ISSN 0973-2721

BULLETIN OF

Kerala Mathematics Association

Special issue dedicated to

International Conference on Linear Algebra and its Applications (ICLAA) 2017

JOURNAL OF THE KERALA MATHEMATICAL ASSOCIATION

Bulletin of Kerala Mathematics Association

Special issue dedicated to

International Conference on Linear Algebra and its Applications (ICLAA) 2017

contents

Registered with Registrar of Newspapers of India under No. KERENG/2004/17381 Printed, Published and owned by Dr. M.S. Samuel and Printed at Penta Offset, Kottayam 686 001, Kerala State, India and Published at Kottayam , Kerala State, Editor, Dr. M.S. Samuel, 15/64, Mattathil, Powath Road, Muttambalam, Kottayam 686 004, Kerala, India .

Contents

Preface

International conference on Linear Algebra and its Applications–ICLAA 2017, third in its sequence, following CMTGIM 2012 and ICLAA 2014, held in Manipal Academy of Higher Education, Manipal, India in December 11-15, 2017.

Like its preceding conferences, ICLAA 2017 is also focused on the theory of Linear Algebra and Matrix Theory, and their applications in Statistics, Network Theory and in other branches of sciences. Study of Covariance Matrices, being part of Matrix Method in Statistics, has applications in various branches of sciences. It plays crucial role in the study of measurement of uncertainty and naturally in the study of Nuclear Data. Theme meeting, which initially planned to be a preconference meeting, further progressed into an independent event parallel to ICLAA 2017, involving discussion on different methodology of generating the covariance information, training modules on different techniques and deliberations on presenting new research.

About 167 delegates have registered for ICLAA 2017 alone (37 Invited + 75 Contributory + 04 Poster) and are from 17 different countries of the world. Interestingly, more than 80% are repeaters from the earlier conference and the remaining 20% are young students or scholars. In spite of a few dropouts due to unavoidable constraints, it is felt evident that the group of scholars with focus area of Linear Algebra, Matrix Methods in Statistics and Matrices and Graphs are not only consolidating, also growing as a society with a strong bond. ICLAA 2017 provided a platform for renowned Mathematicians and Statisticians to come together and discuss research problems, it provided ample of time for young scholars to present their contribution before eminent scholars. Every contributory speaker got not less than thirty minutes to present their results. Also, ICLAA 2017 was with several special lectures from senior scientists aimed at encouraging young scholars.

The sponsors of ICLAA 2017 are NBHM, SERB, CSIR and ICTP. Dr. Ebrahim Ghorbani and Dr. Zheng Bing are the two international participants benifited from ICTP grant for their international travel.

The conference was opened with an informal welcome and opening remark by K. Manjunatha Prasad (Organizing Secretary) and R. B. Bapat (Chairman, Scientific Committee). Invited talks and the special lectures were organized in 13 different sessions and contributory talks in 17 sessions. Poster presentation was arranged on December 12, 2017.

In an informal discussion, it has been consented by the present scientific committee and the organizing committee members that

(i) MAHE would continue to organize ICLAA 2020 in December 2020, the fourth in its sequence

(ii) Manjunatha Prasad would put up a proposal to organize ILAS conference in the earliest possible occasion (2022/23), in consultation with Kirkland

(iii) Manjunatha Prasad to initiate a dialog with the members in the present network to have Indian Society for Linear Algebra and its Application

The organizers are very proud of bringing out two special issues related to the conference, the one with *Bulletin of Kerala Mathematics Association* and the other one with *Special Matrices (De Gruyter).* The organizers are thankful to managerial team of BKMA, particularly Samuel Mattathil, and the chief editor Carlos Martins da Fonseca of Special Matrices for their kind support in bringing up these special issues. They are also thankful to all the authors for submitting their articles and reviewers for sparing their valuable time.

Ravindra B Bapat, *ISI, Delhi, India*
Steve Kirkland, *University of Manitoba, Canada*
K. Manjunatha Prasad, *MAHE, Manipal, India*
Simo Puntanen, *University of Tampere, Finland*

Bulletin of Kerala Mathematics Association, Special Issue
*Vol. 16, No. **1** (2018, June)* 13–45

A REVIEW OF SOME USEFUL PROPERTIES OF THE COVARIANCE MATRIX OF THE BLUE IN THE GENERAL LINEAR MODEL

Jarkko Isotalo*, Simo Puntanen[1]**, George P. H. Styan***

**Department of Forest Sciences,*
University of Helsinki, Finland
`jarkko.isotalo@helsinki.fi`

***Faculty of Natural Sciences,*
University of Tampere, Finland
`simo.puntanen@uta.fi`

****Department of Mathematics and Statistics,*
McGill University, Montréal, Canada
`geostyan@gmail.com`

(Received 05.02.2018; Accepted 01.03.2018)

Abstract. In this paper we consider the linear statistical model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, which can be shortly denoted as the triplet $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$. Here $\mathbf{X}$ is a known $n \times p$ fixed model matrix, the vector $\mathbf{y}$ is an observable n-dimensional random vector, $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed but unknown parameters, and $\boldsymbol{\varepsilon}$ is an unobservable vector of random errors with expectation $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, and covariance matrix $\mathrm{cov}(\boldsymbol{\varepsilon}) = \mathbf{V}$, where the nonnegative definite matrix $\mathbf{V}$ is known. In our considerations it is essential that the covariance matrix $\mathbf{V}$ is known; if this is not the case the statistical considerations become much more complicated.

