambda_2 & \mathbf{g}^\top \\ \mathbf{g} & \mathbf{\Delta} \end{pmatrix})) \mathbf{P}^\top$$

where

$$\mathbf{P} = \left(\begin{pmatrix} \mathbf{V}_1 \\ \mathbf{0}_{4 \times 1} \end{pmatrix}, \begin{pmatrix} \mathbf{V}_2 \mathbf{\Gamma} & \mathbf{0}_{2 \times 4} \\ \mathbf{0}_{4 \times 2} & \mathbf{K}_{2,2} \end{pmatrix} \mathbf{K}_{3,2} \right).$$

□

In the next example, we consider a linear space $\mathcal{T}$ of block-diagonal matrices, such that neither $\mathcal{A}$ nor $\mathcal{B}$ is a commutative quadratic subspace.

Example 3.3

- (a) Let $\mathcal{A} = \mathbb{R}_3^s$ and $\mathcal{B} = \mathbb{R}_2^s$. Then $\mathbf{V} = \mathbf{V}_1 = \mathbf{I}_3$ and $\mathbf{W} = \mathbf{W}_1 = \mathbf{I}_2$, and the only possible choice for $\mathcal{C}$ is $\mathcal{C}_{22} \subseteq \mathbb{R}_{3 \times 2}$ (note that $\mathbf{\Gamma}$ equals ± 1). For $\mathbf{\Lambda} \in \mathcal{A}$, $\mathbf{\Delta} \in \mathcal{B}$ and $\mathbf{G} \in \mathcal{C}_{22} = \mathbb{R}_{3 \times 2}$, we obtain

$$\mathbf{T} = \begin{pmatrix} \mathbf{\Lambda} & \mathbf{G} \\ \mathbf{G}^\top & \mathbf{\Delta} \end{pmatrix},$$

and the only choice for $\mathbf{P}$ in the isomorphic representation (1) is $\mathbf{I}_5$. Note that $\mathbf{T} \in \mathcal{T} = \mathbb{R}_5^s$.

- (b) Let $\mathcal{A} = \{\mathbf{I}_r \otimes \mathbf{\Lambda} : \mathbf{\Lambda} \in \mathbb{R}_3^s\}$ and $\mathcal{B} = \{\mathbf{I}_r \otimes \mathbf{\Delta} : \mathbf{\Delta} \in \mathbb{R}_2^s\}$. Then, $\mathbf{V} = \mathbf{V}_1 = \mathbf{I}_{3r}$ and $\mathbf{W} = \mathbf{W}_1 = \mathbf{I}_{2r}$, and the only possible choice for $\mathcal{C}$ is $\mathcal{C}_{22} = \{\mathbf{\Gamma} \otimes \mathbf{G} \in \mathbb{R}_{3r \times 2r} : \mathbf{G} \in \mathbb{R}_{2 \times 2}, \mathbf{\Gamma} \mathbf{\Gamma}^\top = \mathbf{\Gamma}^\top \mathbf{\Gamma} = \mathbf{I}_r\}$. Therefore, if $\mathbf{G}$ is an arbitrary matrix from $\mathbb{R}_{3 \times 2}$ (rather than from a subspace of $\mathbb{R}_{3 \times 2}$), then

$$\mathbf{T} = \begin{pmatrix} \mathbf{\Gamma} \otimes \mathbf{I}_3 & \mathbf{0}_{3r \times 2r} \\ \mathbf{0}_{2r \times 3r} & \mathbf{I}_{2r} \end{pmatrix} \begin{pmatrix} \mathbf{I}_r \otimes \mathbf{\Lambda} & \mathbf{I}_r \otimes \mathbf{G} \\ \mathbf{I}_r \otimes \mathbf{G}^\top & \mathbf{I}_r \otimes \mathbf{\Delta} \end{pmatrix} \begin{pmatrix} \mathbf{\Gamma}^\top \otimes \mathbf{I}_3 & \mathbf{0}_{3r \times 2r} \\ \mathbf{0}_{2r \times 3r} & \mathbf{I}_{2r} \end{pmatrix}.$$

Using commutation matrices similarly to Example 3.2, we then obtain

$$\mathbf{T} = \mathbf{P} \text{Bdiag}(\mathbf{I}_r \otimes \begin{pmatrix} \mathbf{\Lambda} & \mathbf{G} \\ \mathbf{G}^\top & \mathbf{\Delta} \end{pmatrix}) \mathbf{P}^\top$$

$$\text{with } \mathbf{P} = \begin{pmatrix} (\mathbf{\Gamma} \otimes \mathbf{I}_3) \mathbf{K}_{r,3} & \mathbf{0}_{3r \times 2r} \\ \mathbf{0}_{2r \times 3r} & \mathbf{K}_{2,r} \end{pmatrix} \mathbf{K}_{5,r}.$$

□

To determine a general formula for the block-eigenvalues representation (2.1), we first partition the spaces $\mathcal{A}$ and $\mathcal{B}$ from representation (5) into subspaces corresponding to the commutative and non-commutative components of $\mathbf{A}$ and $\mathbf{B}$, as defined in (7). Specifically, we have

$$\begin{aligned} \sum_{i=1}^{\ell} \lambda_i \mathbf{V}_i \mathbf{V}_i^{\top} &\in \mathcal{A}_1, & \sum_{i=\ell+1}^k \mathbf{V}_i (\mathbf{I}_{r_i} \otimes \mathbf{\Lambda}_i) \mathbf{V}_i^{\top} &\in \mathcal{A}_2, \\ \sum_{i'=1}^{\ell'} \delta_{i'} \mathbf{W}_{i'} \mathbf{W}_{i'}^{\top} &\in \mathcal{B}_1, & \sum_{i'=\ell'+1}^{k'} \mathbf{W}_{i'} (\mathbf{I}_{r_{i'}} \otimes \mathbf{\Delta}_{i'}) \mathbf{W}_{i'}^{\top} &\in \mathcal{B}_2. \end{aligned}$$

Considering each subspace $\mathcal{C}_{vw}$, $v, w \in \{1, 2\}$, from Theorem 3.2 separately, we can correspondingly partition the matrix $\mathbf{T}$ as

$$\mathbf{T} = \sum_{v, w \in \{1, 2\}} \begin{pmatrix} \mathbf{A}_v & \mathbf{C}_{vw} \\ \mathbf{C}_{vw}^{\top} & \mathbf{B}_w \end{pmatrix},$$

where $\mathbf{A}_v \in \mathcal{A}_v$, $\mathbf{B}_w \in \mathcal{B}_w$, and $\mathbf{C}_{vw} \in \mathcal{C}_{vw}$, $v, w \in \{1, 2\}$. Denoting the components of $\mathbf{T}$ by $\mathbf{T}^{(vw)}$, we express them as

$$\begin{aligned} \mathbf{T}^{(11)} &= \sum_{\substack{i=1 \\ (i, i') \notin \mathcal{J}_{11}^*}}^{\ell} \begin{pmatrix} \mathbf{V}_i \\ \mathbf{0}_{q \times r_i} \end{pmatrix} \text{Bdiag}(0, \dots, 0, \lambda_i \mathbf{I}_{r_i}, 0, \dots, 0) (\mathbf{V}_i^{\top}, \mathbf{0}_{r_i \times q}) \\ &+ \sum_{\substack{i'=1 \\ (i, i') \notin \mathcal{J}_{11}^*}}^{\ell'} \begin{pmatrix} \mathbf{0}_{p \times r_{i'}} \\ \mathbf{W}_{i'} \end{pmatrix} \text{Bdiag}(0, \dots, 0, \delta_{i'} \mathbf{I}_{r_{i'}}, 0, \dots, 0) (\mathbf{0}_{r_{i'} \times p}, \mathbf{W}_{i'}^{\top}) \\ &+ \sum_{(i, i') \in \mathcal{J}_{11}^*} \begin{pmatrix} \mathbf{V}_i \mathbf{\Gamma}_{ii'} & \mathbf{0}_{p \times r_i} \\ \mathbf{0}_{q \times r_i} & \mathbf{W}_{i'} \end{pmatrix} \mathbf{K}_{2, r_i} \\ &\quad \times \text{Bdiag}(0, \dots, 0, \mathbf{I}_{r_i} \otimes \begin{pmatrix} \lambda_i & \gamma_{ii'} \\ \gamma_{ii'} & \delta_{i'} \end{pmatrix}, 0, \dots, 0) \\ &\quad \times \mathbf{K}_{r_i, 2} \begin{pmatrix} \mathbf{\Gamma}_{ii'}^{\top} \mathbf{V}_i^{\top} & \mathbf{0}_{r_i \times q} \\ \mathbf{0}_{r_i \times p} & \mathbf{W}_{i'}^{\top} \end{pmatrix}, \end{aligned}$$

$$\begin{aligned}
\mathbf{T}^{(21)} = & \sum_{\substack{i = \ell + 1 \\ (i, i') \notin \mathcal{J}_{21}^*}}^k \begin{pmatrix} \mathbf{V}_i \\ \mathbf{0}_{q \times r_i p_{i'}} \end{pmatrix} \text{Bdiag}(0, \dots, 0, \mathbf{I}_{r_i} \otimes \mathbf{\Lambda}_i, 0, \dots, 0) (\mathbf{V}_i^\top, \mathbf{0}_{r_i p_{i'} \times q}) \\
& + \sum_{\substack{i' = 1 \\ (i, i') \notin \mathcal{J}_{21}^*}}^{\ell'} \begin{pmatrix} \mathbf{0}_{p \times r_{i'}} \\ \mathbf{W}_{i'} \end{pmatrix} \text{Bdiag}(0, \dots, 0, \delta_{i'} \mathbf{I}_{r_{i'}}, 0, \dots, 0) (\mathbf{0}_{r_{i'} \times p}, \mathbf{W}_{i'}^\top) \\
& + \sum_{(i, i') \in \mathcal{J}_{21}^*} \begin{pmatrix} \mathbf{V}_i \mathbf{K}_{p_i, r_i} & \mathbf{0}_{r_i p_i \times r_i} \\ \mathbf{0}_{q \times r_i p_i} & \mathbf{W}_{i'} \mathbf{\Gamma}_{ii'}^\top \end{pmatrix} \mathbf{K}_{1+p_i, r_i} \\
& \quad \times \text{Bdiag}(0, \dots, 0, \mathbf{I}_{r_i} \otimes \begin{pmatrix} \mathbf{\Lambda}_i & \mathbf{g}_{ii'}^\top \\ \mathbf{g}_{ii'}^\top & \delta_{i'} \end{pmatrix}, 0, \dots, 0) \\
& \quad \times \mathbf{K}_{r_i, 1+p_i} \begin{pmatrix} \mathbf{K}_{r_i, p_i} \mathbf{V}_i^\top & \mathbf{0}_{r_i p_i \times q} \\ \mathbf{0}_{r_i \times r_i p_i} & \mathbf{\Gamma}_{ii'} \mathbf{W}_{i'}^\top \end{pmatrix},
\end{aligned}$$

$$\begin{aligned}
\mathbf{T}^{(12)} = & \sum_{\substack{i = 1 \\ (i, i') \notin \mathcal{J}_{12}^*}}^{\ell} \begin{pmatrix} \mathbf{V}_i \\ \mathbf{0}_{q \times r_i} \end{pmatrix} \text{Bdiag}(0, \dots, 0, \lambda_i \mathbf{I}_{r_i}, 0, \dots, 0) (\mathbf{V}_i^\top, \mathbf{0}_{r_i \times q}) \\
& + \sum_{\substack{i' = \ell' + 1 \\ (i, i') \notin \mathcal{J}_{12}^*}}^{k'} \begin{pmatrix} \mathbf{0}_{p \times r_{i'} q_{i'}} \\ \mathbf{W}_{i'} \end{pmatrix} \text{Bdiag}(0, \dots, 0, \mathbf{I}_{r_{i'}} \otimes \mathbf{\Delta}_{i'}, 0, \dots, 0) (\mathbf{0}_{r_{i'} q_{i'} \times p}, \mathbf{W}_{i'}^\top) \\
& + \sum_{(i, i') \in \mathcal{J}_{12}^*} \begin{pmatrix} \mathbf{V}_i \mathbf{\Gamma}_{ii'} & \mathbf{0}_{p \times r_i q_{i'}} \\ \mathbf{0}_{r_i q_{i'} \times r_i} & \mathbf{W}_{i'} \mathbf{K}_{q_{i'}, r_i} \end{pmatrix} \mathbf{K}_{1+q_{i'}, r_i} \\
& \quad \times \text{Bdiag}(0, \dots, 0, \mathbf{I}_{r_i} \otimes \begin{pmatrix} \lambda_i & \mathbf{g}_{ii'}^\top \\ \mathbf{g}_{ii'}^\top & \mathbf{\Delta}_{i'} \end{pmatrix}, 0, \dots, 0) \\
& \quad \times \mathbf{K}_{r_i, 1+q_{i'}} \begin{pmatrix} \mathbf{\Gamma}_{ii'}^\top \mathbf{V}_i^\top & \mathbf{0}_{r_i \times r_i q_{i'}} \\ \mathbf{0}_{r_i q_{i'} \times p} & \mathbf{K}_{r_i, q_{i'}} \mathbf{W}_{i'}^\top \end{pmatrix},
\end{aligned}$$

$$\begin{aligned}
\mathbf{T}^{(22)} &= \sum_{\substack{i=\ell+1 \\ (i,i') \notin \mathcal{J}_{22}^*}}^k \begin{pmatrix} \mathbf{V}_i \\ \mathbf{0}_{q \times r_i p_{i'}} \end{pmatrix} \text{Bdiag}(0, \dots, 0, \mathbf{I}_{r_i} \otimes \mathbf{\Lambda}_i, 0, \dots, 0) (\mathbf{V}_i^\top, \mathbf{0}_{r_i p_{i'} \times q}) \\
&+ \sum_{\substack{i'=\ell'+1 \\ (i,i') \notin \mathcal{J}_{22}^*}}^{k'} \begin{pmatrix} \mathbf{0}_{p \times r_{i'} q_{i'}} \\ \mathbf{W}_{i'} \end{pmatrix} \text{Bdiag}(0, \dots, 0, \mathbf{I}_{r_{i'}} \otimes \mathbf{\Delta}_{i'}, 0, \dots, 0) (\mathbf{0}_{r_{i'} q_{i'} \times p}, \mathbf{W}_{i'}^\top) \\
&+ \sum_{(i,i') \in \mathcal{J}_{22}^*} \begin{pmatrix} \mathbf{V}_i (\mathbf{\Gamma}_{ii'} \otimes \mathbf{I}_{p_i}) \mathbf{K}_{r_i, p_i} & \mathbf{0}_{r_i p_i \times r_i q_{i'}} \\ \mathbf{0}_{r_i q_{i'} \times r_i p_i} & \mathbf{W}_{i'} \mathbf{K}_{q_{i'}, r_i} \end{pmatrix} \mathbf{K}_{p_i + q_{i'}, r_i} \\
&\quad \times \text{Bdiag}(0, \dots, 0, \mathbf{I}_{r_i} \otimes \begin{pmatrix} \mathbf{\Lambda}_i & \mathbf{G}_{ii'} \\ \mathbf{G}_{ii'}^\top & \mathbf{\Delta}_{i'} \end{pmatrix}, 0, \dots, 0) \\
&\quad \times \mathbf{K}_{r_i, p_i + q_{i'}} \begin{pmatrix} \mathbf{K}_{p_i, r_i} (\mathbf{\Gamma}_{ii'}^\top \otimes \mathbf{I}_{p_i}) \mathbf{V}_i^\top & \mathbf{0}_{r_i p_i \times r_i q_{i'}} \\ \mathbf{0}_{r_i q_{i'} \times r_i p_i} & \mathbf{K}_{r_i, q_{i'}} \mathbf{W}_{i'}^\top \end{pmatrix},
\end{aligned}$$

where the respective eigenvalue or block-eigenvalue is located at the first nonzero position of the block-diagonal matrix.

Observe that, since all $\mathbf{V}_i$, $\mathbf{W}_{i'}$, $\mathbf{\Gamma}_{ii'}$, and commutation matrices are orthogonal, each $\mathbf{T}^{(vw)}$, $v, w \in \{1, 2\}$, is isomorphic to a corresponding block-diagonal matrix. Consequently, $\mathbf{T}$ itself is also isomorphic to a block-diagonal matrix

$$\begin{aligned}
&\text{Bdiag} \left(\text{Bdiag}_{(i,i') \in \mathcal{J}_{11}} (\lambda_i \mathbf{I}_{r_i}, \delta_{i'} \mathbf{I}_{r_{i'}}, \mathbf{I}_{r_i} \otimes \begin{pmatrix} \lambda_i & \gamma_{ii'} \\ \gamma_{ii'} & \delta_{i'} \end{pmatrix}), \right. \\
&\quad \text{Bdiag}_{(i,i') \in \mathcal{J}_{21}} (\mathbf{I}_{r_i} \otimes \mathbf{\Lambda}_i, \delta_{i'} \mathbf{I}_{r_{i'}}, \mathbf{I}_{r_i} \otimes \begin{pmatrix} \mathbf{\Lambda}_i & \mathbf{g}_{ii'} \\ \mathbf{g}_{ii'}^\top & \delta_{i'} \end{pmatrix}), \\
&\quad \text{Bdiag}_{(i,i') \in \mathcal{J}_{12}} (\lambda_i \mathbf{I}_{r_i}, \mathbf{I}_{r_{i'}} \otimes \mathbf{\Delta}_{i'}, \mathbf{I}_{r_i} \otimes \begin{pmatrix} \lambda_i & \mathbf{g}_{ii'}^\top \\ \mathbf{g}_{ii'} & \mathbf{\Delta}_{i'} \end{pmatrix}), \\
&\quad \left. \text{Bdiag}_{(i,i') \in \mathcal{J}_{22}} (\mathbf{I}_{r_i} \otimes \mathbf{\Lambda}_i, \mathbf{I}_{r_{i'}} \otimes \mathbf{\Delta}_{i'}, \mathbf{I}_{r_i} \otimes \begin{pmatrix} \mathbf{\Lambda}_i & \mathbf{G}_{ii'}^\top \\ \mathbf{G}_{ii'} & \mathbf{\Delta}_{i'} \end{pmatrix}) \right),
\end{aligned}$$

which is permutationally similar to

$$\begin{aligned}
&\text{Bdiag} \left(\lambda_i \mathbf{I}_{r_i}, \delta_{i'} \mathbf{I}_{r_{i'}}, \mathbf{I}_{r_i} \otimes \mathbf{\Lambda}_i, \mathbf{I}_{r_{i'}} \otimes \mathbf{\Delta}_{i'}, \mathbf{I}_{r_i} \otimes \begin{pmatrix} \lambda_i & \gamma_{ii'} \\ \gamma_{ii'} & \delta_{i'} \end{pmatrix}, \right. \\
&\quad \left. \mathbf{I}_{r_i} \otimes \begin{pmatrix} \mathbf{\Lambda}_i & \mathbf{g}_{ii'} \\ \mathbf{g}_{ii'}^\top & \delta_{i'} \end{pmatrix}, \mathbf{I}_{r_i} \otimes \begin{pmatrix} \lambda_i & \mathbf{g}_{ii'}^\top \\ \mathbf{g}_{ii'} & \mathbf{\Delta}_{i'} \end{pmatrix}, \mathbf{I}_{r_i} \otimes \begin{pmatrix} \mathbf{\Lambda}_i & \mathbf{G}_{ii'}^\top \\ \mathbf{G}_{ii'} & \mathbf{\Delta}_{i'} \end{pmatrix} \right)
\end{aligned}$$

with $i \in \{1, \dots, k\}$, $i' \in \{1, \dots, k'\}$.