Our main focus is to define and introduce, in the general form, without rank conditions, the key properties of the best linear unbiased estimator, BLUE, of $\mathbf{X}\boldsymbol{\beta}$. In particular we consider some specific properties of the covariance matrix of the BLUE. We also deal shortly with the best linear unbiased predictor, BLUP, of $\mathbf{y}_*$, when $\mathbf{y}_*$ is assumed to come from $\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*$, where $\mathbf{X}_*$ is a known matrix, $\boldsymbol{\beta}$ is the same vector of unknown parameters as in $\mathscr{M}$, and $\boldsymbol{\varepsilon}_*$ is a q-dimensional random error vector. This article is of review type, providing easy-to-read collection of useful results concerning specific properties of the covariance matrix of the BLUE. Most results appear in literature but our aim is to create a convenient "package" of some essential results.

Keywords: BLUE, BLUP, covariance matrix, linear statistical model, Löwner partial ordering, generalized inverse.
Classification: 62J05; 62J10

1. Introduction

We will consider the general linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \quad \text{or shortly the triplet } \mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}\,, \tag{1.1}$$

where $\mathbf{X}$ is a known $n \times p$ model matrix, the vector $\mathbf{y}$ is an observable n-dimensional random vector (so-called response vector), $\boldsymbol{\beta}$ is p-dimensional vector of unknown but fixed parameters, and $\boldsymbol{\varepsilon}$ is an unobservable vector of random errors with expectation $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, and covariance matrix $\mathrm{cov}(\boldsymbol{\varepsilon}) = \mathbf{V}$. Often the covariance matrix is of the type $\mathrm{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V}$, where σ^2 is an unknown nonzero constant. However, in most of our considerations σ^2 has no role and in such cases we omit it. The nonnegative definite matrix $\mathbf{V}$ is known and can be singular. The set of nonnegative definite $n \times n$ matrices is denoted as NND_n.

As the covariance matrix is so central concept for our considerations, we might recall that under $\mathscr{M}$ it is defined as

$$\mathbf{V} = \mathrm{cov}(\mathbf{y}) = \mathrm{E}(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})', \quad \text{where } \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} = \mathrm{E}(\mathbf{y})\,, \tag{1.2}$$

and $'$ denotes the transpose of the matrix argument. Thus obviously,

$$\mathrm{cov}(\mathbf{A}\mathbf{y}) = \mathbf{A}\mathbf{V}\mathbf{A}', \tag{1.3}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$, the set of $m \times n$ real matrices. Instead of "covariance matrix", some authors use the name "variance-covariance matrix" or "dispersion matrix". The cross-covariance matrix between random vectors $\mathbf{u}$ and $\mathbf{v}$ is defined as

$$\mathrm{cov}(\mathbf{u}, \mathbf{v}) = \mathrm{E}[\mathbf{u} - \mathrm{E}(\mathbf{u})][\mathbf{v} - \mathrm{E}(\mathbf{v})]'. \tag{1.4}$$

Then some words about the notation. The symbols $\mathbf{A}^-$, $\mathbf{A}^+$, $\mathscr{C}(\mathbf{A})$, and $\mathscr{C}(\mathbf{A})^\perp$, denote, respectively, a generalized inverse, the (unique) Moore–Penrose inverse, the column space, and the orthogonal complement of the column space of the matrix $\mathbf{A}$. The Moore–Penrose inverse $\mathbf{A}^+$ is defined as a unique matrix satisfying the following four conditions:

$$\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}, \quad \mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+, \quad (\mathbf{A}\mathbf{A}^+)' = \mathbf{A}\mathbf{A}^+, \quad (\mathbf{A}^+\mathbf{A})' = \mathbf{A}^+\mathbf{A}. \tag{1.5}$$

[1] Corresponding author

Notation $\mathbf{A}^-$ refers to any matrix satisfying $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$. By $(\mathbf{A} : \mathbf{B})$ we denote the partitioned matrix with $\mathbf{A}_{a\times b}$ and $\mathbf{B}_{a\times c}$ as submatrices. The symbol $\mathbf{A}^\perp$ stands for any matrix satisfying $\mathscr{C}(\mathbf{A}^\perp) = \mathscr{C}(\mathbf{A})^\perp$. Furthermore, we will use $\mathbf{P}_\mathbf{A} = \mathbf{A}\mathbf{A}^+ = \mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}'$ to denote the orthogonal projector (with respect to the standard inner product) onto the column space $\mathscr{C}(\mathbf{A})$, and $\mathbf{Q}_\mathbf{A} = \mathbf{I} - \mathbf{P}_\mathbf{A}$, where $\mathbf{I}$ refers to the identity matrix of conformable dimension. In particular, it appears to be useful to denote

$$\mathbf{H} = \mathbf{P}_\mathbf{X}\,, \quad \mathbf{M} = \mathbf{I}_n - \mathbf{P}_\mathbf{X}\,, \tag{1.6}$$

in which case, for any vector $\mathbf{y} \in \mathbb{R}^n$,

$$\min_{\boldsymbol{\mu}\in\mathscr{C}(\mathbf{X})} \|\mathbf{y} - \boldsymbol{\mu}\|^2 = \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 = \|\mathbf{y} - \hat{\boldsymbol{\mu}}\|^2 = \mathbf{y}'\mathbf{M}\mathbf{y}\,, \tag{1.7}$$

where $\hat{\boldsymbol{\mu}} = \mathbf{H}\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}}$, with $\hat{\boldsymbol{\beta}}$ being any (least-squares) solution to so-called normal equation

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}\,. \tag{1.8}$$

Notice that in (1.7) and (1.8) we use $\mathbf{y}$, $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\mu}$ as "merely mathematical" vectors, not random vectors nor parameters of the model $\mathscr{M}$.

The notation $\mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}}$, where $\mathbf{V}$ is positive definite, refers to the orthogonal projector onto $\mathscr{C}(\mathbf{X})$ with respect to the inner product matrix $\mathbf{V}^{-1}$, i.e.,

$$\min_{\boldsymbol{\mu}\in\mathscr{C}(\mathbf{X})} (\mathbf{y} - \boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) = \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2_{\mathbf{V}^{-1}} = \|\mathbf{y} - \tilde{\boldsymbol{\mu}}\|^2_{\mathbf{V}^{-1}}\,, \tag{1.9}$$

where $\|\mathbf{a}\|^2_{\mathbf{V}^{-1}} = \mathbf{a}'\mathbf{V}^{-1}\mathbf{a}$ for $\mathbf{a} \in \mathbb{R}^n$, and

$$\tilde{\boldsymbol{\mu}} = \mathbf{X}\tilde{\boldsymbol{\beta}} = \mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}}\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^-\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\,, \tag{1.10}$$

with $\tilde{\boldsymbol{\beta}}$ being any solution to the generalized normal equation

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\,. \tag{1.11}$$

We shall concentrate on the linear unbiased estimators, LUEs, and hence we need the concept of estimability. The parametric function $\boldsymbol{\eta} = \mathbf{K}\boldsymbol{\beta}$, where $\mathbf{K} \in \mathbb{R}^{q\times p}$, is estimable under $\mathscr{M}$ if and only if there exists a matrix $\mathbf{B} \in \mathbb{R}^{q\times n}$ such that

$$\mathrm{E}(\mathbf{B}\mathbf{y}) = \mathbf{B}\mathbf{X}\boldsymbol{\beta} = \mathbf{K}\boldsymbol{\beta} \quad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^p, \ \text{ i.e., } \ \mathbf{B}\mathbf{X} = \mathbf{K}\,. \tag{1.12}$$

Such a matrix $\mathbf{B}$ exists only when

$$\mathscr{C}(\mathbf{K}') \subset \mathscr{C}(\mathbf{X}')\,, \tag{1.13}$$

which, therefore, is the condition for $\boldsymbol{\eta} = \mathbf{K}\boldsymbol{\beta}$ to be estimable. The LUE $\mathbf{B}\mathbf{y}$ is the best linear unbiased estimator, BLUE, of estimable $\mathbf{K}\boldsymbol{\beta}$ if $\mathbf{B}\mathbf{y}$ has the smallest covariance matrix in the Löwner sense among all linear unbiased estimators of $\mathbf{K}\boldsymbol{\beta}$:

$$\mathrm{cov}(\mathbf{B}\mathbf{y}) \leq_{\mathrm{L}} \mathrm{cov}(\mathbf{B}_\#\mathbf{y}) \quad \text{for all } \mathbf{B}_\# : \mathbf{B}_\#\mathbf{X} = \mathbf{K}\,, \tag{1.14}$$

that is, $\operatorname{cov}(\mathbf{B}_{\#}\mathbf{y}) - \operatorname{cov}(\mathbf{B}\mathbf{y})$ is nonnegative definite for all $\mathbf{B}_{\#} : \mathbf{B}_{\#}\mathbf{X} = \mathbf{K}$.

We assume the model $\mathscr{M}$ to be consistent in the sense that the observed value of $\mathbf{y}$ lies in $\mathscr{C}(\mathbf{X} : \mathbf{V})$ with probability 1. Hence we assume that under the model $\mathscr{M}$,

$$\mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = \mathscr{C}(\mathbf{X} : \mathbf{V}\mathbf{M}) = \mathscr{C}(\mathbf{X}) \oplus \mathscr{C}(\mathbf{V}\mathbf{M}) , \tag{1.15}$$

where $\oplus$ refers to the direct sum. For the equality $\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{X} : \mathbf{V}\mathbf{M})$, see, e.g., Rao [61, Lemma 2.1]. All models that we consider are assumed to be consistent in the sense of (1.15).

Let $\mathbf{A}$ and $\mathbf{B}$ be $m \times n$ matrices. Then, in the consistent linear model $\mathscr{M}$, the estimators $\mathbf{A}\mathbf{y}$ and $\mathbf{B}\mathbf{y}$ are said to be equal with probability 1 if

$$\mathbf{A}\mathbf{y} = \mathbf{B}\mathbf{y} \quad \text{for all } \mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V}) , \tag{1.16}$$

which will be a crucial property in our considerations. Sometimes, when talking about the equality of estimators, we drop off the phrase “with probability 1”. For the equality of two estimators, see, e.g., Groß & Trenkler [22].

As for the structure of this article, in Section 2 we consider some properties of the ordinary least squares estimator, OLSE. We introduce a simple version of the Gauss–Markov theorem and use that to find the BLUE when $\mathbf{V}$ is positive definite. While doing that we touch the concept of linear sufficiency. The covariance matrix of OLSE is studied in Section 3. The fundamental BLUE equation is given Section 4 and it is utilized for finding general expressions for the BLUE in Section 5. We study the relative efficiency of OLSE with respect to BLUE in Section 6. The further sections deal with weighted sum of squares of errors (needed in particular in hypothesis testing), peculiar connection between the BLUE’s covariance matrix and specific proper eigenvalues, and the shorted matrix. The paper is completed with a short section on the best linear unbiased prediction. Our aim is to call main results Theorems while Lemmas refer to more technical results. This is a review-type article attempting to provide a readable summary of some useful properties related to the concept of BLUE and in particular, to the covariance matrix of the BLUE.