Finally, note that the construction of partitioned matrices belonging to quadratic subspaces can be extended to block partitions with more than two blocks. For example, we first construct a matrix $\mathbf{T} \in \mathcal{T}$, as defined in (6), partitioned into four blocks. We then treat this matrix as the new first diagonal block, $\mathbf{A}_1$, of a larger block-diagonal matrix and introduce an additional diagonal block, $\mathbf{B}_1$, of an appropriate order, r . Next, we repeat the procedure for selecting the off-diagonal block to obtain a new matrix belonging to the quadratic subspace of matrices of order $p+q+r$. This process can be continued until the required partition is achieved. The following example illustrates this method.

Example 3.4 Let

$$\begin{aligned}\mathcal{A} &= \{a_1 \mathbf{v}_1 \mathbf{v}_1^\top + a_2 \mathbf{v}_2 \mathbf{v}_2^\top + a_3 \mathbf{v}_3 \mathbf{v}_3^\top : a_1, a_2, a_3 \in \mathbb{R}\}, \\ \mathcal{B} &= \{\delta_1 \mathbf{w}_1 \mathbf{w}_1^\top + \delta_2 \mathbf{w}_2 \mathbf{w}_2^\top : \delta_1, \delta_2 \in \mathbb{R}\},\end{aligned}$$

where

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \mathbf{w}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{w}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

We first consider a block-diagonal matrix with $\mathbf{A} \in \mathcal{A}$ and $\mathbf{B} \in \mathcal{B}$ on the diagonal. Then, according to Theorem 3.2, if the off-diagonal block is taken from $\mathcal{C}_{11} = \{\gamma_1 \mathbf{v}_1 \mathbf{w}_1^\top : \gamma_1 \in \mathbb{R}\}$, leading to

$$\mathbf{A}_1 = \begin{pmatrix} a_1 & 0 & 0 & \gamma_1^* & \gamma_1^* \\ 0 & a_2 & 0 & 0 & 0 \\ 0 & 0 & a_3 & 0 & 0 \\ \gamma_1^* & 0 & 0 & b_1 & b_2 \\ \gamma_1^* & 0 & 0 & b_2 & b_1 \end{pmatrix},$$

where $\gamma_1^* = \frac{1}{\sqrt{2}}\gamma_1$, then $\mathbf{A}_1$ belongs to a quadratic space, say $\mathcal{A}_1$. Note that

$$\mathbf{A}_1 = \mathbf{P} \text{Bdiag}\left(\begin{pmatrix} a_1 & \gamma_1 \\ \gamma_1 & \delta_1 \end{pmatrix}, a_2, a_3, \delta_2\right) \cdot \mathbf{P}^\top$$

with

$$\mathbf{P} = \left(\begin{pmatrix} \mathbf{v}_1 & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{2 \times 1} & \mathbf{w}_1 \end{pmatrix}, \begin{pmatrix} \mathbf{v}_2 \\ \mathbf{0}_{2 \times 1} \end{pmatrix}, \begin{pmatrix} \mathbf{v}_3 \\ \mathbf{0}_{2 \times 1} \end{pmatrix}, \begin{pmatrix} \mathbf{0}_{3 \times 1} \\ \mathbf{w}_2 \end{pmatrix} \right).$$

Consider now a block-diagonal matrix with $\mathbf{A}_1 \in \mathcal{A}_1$ and $\mathbf{B}_1 \in \mathcal{B}$ on the diagonal. Then, if the off-diagonal block is taken from $\mathcal{C}_{11} = \{\gamma_2 \begin{pmatrix} \mathbf{v}_2 \\ \mathbf{0}_2 \end{pmatrix} \mathbf{w}_1^\top + \gamma_3 \begin{pmatrix} \mathbf{0}_3 \\ \mathbf{w}_2 \end{pmatrix} \mathbf{w}_2^\top : \gamma_2, \gamma_3 \in \mathbb{R}\}$, which results in

$$\mathbf{T} = \begin{pmatrix} a_1 & 0 & 0 & \gamma_1^* & \gamma_1^* & 0 & 0 \\ 0 & a_2 & 0 & 0 & 0 & \gamma_2^* & \gamma_2^* \\ 0 & 0 & a_3 & 0 & 0 & 0 & 0 \\ \gamma_1^* & 0 & 0 & b_1 & b_2 & \gamma_3^* & -\gamma_3^* \\ \gamma_1^* & 0 & 0 & b_2 & b_1 & -\gamma_3^* & \gamma_3^* \\ 0 & \gamma_2^* & 0 & \gamma_3^* & -\gamma_3^* & c_1 & c_2 \\ 0 & \gamma_2^* & 0 & -\gamma_3^* & \gamma_3^* & c_2 & c_1 \end{pmatrix},$$

where $\gamma_2^* = \frac{1}{\sqrt{2}}\gamma_2$ and $\gamma_3^* = \frac{1}{2}\gamma_3$, then $\mathbf{T}$ belongs to a quadratic space. $\square$

4. Applications

4.1. Statistical analysis

Consider an experiment with n observations, each represented as an m -dimensional random vector $\mathbf{x}_i$, $i \in \{1, \dots, n\}$. Assume that the $\mathbf{x}_i$ are mutually independent and normally distributed with mean $\mathbb{E}(\mathbf{x}_i) = \boldsymbol{\mu}$, an m -dimensional vector, and covariance matrix $\mathbb{D}(\mathbf{x}_i) = \boldsymbol{\Sigma}$, a symmetric p.d. matrix of order m . In modern applications, the dimension m often exceeds n , and both values tend to be large (so-called high-dimensional settings). This results in a significant challenge due to the large number of parameters that need to be estimated. A common approach to circumvent this issue is to impose structure on the covariance matrix, which reduces the number of free parameters.

Suppose each observation $\mathbf{x}_i$, $i \in \{1, \dots, n\}$, consists of multiple subvectors, corresponding to natural groupings of observed variables. These groups may arise from different feature types or repeated measurements across various conditions (e.g., locations, time points, depths). In such cases, the partitioning of $\mathbf{x}_i$ should be reflected in the covariance matrix.

In two-level multivariate models several structured covariance models have been extensively studied, including separable covariance matrices (see, e.g., [25], [26], [27], [28]), block compound symmetry (BCS, see, e.g., [10], [29],

[11], [30], [28]), and the block circular Toeplitz structure (see, e.g., [7], [8]). For separable covariance matrices, the maximum likelihood estimator (MLE) is typically computed numerically. In contrast, for BCS and block circular Toeplitz structures, explicit formulas for the MLEs exist, since both belong to quadratic subspaces; cf. [12, Proposition 4.1]. The general formula for the MLE of any structure belonging to a quadratic subspace is provided in [9]. By [12, Theorem 4.1], let $\mathbf{S} = \mathbf{X}(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\mathbf{X}^\top$, where $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$. Then, the MLE is given by the projection of $\frac{1}{n}\mathbf{S}$ onto the space of the relevant covariance structures. Additionally, the unbiased estimator for a BCS structure, obtained as the projection of the sample covariance matrix $\frac{1}{n-1}\mathbf{S}$ has been studied in [10, 11, 31].

In three-level multivariate models, [32] presented the best unbiased estimators for a doubly exchangeable structure, that is, a BCS structure with structured blocks, while [33] considered maximum likelihood estimation for this structure and proposed a hypothesis test for this structure. As can be observed, in two- or multi-level multivariate models, the “levels” can be reflected in the covariance matrix by appropriately structuring its subblocks. Another approach to modeling specific dependencies in multi-level multivariate models is via separability; cf. [34]; however, such a structure is nonlinear, and therefore the results presented in this paper are not applicable.

A key limitation of the aforementioned structures is the assumption that each block has the same dimension across all subvectors. However, in many practical applications, subvector dimensions vary (see, e.g., [13, Figure 1], which illustrates covariance matrices from real-world datasets). To accommodate this variability, [13] proposed MLEs for covariance matrices with blocks of unequal sizes and a specific structure, in which the diagonal blocks exhibit compound symmetry, while the off-diagonal blocks are proportional to a matrix of ones. Moreover, in a corollary, they showed that the inverse of the considered covariance structure is given in closed form. Notably, this structure belongs to a quadratic subspace.

The results presented in this paper extend the estimation approach of [13] by allowing for more general covariance structures within quadratic subspaces. An explicit formula for the MLE of any such structure is given below.

Proposition 4.1. *Let $\mathbf{S} = \mathbf{X}(\mathbf{I}_n - n^{-1}\mathbf{1}_n\mathbf{1}_n^\top)\mathbf{X}^\top$, where $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n) \sim \mathcal{N}_{m,n}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{I}_n)$, and let $\boldsymbol{\Sigma} \in \mathcal{Q}$ with $\mathcal{Q}$ denoting a quadratic subspace. Then, the projection $\boldsymbol{\Omega}_{\mathcal{Q}}((n-1)^{-1}\mathbf{S})$, given in Proposition 2.2, is the best unbiased quadratic estimator of the covariance structure $\boldsymbol{\Sigma}$. Moreover, the projection*

$\Omega_Q(n^{-1}\mathbf{S})$ is the maximum likelihood estimator of Σ .

Proof. The first part follows from [32], and the second follows from [12]. $\square$

Remark 4.2. The projection $\Omega_Q((n-1)^{-1}\mathbf{S})$ is the best unbiased quadratic estimator of the covariance structure Σ even when $\mathbf{X}$ follows a quasi-normal distribution, that is, a distribution whose first four moments coincide with those of the normal distribution; cf. [32].

Remark 4.3. The determination of MLEs relies on specific distributional assumptions, with normality playing a crucial role. In this case, explicit solutions for MLEs exist if and only if Σ belongs to a quadratic subspace; cf. [12, Proposition 4.1]. When the underlying distribution is non-normal, the determination of MLEs typically requires numerical algorithms, regardless of the structure of Σ . For example, if $\mathbf{X}$ follows a multivariate t distribution with location parameter $\boldsymbol{\mu}$ and scale matrix Σ , the MLEs are not available in closed form, even though the space of symmetric matrices is quadratic; cf. [35]. In such cases, estimation is usually performed using the EM algorithm; cf. [36].

In the high-dimensional setting, the matrix $\mathbf{S}$ defined above is singular, which makes statistical analysis challenging. Projecting $\mathbf{S}$ onto a quadratic space typically improves its conditioning (see Proposition 2.2(a)), resulting in a well-conditioned estimator of Σ . For instance, under the assumption of a spherical covariance structure, the resulting estimator is always positive definite, since the projection of any (nonzero) nonnegative definite matrix onto the space of matrices proportional to the identity is always positive definite. Clearly, assuming a specific covariance structure reduces the number of parameters, and selecting such a structure from a quadratic space facilitates estimation, thereby enabling more efficient data analysis. For example, to address the high-dimensionality problem in certain metabolomic datasets, [37] considered fitting several commonly used covariance structures inferred from heatmaps and proposed corresponding estimators. The results presented in this paper indicate that restricting attention to structures belonging to quadratic subspaces may lead to more efficient estimation procedures. In particular, for such structures the MLEs can be obtained in a straightforward and computationally efficient manner, suggesting that exploring quadratic subspace models could be especially advantageous in high-dimensional applications. A similar observation was made indirectly by [13],

who addressed the high-dimensionality problem in a biomedical dataset by assuming a particular covariance structure: compound-symmetry matrices on the diagonal and matrices proportional to all-ones matrices as off-diagonal blocks. Although the authors did not explicitly refer to the concept of a quadratic subspace, this structure indeed is a quadratic subspace (and can be constructed using the method described in Theorem 3.2). Consequently, the MLEs of the covariance parameters given in [13, formula (5)] can be obtained by projecting $n^{-1}\mathbf{S}$ onto the corresponding quadratic subspace.

Finally, according to [12], the quadraticity of the covariance space is a necessary and sufficient condition for determining the explicit form of the MLE of the covariance matrix. On the other hand, the projection of $n^{-1}\mathbf{S}$ onto a linear space is simply the least-squares estimator of Σ ; cf. [38]. However, note that since projection onto a non-quadratic space does not preserve definiteness (cf. [6]), the resulting estimator may be unrealistic. To restore the positive definiteness of the estimator, one can apply the shrinkage method proposed by [39] or its modifications, which yield a structured estimator; see, e.g., [40], [41]. Nonetheless, as shown in [41], a quadratic space plays a crucial role even in the shrinkage method, since improving the algebraic properties of the resulting estimate requires the target matrix to belong to a quadratic space.

The identification of the most suitable covariance structure is an important topic in statistical analysis. Various approaches to addressing this problem can be found in the literature, including graphical methods; see, e.g., [15], [16], [17]). However, in graphical models, researchers typically focus on precision matrix (the inverse of the covariance matrix) and are often interested in identifying strongly dependent variables while discarding weak dependencies, leading to sparse representations. The difficulty is that even if the precision matrix is sparse, the corresponding covariance matrix may not be, which can complicate interpretation. In contrast, when the assumed structure belongs to a quadratic subspace, this issue does not arise, since the inverse of a matrix from a quadratic space also belongs to the same space.

On the other hand, a covariance matrix belonging to a quadratic space does not necessarily have to be sparse, yet its inverse still preserves the same structure. Therefore, characterizing the quadratic space of partitioned matrices is of great importance, particularly when estimating dependencies between subvectors, which are represented by the off-diagonal blocks of the covariance matrix (commonly referred to as the cross-covariance matrix).

Nevertheless, since graphical models used in pattern recognition often

involve concepts from graph theory, in the following section we employ adjacency matrices of undirected graphs to illustrate the main results of this paper.

4.2. Graph theory

Consider a disconnected graph consisting of two separate subgraphs with p and q vertices, respectively. Clearly, its adjacency matrix can be expressed as a block-diagonal matrix with $p \times p$ and $q \times q$ blocks on the diagonal. It follows that such an adjacency matrix belongs to a quadratic subspace, meaning that its square retains the same block-diagonal structure. Interpreting this result, since the square of the adjacency matrix indicates the number of length 2 paths between vertices, it is evident that if no path exists between a vertex in the first subgraph and a vertex in the second subgraph, then no length 2 path can exist either. Now, one may ask whether it is possible to add one or more edges between the subgraphs in such a way that the number of paths of length 2 (or higher) remains unchanged. In other words, we search for a nonzero off-diagonal block of the adjacency matrix while preserving the overall structure. According to Theorem 3.2, this off-diagonal block must be constructed based on the eigenvectors of diagonal blocks.

The adjacency matrix of an undirected graph is usually a binary matrix; however, when nonbinary entries are allowed, the graph is referred to as a weighted graph. In this context, one may be interested in graphs whose structure is preserved under powers of the adjacency matrix (see also [1], where it is shown that the inverse also preserves the structure). This means that the resulting graph does not acquire new edges, and the weights – whether equal or proportional in the original graph – remain unchanged or proportional in the transformed graph.

Clearly, this property holds for the block-diagonal adjacency matrix of a weighted graph when its diagonal blocks belong to quadratic subspaces. Theorem 3.2 further shows that additional edges can be introduced between vertices associated with the spaces of diagonal blocks while still preserving the overall structure. Examples 4.1 and 4.2 illustrate the application of these results.

Example 4.1 Let

$$\begin{aligned}\mathcal{A} &= \{\text{Bdiag}(a_0 \mathbf{I}_3, \begin{pmatrix} a_1 & a_2 \\ a_2 & a_1 \end{pmatrix}) : a_0, a_1, a_2 \in \mathbb{R}\}, \\ \mathcal{B} &= \{\text{Bdiag}(b_0, \begin{pmatrix} b_1 & b_2 \\ b_2 & b_1 \end{pmatrix}) : b_0, b_1, b_2 \in \mathbb{R}\},\end{aligned}$$

where

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{v}_4 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \mathbf{v}_5 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}$$

and

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{w}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \mathbf{w}_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}.$$

The graph represented by the adjacency matrix $\text{Bdiag}(\mathbf{A}, \mathbf{B})$, where $\mathbf{A} \in \mathcal{A}$, $\mathbf{B} \in \mathcal{B}$, is illustrated in Figure 1(a), where vertices of the same shape correspond to identical diagonal entries of the given matrix, and the edge colors indicate the values of the off-diagonal entries. The upper row depicts the graph associated with the adjacency matrix $\mathbf{A}$, whereas the lower row corresponds to that associated with $\mathbf{B}$.

Now, assume that $\mathbf{C} \in \mathcal{C} = \{g \mathbf{v}_4 \mathbf{w}_2^\top : g \in \mathbb{R}\}$, and consider $\mathbf{T} \in \mathcal{T}$ as defined in (6). Since the adjacency matrix is no longer block-diagonal, new edges appear in the graph shown in Figure 1(a), resulting in the graph depicted in Figure 1(b).

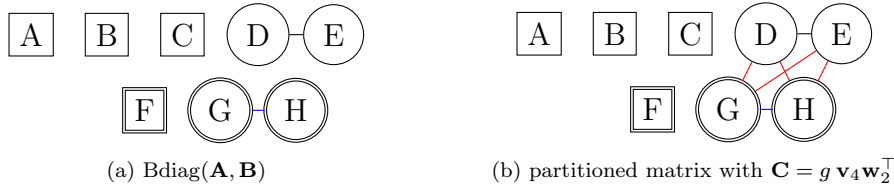


Figure 1: Graphs represented by their respective adjacency matrices.

Since $\mathcal{T}$ is a quadratic subspace, the matrices $\mathbf{T}^2$ and $\mathbf{T}^{-1}$ also belong to this space; cf. [1]. This implies that the graphs corresponding to the

adjacency matrices $\mathbf{T}$, $\mathbf{T}^2$ and $\mathbf{T}^{-1}$ preserve both the vertex shapes and the edge colors. While the edge weights may change, edges that are equal in weights remain proportional. Finally, as shown in Figure 1(b), the matrix $\mathbf{T}$ is isomorphic to

$$\text{Bdiag}(a_0 \mathbf{I}_3, b_0, \begin{pmatrix} a_1 & a_2 & g & g \\ a_2 & a_1 & g & g \\ g & g & b_1 & b_2 \\ g & g & b_2 & b_1 \end{pmatrix}),$$

where g represents the weight of the edge shown in red. $\square$

Remark. Consider the same block-diagonal matrix as in Example 4.1 assuming $b_0 = a_0$. One can then permute this matrix so that the first diagonal entry becomes the fifth diagonal entry. As a consequence, we obtain the following partition:

$$\left(\begin{array}{cccc|cccc} a_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & a_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_1 & a_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_2 & a_1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & a_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & a_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & b_1 & b_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & b_2 & b_1 \end{array} \right).$$

The off-diagonal block may then be taken from the block-eigenspace consisting on the block-eigenvectors corresponding to the eigenvalue a_0 (which has multiplicity 2) and the block-eigenvectors corresponding to the 2×2 block-eigenvalue (with multiplicity 1), i.e.,

$$\mathbf{C} = \gamma \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} + g \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} (0 \ 0 \ 1 \ 1)$$

resulting in

$$\left(\begin{array}{cccc|cccc} a_0 & 0 & 0 & 0 & \gamma & 0 & 0 & 0 \\ 0 & a_0 & 0 & 0 & 0 & \gamma & 0 & 0 \\ 0 & 0 & a_1 & a_2 & 0 & 0 & g & g \\ 0 & 0 & a_2 & a_1 & 0 & 0 & g & g \\ \hline \gamma & 0 & 0 & 0 & a_0 & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 & 0 & a_0 & 0 & 0 \\ 0 & 0 & g & g & 0 & 0 & b_1 & b_2 \\ 0 & 0 & g & g & 0 & 0 & b_2 & b_1 \end{array} \right).$$

Reverting to the original partition, we obtain

$$\left(\begin{array}{ccccc|cccc} a_0 & \gamma & 0 & 0 & 0 & 0 & 0 & 0 \\ \gamma & a_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_0 & 0 & 0 & \gamma & 0 & 0 \\ 0 & 0 & 0 & a_1 & a_2 & 0 & g & g \\ 0 & 0 & 0 & a_2 & a_1 & 0 & g & g \\ \hline 0 & 0 & \gamma & 0 & 0 & a_0 & 0 & 0 \\ 0 & 0 & 0 & g & g & 0 & b_1 & b_2 \\ 0 & 0 & 0 & g & g & 0 & b_2 & b_1 \end{array} \right),$$

whose square retains the same structure. However, note that the above matrix does not belong to the same quadratic subspace as the original matrix, since the structure of the first diagonal block, $\mathbf{A}$, has collapsed. Moreover, because we are working with subspaces defined in (5) rather than individual matrices, we cannot arbitrarily change the partition of a matrix, as was done in the considerations above.

Example 4.2 Consider the block-diagonal matrix from Example 3.1 with $\mathbf{C} \in \mathcal{C}_{11}^{(\text{II})}$. The graph associated with its adjacency matrix is shown in Figure 2, where a dashed line indicates negative edge weights corresponding to edges of the same color. Since $\mathcal{T}$, defined as (6), is a quadratic subspace, the graphs associated with the adjacency matrices $\mathbf{T}^2$ and $\mathbf{T}^{-1}$ also preserve the same structure.

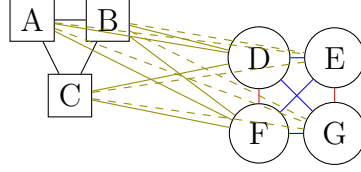


Figure 2: The graph associated with the partitioned adjacency matrix from Example 3.1 with $\mathbf{C} = \mathbf{C}_2$.

Observe that the graph in Figure 2 is connected, which suggests that its adjacency matrix $\mathbf{T}$ cannot be represented in block-diagonal form. However, using the results presented in Section 3 (Example 3.1 - cont.), $\mathbf{T}$ can be expressed as a matrix isomorphic to $\text{Bdiag}(\begin{pmatrix} \lambda_1 & \gamma_2 \\ \gamma_2 & \delta_3 \end{pmatrix}, \lambda_2 \mathbf{I}_2, \delta_1, \delta_2 \mathbf{I}_2)$. $\square$

Finally, note that the graph structure plays a crucial role in graphical methods for pattern recognition, particularly when the primary objective is to recover dependency structures among variables or to analyze correlation patterns.

4.3. Systems of linear equations

The system of linear equations $\mathbf{Ax} = \mathbf{b}$ ($\det \mathbf{A} \neq 0$) plays a crucial role in nearly all areas of science. While the solution is given by $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$, computing $\mathbf{A}^{-1}$ can be computationally expensive for large matrices. However, if $\mathbf{A}$ belongs to a quadratic space $\mathcal{Q}$, it can be represented in a block-diagonal form $\mathcal{D}$, allowing for more efficient computation by inverting smaller matrices. Specifically, the solution can be expressed as

$$\mathbf{x} = \mathbf{P} \cdot \text{Bdiag}_{i=1}^k(\mathbf{\Lambda}_i^{-1}) \cdot \mathbf{P}^\top \mathbf{b},$$

where $\mathbf{P}$ is an orthogonal matrix whose columns are the block-eigenvectors of $\mathbf{A}$, and $\mathbf{\Lambda}_i$ are the corresponding block-eigenvalues.

Appendix A. Frobenius norm minimizers

Assume that $\mathcal{D}$ consists of different blocks on the diagonal. Then, for $i \in \{1, \dots, k\}$, we seek

$$\arg \min_{\Psi_i \in \mathbb{R}_{p_i}^s} \|\mathbf{Y} - \text{Bdiag}_{i=1}^k(\Psi_i)\|_F^2.$$

Defining

$$f_i(\mathbf{Y}, \mathbf{\Psi}_i) = -2 \text{Tr}(\mathbf{Y}_{ii} \mathbf{\Psi}_i) + \text{Tr}(\mathbf{\Psi}_i^2), \quad i \in \{1, \dots, k\},$$

where $\mathbf{Y}_{ii}$ is the i th $p_i \times p_i$ diagonal block of $\mathbf{Y}$, the Frobenius norm can be rewritten as

$$\|\mathbf{Y} - \text{Bdiag}_{i=1}^k(\mathbf{\Psi}_i)\|_F^2 = \text{Tr}(\mathbf{Y}^2) + \sum_{i=1}^k f_i(\mathbf{Y}, \mathbf{\Psi}_i) := f(\mathbf{Y}, \text{Bdiag}_{i=1}^k(\mathbf{\Psi}_i)).$$

Differentiating with respect to $\mathbf{\Psi}_i$, $i \in \{1, \dots, k\}$, following the rules given in [42], yields

$$\begin{aligned} \frac{\partial f}{\partial \mathbf{\Psi}_i} &= \left\{ -2 \text{vec}^\top \mathbf{Y}_{ii} + \text{vec}^\top \mathbf{I}_{p_i} [\mathbf{I}_{p_i} \otimes \mathbf{\Psi}_i + \mathbf{\Psi}_i \otimes \mathbf{I}_{p_i}] \right\} \cdot \mathbf{D}_{p_i} \\ &= \left\{ -2 \text{vec}^\top \mathbf{Y}_{ii} + 2 \text{vec}^\top \mathbf{\Psi}_i \right\} \cdot \mathbf{D}_{p_i}, \end{aligned}$$

where $\mathbf{D}_{p_i}$ is the $p_i \times p_i(p_i+1)/2$ duplication matrix; see e.g. [24]. Setting the derivative to zero (noting that $\mathbf{D}_{p_i}$ does not affect the solution), we obtain

$$\mathbf{\Psi}_i = \mathbf{Y}_{ii}, \quad i \in \{1, \dots, k\}.$$

Now, assume that $\mathcal{D}$ consists of matrices with identical diagonal blocks. In this case, we seek

$$\arg \min_{\mathbf{\Psi} \in \mathbf{R}_{m/k}^s} \|\mathbf{Y} - (\mathbf{I}_k \otimes \mathbf{\Psi})\|_F^2.$$

Differentiating

$$f(\mathbf{Y}, \mathbf{\Psi}) = \|\mathbf{Y} - (\mathbf{I}_k \otimes \mathbf{\Psi})\|_F^2 = \text{Tr}(\mathbf{Y}^2) - 2 \text{Tr}(\mathbf{Y}(\mathbf{I}_k \otimes \mathbf{\Psi})) + k \text{Tr}(\mathbf{\Psi}^2)$$

with respect to $\mathbf{\Psi}$ gives

$$\begin{aligned} \frac{\partial f}{\partial \mathbf{\Psi}} &= \left[-2 \text{vec}^\top \mathbf{Y} (\mathbf{I}_k \otimes \mathbf{K}_{m/k,k} \otimes \mathbf{I}_{m/k}) (\text{vec} \mathbf{I}_k \otimes \mathbf{I}_{(m/k)^2}) \right. \\ &\quad \left. + k \text{vec}^\top \mathbf{I}_{m/k} [\mathbf{I}_{m/k} \otimes \mathbf{\Psi} + \mathbf{\Psi} \otimes \mathbf{I}_{m/k}] \right] \mathbf{D}_{m/k} \\ &= \left[-2 \text{vec}^\top \mathbf{Y} (\mathbf{I}_k \otimes \mathbf{K}_{m/k,k} \otimes \mathbf{I}_{m/k}) (\text{vec} \mathbf{I}_k \otimes \mathbf{I}_{(m/k)^2}) + 2k \text{vec}^\top \mathbf{\Psi} \right] \mathbf{D}_{m/k}. \end{aligned}$$

Setting the derivative to zero and applying [20, Lemma 2.9] we obtain the minimizer

$$\mathbf{\Psi} = \frac{1}{k} \sum_{i=1}^k \mathbf{Y}_{ii},$$

where $\mathbf{Y}_{ii}$, $i \in \{1, \dots, k\}$, are the $m/k \times m/k$ diagonal blocks of $\mathbf{Y}$.

Acknowledgments

This work was supported by the Poznan University of Technology under Grant no. 0213/SBAD/0119 (K. Filipiak and M. Mrowińska).

References

- [1] J. Seely, Quadratic subspaces and completeness, *Ann. Math. Statist.* 42 (2) (1971) 710–721.
- [2] S. T. Jensen, Covariance hypotheses which are linear in both the covariance and the inverse of covariance, *Ann. Statist.* 16 (1988) 302–322.
- [3] J. E. Denis, J. F. Traub, R. P. Weber, On the matrix polynomial, lambda matrix and block-eigenvalue problems, Cornell University Technical Report TR71-147, 1971.
- [4] J. Eisenfeld, Block diagonalization and eigenvalues, *Linear Algebra Appl.* 15 (1976) 205–215.
- [5] J. Vitoria, Block eigenvalues of block compound matrices, *Linear Algebra Appl.* 47 (1982) 23–34.
- [6] B. Fuglede, T. S. Jensen, Positive projections of symmetric matrices and Jordan algebras, *Expo. Math.* 31 (2013) 295–303.
- [7] Y. Liang, D. von Rosen, T. von Rosen, On estimation in hierarchical models with block circular covariance structures, *Ann. Inst. Statist. Math.* 67 (4) (2015) 773–791.
- [8] Y. Liang, C. A. Coelho, T. von Rosen, Hypothesis testing in multivariate normal models with block circular covariance structures, *Biom. J.* 64 (2022) 557–576.
- [9] K. Filipiak, M. John, Y. Liang, Testing covariance structures belonging to a quadratic subspace under a doubly multivariate model, *TEST* 33 (2024) 847–876.
- [10] A. Roy, R. Zmysłony, M. Fonseca, R. Leiva, Optimal estimation for doubly multivariate data in blocked compound symmetric covariance structure, *J. Multivariate Anal.* 144 (2016) 81–90.

- [11] M. Fonseca, A. Koziol, R. Zmyslony, Testing hypotheses of covariance structure in multivariate data, *Electron. J. Linear Algebra* 33 (2018) 53–62.
- [12] K. Filipiak, M. John, A. Markiewicz, Comments on maximum likelihood estimation and projections under multivariate statistical models, in: T. Holgersson, M. Singull (Eds.), *Recent Developments in Multivariate and Random Matrix Analysis*, Springer, Cham, 2020, pp. 51–66.
- [13] Y. Yang, C. Chen, S. Chen, Covariance matrix estimation for high-throughput biomedical data with interconnected communities, *Amer. Statist.* 78 (4) (2024) 401–411.
- [14] I. Archakov, P. R. Hansen, A canonical representation of block matrices with applications to covariance and correlation matrices, *Rev. Econ. Stat.* 106 (4) (2024) 1099–1113.
- [15] E. Devijver, M. Gallopin, Block-diagonal covariance selection for high-dimensional Gaussian graphical models, *J. Amer. Statist. Assoc.* 113 (521) (2018) 306–314.
- [16] R. Riccobello, M. Bogdan, G. Bonaccolto, P. J. Kremer, S. Paterlini, P. Sobczyk, Sparse graphical modelling via the sorted l_1 -norm, *arXiv:2204.10403* (2022).
- [17] E. Ben-David, T. Li, H. Massam, B. Rajaratnam, High dimensional Bayesian inference for Gaussian directed acyclic graph models, *arXiv:1109.4371* (2011).
- [18] K. Filipiak, D. Klein, M. Markiewicz, M. Mokrzycka, Approximation with a Kronecker product structure with one component as compound symmetry or autoregression via entropy loss function, *Linear Algebra Appl.* 610 (2021) 625–646.
- [19] K. Filipiak, D. Klein, M. Mokrzycka, Separable covariance structure identification for doubly multivariate data, in: K. Filipiak, A. Markiewicz, D. von Rosen (Eds.), *Multivariate, Multilinear and Mixed Linear Models*, Springer, 2021, pp. 131–155.

- [20] K. Filipiak, D. Klein, E. Vojtkova, The properties of partial trace and block trace operators of partitioned matrices, *Electron. J. Linear Algebra* 33 (2018b) 3–15.
- [21] A. W. Marshall, I. Olkin, B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, 2nd Edition, Springer, 2011.
- [22] I. Olkin, S. Press, Testing and estimation for a circular stationary model, *Ann. Math. Statist.* 40 (1969) 1358–1373.
- [23] A. Basilevsky, *Applied Matrix Algebra in the Statistical Sciences*, Elsevier Science Publishers B.V., North-Holland, 1983.
- [24] J. Magnus, H. Neudecker, Symmetry, 0-1 matrices and Jacobians: a review, *Econom. Theory* 2 (1986) 157–190.
- [25] N. Lu, D. Zimmerman, The likelihood ratio test for a separable covariance matrix, *Statist. Probab. Lett.* 73 (2005) 449–457.
- [26] K. Filipiak, D. Klein, A. Roy, A comparison of likelihood ratio tests and Rao’s score test for three separable covariance matrix structures, *Biom. J.* 59 (2017) 192–215.
- [27] K. Filipiak, D. Klein, M. Mokrzycka, Estimators comparison of separable covariance structure with one component as compound symmetry matrix, *Electron. J. Linear Algebra* 22 (2018a) 83–98.
- [28] K. Filipiak, D. Klein, Estimation and testing of the covariance structure of doubly multivariate data, in: K. Filipiak, A. Markiewicz, D. von Rosen (Eds.), *Multivariate, Multilinear and Mixed Linear Model*, Springer, 2021, pp. 113–130.
- [29] C. Coelho, A. Roy, Testing the hypothesis of a block compound symmetric covariance matrix for elliptically contoured distributions, *TEST* 26 (2017) 308–330.
- [30] A. Roy, K. Filipiak, D. Klein, Testing a block exchangeable covariance matrix, *Statistics* 52 (2) (2018) 393–408.
- [31] A. Koziol, A. Roy, R. Zmyslony, I. Žežula, M. Fonseca, Estimation and testing hypotheses in two-level and three-level multivariate data

- with block compound symmetric covariance structure., in: K. Filipiak, A. Markiewicz, D. von Rosen (Eds.), *Multivariate, Multilinear and Mixed Linear Model*, Springer, 2021, pp. 203–232.
- [32] A. Koziol, A. Roy, R. Zmyslony, R. Leiva, M. Fonseca, Best unbiased estimates for parameters of three-level multivariate data with doubly exchangeable covariance structure, *Linear Algebra Appl.* 535 (2017) 87–104.
 - [33] D. Klein, A. Roy, D. von Rosen, K. Filipiak, Estimating and testing a doubly exchangeable covariance matrix using Rao’s score test, *Sankhyā B* (2025). doi:10.1007/s13571-025-00366-y.
 - [34] K. Filipiak, D. Klein, Estimation of parameters under a generalized growth curve model, *J. Multivariate Anal.* 158 (2017) 73–86.
 - [35] K. Filipiak, T. Kollo, Covariance structure tests for multivariate t-distribution, *Statist. Papers* 65 (2024) 4537–4566.
 - [36] C. Liu, D. Rubin, ML estimation of the t distribution using EM and its extensions, ECM and ECME, *Statist. Sinica* 5 (1995) 19–39.
 - [37] A. Mieldzioc, M. Mokrzycka, A. Sawkowska, Identification of block-structured covariance matrix on an example of metabolomic data, *Separations* 8 (2021).
 - [38] M. Ohlson, D. von Rosen, Explicit estimators of parameters in the growth curve model with linearly structured covariance matrices, *J. Multivariate Anal.* 101 (2010) 1284–1295.
 - [39] O. Ledoit, M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices, *J. Multivariate Anal.* 88 (2) (2004) 365–411.
 - [40] A. Markiewicz, M. Mokrzycka, M. Mrowińska, Quasi shrinkage estimation of a block-structured covariance matrix, *J. Stat. Comput. Simul.* 94 (14) (2024) 3093–3110.
 - [41] K. Filipiak, A. Markiewicz, A. Mieldzioc, M. Mrowińska, On estimation of a partitioned covariance matrix with linearly structured blocks, *Statist. Papers* 66 (2025).

- [42] J. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, Wiley, Chichester, UK, 1999.

[E] On estimation of partitioned covariance matrix with linearly structured blocks

by Katarzyna Filipiak, Augustyn Markiewicz, Adam Mieldzioc,
and Malwina Mrowińska

Statistical Papers, **66**, 98, 2025.