2. Ordinary least squares estimator and the Gauss–Markov theorem

Consider now the model $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$. Then the ordinary least squares estimator, OLSE, for $\boldsymbol{\beta}$ is the solution minimizing the quantity $\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$ with respect to $\boldsymbol{\beta}$ yielding to the normal equation $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$. Thus, if $\mathbf{X}$ has full column rank, the OLSE of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}^{+}\mathbf{y} , \tag{2.1}$$

while its covariance matrix is

$$\operatorname{cov}(\hat{\boldsymbol{\beta}}) = \operatorname{cov}(\mathbf{X}^{+}\mathbf{y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \tag{2.2}$$

Notice that in (2.1) we keep $\mathbf{y}$ as a random vector. The set of *all* vectors $\hat{\boldsymbol{\beta}}$ satisfying $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$, can be written as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} + [\mathbf{I}_p - (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}]\mathbf{t}\,, \tag{2.3}$$

where $(\mathbf{X}'\mathbf{X})^{-}$ is an arbitrary (but fixed) generalized inverse of $\mathbf{X}'\mathbf{X}$ and $\mathbf{t} \in \mathbb{R}^p$ is free to vary. On the other hand, every solution to the normal equation can be written as $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$ for some $(\mathbf{X}'\mathbf{X})^{-}$. Of course, it is questionable whether it is quite correct to call $\hat{\boldsymbol{\beta}}$ an estimator when it is not unique (after $\mathbf{y}$ is being observed); might be then better to call it a least-squares-solution.

If $\mathbf{K}\boldsymbol{\beta}$, where $\mathbf{K} \in \mathbb{R}^{q\times p}$, is estimable, then $\mathbf{K}\hat{\boldsymbol{\beta}}$, i.e., the OLSE of $\mathbf{K}\boldsymbol{\beta}$ is unique whatever choice of $\hat{\boldsymbol{\beta}}$ we use. This can be seen by premultiplying (2.3) by $\mathbf{K}$ yielding

$$\mathbf{K}\hat{\boldsymbol{\beta}} = \mathbf{K}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}\,, \tag{2.4}$$

and utilizing Lemma 2.2.4 of Rao & Mitra [64] saying the following:

LEMMA 2.1. *For nonnull matrices* $\mathbf{A}$ *and* $\mathbf{C}$ *the following holds:*

(a) $\mathbf{A}\mathbf{B}^{-}\mathbf{C} = \mathbf{A}\mathbf{B}^{+}\mathbf{C}$ *for all* $\mathbf{B}^{-} \iff \mathscr{C}(\mathbf{C}) \subset \mathscr{C}(\mathbf{B})\ \&\ \mathscr{C}(\mathbf{A}') \subset \mathscr{C}(\mathbf{B}')$.

(b) $\mathbf{A}\mathbf{A}^{-}\mathbf{C} = \mathbf{C}$ *or some (and hence for all)* $\mathbf{A}^{-} \iff \mathscr{C}(\mathbf{C}) \subset \mathscr{C}(\mathbf{A})$.

One obvious choice in (2.4) is $\mathbf{K} = \mathbf{X}$ yielding

$$\text{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} = \mathbf{P}_{\mathbf{X}}\mathbf{y} = \mathbf{H}\mathbf{y} = \hat{\boldsymbol{\mu}}\,, \tag{2.5}$$

and

$$\text{cov}(\hat{\boldsymbol{\mu}}) = \mathbf{HVH} \quad \text{under } \mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}. \tag{2.6}$$

Obviously $\mathbf{Hy}$ is a linear unbiased estimator for $\mathbf{X}\boldsymbol{\beta}$ as $\text{E}(\mathbf{Hy}) = \mathbf{HX}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta}$. Let $\mathbf{By}$ be another LUE of $\mathbf{X}\boldsymbol{\beta}$, i.e., $\mathbf{B}$ satisfies $\mathbf{BX} = \mathbf{X}$ and thereby $\mathbf{BH} = \mathbf{H} = \mathbf{HB}'$. Thus, under the model $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{I}_n\}$:

$$\begin{aligned}
\text{cov}(\mathbf{By} - \mathbf{Hy}) &= \text{cov}(\mathbf{By}) + \text{cov}(\mathbf{Hy}) - \text{cov}(\mathbf{By}, \mathbf{Hy}) - \text{cov}(\mathbf{Hy}, \mathbf{By}) \\
&= \mathbf{BB}' + \mathbf{H} - \mathbf{BH} - \mathbf{HB}' \\
&= \mathbf{BB}' - \mathbf{H}\,,
\end{aligned} \tag{2.7}$$

which implies

$$\mathbf{BB}' - \mathbf{H} = \text{cov}(\mathbf{By}) - \text{cov}(\hat{\boldsymbol{\mu}}) = \text{cov}(\mathbf{By} - \hat{\boldsymbol{\mu}}) \geq_{\text{L}} \mathbf{0}\,, \tag{2.8}$$

where the Löwner inequality follows from the fact that every covariance matrix is nonnegative definite. Now (2.8) means that under $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{I}_n\}$ we have the Löwner ordering

$$\text{cov}(\hat{\boldsymbol{\mu}}) \leq_{\text{L}} \text{cov}(\mathbf{By}) \quad \text{for all } \mathbf{B}\text{: } \mathbf{BX} = \mathbf{X}\,. \tag{2.9}$$