Contribution:

- Katarzyna Filipiak (15%) – conceptualization and methodological supervision, writing – review and editing
- Augustyn Markiewicz (15%) – conceptualization and methodological supervision
- Adam Mieldzioc (30%) – methodology, literature review, conceptual contribution to the choice of quadratic target spaces (proof of Proposition 2), formulation and proof of the properties of proposed estimators (Theorem 3)
- Malwina Mrowińska (40%) – methodology, literature review, development of theoretical results and mathematical proofs (Theorem 1, Lemma 1), introducing a toy and real-world data examples, simulation studies, writing – original draft preparation and editing



On estimation of a partitioned covariance matrix with linearly structured blocks

Katarzyna Filipiak¹ · Augustyn Markiewicz² · Adam Mieldzioc² · Malwina Mrowińska¹

Received: 7 November 2024 / Revised: 26 February 2025
© The Author(s) 2025

Abstract

The aim of this paper is to introduce an estimation method for a linearly structured partitioned covariance matrix. In contrast to well known linear structures of partitioned matrices, for example block compound symmetry, we allow the diagonal blocks of the covariance matrix to be of different dimensions. We adapt the shrinkage method to improve the properties of the projection of the sample covariance matrix onto the linear structure space. As spaces of target matrices, we choose various quadratic subspaces of structure space. This is a novel approach in the context of the structure space under consideration, and as a result a positive definite and well-conditioned estimator having the desired structure is determined. It is also shown that the statistical and algebraic properties of the estimator depend on the choice of target space.

Keywords Multivariate model · Covariance structure · Least squares estimation · Projection · Shrinking · Quadratic subspace

Mathematics Subject Classification 62H10 · 62H12

Augustyn Markiewicz, Adam Mieldzioc, and Malwina Mrowińska have contributed equally to this work.

Katarzyna Filipiak
katarzyna.filipiak@put.poznan.pl

Augustyn Markiewicz
augustyn.markiewicz@up.poznan.pl

Adam Mieldzioc
adam.mieldzioc@up.poznan.pl

Malwina Mrowińska
malwina.mrowinska@put.poznan.pl

¹ Institute of Mathematics, Poznan University of Technology, Poznań, Poland

² Department of Mathematical and Statistical Methods, Poznan University of Life Sciences, Poznań, Poland

1 Introduction

Estimation of the covariance matrix is one of the basic problems in multivariate analysis, as the dependence structure between variables is of interest in almost every area of science. For example, in all biological experiments, including agricultural ones, gene expressions and the correlations between genes are of particular interest. To discover possible structures of a covariance matrix, researchers often use heatmaps (see e.g. Mieldzioc et al. 2021 and Carrión-García et al. 2020), neural networks or mapping (Gilson et al. 2020), a graphical lasso algorithm (Devijver and Gallopin 2018), as well as algebraic methods (Filipiak et al. 2018) or statistical testing; the latter is used when the researcher already has knowledge about the most reasonable possible covariance structure. It should also be pointed out that in many applied problems in multivariate analysis, estimators of the covariance matrix as well as its inverse are required. When the sample size is too small relative to the number of variables, the unstructured sample covariance matrix is singular or ill-conditioned. In such a case, the underlying covariance structure allows us to reduce the dimension of the parameter space and makes the estimation effective.

Several applications in mathematical statistics and data analysis involve linear structures of the covariance matrix defined as follows; cf. Anderson (1973).

Definition 1 The covariance matrix Σ has a linear structure if it can be expressed as a linear combination of known symmetric $m \times m$ matrices $\mathbf{G}_0, \dots, \mathbf{G}_z$ that are assumed to be linearly independent, i.e.,

$$\Sigma = \sum_{i=0}^z \sigma_i \mathbf{G}_i,$$

where $\sigma_i \in \mathbb{R}$ are unknown parameters. It is also assumed that there exists at least one combination of $\sigma_0, \dots, \sigma_z$ such that Σ is a positive definite matrix.

The linear space $\mathcal{L}$ with the basis $\mathbf{G}_0, \dots, \mathbf{G}_z$ will be called a *linear structure space*.

There is a wide literature on the estimation of covariance structures belonging to quadratic subspaces (see Seely 1971 or Sect. 2 for definition), as it is known that the projection of a sample covariance matrix onto the space of covariance matrices belonging to a quadratic subspace is the best quadratic unbiased estimator of a covariance structure; cf. Koziol et al. (2017). Moreover, Filipiak et al. (2021) showed that the maximum likelihood estimator (MLE) of such a covariance matrix can be obtained as the projection of the MLE of an unstructured covariance matrix onto this quadratic space. Both estimators are positive definite, since the orthogonal projection of a positive definite matrix onto a quadratic subspace preserves positive definiteness; cf. Fuglede and Jensen (2013).

When the structure space is not quadratic, the determination of the MLE requires iterative methods (cf. Anderson 1973, Szatrowski 1980). These are usually time-consuming, and the time needed to derive the estimator increases with an increase in the dimension. Therefore, it is vital to develop fast methods for estimating structured covariance matrices.

Following Ohlson and von Rosen (2010), a computationally simpler least squares method based on the projection of an unstructured estimator onto a structure space can be used. However, this method can lead to estimators that are not positive definite (p.d.) or well-conditioned, and may even be indefinite; cf. Fuglede and Jensen (2013). The projection onto the convex cone of non-negative definite matrices from the structure space preserves nonnegative definiteness; however, this method is also time consuming; cf. Filipiak et al. (2018). An appropriate method to restore the required properties of the sample covariance matrix (estimator of the unstructured covariance matrix) is the shrinkage method; cf. Ledoit and Wolf (2004). Nevertheless, in this paper we work with structured matrices, and thus the shrinkage estimation method has to be modified.

Note that all of the papers mentioned in the previous paragraph deal with structured covariance matrices related to a vector of observations; however, in numerous experiments the observation vector is partitioned into several subvectors, related to grouping of observed variables into subgroups (for example when various kinds of variables are observed) or to vectorization of doubly multivariate data, when observations are additionally repeated at several times or under several conditions; for more details on doubly multivariate models, see, e.g., Arnold (1979), Crowder and Hand (1990) or Lu and Zimmerman (2005). Then in the structure of the covariance matrix, the partition of the observation vector should be taken into account, and thus the covariance matrix should be assumed to be partitioned into respective subblocks. Note also that in the literature on doubly multivariate data the subblocks considered are usually of equal dimensions and have a particular structure (for example the Kronecker product of two matrices or a block exchangeable structure); nevertheless, in many cases it is more reasonable to relax the equal dimensions condition, as in real-life applications the number of repetitions may change together with the observed variables, or the sizes of groups of variables may differ from each other.

For matrices partitioned into square subblocks such that they belong to a quadratic subspace, the estimation techniques are the same as in the case of non-partitioned blocks. In the case of rectangular off-diagonal blocks of a covariance matrix belonging to quadratic subspaces, Yang et al. (2024) recently observed that taking the average of the respective entries of the sample covariance matrix leads to uniformly minimum-variance unbiased estimators of a covariance matrix with block-uniform structure partitioned into four subblocks. Note that this technique is nothing else than the projection of the sample covariance matrix into the respective quadratic subspace.

Estimation of partitioned covariance structures that do not belong to a quadratic subspace and can have rectangular off-diagonal blocks has recently been studied by Markiewicz et al. (2024). However, since there the maximum likelihood estimator was of interest, the method presented can be improved. Therefore, in this paper we give an estimation technique based on another estimator of relevant structure. Following Ohlson and von Rosen (2010) we consider a least squares estimator of a structured covariance matrix, which is simply the projection of the sample covariance matrix onto the space of covariance structures, and we adapt the shrinkage method to improve this estimator such that the final estimator will be positive definite and well-conditioned and will have good statistical properties. We show that an arbitrary quadratic subspace of the space of covariance structures can be chosen as the space of target matrices (in

short: *target space*); however, this choice influences the statistical properties of the final estimator and it is therefore crucial to find a balance between the (well-)conditionality of the result and its statistical properties.

The paper is organized as follows. In Sect. 2 we present the model in which a structured partitioned covariance matrix plays a crucial role. The estimation method and examples illustrating this method for particular covariance structure spaces are given in Sect. 3. A comparison of the estimators obtained with the use of four different target spaces based on a simulation study is presented in Sect. 4. Conclusions and recommendations are given in the final section.

2 Model

Let us consider an experiment with n observations being m -dimensional random vectors, $\mathbf{x}_i$, $i = 1, \dots, n$, which are mutually independent and identically distributed with $\mathbb{E}(\mathbf{x}_i) = \boldsymbol{\mu}$, an m -dimensional mean vector, and $\mathbb{D}(\mathbf{x}_i) = \boldsymbol{\Sigma}$, a symmetric p.d. covariance matrix of order m .

Assume that each $\mathbf{x}_i$, $i = 1, \dots, n$, may consist of several subvectors, related to the grouping of observed variables or to the vectorization of doubly multivariate data. Then in the covariance matrix, the partition of the observation vector should be taken into account, and thus $\boldsymbol{\Sigma}$ should be assumed to have a partitioned structure. Since in many practical problems the dimensions of subvectors can differ from each other, we assume that the diagonal blocks of the covariance matrix are p.d. matrices of different orders, and as a consequence, the off-diagonal blocks are rectangular matrices.

For simplicity we assume that each $\mathbf{x}_i$, $i = 1, \dots, n$, consists of two subvectors, that is, $\mathbf{x}_i = (\mathbf{y}'_i, \mathbf{z}'_i)'$, where $\mathbf{y}_i$ and $\mathbf{z}_i$ are, respectively, p - and q -dimensional vectors and hence $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$, where $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ are p.d. matrices of order p and q , respectively, while $\boldsymbol{\Sigma}_{12}$ is of dimension $p \times q$. However, it is worth noting that the estimation technique presented in the next section can be generalized to larger numbers of subvectors.

Note that study of the dependence structure between two subvectors, that is the specific structure of $\boldsymbol{\Sigma}_{12}$, might also be of interest; this problem has been studied recently by, e.g., Janiszewska et al. (2020), Markiewicz et al. (2024), and Filipiak et al. (2024). However, in this paper, we assume only that the structure of $\boldsymbol{\Sigma}$ and its subblocks is linear, since linear structures have numerous applications in mathematical statistics and data analysis. Therefore, without loss of generality, we assume that the matrix $\boldsymbol{\Sigma}$ belongs to a linear space $\mathcal{S}$ which can be decomposed into the sum of orthogonal subspaces, i.e.,

$$\mathcal{S} = \mathcal{S}_1 \boxplus \mathcal{S}_2 \boxplus \mathcal{S}_3, \quad (1)$$

where

$$\begin{aligned}\mathcal{S}_1 &= \left\{ \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \in \mathbb{R}_m^s : \mathbf{A} \in \mathcal{A} \subseteq \mathbb{R}_p^s \right\}, \\ \mathcal{S}_2 &= \left\{ \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix} \in \mathbb{R}_m^s : \mathbf{B} \in \mathcal{B} \subseteq \mathbb{R}_q^s \right\}, \\ \mathcal{S}_3 &= \left\{ \begin{pmatrix} \mathbf{0} & \mathbf{C} \\ \mathbf{C}' & \mathbf{0} \end{pmatrix} \in \mathbb{R}_m^s : \mathbf{C} \in \mathcal{C} \subseteq \mathbb{R}_{p \times q} \right\},\end{aligned}\quad (2)$$

and $\mathcal{A}$, $\mathcal{B}$ are linear subspaces of the spaces of real symmetric matrices of order p , $\mathbb{R}_p^s$, and q , $\mathbb{R}_q^s$, respectively, and $\mathcal{C}$ is a linear subspace of the space of $p \times q$ real matrices, $\mathbb{R}_{p \times q}$.

The estimate of $\boldsymbol{\Sigma} \in \mathcal{S}$ is required to have the assumed structure, which can be obtained by projection of the sample covariance matrix $\mathbf{S} = \frac{1}{n-1} \mathbf{X} (\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n') \mathbf{X}'$ onto $\mathcal{S}$, where $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ is an $m \times n$ observation matrix, $\mathbf{I}_n$ is the identity matrix of order n , and $\mathbf{1}_n$ is an n -dimensional vector of ones; cf. Ohlson and von Rosen (2010). Note that the projection is simply the minimizer of the respective Frobenius norm, i.e.,

$$\mathbf{P}_{\mathcal{S}}(\mathbf{S}) = \operatorname{argmin}_{\mathbf{V} \in \mathcal{S}} \|\mathbf{V} - \mathbf{S}\|^2,$$

where $\|\mathbf{A}\|^2 = \operatorname{tr}(\mathbf{A}\mathbf{A}')$. If $\{\mathbf{V}_1, \dots, \mathbf{V}_v\}$ is an orthonormal basis of $\mathcal{S}$, then the projection can also be expressed as

$$\mathbf{P}_{\mathcal{S}}(\mathbf{S}) = \sum_{i=1}^v \operatorname{tr}(\mathbf{V}_i \mathbf{S}) \mathbf{V}_i. \quad (3)$$

Filipiak et al. (2024) considered a particular linear space which is quadratic. Following (Seely 1971) a space $\mathcal{S}$ is a quadratic subspace of a linear space if for an arbitrary matrix $\mathbf{T} \in \mathcal{S}$ also $\mathbf{T}^2 \in \mathcal{S}$. Assuming $p = 3$ and $q = 4$, examples of quadratic subspaces $\mathcal{S}$ of linearly structured partitioned matrices of the type under consideration are

$$\left\{ \left(\begin{array}{ccc|cccc} a_0 & a & a & 0 & 0 & 0 & 0 \\ a & a_0 & a & 0 & 0 & 0 & 0 \\ a & a & a_0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & b_0 & b & b & b \\ 0 & 0 & 0 & b & b_0 & b & b \\ 0 & 0 & 0 & b & b & b_0 & b \\ 0 & 0 & 0 & b & b & b & b_0 \end{array} \right) : a_0, a, b_0, b \in \mathbb{R} \right\},$$

$$\left\{ \left(\begin{array}{ccc|cccc} a_0 & a & a & c & c & c & c \\ a & a_0 & a & c & c & c & c \\ a & a & a_0 & c & c & c & c \\ \hline c & c & c & b_0 & b & b & b \\ c & c & c & b & b_0 & b & b \\ c & c & c & b & b & b_0 & b \\ c & c & c & b & b & b & b_0 \end{array} \right) : a_0, a, b_0, b, c \in \mathbb{R} \right\},$$

and

$$\left\{ \left(\begin{array}{ccc|cccc} a_0 & a & a & \sqrt{3}c & c - \sqrt{3}c & -c \\ a & a_0 & a & 0 & -2c & 0 & 2c \\ a & a & a_0 & -\sqrt{3}c & c & \sqrt{3}c & -c \\ \hline \sqrt{3}c & 0 & -\sqrt{3}c & b_0 & b_1 & b_2 & b_1 \\ c & -2c & c & b_1 & b_0 & b_1 & b_2 \\ -\sqrt{3}c & 0 & \sqrt{3}c & b_2 & b_1 & b_0 & b_1 \\ -c & 2c & -c & b_1 & b_2 & b_1 & b_0 \end{array} \right) : a_0, a, b_0, b_1, b_2, c \in \mathbb{R} \right\},$$

to mention a few.

The advantage of considering a quadratic space is that the best unbiased estimator of $\Sigma \in \mathcal{S}$ is the projection of the sample covariance matrix onto this space; cf. Koziol et al. (2017), Filipiak et al. (2021). Moreover, under the normality assumption, the maximum likelihood estimator of $\Sigma \in \mathcal{S}$ is the projection of $\mathbf{S}_{\text{ML}} = \frac{1}{n} \mathbf{X} (\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n') \mathbf{X}'$, onto this space, that is, $\mathbf{P}_{\mathcal{S}}(\mathbf{S}_{\text{ML}})$; cf. Filipiak et al. (2021). It should also be mentioned that the projection onto the quadratic subspace preserves definiteness (cf. Fuglede and Jensen 2013), and hence the estimates obtained are positive definite (for $n > m$).

In the case of a non-quadratic linear structure space, the MLE is not given in an explicit form and its determination is challenging and time-consuming; cf. Filipiak et al. (2021), Markiewicz et al. (2024). In the case of unbiased estimation, Ohlson and von Rosen (2010) studied the least squares method. This approach provides an explicit, unbiased and consistent estimator, which is the projection of the sample covariance matrix $\mathbf{S}$ onto the linear space $\mathcal{S}$ given in (2). However, if the structure space is not quadratic, then the projection does not preserve definiteness; cf. Fuglede and Jensen (2013). The estimator obtained by projection may be positive definite, but also singular or even indefinite; see the example in Markiewicz et al. (2024). Therefore, to restore the positive definiteness of the covariance matrix estimator, a method based on shrinking can be used. A shrinkage estimation method for an unstructured covariance matrix was

proposed by Ledoit and Wolf (2004); subsequently Markiewicz et al. (2024) proposed its modification for structured matrices. Recall that the original shrinkage estimator is defined as a convex combination of some target matrix and the estimator of an unstructured matrix (usually $\mathbf{S}$ or $\mathbf{S}_{\text{ML}}$), and the aim of this procedure is to improve the properties of the estimator under consideration. In the literature, commonly used target matrices are matrices proportional to the identity matrix or diagonal matrices; cf. Hannart and Naveau (2014), Ikeda et al. (2016). The target matrix and shrinkage intensity are chosen using the optimization criterion given by Ledoit and Wolf (2004). However, it should be noted that when structured covariance matrices are considered, this procedure does not produce a structured estimator. Therefore, the proposed modification involves replacing the estimator of a structured matrix by a relevant estimator having the required structure, whose properties then have to be improved. Moreover, the target matrix is not arbitrary but must belong to a subspace of the space of the considered structure; note that the most common target matrices—matrices proportional to identity and diagonal matrices—belong to the quadratic subspace.

The detailed procedure for determining the positive definite estimator of Σ is given in the next section.

3 Estimation of covariance matrix

As suggested by Ohlson and von Rosen (2010), we consider the least squares estimator of a linearly structured partitioned Σ , which is the projection of a sample covariance matrix $\mathbf{S}$ onto the space $\mathcal{S}$, that is,

$$\widehat{\Sigma}_{\text{LS}} = \mathbf{P}_{\mathcal{S}}(\mathbf{S}),$$

where the projection is as defined in (3). To improve the definiteness of $\widehat{\Sigma}_{\text{LS}}$, the structured shrinkage method will be used. This means that the improved estimator of Σ , denoted by $\widehat{\Sigma}$, can be represented as

$$\widehat{\Sigma} = \lambda \mathbf{T} + (1 - \lambda) \widehat{\Sigma}_{\text{LS}}, \quad (4)$$

where $\mathbf{T}$ is a positive definite target matrix and $\lambda \in [0, 1]$ is a shrinkage intensity. Moreover, since $\widehat{\Sigma}$ must have the same structure as $\widehat{\Sigma}_{\text{LS}}$, the target matrix must be chosen from a subspace of $\mathcal{S}$, say $\mathcal{T} \subseteq \mathcal{S}$.

To determine the optimal shrinkage intensity λ and the optimal form of the target matrix $\mathbf{T} \in \mathcal{T}$, following Ledoit and Wolf (2004), the following quadratic risk function will be minimized:

$$f(\mathbf{T}, \lambda) = \mathbb{E} [\|\widehat{\Sigma} - \Sigma\|^2] \quad (5)$$

subject to $\widehat{\Sigma} \in \mathcal{S}$ as given in (4).

We can now prove the following theorem, which is both, a modification and extension of Theorem 2.1 of Ledoit and Wolf (2004), as it makes it possible to determine the structured estimator of the covariance matrix and to relax the assumption that

the target matrix is proportional to identity. Clearly, Theorem 2.1 in Ledoit and Wolf (2004) is a special case of the following theorem.

Theorem 1 *Let $\mathbf{S}$ be an $m \times m$ sample covariance matrix obtained from a random sample of size n from the distribution with mean μ and covariance matrix $\Sigma \in \mathcal{S}$. The minimum of the quadratic risk, i.e.,*

$$\min_{\mathbf{T} \in \mathcal{T} \subseteq \mathcal{S}, \lambda \in \mathbb{R}} \mathbb{E} \left[\|\widehat{\Sigma} - \Sigma\|^2 \right] \quad \text{subject to} \quad \widehat{\Sigma} = \lambda \mathbf{T} + (1 - \lambda) \mathbf{P}_{\mathcal{S}}(\mathbf{S}) \in \mathcal{S}$$

is attained at

$$\mathbf{T} = \mathbf{P}_{\mathcal{T}}(\Sigma), \quad \lambda = \frac{\mathbb{E}[\operatorname{tr}(\mathbf{P}_{\mathcal{S}}^2(\mathbf{S}))] - \|\mathbf{P}_{\mathcal{S}}(\Sigma)\|^2}{\mathbb{E}[\operatorname{tr}(\mathbf{P}_{\mathcal{S}}^2(\mathbf{S}))] - \|\mathbf{P}_{\mathcal{T}}(\Sigma)\|^2}.$$

Moreover, $\lambda \in [0, 1]$.

Proof Let $\{\mathbf{V}_1, \dots, \mathbf{V}_v\}$, $\mathbf{V}_i \in \mathbb{R}_{m \times m}$, be an orthonormal basis of the space $\mathcal{S}$ of dimension v . Without loss of generality, we assume that the basis of $\mathcal{T} \subseteq \mathcal{S}$ consists of the first $k \leq v$ basis matrices. Then the target matrix $\mathbf{T} \in \mathcal{T}$ can be expressed as $\mathbf{T} = \sum_{i=1}^k \alpha_i \mathbf{V}_i$, and the quadratic risk given in (5) is a function of α_i , $i = 1, \dots, k$, and $\lambda \in \mathbb{R}$, i.e.,

$$\begin{aligned} f(\mathbf{T}, \lambda) &= f(\alpha_1, \dots, \alpha_k, \lambda) \\ &= \lambda^2 \operatorname{tr}(\mathbf{T}^2) + 2\lambda(1 - \lambda) \operatorname{tr}(\mathbf{T} \mathbb{E}[\mathbf{P}_{\mathcal{S}}(\mathbf{S})]) + (1 - \lambda)^2 \operatorname{tr}(\mathbb{E}[\mathbf{P}_{\mathcal{S}}^2(\mathbf{S})]) \\ &\quad - 2\lambda \operatorname{tr}(\mathbf{T} \Sigma) - 2(1 - \lambda) \operatorname{tr}(\Sigma \mathbb{E}[\mathbf{P}_{\mathcal{S}}(\mathbf{S})]) + \operatorname{tr}(\Sigma^2). \end{aligned}$$

Since it is a convex function, its minimum is attained at the stationary point. Differentiation of $f(\mathbf{T}, \lambda)$ gives the following system of equations

$$\begin{aligned} \lambda \alpha_i + (1 - \lambda) \mathbb{E}[\operatorname{tr}(\mathbf{V}_i \mathbf{P}_{\mathcal{S}}(\mathbf{S}))] - \operatorname{tr}(\mathbf{V}_i \Sigma) &= 0, \\ \lambda \operatorname{tr}(\mathbf{T}^2) + (1 - 2\lambda) \mathbb{E}[\operatorname{tr}(\mathbf{T} \mathbf{P}_{\mathcal{S}}(\mathbf{S}))] - (1 - \lambda) \mathbb{E}[\operatorname{tr}(\mathbf{P}_{\mathcal{S}}^2(\mathbf{S}))] \\ - \operatorname{tr}(\mathbf{T} \Sigma) + \mathbb{E}[\operatorname{tr}(\mathbf{P}_{\mathcal{S}}(\mathbf{S}) \Sigma)] &= 0. \end{aligned} \quad (6)$$

Since $\mathbb{E}(\mathbf{S}) = \Sigma$ we obtain that

$$\mathbb{E}[\mathbf{P}_{\mathcal{S}}(\mathbf{S})] = \sum_{i=1}^v \operatorname{tr}(\mathbf{V}_i \mathbb{E}[\mathbf{S}]) \mathbf{V}_i = \sum_{i=1}^v \operatorname{tr}(\mathbf{V}_i \Sigma) \mathbf{V}_i = \mathbf{P}_{\mathcal{S}}(\Sigma) = \Sigma, \quad (7)$$

and hence

$$\mathbb{E}[\operatorname{tr}(\mathbf{V}_i \mathbf{P}_{\mathcal{S}}(\mathbf{S}))] = \operatorname{tr}(\mathbf{V}_i \Sigma), \quad \mathbb{E}[\operatorname{tr}(\mathbf{T} \mathbf{P}_{\mathcal{S}}(\mathbf{S}))] = \operatorname{tr}(\mathbf{T} \Sigma), \quad \mathbb{E}[\operatorname{tr}(\mathbf{P}_{\mathcal{S}}(\mathbf{S}) \Sigma)] = \operatorname{tr}(\Sigma^2).$$

Applying respective expectation to the first equation of (6) we obtain $\alpha_i = \operatorname{tr}(\mathbf{V}_i \Sigma)$, $i = 1, \dots, k$, and hence $\mathbf{T} = \sum_{i=1}^k \operatorname{tr}(\mathbf{V}_i \Sigma) \mathbf{V}_i = \mathbf{P}_{\mathcal{T}}(\Sigma)$. Substituting this result into

the second equation of (6) gives

$$\lambda = \frac{\mathbb{E}[\operatorname{tr}(\mathbf{P}_{\mathcal{S}}^2(\mathbf{S}))] - \sum_{i=1}^v \operatorname{tr}^2(\mathbf{V}_i \boldsymbol{\Sigma})}{\mathbb{E}[\operatorname{tr}(\mathbf{P}_{\mathcal{S}}^2(\mathbf{S}))] - \sum_{i=1}^k \operatorname{tr}^2(\mathbf{V}_i \boldsymbol{\Sigma})}.$$

Since $\sum_{i=1}^v \operatorname{tr}^2(\mathbf{V}_i \boldsymbol{\Sigma}) = \|\mathbf{P}_{\mathcal{S}}(\boldsymbol{\Sigma})\|^2$ and $\sum_{i=1}^k \operatorname{tr}^2(\mathbf{V}_i \boldsymbol{\Sigma}) = \|\mathbf{P}_{\mathcal{T}}(\boldsymbol{\Sigma})\|^2$, we obtain the form of λ given in the statement of the theorem. Moreover, because $\mathbf{T} \in \mathcal{T} \subseteq \mathcal{S}$ and $\mathbf{P}_{\mathcal{S}}(\mathbf{S}) \in \mathcal{S}$, also $\widehat{\boldsymbol{\Sigma}} = \lambda \mathbf{T} + (1 - \lambda) \mathbf{P}_{\mathcal{S}}(\mathbf{S})$ belongs to $\mathcal{S}$.

It is enough to show now that $\lambda \in [0, 1]$. From the orthonormality of $\mathbf{V}_i$, $i = 1, \dots, v$, we have

$$\mathbb{E}[\operatorname{tr}(\mathbf{P}_{\mathcal{S}}^2(\mathbf{S}))] = \mathbb{E}\left[\operatorname{tr}\left(\sum_{i=1}^v \operatorname{tr}(\mathbf{V}_i \mathbf{S}) \mathbf{V}_i \cdot \sum_{j=1}^v \operatorname{tr}(\mathbf{V}_j \mathbf{S}) \mathbf{V}_j\right)\right] = \mathbb{E}\left[\sum_{i=1}^v \operatorname{tr}^2(\mathbf{V}_i \mathbf{S})\right], \quad (8)$$

and from Jensen's inequality, since the function under the expectation is convex,

$$\mathbb{E}\left[\sum_{i=1}^v \operatorname{tr}^2(\mathbf{V}_i \mathbf{S})\right] \geq \sum_{i=1}^v \operatorname{tr}^2(\mathbf{V}_i \mathbb{E}[\mathbf{S}]) = \sum_{i=1}^v \operatorname{tr}^2(\mathbf{V}_i \boldsymbol{\Sigma}) = \|\mathbf{P}_{\mathcal{S}}(\boldsymbol{\Sigma})\|^2.$$

Thus, the nominator in λ is nonnegative. Moreover, since in the nominator we subtract the norm of the projection onto $\mathcal{S}$, while in the denominator we subtract the norm of the projection onto $\mathcal{T} \subseteq \mathcal{S}$, we have $\lambda \in [0, 1]$. $\square$

It is worth mentioning here that if $\mathcal{T} = \mathcal{S}$, then in the above theorem $\lambda = 1$ and hence $\widehat{\boldsymbol{\Sigma}} = \mathbf{P}_{\mathcal{S}}(\mathbf{S})$, which does not need to be positive definite (or well-conditioned) – unless $\mathcal{S}$ is a quadratic space, since projection of a p.d. matrix onto a quadratic space preserves definiteness; cf. Fuglede and Jensen (2013).

To avoid the problem of the unknown expectation of $\operatorname{tr}(\mathbf{P}_{\mathcal{S}}^2(\mathbf{S}))$ in Theorem 1, from now on we will assume the normality of the vector of observations. Therefore, assume that $\mathbf{S} = \frac{1}{n-1} \mathbf{X} (\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n') \mathbf{X}'$, where $\mathbf{X}$ is a normally distributed matrix of observations, i.e.,

$$\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n) \sim N_{m,n}(\mu \mathbf{1}_n', \boldsymbol{\Sigma}, \mathbf{I}_n).$$

Then we can compute the expectation from the formula for optimal λ and thus prove the following.

Proposition 2 *Let $\mathbf{S}$ be the $m \times m$ sample covariance matrix obtained from a random sample of size n from the multivariate normal distribution with mean μ and covariance*

matrix $\Sigma \in \mathcal{S}$. For $\mathcal{T} \subseteq \mathcal{S}$ the minimum of the quadratic risk, i.e.,

$$\min_{\lambda \in \mathbb{R}} \mathbb{E} \left[\|\widehat{\Sigma} - \Sigma\|^2 \right] \quad \text{subject to} \quad \widehat{\Sigma} = \lambda \mathbf{P}_{\mathcal{T}}(\Sigma) + (1 - \lambda) \mathbf{P}_{\mathcal{S}}(\mathbf{S}) \in \mathcal{S}$$

is attained at

$$\lambda = \frac{2 \sum_{i=1}^v \text{tr}((\mathbf{V}_i \Sigma)^2)}{2 \sum_{i=1}^v \text{tr}((\mathbf{V}_i \Sigma)^2) + (n-1)(\|\mathbf{P}_{\mathcal{S}}(\Sigma)\|^2 - \|\mathbf{P}_{\mathcal{T}}(\Sigma)\|^2)},$$

where v is the dimension of $\mathcal{S}$.

Proof Under the normality assumption, matrix $\mathbf{S}$ has a Wishart distribution with m degrees of freedom, i.e.,

$$\mathbf{S} \sim W_m \left(n-1, \frac{1}{n-1} \Sigma \right).$$

Therefore, since (8) holds, we may use the formula

$$\mathbb{E} [\text{tr}(\mathbf{A}\mathbf{W}) \text{tr}(\mathbf{B}\mathbf{W})] = n \text{tr}(\mathbf{A}\mathbf{\Omega}\mathbf{B}\mathbf{\Omega}) + n \text{tr}(\mathbf{A}'\mathbf{\Omega}\mathbf{B}\mathbf{\Omega}) + n^2 \text{tr}(\mathbf{A}\mathbf{\Omega}) \text{tr}(\mathbf{B}\mathbf{\Omega})$$

for $\mathbf{W} \sim W_q(n, \mathbf{\Omega})$, $q < n$ (cf. Gupta and Nagar 2000, Theorem 3.3.15), and then

$$\mathbb{E} [\text{tr}(\mathbf{P}_{\mathcal{S}}^2(\mathbf{S}))] = \frac{2}{n-1} \sum_{i=1}^v \text{tr}((\mathbf{V}_i \Sigma)^2) + \sum_{i=1}^v \text{tr}^2(\mathbf{V}_i \Sigma).$$

Substituting this expectation into the formula for λ given in Theorem 1 and observing that $\sum_{i=1}^v \text{tr}^2(\mathbf{V}_i \Sigma) - \sum_{i=1}^k \text{tr}^2(\mathbf{V}_i \Sigma) = \|\mathbf{P}_{\mathcal{S}}(\Sigma)\|^2 - \|\mathbf{P}_{\mathcal{T}}(\Sigma)\|^2$, we obtain the form of λ given in the proposition. $\square$

Note that both the target matrix $\mathbf{P}_{\mathcal{T}}(\Sigma)$ and the optimal λ depend on unknown Σ . Therefore, to obtain a $\widehat{\Sigma}$ in the form (4) which is an estimator of Σ , we replace Σ by the sample covariance matrix $\mathbf{S}$. Then

$$\widehat{\Sigma} = \widehat{\lambda} \mathbf{P}_{\mathcal{T}}(\mathbf{S}) + (1 - \widehat{\lambda}) \mathbf{P}_{\mathcal{S}}(\mathbf{S}) \quad (9)$$

with

$$\widehat{\lambda} = \frac{2 \sum_{i=1}^v \text{tr}[(\mathbf{V}_i \mathbf{S})^2]}{2 \sum_{i=1}^v \text{tr}[(\mathbf{V}_i \mathbf{S})^2] + (n-1)(\|\mathbf{P}_{\mathcal{S}}(\mathbf{S})\|^2 - \|\mathbf{P}_{\mathcal{T}}(\mathbf{S})\|^2)} \quad (10)$$

will be called a *structured shrinkage estimator* of Σ .

In the following result, we show the statistical properties of the estimator given in (9) with $\hat{\lambda}$ as in (10).

Theorem 3 *Let $\hat{\Sigma}$ as given in (9), with $\mathcal{T} \subseteq \mathcal{S}$ and $\hat{\lambda}$ as defined in (10), be the estimator of $\Sigma \in \mathcal{S}$ under normality. Then it is consistent and asymptotically unbiased. Moreover, if $\mathcal{T} = \mathcal{S}$, then $\hat{\Sigma}$ is an unbiased estimator of Σ .*

Proof Consistency.

Let us consider the sample covariance matrix $\mathbf{S}$ as a function of sample size, n , and let us denote $\mathbf{S}(n)$ by $\mathbf{S}_n$. It is known that it is a positive semi-definite matrix and a consistent estimator of the covariance matrix Σ , that is $\mathbf{S}_n \xrightarrow[n \rightarrow \infty]{P} \Sigma$, where $\xrightarrow{P}$ denotes convergence in probability [cf. Ohlson and von Rosen (2010)]. Therefore $\mathbf{P}_{\mathcal{T}}(\mathbf{S}_n) \xrightarrow[n \rightarrow \infty]{P} \mathbf{P}_{\mathcal{T}}(\Sigma)$ and $\mathbf{P}_{\mathcal{S}}(\mathbf{S}_n) \xrightarrow[n \rightarrow \infty]{P} \mathbf{P}_{\mathcal{S}}(\Sigma) = \Sigma$ (the latter equality follows from the structure of $\Sigma \in \mathcal{S}$), and $\|\mathbf{P}_{\mathcal{S}}(\Sigma)\|^2 < \infty$, $\|\mathbf{P}_{\mathcal{T}}(\Sigma)\|^2 < \infty$.

Assume first that $\Sigma \in \mathcal{T} \subseteq \mathcal{S}$. Then $\hat{\Sigma} = \mathbf{P}_{\mathcal{T}}(\mathbf{S}_n) \xrightarrow[n \rightarrow \infty]{P} \mathbf{P}_{\mathcal{T}}(\Sigma) = \Sigma$ and hence $\hat{\Sigma}$ is a consistent estimator of Σ .

Assume now that $\Sigma \notin \mathcal{T}$. This implies that

$$\|\mathbf{P}_{\mathcal{S}}(\mathbf{S}_n)\|^2 - \|\mathbf{P}_{\mathcal{T}}(\mathbf{S}_n)\|^2 \xrightarrow[n \rightarrow \infty]{P} \|\mathbf{P}_{\mathcal{S}}(\Sigma)\|^2 - \|\mathbf{P}_{\mathcal{T}}(\Sigma)\|^2 > 0.$$

Moreover, $\text{tr}((\mathbf{V}_i \mathbf{S}_n)^2) \xrightarrow[n \rightarrow \infty]{P} \text{tr}((\mathbf{V}_i \Sigma)^2) < \infty$. Hence,

$$\hat{\lambda}_n = \frac{2 \sum_{i=1}^v \text{tr}[(\mathbf{V}_i \mathbf{S}_n)^2]}{2 \sum_{i=1}^v \text{tr}[(\mathbf{V}_i \mathbf{S}_n)^2] + (n-1)(\|\mathbf{P}_{\mathcal{S}}(\mathbf{S}_n)\|^2 - \|\mathbf{P}_{\mathcal{T}}(\mathbf{S}_n)\|^2)} \xrightarrow[n \rightarrow \infty]{P} 0 \quad (11)$$

and the consistency of $\hat{\Sigma}$ follows.

Unbiasedness.

Let $\mathcal{T} = \mathcal{S}$. Then $\hat{\Sigma} = \mathbf{P}_{\mathcal{S}}(\mathbf{S}_n)$ and unbiasedness follows from (7).