Thus we have proved a simple version of the *Gauss–Markov theorem*:

THEOREM 2.2. *Under the model* $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{I}_n\}$*,*

$$\mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})\,, \quad \textit{i.e.,} \quad \hat{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}} \ \textit{with probability}\ 1, \tag{2.10}$$

and for any estimable $\boldsymbol{\eta} = \mathbf{K}\boldsymbol{\beta}$*,*

$$\mathrm{OLSE}(\mathbf{K}\boldsymbol{\beta}) = \mathrm{BLUE}(\mathbf{K}\boldsymbol{\beta})\,, \quad \textit{i.e.,} \quad \hat{\boldsymbol{\eta}} = \tilde{\boldsymbol{\eta}} \ \textit{with probability}\ 1. \tag{2.11}$$

Consider now the model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, where $\mathbf{V}$ is positive definite, and suppose that $\mathbf{V}^{1/2}$ is the positive definite square root of $\mathbf{V}$. Premultiplying $\mathscr{M}$ by $\mathbf{V}^{-1/2}$ gives the transformed model

$$\mathscr{M}_{\#} = \{\mathbf{V}^{-1/2}\mathbf{y},\, \mathbf{V}^{-1/2}\mathbf{X}\boldsymbol{\beta},\, \mathbf{I}_n\} = \{\mathbf{y}_{\#},\, \mathbf{X}_{\#}\boldsymbol{\beta},\, \mathbf{I}_n\}\,. \tag{2.12}$$

Now, in light of (2.10), the BLUE of $\mathbf{X}\boldsymbol{\beta}$ under $\mathscr{M}_{\#}$ equals the OLSE under $\mathscr{M}_{\#}$:

$$\mathrm{BLUE}(\mathbf{X}_{\#}\boldsymbol{\beta} \mid \mathscr{M}_{\#}) = \mathrm{OLSE}(\mathbf{X}_{\#}\boldsymbol{\beta} \mid \mathscr{M}_{\#}) = \mathbf{V}^{-1/2}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\,, \tag{2.13}$$

so that

$$\mathbf{V}^{-1/2}\,\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_{\#}) = \mathbf{V}^{-1/2}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\,, \tag{2.14}$$

and thus

$$\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_{\#}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \mathbf{P}_{\mathbf{X},\mathbf{V}^{-1}}\mathbf{y}\,, \tag{2.15}$$

where $\mathbf{P}_{\mathbf{X},\mathbf{V}^{-1}}$, as in (1.10), is the orthogonal projector onto $\mathscr{C}(\mathbf{X})$ when the inner product matrix is $\mathbf{V}^{-1}$. Here is a crucial question: is the BLUE of $\mathbf{X}\boldsymbol{\beta}$ under $\mathscr{M}_{\#}$ the same as under $\mathscr{M}$, in other words, has the transformation done via $\mathbf{V}^{-1/2}$ any effect on the BLUE of $\mathbf{X}\boldsymbol{\beta}$? The answer is that indeed there is no effect and that

$$\mathbf{P}_{\mathbf{X},\mathbf{V}^{-1}}\mathbf{y} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}) = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_{\#}) = \mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_{\#})\,. \tag{2.16}$$

The result (2.16), sometimes referred to as the Aitken-approach, see Aitken [1, 1936], is well known in statistical textbooks. However, usually the textbooks give the proof by assuming that the expression $\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}) = \mathbf{P}_{\mathbf{X},\mathbf{V}^{-1}}\mathbf{y}$ is known. Interestingly, a more common approach would be to consider whether $\mathbf{V}^{-1/2}\mathbf{y}$ would be a *linearly sufficient* statistics for $\mathbf{X}\boldsymbol{\beta}$. We describe this concept now briefly. Let $\mathbf{F}$ be an $f \times n$ matrix. Then $\mathbf{F}\mathbf{y}$ is called linearly sufficient for $\mathbf{X}\boldsymbol{\beta}$ under $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ if there exists a matrix $\mathbf{A}_{q\times f}$ such that $\mathbf{A}\mathbf{F}\mathbf{y}$ is the BLUE for $\mathbf{X}\boldsymbol{\beta}$. The transformed model

$$\mathscr{M}_t = \{\mathbf{F}\mathbf{y}, \mathbf{F}\mathbf{X}\boldsymbol{\beta}, \mathbf{F}\mathbf{V}\mathbf{F}'\} \tag{2.17}$$

has very strong connection with the concept of linear sufficiency. The equality of BLUEs under the original model and the transformed model can be characterized via linear sufficiency. The following Lemma 2.3 collects some useful related results. For proofs, see, e.g., Baksalary & Kala [5,6], Drygas [19], Tian & Puntanen [69, Th. 2.8], and Kala et al. [34, Th. 2].

LEMMA 2.3. *Let $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ be estimable under $\mathscr{M}_t$. Then the following statements are equivalent:*

(a) $\mathbf{Fy}$ *is linearly sufficient for* $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$.

(b) $\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}) = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_t)$*, or shortly,* $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}_t$ *with probability* 1.

(c) $\mathrm{cov}(\tilde{\boldsymbol{\mu}}) = \mathrm{cov}(\tilde{\boldsymbol{\mu}}_t)$.

(d) $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WF}')$*, where* $\mathbf{W} \in \mathcal{W}$*, with* $\mathcal{W}$ *being defined as in* (4.6).