Let $\mathcal{T} \subset \mathcal{S}$. Note that

$$\mathbf{P}_{\mathcal{S}}(\mathbf{S}_n) = \sum_{i=1}^k \text{tr}(\mathbf{V}_i \mathbf{S}_n) \mathbf{V}_i + \sum_{i=k+1}^v \text{tr}(\mathbf{V}_i \mathbf{S}_n) \mathbf{V}_i = \mathbf{P}_{\mathcal{T}}(\mathbf{S}_n) + \sum_{i=k+1}^v \text{tr}(\mathbf{V}_i \mathbf{S}_n) \mathbf{V}_i,$$

and hence

$$\lim_{n \rightarrow \infty} \mathbb{E}[\hat{\Sigma}] = \mathbb{E} \left[\lim_{n \rightarrow \infty} \sum_{i=1}^k \text{tr}(\mathbf{V}_i \mathbf{S}_n) \mathbf{V}_i + \lim_{n \rightarrow \infty} (1 - \hat{\lambda}_n) \lim_{n \rightarrow \infty} \sum_{i=k+1}^v \text{tr}(\mathbf{V}_i \mathbf{S}_n) \mathbf{V}_i \right].$$

Moreover, from (11) we can further write

$$\begin{aligned}\mathbb{E}[\widehat{\Sigma}] &\xrightarrow[n \rightarrow \infty]{P} \mathbb{E}\left[\lim_{n \rightarrow \infty} \sum_{i=1}^k \text{tr}(\mathbf{V}_i \mathbf{S}_n) \mathbf{V}_i + \lim_{n \rightarrow \infty} \sum_{i=k+1}^v \text{tr}(\mathbf{V}_i \mathbf{S}_n) \mathbf{V}_i\right] \\ &= \mathbb{E}\left[\lim_{n \rightarrow \infty} \sum_{i=1}^v \text{tr}(\mathbf{V}_i \mathbf{S}_n) \mathbf{V}_i\right] = \lim_{n \rightarrow \infty} \mathbb{E}[\mathbf{P}_{\mathcal{S}}(\mathbf{S}_n)]\end{aligned}$$

and asymptotic unbiasedness follows from (7). $\square$

It can easily be seen from (9) that to achieve the positive definiteness of $\widehat{\Sigma}$ or to improve the conditioning of the positive definite $\mathbf{P}_{\mathcal{S}}(\mathbf{S})$, the space $\mathcal{T}$ should have strong algebraic properties. Recall that a matrix is ill-conditioned when its condition number, which is simply the ratio between the largest and the smallest eigenvalue, is higher than some predefined κ_0 , depending, among other things, on the matrix dimension. Therefore, we prove the following lemma.

Lemma 1 *Let $\widehat{\Sigma}$ as given in (9) be the estimator of Σ . A necessary condition to achieve positive definiteness or improve the conditioning of $\widehat{\Sigma}$ is for $\mathcal{T}$ to be a quadratic subspace.*

Proof Assume that $\mathbf{P}_{\mathcal{S}}(\mathbf{S})$ is indefinite. To achieve positive definiteness of $\widehat{\Sigma}$, the target matrix $\mathbf{P}_{\mathcal{T}}(\mathbf{S})$ must be positive definite. Since a projection preserves definiteness if and only if the space onto which we project is quadratic (cf. Fuglede and Jensen 2013, Theorem 1), the claim follows.

Assume that $\mathbf{P}_{\mathcal{S}}(\mathbf{S})$ is p.d. and is ill-conditioned. Since the condition number of the projection onto the quadratic space is not lower than the condition number of the projected matrix (Filipiak et al. 2024, Proposition 2.2), a necessary condition to improve the conditioning of $\widehat{\Sigma}$ is for $\mathcal{T}$ to be a quadratic subspace of $\mathcal{S}$. $\square$

It is worth mentioning here that if $\mathcal{T} = \mathcal{S}$ and $\mathcal{S}$ is a quadratic space, then $\widehat{\Sigma} = \mathbf{P}_{\mathcal{T}}(\mathbf{S})$ is the best unbiased quadratic estimator of the covariance structure Σ ; cf. Koziol et al. (2017). Note that the structure space contains at least one quadratic subspace, namely the space of matrices proportional to the identity matrix. However, there may be more quadratic subspaces available (for example diagonal matrices, block-diagonal matrices, a quadratic space of block matrices with nonzero off-diagonal blocks—see examples in Sect. 2); a general characterization of quadratic subspaces of partitioned, linearly structured matrices is given in Filipiak et al. (2024).

It is worth recalling that in the original shrinking method proposed by Ledoit and Wolf (2004) a matrix proportional to the identity matrix was chosen as the target matrix. Note that the space of matrices proportional to identity is the smallest quadratic space. Moreover, it is a subset of every $\mathcal{S}$, and thus such a choice of target space can be seen as a special case of $\mathcal{T}$ as given in this paper.

The main goal of the estimation method presented was to achieve positive definiteness of $\widehat{\Sigma}$; note, however, that even if $\mathcal{T}$ is a quadratic subspace of $\mathcal{S}$, the estimate obtained may still be indefinite or ill-conditioned (in Lemma 1, the quadraticity of $\mathcal{T}$ is only a necessary condition). Since λ is a shrinkage intensity, its estimate $\widehat{\lambda}$ must belong to the interval $[0, 1]$. If $\widehat{\lambda} = 1$, then $\widehat{\Sigma} = \mathbf{P}_{\mathcal{T}}(\mathbf{S})$, which is a p.d. matrix with a condition

number that does not exceed the condition number of $\mathbf{S}$. If $\hat{\lambda}$ is relatively small, it may be that the indefiniteness of $\mathbf{P}_{\mathcal{S}}(\mathbf{S})$ significantly influences the definiteness of $\hat{\Sigma}$. In such cases, we propose to rescale $\hat{\lambda}$ by $(1 + \eta)$, where η is chosen as small as possible to guarantee the positive definiteness and well-conditioning of the structured estimate of Σ . This modification improves $\hat{\Sigma}$ while maintaining the balance between the p.d. $\mathbf{P}_{\mathcal{T}}(\mathbf{S})$ and indefinite $\mathbf{P}_{\mathcal{S}}(\mathbf{S})$. Summing up, for $\mathcal{T}$ being a quadratic subspace of $\mathcal{S}$, the final estimate has the form

$$\tilde{\Sigma} = \tilde{\lambda} \mathbf{P}_{\mathcal{T}}(\mathbf{S}) + (1 - \tilde{\lambda}) \mathbf{P}_{\mathcal{S}}(\mathbf{S}) \quad (12)$$

with $\tilde{\lambda} = \hat{\lambda}$ if $\hat{\Sigma} = \hat{\lambda} \mathbf{P}_{\mathcal{T}}(\mathbf{S}) + (1 - \hat{\lambda}) \mathbf{P}_{\mathcal{S}}(\mathbf{S})$ is a well-conditioned p.d. matrix and $\tilde{\lambda} = (1 + \eta) \hat{\lambda}$, $\eta \in (0, 1)$, otherwise. In our simulation studies presented in Sect. 4, we are looking for $\eta = u \cdot 10^{-3}$ with the smallest integer u that guarantees the required maximal condition number ($\kappa_0 = 300$). It has been observed that this additional adjustment depends on the condition number of the true Σ : a higher $\kappa(\Sigma)$ implies more cases when additional improvement is applied.

In the following two examples we illustrate the proposed estimation technique with the use of four quadratic subspaces. The same examples will be used in simulation studies in Sect. 4. Note that the second example is based on the real data dependence structure.

Example 1 Let us consider a vector of observations which is divided into two subvectors of dimensions p and q respectively. Such a situation may occur, for example, in experiments where two types of variables are observed. It is not necessary to have independent subvectors and thus the off-diagonal block of partitioned Σ may be nonzero. Assume, moreover, that the dependence between the components of each subvector is the same for each pair; that is, the off-diagonal block is proportional to $\mathbf{1}_p \mathbf{1}_q'$. Additionally, assume that within each group the dependence between every pair of components is the same. Such a structure is illustrated in Fig. 1 (first column), and the space of such structures will be denoted by $\mathcal{S}_{(a)}$. Note, however, that a stronger dependence between “nearest neighbors” is a reasonable assumption which may appear in many applications. Then, the covariance structure can be represented as in Fig. 1 (second column) and the space of such structures will be denoted by $\mathcal{S}_{(b)}$. Moreover, a stronger dependence between the last variable of the first subvector and the first variable of the second vector (see Fig. 1, third column) may be of interest, and the space of such structures will be denoted by $\mathcal{S}$. Note also that $\mathcal{S}_{(a)} \subset \mathcal{S}_{(b)} \subset \mathcal{S}$.

Due to (1), the spaces under consideration can be expressed as the sum of orthogonal subspaces as given in (2) with

$$\begin{aligned} \mathcal{A} &= \left\{ a_0 \mathbf{I}_p + a_1 \mathbf{H}_{1,p} + a \sum_{i=2}^{p-1} \mathbf{H}_{i,p} : a_0, a_1, a \in \mathbb{R} \right\}, \\ \mathcal{B} &= \left\{ b_0 \mathbf{I}_q + b_1 \mathbf{H}_{1,q} + b \sum_{i=2}^{q-1} \mathbf{H}_{i,q} : b_0, b_1, b \in \mathbb{R} \right\}, \\ \mathcal{C} &= \left\{ c \mathbf{1}_p \mathbf{1}_q' + (c_1 - c) \mathbf{e}_{p,p} \mathbf{e}_{1,q}' : c, c_1 \in \mathbb{R} \right\}, \end{aligned} \quad (13)$$

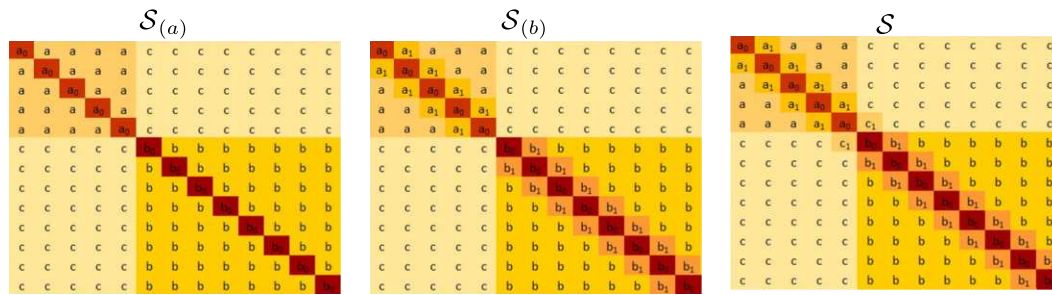


Fig. 1 Visualization of covariance structures: $\mathcal{S}_{(a)}$ the same dependence between the pairs of components of subvectors and the same dependence between components of different subvectors; $\mathcal{S}_{(b)}$ additional noise on the first sub- and super-diagonals within each block; $\mathcal{S}$ additional noise on the first sub- and super-diagonals of the whole covariance matrix

where $\mathbf{H}_{i,m}$ is an $m \times m$ binary matrix with ones on the i -th super-diagonal and sub-diagonal, and $\mathbf{e}_{u,m}$ is the u th column of $\mathbf{I}_m$. Obviously, $c_1 = c$ in spaces $\mathcal{S}_{(a)}$ and $\mathcal{S}_{(b)}$, and $a_1 = a$, $b_1 = b$ in space $\mathcal{S}_{(a)}$. The basis of $\mathcal{S}$ consists of 8 orthonormal matrices, $\{\mathbf{V}_1, \dots, \mathbf{V}_8\}$, in particular

$$\begin{aligned}
 \mathbf{V}_1 &= \begin{pmatrix} \mathbf{Q}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, & \mathbf{Q}_1 &= \frac{1}{p} \mathbf{1}_p \mathbf{1}_p', \\
 \mathbf{V}_2 &= \begin{pmatrix} \mathbf{Q}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, & \mathbf{Q}_2 &= \frac{1}{\sqrt{p-1}} \left(\mathbf{I}_p - \frac{1}{p} \mathbf{1}_p \mathbf{1}_p' \right), \\
 \mathbf{V}_3 &= \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_3 \end{pmatrix}, & \mathbf{Q}_3 &= \frac{1}{q} \mathbf{1}_q \mathbf{1}_q', \\
 \mathbf{V}_4 &= \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_4 \end{pmatrix}, & \mathbf{Q}_4 &= \frac{1}{\sqrt{q-1}} \left(\mathbf{I}_q - \frac{1}{q} \mathbf{1}_q \mathbf{1}_q' \right), \\
 \mathbf{V}_5 &= \begin{pmatrix} \mathbf{0} & \mathbf{Q}_5 \\ \mathbf{Q}_5' & \mathbf{0} \end{pmatrix}, & \mathbf{Q}_5 &= \frac{1}{\sqrt{2pq}} \mathbf{1}_p \mathbf{1}_q', \\
 \mathbf{V}_6 &= \begin{pmatrix} \mathbf{Q}_6 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, & \mathbf{Q}_6 &= \frac{1}{\sqrt{2p + \frac{4}{p} - 6}} \left(\mathbf{H}_{1,p} + \frac{2}{p} (\mathbf{I}_p - \mathbf{1}_p \mathbf{1}_p') \right), \\
 \mathbf{V}_7 &= \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_7 \end{pmatrix}, & \mathbf{Q}_7 &= \frac{1}{\sqrt{2q + \frac{4}{q} - 6}} \left(\mathbf{H}_{1,q} + \frac{2}{q} (\mathbf{I}_q - \mathbf{1}_q \mathbf{1}_q') \right), \\
 \mathbf{V}_8 &= \begin{pmatrix} \mathbf{0} & \mathbf{Q}_8 \\ \mathbf{Q}_8' & \mathbf{0} \end{pmatrix}, & \mathbf{Q}_8 &= \frac{1}{\sqrt{2 - \frac{2}{pq}}} \left(\mathbf{e}_{p,p} \mathbf{e}_{1,q}' - \frac{1}{pq} \mathbf{1}_p \mathbf{1}_q' \right).
 \end{aligned} \tag{14}$$

Observe that $\frac{1}{m} \mathbf{1}_m \mathbf{1}_m'$ and $\mathbf{I}_m - \frac{1}{m} \mathbf{1}_m \mathbf{1}_m'$ are idempotent matrices. Moreover, the matrices $\mathbf{V}_i$, $i = 1, \dots, 5$, are strongly orthogonal (i.e., $\mathbf{V}_i \mathbf{V}_j = \mathbf{0}$, $i \neq j$). Due to Seely (1971), such matrices span a quadratic subspace (more precisely, they span the space $\mathcal{S}_{(a)}$). Thus, any subset of the matrices $\mathbf{V}_i$, $i = 1, \dots, 5$, also spans a quadratic subspace, and the following two quadratic subspaces can be taken as the quadratic

target spaces in the structured shrinkage estimation method:

$$\begin{aligned}\mathcal{T}_4 &= \left\{ \sum_{i=1}^4 \alpha_i \mathbf{V}_i : \alpha_i \in \mathbb{R}, i = 1, 2, 3, 4 \right\}, \\ \mathcal{T}_5 &= \left\{ \sum_{i=1}^5 \alpha_i \mathbf{V}_i : \alpha_i \in \mathbb{R}, i = 1, 2, 3, 4, 5 \right\}.\end{aligned}$$

Another quadratic subspace of $\mathcal{S}$ is

$$\mathcal{T}_{\text{II}} = \{\alpha_1 \mathbf{W}_1 + \alpha_3 \mathbf{W}_3 : \alpha_i \in \mathbb{R}, i = 1, 3\},$$

where

$$\mathbf{W}_1 = \begin{pmatrix} \frac{1}{\sqrt{p}} \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad \text{and} \quad \mathbf{W}_3 = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sqrt{q}} \mathbf{I}_q \end{pmatrix}.$$

Then the basis $\{\mathbf{V}_1, \dots, \mathbf{V}_8\}$ of $\mathcal{S}$ given in (14) must be transformed accordingly, and it will now be represented by $\{\mathbf{W}_1, \dots, \mathbf{W}_4, \mathbf{V}_5, \dots, \mathbf{V}_8\}$ with

$$\mathbf{W}_2 = \begin{pmatrix} \frac{1}{\sqrt{p}}(\sqrt{p-1}\mathbf{Q}_1 - \mathbf{Q}_2) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad \text{and} \quad \mathbf{W}_4 = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sqrt{q}}(\sqrt{q-1}\mathbf{Q}_3 - \mathbf{Q}_4) \end{pmatrix},$$

where $\mathbf{Q}_i, i = 1, \dots, 4$, are the same as defined for the corresponding matrices $\mathbf{V}_i$.

In the literature on the shrinkage estimation of covariance matrices, the most common choice of target matrix is a matrix proportional to the identity matrix; see e.g. Ledoit and Wolf (2004). Therefore, let us now assume the target space to be

$$\mathcal{T}_{\text{I}} = \{\alpha \mathbf{I}_m : \alpha \in \mathbb{R}\}.$$

Transforming the basis of $\mathcal{S}$ given in (14) accordingly, we can propose another basis of $\mathcal{S}$ as $\{\mathbf{U}_1, \mathbf{W}_2, \mathbf{U}_3, \mathbf{W}_4, \mathbf{V}_5, \mathbf{V}_6, \mathbf{V}_7, \mathbf{V}_8\}$, where

$$\mathbf{U}_1 = \frac{1}{\sqrt{m}} \mathbf{I}_m \quad \text{and} \quad \mathbf{U}_3 = \sqrt{\frac{q}{pq+p^2}} \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & -\frac{p}{q} \mathbf{I}_q \end{pmatrix}.$$

Projecting the sample covariance matrix $\mathbf{S}$ onto $\mathcal{S}$ and $\mathcal{T}_i, i \in \{4, 5, \text{II}, \text{I}\}$, and computing optimal intensities $\hat{\lambda}_i$ following (10), we obtain four shrinkage estimators,

$$\hat{\Sigma}_i = \hat{\lambda}_i \mathbf{P}_{\mathcal{T}_i}(\mathbf{S}) + (1 - \hat{\lambda}_i) \mathbf{P}_{\mathcal{S}}(\mathbf{S}), \quad i \in \{4, 5, \text{II}, \text{I}\}.$$

If for a given data set the estimate $\hat{\Sigma}_i, i \in \{4, 5, \text{II}, \text{I}\}$, is indefinite or ill-conditioned, the final estimate should be calculated by numerical adjustment of the shrinkage intensity, as given in (12).

Finally, note that $\mathcal{T}_{\text{I}} \subset \mathcal{T}_{\text{II}} \subset \mathcal{T}_4 \subset \mathcal{T}_5$. □

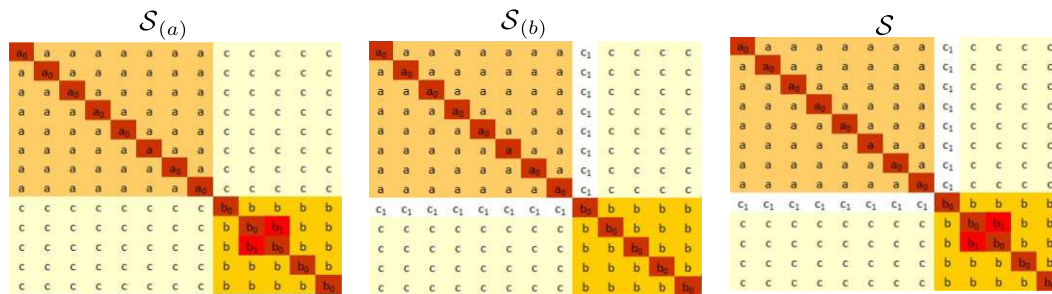


Fig. 2 Visualization of three structure spaces such that $S_{(a)} \subset S$ and $S_{(b)} \subset S$

Example 2 Ritchie et al. (2023) considered metabolic biomarker data quantified by nuclear magnetic resonance (NMR) spectroscopy for approximately 121,000 UK Biobank participants, comprising absolute concentrations and ratios of 249 circulating metabolites, lipids, and lipoprotein subfractions. The authors identified and characterized additional sources of unwanted technical variation influencing individual biomarkers in the data available to download from UK Biobank, developed a procedure to remove this unwanted technical variation, and demonstrated that it increases the signal for genetic and epidemiological studies of the NMR metabolic biomarker data in the UK Biobank. The pairwise Pearson correlation coefficients between the 325 biomarkers after removing technical variation are presented in Fig. 3(C), p. 9, of the afore mentioned paper. Focusing on correlation structure between biomarkers, particularly among lipid percentages and lipids in VLDL, we can observe strong correlation, which is a typical property for this type of data; cf. Ritchie et al. (2023). Yang et al. (2024) used the correlation matrix presented by Ritchie et al. (2023) as a starting point for graphical identification of the covariance structure and, assuming particular structure (which in our notation belongs to the quadratic subspace $\mathcal{T}_5$), determined the maximum likelihood estimator of this structure using the so-called *block Hadamard product*; however, this estimator can also be obtained as the projection of $\frac{n-1}{n}\mathbf{S}$ onto $\mathcal{T}_5$; cf. Filipiak et al. (2021).

Motivated by this example, we take into account the correlation structure between lipid percentages and lipids in VLDL to assume particular covariance structures (see Fig. 2) and then we illustrate the estimation method presented using four quadratic target spaces.

Due to (1), the spaces under consideration can be expressed as the sum of orthogonal subspaces as given in (2) with

$$\begin{aligned} \mathcal{A} &= \left\{ a_0 \mathbf{I}_p + a \sum_{i=1}^{p-1} \mathbf{H}_{i,p} : a_0, a \in \mathbb{R} \right\}, \\ \mathcal{B} &= \left\{ b_0 \mathbf{I}_q + b_1 (\mathbf{e}_{q_i,q} \mathbf{e}_{q_i+1,q}' + \mathbf{e}_{q_i+1,q} \mathbf{e}_{q_i,q}') + b \sum_{i=2}^{q-1} \mathbf{H}_{i,q} : b_0, b_1, b \in \mathbb{R} \right\}, \\ \mathcal{C} &= \left\{ c \mathbf{1}_p \mathbf{1}_q' - c_1 \mathbf{1}_p \mathbf{e}_{1,q}' : c, c_1 \in \mathbb{R} \right\}, \end{aligned} \quad (15)$$

with $\mathbf{H}_{i,p}$ and $\mathbf{e}_{1,p}$ defined as in Example 1, and with q_i being the variable index such that the dependence between q_i and $q_i + 1$ is stronger than between other variables (in Fig. 2 we have $q_i = 2$). Note also that $c_1 = 0$ in $\mathcal{S}_{(a)}$ and $b_1 = b$ in $\mathcal{S}_{(b)}$.

It can be seen that the dimension of $\mathcal{S}$ is 7, while both considered subspaces have dimension 6. In particular, the basis of all these spaces contains the matrices $\mathbf{V}_i$, $i = 1, \dots, 5$, defined in (14), and the remaining two basis matrices of $\mathcal{S}$ have the form

$$\mathbf{V}_6 = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_6 \end{pmatrix}, \quad \mathbf{Q}_6 = \frac{1}{\sqrt{2 - \frac{4}{q(q-1)}}} \left(\mathbf{e}_{q_i,q} \mathbf{e}'_{q_i+1,q} + \mathbf{e}_{q_i+1,q} \mathbf{e}'_{q_i,q} - \frac{2}{q(q-1)} (\mathbf{1}_q \mathbf{1}'_q - \mathbf{I}_q) \right),$$

$$\mathbf{V}_7 = \begin{pmatrix} \mathbf{0} & \mathbf{Q}_7 \\ \mathbf{Q}'_7 & \mathbf{0} \end{pmatrix}, \quad \mathbf{Q}_7 = \frac{1}{\sqrt{\frac{2p(q-1)}{q}}} \left(\mathbf{1}_p \mathbf{e}'_{1,q} - \frac{1}{q} \mathbf{1}_p \mathbf{1}'_q \right).$$

Clearly, the basis of $\mathcal{S}_{(a)}$ ($\mathcal{S}_{(b)}$) is the same as the basis of $\mathcal{S}$ without $\mathbf{V}_7$ ($\mathbf{V}_6$).

Since $\mathbf{V}_i$, $i = 1, \dots, 5$ span the quadratic subspace, we can consider the same target spaces ($\mathcal{T}_4$, $\mathcal{T}_5$, $\mathcal{T}_{\text{II}}$ and $\mathcal{T}_{\text{I}}$) as in Example 1.

Projecting the sample covariance matrix $\mathbf{S}$ onto $\mathcal{S}_j$, where $j \in \{0, (a), (b)\}$ and $\mathcal{S}_0 = \mathcal{S}$, and onto $\mathcal{T}_i$, $i \in \{4, 5, \text{II}, \text{I}\}$, and computing optimal intensities $\hat{\lambda}_i^j$ following (10), for each space $\mathcal{S}_j$ we obtain four shrinkage estimators of $\boldsymbol{\Sigma} \in \mathcal{S}_j$ of the form

$$\hat{\boldsymbol{\Sigma}}_i^j = \hat{\lambda}_i^j \mathbf{P}_{\mathcal{T}_i}(\mathbf{S}) + (1 - \hat{\lambda}_i^j) \mathbf{P}_{\mathcal{S}_j}(\mathbf{S}).$$

Similarly as in Example 1, if for a given data set the estimate $\hat{\boldsymbol{\Sigma}}_i^j$, $i \in \{4, 5, \text{II}, \text{I}\}$, $j \in \{0, (a), (b)\}$, is indefinite or ill-conditioned, the final estimate should be calculated by numerical adjustment of the shrinkage intensity, as given in (12). $\square$

4 Simulation study

The main objective of this section is to verify how the choice of the target space affects the properties of the structured shrinkage estimator. For various spaces of covariance structures and four quadratic target spaces, we compare conditioning of the estimates, with particular emphasis on the additional adjustments resulting in a $\tilde{\boldsymbol{\Sigma}}$ different from $\hat{\boldsymbol{\Sigma}}$, and we compare the empirical risk (average loss) based on three loss functions: Stein loss, inverse Stein loss, and quadratic loss (cf. Ledoit and Wolf 2021), defined respectively as

$$L_S(\tilde{\boldsymbol{\Sigma}}) = \frac{1}{m} [\text{tr}(\boldsymbol{\Sigma}^{-1} \tilde{\boldsymbol{\Sigma}}) - \log |\boldsymbol{\Sigma}^{-1} \tilde{\boldsymbol{\Sigma}}|] - 1,$$

$$L_{\text{IS}}(\tilde{\boldsymbol{\Sigma}}) = \frac{1}{m} [\text{tr}(\boldsymbol{\Sigma} \tilde{\boldsymbol{\Sigma}}^{-1}) - \log |\boldsymbol{\Sigma} \tilde{\boldsymbol{\Sigma}}^{-1}|] - 1,$$

$$L_Q(\tilde{\boldsymbol{\Sigma}}) = \text{tr}[(\boldsymbol{\Sigma}^{-1} \tilde{\boldsymbol{\Sigma}} - \mathbf{I}_m)^2]$$

(cf. Anderson 2003), as well as the squared Frobenius norm $f(\boldsymbol{\Sigma}) = \|\tilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|^2$.

We perform Monte Carlo simulation studies in which we generate 1000 datasets from $N_{m,100}(\mathbf{0}, \boldsymbol{\Sigma}, \mathbf{I}_{100})$, where $\boldsymbol{\Sigma} \in \mathcal{S}$ as given in Example 1 of the previous section

or $\Sigma \in \mathcal{S}_j$, $j \in \{0, (a), (b)\}$ as given in Example 2. As target spaces, we consider $\mathcal{T}_I$, $\mathcal{T}_{II}$, $\mathcal{T}_4$, and $\mathcal{T}_5$, as they are quadratic subspaces of each considered linear structure space. Recall, moreover, that $\mathcal{T}_I \subset \mathcal{T}_{II} \subset \mathcal{T}_4 \subset \mathcal{T}_5$. The values of the parameters p and q (such that $p + q = m$) as well as the exact forms of Σ are specified in the following examples.

In the first example, our main goal is to show that the strength of the noise on the first sub- and super-diagonals does not have a significant influence on the analyzed functions, unless the covariance matrix Σ is close to a singular matrix. Moreover, as expected, the increase in the dimension of the matrix also increases the values of all of the functions analyzed.

In the second example, we consider three linear structure spaces, and we show that if the noise is present only in the off-diagonal block of the covariance matrix, then the values of all analyzed functions do not differ significantly depending on the choice of target space.

Our aim is to identify the target space that gives the estimate with the lowest condition number and the smallest value of empirical risk and squared Frobenius norm. We observe the behavior of the estimates with respect to the dimension of the data set (p, q) , and the conditioning of Σ .

Example 1

Consider $(p, q) \in \{(3, 4), (6, 10)\}$ and $\mathcal{S}$ of the form shown in (1) and (2), where $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are as given in (13). We assume $a_0 = 15$, $a = 2$, $b_0 = 11$, $b = 1.5$, $c = 3$ and various levels of noise on the first sub- and super-diagonals as presented in the following three setups:

- (a) $b_1 = 3$, $c_1 = 0.5$ and various $a_1 \in \{6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5\}$ for $(p, q) = (3, 4)$ and $a_1 \in \{4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5\}$ for $(p, q) = (6, 10)$,
- (b) $a_1 = 8$, $c_1 = 0.5$ and various $b_1 \in \{2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6\}$ for $(p, q) = (3, 4)$ and $b_1 \in \{1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5\}$ for $(p, q) = (6, 10)$,
- (c) $a_1 = 8$, $b_1 = 3$ and various $c_1 \in \{5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9\}$ for $(p, q) = (3, 4)$ and $c_1 \in \{5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5\}$ for $(p, q) = (6, 10)$.

Then for each dataset we determine $\hat{\Sigma}_i$ using the formula (9) with $\hat{\lambda}_i$ as given in (10), $i \in \{I, II, 4, 5\}$, and we compute its condition number. If it exceeds 100, or if $\hat{\Sigma}_i$ is indefinite, an additional adjustment is used. As a result, we obtain 1000 structured shrinkage estimators $\tilde{\Sigma}_i$ computed via (12). Finally, for each set of parameters and each target space, the averaged condition number is calculated, as well as the empirical risks related to three loss functions (the values of the loss functions averaged over 1000 trials) and the averaged squared Frobenius norm.

Figure 3 show the averaged condition number for $(p, q) = (3, 4)$ (first column) and $(p, q) = (6, 10)$ (second column), with respect to the values of a_1 , b_1 and c_1 in the setup of the corresponding parameters [(a), (b) and (c)], and the target space used in the structured shrinkage estimation method. It can be observed that choosing the target matrix from a larger space significantly reduces the condition number of the estimate (the pink line is always the lowest). Moreover, there is always a significant difference between the pink line and the remaining ones, which means that taking into account a nonzero off-diagonal block of the target matrix significantly improves the conditioning of the final estimate. On the other hand, the condition number for $\mathcal{T}_5$ is

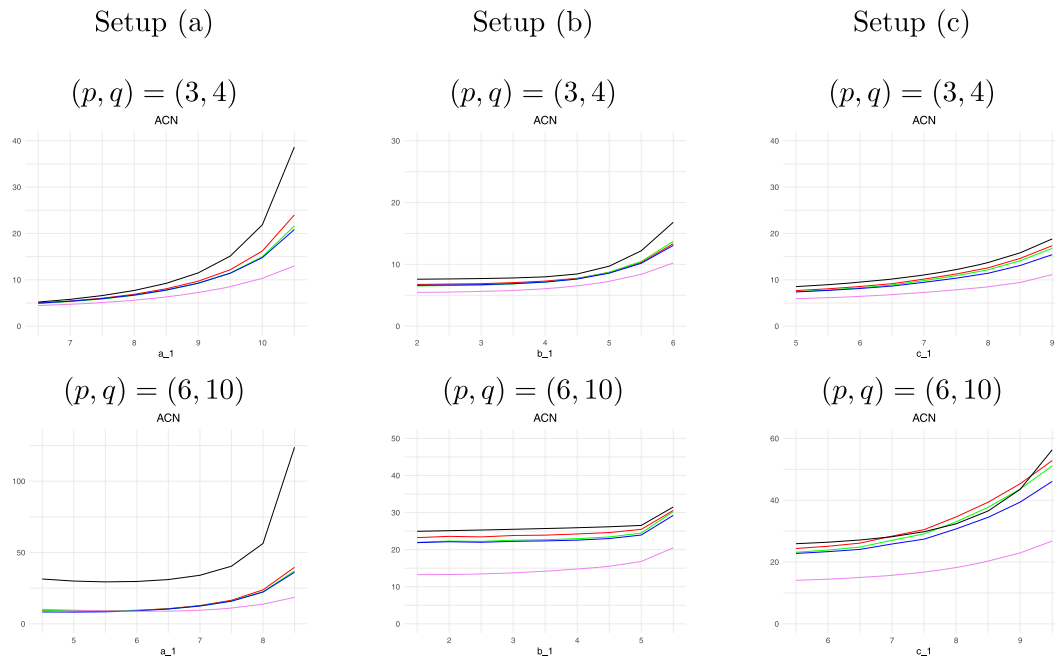


Fig. 3 The averaged condition number (ACN) of the structured shrinkage estimator with respect to the parameters a_1 (setup (a)), b_1 (setup (b)), and c_1 (setup (c)) (in columns), parameters $(p, q) \in \{(3, 4), (6, 10)\}$ (in rows) and the choice of the target space: $\mathcal{T}_I$ —red, $\mathcal{T}_{II}$ —green, $\mathcal{T}_4$ —blue, $\mathcal{T}_5$ —pink line. The black line depicts the condition number of Σ

farthest from the condition number of Σ , which is in fact not surprising, as projecting into a quadratic subspace decreases the condition number of the projected matrix, so a larger quadratic subspace implies a larger decrease in the condition number.

We can also observe that the choice of parameter setup (related to (p, q) as well as a_1, b_1, c_1) does not affect the results significantly, but increasing a_1, b_1 and c_1 in the corresponding setups increases the conditioning of the estimates. This may be caused by the conditioning of Σ , which also increases in the case considered (black line). From the graph for $(p, q) = (3, 4)$ and the setups (a) and (b), it can also be observed that the choice of the largest target space is the least sensitive to conditioning of Σ .

Finally, during the simulations, we observed that the frequency of adjusted estimates increases with the conditioning of Σ , which may explain the increase in the averaged condition number at the end of each domain. It is also noteworthy that the smallest number of adjusted estimates $\tilde{\Sigma}_i$ was observed for target matrices belonging to $\mathcal{T}_5$.

In Fig. 4 the empirical Stein risk, the empirical inverse Stein risk and the empirical quadratic risk are presented together with the averaged Frobenius norm of the difference $\tilde{\Sigma}_i - \Sigma$ depending on the parameters (p, q) (in columns) and setups (a), (b) and (c) (in rows). The same colors as previously are used to denote estimates obtained via the target matrix from the corresponding quadratic subspace. Similarly as for the conditioning of the estimate, it can be observed that the choice of parameter setup does not influence the statistical properties significantly, while the ordering of the color lines is reversed. Moreover, similarly as in the previous case, the pink line related to the target matrix from the space $\mathcal{T}_5$ differs significantly from the remaining lines in each graph, while the remaining functions almost overlap if risks are taken into account.

This similarity may be caused by the commutativity of these three quadratic spaces; for more details about commutative quadratic spaces see, e.g., Seely (1971). Nevertheless, since all of the lines are almost parallel, we may conclude that the increase in the risk and the distance between the estimates and the true covariance matrix does not depend on the choice of target space. Note also that the conditioning of Σ influences the risk.

Finally, observe that the average Frobenius norm of the difference between the estimate and the true covariance matrix is resistant to changes in the values of the parameters a_1 , b_1 , and c_1 and increases slightly with the dimensions.

The variance of the resulting condition numbers, losses and squared Frobenius norms was also examined. The results for each parameter setup and for $(p, q) = (3, 4)$ are presented in Fig. 5. It can be observed that the variability of matrix conditioning for $\mathcal{T}_5$ is always the lowest, while for the loss functions and the squared Frobenius norm, the variance is usually the highest. For the remaining target spaces, there is no significant difference in variability. Similar results were obtained for $(p, q) = (6, 10)$ and thus are not reported here. $\square$

Example 2

Consider $(p, q) = (8, 5)$ and $\mathcal{S}$ of the form shown in (1) and (2), where $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are as given in (15). In other words, we study the structures illustrated in Fig. 2 with $a_0 = 1$, $a = 0.3$, $b_0 = 1$, $b_1 = 0.7$, $b = 0.4$, $c_1 = 0$ and $c = 0.2$.

For each dataset we determine $\widehat{\Sigma}_i^j$ using formula (9) with $\widehat{\lambda}_i^j$ as given in (10), $i \in \{\text{I}, \text{II}, 4, 5\}$, $j \in \{0, (a), (b)\}$ and we compute its condition number. If it exceeds 100, or if $\widehat{\Sigma}_i^j$ is indefinite, an additional adjustment is necessary. As a result, we obtain 1000 structured shrinkage estimators $\widetilde{\Sigma}_i^j$ computed via (12). Finally, for each $\mathcal{S}_j$ and $\mathcal{T}_i$ the averaged condition number is computed, as well as the empirical risks related to three loss functions (the values of the loss functions averaged over 1000 trials) and the average squared Frobenius norm of the difference between $\widetilde{\Sigma}_i^j$ and Σ ; these are given in Table 1.

Similar conclusions can be drawn here as from Example 1. For all three structures of Σ the conditioning of the structured shrinkage estimate is the lowest if the largest quadratic target space is considered. However, the condition number of $\mathcal{T}_5$ is closest to the condition number of Σ (13.97 for $\mathcal{S}_{(a)}$, 8.21 for $\mathcal{S}_{(b)}$, and 13.32 for $\mathcal{S}$), which is not surprising, as projecting onto a quadratic subspace decreases the condition number of the projected matrix, so a larger quadratic subspace implies greater reduction. For the risks considered and the Frobenius norm, the lowest values are obtained for the smallest quadratic target spaces, $\mathcal{T}_\text{I}$ and $\mathcal{T}_\text{II}$.