For the class $\mathcal{W}$ of nonnegative definite matrices of type $\mathbf{W} = \mathbf{V} + \mathbf{XUU}'\mathbf{X}'$, where $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$, see Lemma 4.2 in Section 4. It is clear from part (d) of Lemma 2.3 that for any nonsingular $\mathbf{F}$, the statistics $\mathbf{Fy}$ is linearly sufficient, in particular, this holds for $\mathbf{V}^{-1/2}\mathbf{y}$. We may mention that according to Kala et al. [33, Th. 4], in (d) above, $\mathcal{W}$ can be replaced with $\mathcal{W}_*$; see (4.7) in Section 4.

In view of (1.13), the vector $\boldsymbol{\beta}$ itself is estimable if and only if $\mathscr{C}(\mathbf{I}_p) \subset \mathscr{C}(\mathbf{X}')$, i.e., $\mathbf{X}$ has full column rank. We denote the BLUE of $\boldsymbol{\beta}$ as $\tilde{\boldsymbol{\beta}} = (\tilde{\beta}_1, \dots, \tilde{\beta}_p)'$. Because the Löwner ordering is so strong ordering, see, e.g., Puntanen et al. [56, p. 12], we have the following inequalities:

$$\mathrm{var}(\tilde{\beta}_i) \le \mathrm{var}(\beta_i^{\#}), \quad i = 1, \dots, p, \tag{2.18a}$$

$$\mathrm{trace}\,\mathrm{cov}(\tilde{\boldsymbol{\beta}}) \le \mathrm{trace}\,\mathrm{cov}(\boldsymbol{\beta}^{\#}), \tag{2.18b}$$

$$\det \mathrm{cov}(\tilde{\boldsymbol{\beta}}) \le \det \mathrm{cov}(\boldsymbol{\beta}^{\#}), \tag{2.18c}$$

$$\mathrm{ch}_i[\mathrm{cov}(\tilde{\boldsymbol{\beta}})] \le \mathrm{ch}_i[\mathrm{cov}(\boldsymbol{\beta}^{\#})], \quad i = 1, \dots, p, \tag{2.18d}$$

$$\|\mathrm{cov}(\tilde{\boldsymbol{\beta}})\|_F \le \|\mathrm{cov}(\boldsymbol{\beta}^{\#})\|_F, \tag{2.18e}$$

$$\|\mathrm{cov}(\tilde{\boldsymbol{\beta}})\|_2 \le \|\mathrm{cov}(\boldsymbol{\beta}^{\#})\|_2, \tag{2.18f}$$

for any $\boldsymbol{\beta}^{\#}$ which is a linear unbiased estimator of $\boldsymbol{\beta}$. Above $\mathrm{var}(\cdot)$ refers to the variance of a random variable, $\|\cdot\|_F$ and $\|\cdot\|_2$ refer to the Frobenius norm and the spectral norm, respectively, $\det(\cdot)$ refers to determinant, and $\mathrm{ch}_i(\cdot)$ refers to the ith largest eigenvalue.

3. OLSE's covariance matrix in the full rank model

Let us first recall that when $\mathbf{X}$ has a full column rank and $\mathbf{V}$ is positive definite, in which case we say that $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ is a full rank model, then the OLSE and

BLUE of $\boldsymbol{\beta}$ are, respectively, $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, and $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$, while the corresponding covariance matrices are

$$\operatorname{cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \quad \operatorname{cov}(\tilde{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \tag{3.1}$$

Hence we have the Löwner ordering

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \geq_{\mathrm{L}} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}, \tag{3.2}$$

i.e., the matrix

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = \mathbf{D} \tag{3.3}$$

is nonnegative definite. If $\mathbf{X}$ does not have a full column rank (but $\mathbf{V}$ is positive definite) then

$$\mathbf{X}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\mu}} = \mathbf{H}\mathbf{y}, \quad \mathbf{X}\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\mu}} = \mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}}\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \tag{3.4a}$$

$$\operatorname{cov}(\hat{\boldsymbol{\mu}}) = \mathbf{H}\mathbf{V}\mathbf{H} \geq_{\mathrm{L}} \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}' = \operatorname{cov}(\tilde{\boldsymbol{\mu}})\,, \tag{3.4b}$$

and

$$\mathbf{H}\mathbf{V}\mathbf{H} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X} = \mathbf{E}\,, \tag{3.5}$$

where $\mathbf{E}$ is nonnegative definite. What is very interesting here is that there is an alternative useful expression for $\mathbf{D}$ (as well as for $\mathbf{E}$) available as shown in Theorem 3.1.

Among the first places where Theorem 3.1 occurs are probably the papers by Khatri [35, Lemma 1] and Rao [57, Lemmas 2a–2c]; see also Rao [59, Problem 33, p. 77].

THEOREM 3.1. *Consider the linear model* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, *where* $\mathbf{X}$ *has full column rank and* $\mathbf{V}$ *is positive definite, and denote* $\mathbf{H} = \mathbf{P}_{\mathbf{X}}$, $\mathbf{M} = \mathbf{I}_n - \mathbf{H}$. *Then*

$$\begin{aligned}\operatorname{cov}(\tilde{\boldsymbol{\beta}}) &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}[\mathbf{X}'\mathbf{V}\mathbf{X} - \mathbf{X}'\mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}\mathbf{X}](\mathbf{X}'\mathbf{X})^{-1} \\ &= \mathbf{X}^{+}[\mathbf{V} - \mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}](\mathbf{X}^{+})' \\ &= \operatorname{cov}(\hat{\boldsymbol{\beta}}) - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},\end{aligned} \tag{3.6}$$

and hence

(a) $\operatorname{cov}(\hat{\boldsymbol{\beta}}) - \operatorname{cov}(\tilde{\boldsymbol{\beta}}) = \mathbf{X}^{+}\mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}(\mathbf{X}^{+})'$,

(b) $\operatorname{cov}(\mathbf{X}\hat{\boldsymbol{\beta}}) - \operatorname{cov}(\mathbf{X}\tilde{\boldsymbol{\beta}}) = \mathbf{H}\mathbf{V}\mathbf{H} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}' = \mathbf{H}\mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}\mathbf{H}$,

(c) $\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1} = \mathbf{I}_n - \mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}$,

(d) $\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1} = \mathbf{H} - \mathbf{H}\mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}$.

In (b)–(d) *the matrix* $\mathbf{X}$ *does not need to have full column rank.*

Proof. To prove (3.6) we first observe following:

$$\begin{aligned}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}' &= \mathbf{V}^{1/2}\mathbf{V}^{-1/2}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1/2}\mathbf{V}^{1/2} \\ &= \mathbf{V}^{1/2}\mathbf{P}_{\mathbf{V}^{-1/2}\mathbf{X}}\mathbf{V}^{1/2} \\ &= \mathbf{V}^{1/2}(\mathbf{I}_n - \mathbf{P}_{(\mathbf{V}^{-1/2}\mathbf{X})^{\perp}})\mathbf{V}^{1/2} \\ &= \mathbf{V}^{1/2}(\mathbf{I}_n - \mathbf{P}_{\mathbf{V}^{1/2}\mathbf{M}})\mathbf{V}^{1/2} \\ &= \mathbf{V} - \mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}. \end{aligned} \tag{3.7}$$

Above we have used the fact that for a positive definite $\mathbf{V}$ we have

$$\mathscr{C}(\mathbf{V}^{-1/2}\mathbf{X})^{\perp} = \mathscr{C}(\mathbf{V}^{1/2}\mathbf{M})\,. \tag{3.8}$$

For properties of $^{\perp}$, we refer to [39]. Supposing that $\mathbf{X}$ has full column rank, then post- and premultiplying (3.7) by $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ gives (3.6). Claim (c) follows from postmultiplying (3.7) by $\mathbf{V}^{-1}$ and (d) comes from premultiplying (c) by $\mathbf{H}$. □

Notice that if $\mathbf{V}$ is positive definite, then by Theorem 3.1, we have

$$\tilde{\boldsymbol{\mu}} = \mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}}\mathbf{y} = \mathbf{H} - \mathbf{H}\mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y} = [\mathbf{I}_n - \mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}]\mathbf{y}\,, \tag{3.9}$$

and the covariance matrix of $\tilde{\boldsymbol{\mu}}$ has representations

$$\begin{aligned}\operatorname{cov}(\tilde{\boldsymbol{\mu}}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}' &= \mathbf{H}\mathbf{V}\mathbf{H} - \mathbf{H}\mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}\mathbf{H} \\ &= \mathbf{V} - \mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}, \end{aligned} \tag{3.10}$$

and

$$\operatorname{cov}(\hat{\boldsymbol{\mu}}) - \operatorname{cov}(\tilde{\boldsymbol{\mu}}) = \mathbf{H}\mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}\mathbf{H} = \operatorname{cov}(\hat{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}})\,. \tag{3.11}$$

Because $\hat{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\mu}}$ coincide if (and only if) their covariance matrices coincide, the "size" of the matrix $\mathbf{H}\mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}\mathbf{H}$ in a way describes the goodness of $\hat{\boldsymbol{\mu}}$ with respect to $\tilde{\boldsymbol{\mu}}$. We will later see that the last two presentations in (3.10) are actually valid even if $\mathbf{V}$ is singular and the same concerns (3.11).

We may say a few words about the matrix $\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}$ which we denote as

$$\dot{\mathbf{M}} = \mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\,. \tag{3.12}$$

If $\mathbf{V}$ is positive definite and $\mathbf{Z}$ is a matrix with the property $\mathscr{C}(\mathbf{Z}) = \mathscr{C}(\mathbf{M})$, then we obviously have

$$\begin{aligned}\dot{\mathbf{M}} = \mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M} &= \mathbf{V}^{-1/2}\mathbf{P}_{\mathbf{V}^{1/2}\mathbf{M}}\mathbf{V}^{-1/2} \\ &= \mathbf{V}^{-1/2}\mathbf{P}_{\mathbf{V}^{1/2}\mathbf{Z}}\mathbf{V}^{-1/2} = \mathbf{Z}(\mathbf{Z}'\mathbf{V}\mathbf{Z})^{-}\mathbf{Z}', \end{aligned} \tag{3.13}$$

which is clearly unique. In general, when $\mathbf{V}$ is nonnegative definite, the matrix $\dot{\mathbf{M}}$ is not necessarily unique with respect to the choice of $(\mathbf{MVM})^-$. It can be shown that

$$\mathbf{M}(\mathbf{MVM})^-\mathbf{M} = \mathbf{M}(\mathbf{MVM})^+\mathbf{M} \quad \iff \quad \operatorname{rank}(\mathbf{X} : \mathbf{V}) = n. \tag{3.14}$$

However, we always have

$$\mathbf{M}(\mathbf{MVM})^+\mathbf{M} = (\mathbf{MVM})^+\mathbf{M} = \mathbf{M}(\mathbf{MVM})^+ = (\mathbf{MVM})^+. \tag{3.15}$$