When considering the variability of the computed estimators, for conditioning the variability for $\mathcal{T}_5$ is the lowest, and the same holds for the inverse Stein risk, while for the remaining functions there are no significant differences in variability (except for spaces $\mathcal{S}_{(a)}$, $\mathcal{S}$ with target space $\mathcal{T}_5$, where the quadratic risk variability is significantly higher than for the remaining target spaces).

Finally, note that in the case of space $\mathcal{S}_{(b)}$, the values of all analyzed functions do not differ significantly depending on the choice of the target space. However, in this space, the choice of $\mathcal{T}_5$ is optimal, as all of the values are then the smallest. This may be because in this case the covariance matrix Σ seems to be the least divergent from

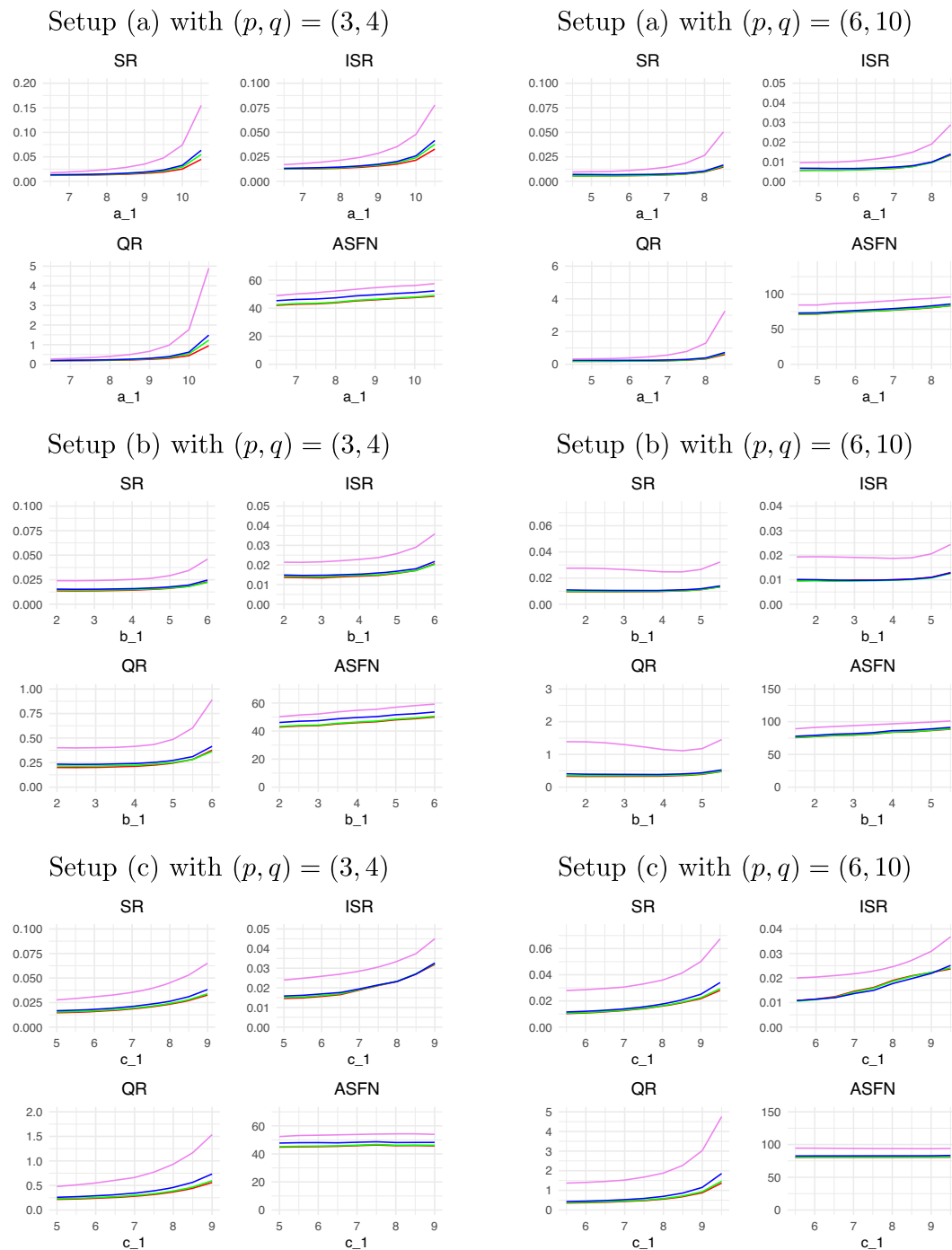


Fig. 4 Empirical Stein risk (SR), inverse Stein risk (ISR), quadratic risk (QR) and average squared Frobenius norm (ASFN) for $(p, q) = (3, 4)$ (left) and $(p, q) = (6, 10)$ (right) depending on the value of (a) a_1 , (b) b_1 , (c) c_1 , and the target space used in the shrinkage method: red line—target matrix from $\mathcal{T}_I$, green line—target matrix from $\mathcal{T}_{II}$, blue line—target matrix from $\mathcal{T}_4$, pink line—target matrix from $\mathcal{T}_5$

Table 1 Averaged condition number, empirical Stein risk, inverse Stein risk, quadratic risk, and averaged squared Frobenius norm (with respective standard deviations in parentheses) depending on structure space and target space

		$\bar{\kappa}(\tilde{\Sigma})$	$\bar{L}_S(\tilde{\Sigma})$	$\bar{L}_{IS}(\tilde{\Sigma})$	$\bar{L}_Q(\tilde{\Sigma})$	$\bar{f}(\tilde{\Sigma})$
$S_{(a)}$	$\mathcal{T}_I$	13.443 (6.347)	0.007 (0.007)	0.008 (0.015)	0.195 (0.150)	0.567 (0.491)
	$\mathcal{T}_{II}$	13.597 (9.479)	0.008 (0.008)	0.009 (0.030)	0.197 (0.152)	0.570 (0.492)
	$\mathcal{T}_4$	12.824 (8.167)	0.008 (0.006)	0.008 (0.024)	0.209 (0.151)	0.642 (0.533)
	$\mathcal{T}_5$	8.512 (1.544)	0.019 (0.006)	0.016 (0.005)	0.622 (0.245)	0.643 (0.513)
$S_{(b)}$	$\mathcal{T}_I$	6.683 (1.232)	0.021 (0.004)	0.018 (0.003)	0.665 (0.191)	1.043 (0.437)
	$\mathcal{T}_{II}$	6.687 (1.237)	0.021 (0.004)	0.019 (0.003)	0.667 (0.191)	1.046 (0.438)
	$\mathcal{T}_4$	6.714 (1.256)	0.022 (0.004)	0.018 (0.003)	0.727 (0.221)	1.144 (0.501)
	$\mathcal{T}_5$	6.570 (1.014)	0.020 (0.003)	0.017 (0.003)	0.629 (0.175)	1.030 (0.453)
S	$\mathcal{T}_I$	12.193 (4.863)	0.022 (0.006)	0.020 (0.009)	0.673 (0.214)	1.068 (0.450)
	$\mathcal{T}_{II}$	12.195 (4.824)	0.022 (0.006)	0.020 (0.009)	0.675 (0.215)	1.071 (0.452)
	$\mathcal{T}_4$	11.052 (3.850)	0.023 (0.006)	0.020 (0.006)	0.732 (0.251)	1.168 (0.508)
	$\mathcal{T}_5$	7.978 (1.430)	0.032 (0.007)	0.028 (0.005)	1.054 (0.313)	1.149 (0.476)

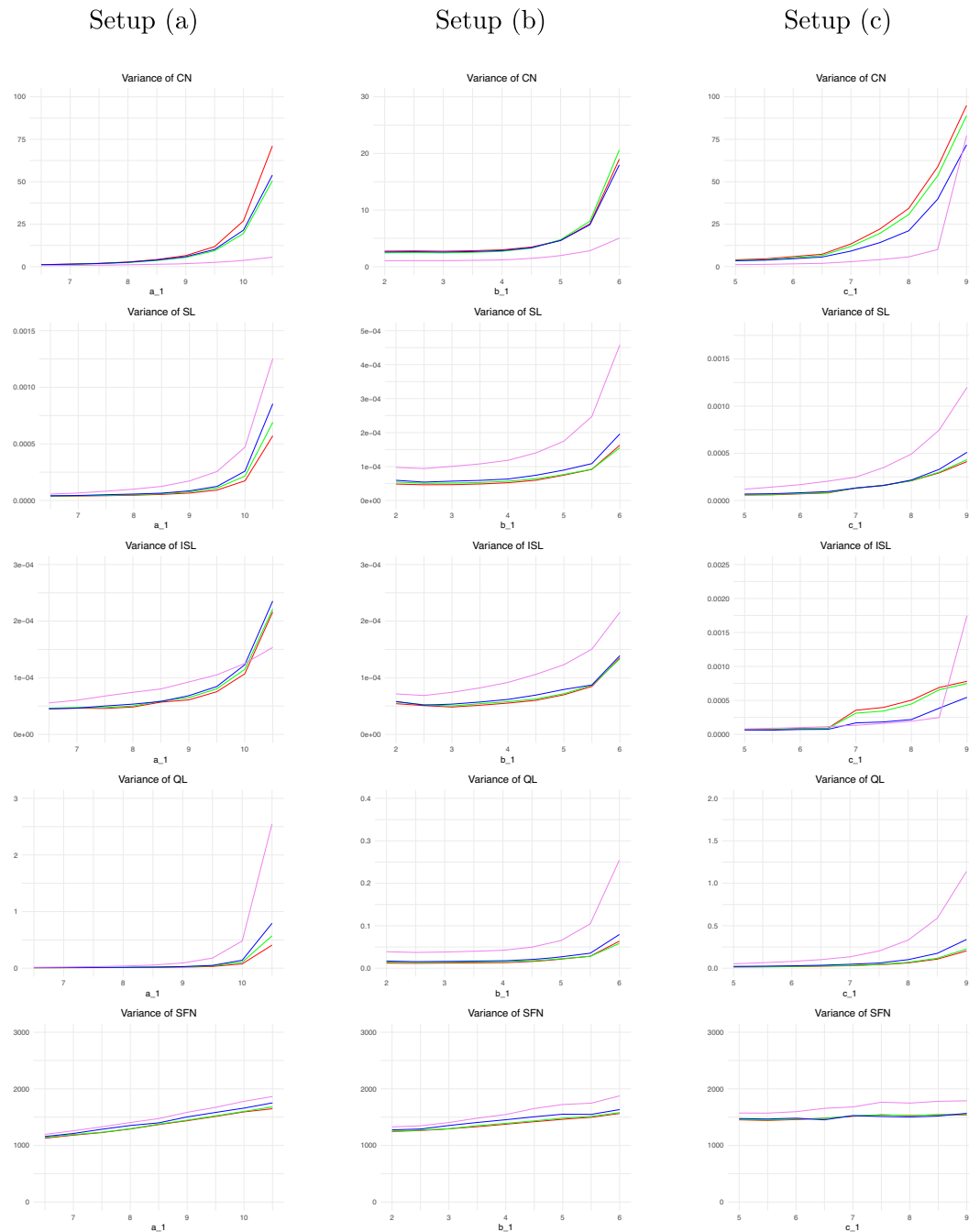


Fig. 5 Variance of: Stein loss (SL), inverse Stein loss (ISL), quadratic loss (QL) and squared Frobenius norm (SFN) for $(p, q) = (3, 4)$ depending on the value of (a) a_1 , (b) b_1 , (c) c_1 , and the target space used in the shrinkage method: red line—target matrix from $\mathcal{T}_I$, green line—target matrix from $\mathcal{T}_{II}$, blue line—target matrix from $\mathcal{T}_4$, pink line—target matrix from $\mathcal{T}_5$

$\mathcal{T}_5$. This observation is also confirmed if we consider space $\mathcal{S}$ with little noise, that is, $b_1 = 0.5$, $c_1 = 0.1$. The corresponding results are given in Table 2.

Table 2 Averaged condition number, empirical Stein risk, inverse Stein risk, quadratic risk, and averaged squared Frobenius norm (with respective standard deviations in parentheses) depending on target space for $\Sigma \in \mathcal{S}$ with $b_1 = 0.5$, $c_1 = 0.1$

	$\bar{\kappa}(\tilde{\Sigma})$	$\bar{L}_S(\tilde{\Sigma})$	$\bar{L}_{IS}(\tilde{\Sigma})$	$\bar{L}_Q(\tilde{\Sigma})$	$\bar{f}(\tilde{\Sigma})$
$\mathcal{T}_I$	7.810 (1.807)	0.009 (0.003)	0.009 (0.004)	0.236 (0.098)	0.675 (0.462)
$\mathcal{T}_{II}$	7.813 (1.808)	0.009 (0.003)	0.009 (0.004)	0.239 (0.099)	0.678 (0.463)
$\mathcal{T}_4$	7.771 (1.748)	0.009 (0.003)	0.009 (0.003)	0.252 (0.106)	0.765 (0.516)
$\mathcal{T}_5$	6.997 (1.100)	0.008 (0.003)	0.008 (0.003)	0.225 (0.077)	0.674 (0.481)

□

5 Conclusions

In this paper, we have proposed a method for the estimation of a partitioned covariance matrix with different block sizes, which does not belong to a quadratic subspace. The method is quick and gives an estimate which has the required structure and which is positive definite and well-conditioned. Moreover, it can be easily extended to matrices suitable for the partition of observation vectors into more subvectors than two.

It should also be emphasized that the choice of quadratic target space plays an important role in estimation. As can be observed, there is a difference between statistical and algebraic properties depending on the target space: the quadratic subspace of block-diagonal matrices, or the quadratic subspace of partitioned matrices with non-zero off-diagonal blocks. Moreover, for block diagonal target spaces the statistical properties of the proposed estimators do not differ significantly, while the conditioning might improve slightly. Nevertheless, when the structure of the true covariance matrix is close to the structure from a quadratic subspace, then the choice of target matrix from that quadratic subspace seems to be the most suitable.

Summing up, if the researcher is interested in an estimator which is well-conditioned, then the quadratic space closest to the considered structure should be selected as the target space. On the other hand, if statistical properties are of main importance, a target space being a commutative quadratic subspace of the relevant structure space should be considered.

Acknowledgements The authors are extremely grateful to the Referees for all of their valuable comments and suggestions.

Funding This work was supported by the Poznań University of Technology under Grant No. 0213/SBAD/0119 (K. Filipiak and M. Mrowińska).

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted

use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Anderson TW (1973) Asymptotically efficient estimation of covariance matrices with linear structure. *Ann Stat* 1(1):135–141
- Anderson TW (2003) An introduction to multivariate statistical analysis. Wiley, New York
- Arnold SF (1979) Linear models with exchangeably distributed errors. *J Am Stat Assoc* 74:194–199
- Carrión-García C, Eduardo Jesús Guerra-Hernández E, García-Villanova B, Serafini M, Sánchez M, Amiano P, Molina-Montes E (2020) Plasma non-enzymatic antioxidant capacity (NEAC) in relation to dietary NEAC, nutrient antioxidants and inflammation-related biomarkers. *Antioxidants* 9(4):301
- Crowder MJ, Hand DJ (1990) Analysis of repeated measures, 1st edn. Chapman & Hall/CRC, New York
- Devijver E, Gallopin M (2018) Block-diagonal covariance selection for high-dimensional Gaussian graphical models. *J Am Stat Assoc* 113(521):306–314
- Filipiak K, Markiewicz A, Mieldzioc A, Sawikowska A (2018) On projection of a positive definite matrix on a cone of nonnegative definite Toeplitz matrices. *Electron J Linear Algebra* 33:74–82
- Filipiak K, John M, Markiewicz A (2021) Comments on maximum likelihood estimation and projections under multivariate statistical models. In: Holgersson T, Singull M (eds) Recent developments in multivariate and random matrix analysis. Springer, Cham, pp 51–66
- Filipiak K, Markiewicz A, Mrowinska M (2024) Characterization of the quadratic space of partitioned matrices Submitted
- Fuglede B, Jensen TS (2013) Positive projections of symmetric matrices and Jordan algebras. *Expo Math* 31:295–303
- Gilson M, Dahmen D, Moreno-Bote R, Insabato A, Helias M (2020) The covariance perceptron: a new framework for classification and processing of time series in recurrent neural networks. *PLoS Comput Biol* 16(10):e1008127
- Gupta A, Nagar D (2000) Matrix variate distributions. Chapman & Hall/CRC, Boca Raton
- Hannart A, Naveau P (2014) Estimating high dimensional covariance matrices: a new look at the Gaussian conjugate framework. *J Multivar Anal* 131:149–162
- Ikedo Y, Kubokawa T, Srivastava MS (2016) Comparison of linear shrinkage estimators of a large covariance matrix in normal and non-normal distributions. *Comput Stat Data Anal* 95:95–108
- Janiszewska M, Markiewicz A, Mokrzycka M (2020) Block matrix approximation via entropy loss function. *Appl Math* 65(6):829–844
- Kozioł A, Roy A, Zmyślony R, Leiva R, Fonseca M (2017) Best unbiased estimates for parameters of three-level multivariate data with doubly exchangeable covariance structure. *Linear Algebra Appl* 535:87–104
- Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. *J Multivar Anal* 88(2):365–411
- Ledoit O, Wolf M (2021) Shrinkage estimation of large covariance matrices: keep it simple, statistician? *J Multivar Anal* 186:104796
- Lu N, Zimmerman D (2005) The likelihood ratio test for a separable covariance matrix. *Stat Probab Lett* 73:449–457
- Markiewicz A, Mokrzycka M, Mrowińska M (2024) Quasi shrinkage estimation of a block-structured covariance matrix. *J Stat Comput Simul* 94(14):1–18
- Mieldzioc A, Mokrzycka M, Sawikowska A (2021) Identification of block-structured covariance matrix on an example of metabolomic data. *Separations* 8(11):205
- Ohlson M, von Rosen D (2010) Explicit estimators of parameters in the growth curve model with linearly structured covariance matrices. *J Multivar Anal* 101:1284–1295
- Ritchie S, Surendran P, Karthikeyan S, Lambert S, Bolton T, Pennells L, Danesh J, Di Angelantonio E, Butterworth A, Inouye M (2023) Quality control and removal of technical variation of NMR metabolic biomarker data in 120,000 UK Biobank participants. *Sci Data* 10:1–15
- Seely J (1971) Quadratic subspaces and completeness. *Ann Math Stat* 42(2):710–721
- Szatrowski TH (1980) Necessary and sufficient conditions for explicit solutions in the multivariate normal estimation problem for patterned means and covariances. *Ann Stat* 8:802–810

Yang Y, Chen C, Chen S (2024) Covariance matrix estimation for high-throughput biomedical data with interconnected communities. *Am Stat* 78(4):401–411

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.