The matrix $\dot{\mathbf{M}}$ and its versions appear to be very handy in many ways related to linear model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$. For example, consider the partitioned linear model $\mathscr{M}_{12} = \{\mathbf{y}, \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2, \mathbf{V}\}$, where $\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2)$ has full column rank and $\mathbf{V}$ is positive definite. Premultiplying $\mathscr{M}_{12}$ by $\mathbf{M}_1 = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}_1}$ yields a reduced model

$$\mathscr{M}_{12\cdot 1} = \{\mathbf{M}_1\mathbf{y},\ \mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2,\ \mathbf{MVM}\}\,. \tag{3.16}$$

Now it appers, see, e.g., Gross & Puntanen [21], that

$$\tilde{\boldsymbol{\beta}}_2(\mathscr{M}_{12}) = \tilde{\boldsymbol{\beta}}_2(\mathscr{M}_{12\cdot 1}) = (\mathbf{X}_2'\dot{\mathbf{M}}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\dot{\mathbf{M}}_1\mathbf{y}, \tag{3.17}$$

and $\operatorname{cov}(\tilde{\boldsymbol{\beta}}_2) = (\mathbf{X}_2'\dot{\mathbf{M}}_1\mathbf{X}_2)^{-1}$, where $\dot{\mathbf{M}}_1 = \mathbf{M}_1(\mathbf{M}_1\mathbf{V}\mathbf{M}_1)^-\mathbf{M}_1$.

For a review of the properties of $\dot{\mathbf{M}}$, see Puntanen et al. [56, Ch. 15].

4. The fundamental BLUE equation

Theorem 4.1 below provides so-called fundamental BLUE equations. For the proofs, see, e.g., Drygas [18, p. 55], Rao [60, p. 282], and Puntanen et al. [56, Th. 10].

THEOREM 4.1. *Consider the linear model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ and let $\boldsymbol{\eta} = \mathbf{K}\boldsymbol{\beta}$, where $\mathbf{K} \in \mathbb{R}^{q\times p}$, be estimable, so that $\mathscr{C}(\mathbf{K}') \subset \mathscr{C}(\mathbf{X}')$. Then the linear estimator $\mathbf{By}$ is the* BLUE *of $\boldsymbol{\eta} = \mathbf{K}\boldsymbol{\beta}$ if and only if $\mathbf{B} \in \mathbb{R}^{q\times n}$ satisfies the equation*

$$\mathbf{B}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{K} : \mathbf{0})\,. \tag{4.1}$$

In particular, $\mathbf{Cy}$ is the BLUE *for $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ if and only if $\mathbf{C} \in \mathbb{R}^{n\times n}$ satisfies the equation*

$$\mathbf{C}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{X} : \mathbf{0})\,. \tag{4.2}$$

Of course, in (4.1) and (4.2) we can replace $\mathbf{X}^{\perp}$ with $\mathbf{M}$. Equation (4.2) is always solvable for $\mathbf{C}$ while (4.1) is solvable whenever $\mathbf{K}\boldsymbol{\beta}$ is estimable.

It is clear that $\mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^-\mathbf{X}'\mathbf{V}^{-1}$ is one choice for $\mathbf{C}$ in (4.2) and thus we have the well-known result, see (2.16):

$$\tilde{\boldsymbol{\mu}} = \mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}}\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^-\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\,. \tag{4.3}$$

Without rank conditions, one well-known solution for $\mathbf{C}$ in (4.2) appears to be (see Theorem 5.1 below)

$$\mathbf{P}_{\mathbf{X};\mathbf{W}^-} := \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^-, \tag{4.4}$$

and for $\mathbf{B}$ in (4.1),

$$\mathbf{B} = \mathbf{K}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^-, \tag{4.5}$$

where $\mathbf{W}$ is a matrix belonging to the set of nonnegative definite matrices defined as

$$\mathcal{W} = \left\{\mathbf{W} \in \mathrm{NND}_n : \mathbf{W} = \mathbf{V} + \mathbf{XUU}'\mathbf{X}',\ \mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})\right\}. \tag{4.6}$$

We could replace $\mathcal{W}$ with an extended set of matrices of the type

$$\mathcal{W}_* = \left\{\mathbf{W} \in \mathbb{R}^{n\times n} : \mathbf{W} = \mathbf{V} + \mathbf{XTX}',\ \mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})\right\}. \tag{4.7}$$

Notice that $\mathbf{W}$ that belongs to $\mathcal{W}_*$ is not necessarily nonnegative definite and it can be nonsymmetric. In light of part (a) of Lemma 2.1, the matrix $\mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'$ is invariant for any choices of the generalized inverses involved and the same concerns

$$\mathbf{P}_{\mathbf{X};\mathbf{W}^-}\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^-\mathbf{y} \quad \text{for } \mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V}). \tag{4.8}$$

Observe that

$$\mathbf{P}_{\mathbf{X};\mathbf{W}^+} = \mathbf{X}(\mathbf{X}'\mathbf{W}^+\mathbf{X})^-\mathbf{X}'\mathbf{W}^+ = \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^+, \tag{4.9}$$

which is unique for any generalized inverses denoted as superscript $^-$. It is worth noting that $\mathbf{P}_{\mathbf{X};\mathbf{W}^+}$ may not be a "regular" orthogonal projector with respect to inner product matrix $\mathbf{W}^+$ as $\mathbf{W}^+$ may not be positive definite. However, it can be shown that for all $\mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V})$ and for all $\boldsymbol{\beta} \in \mathbb{R}^p$ the following holds:

$$(\mathbf{y} - \mathbf{P}_{\mathbf{X};\mathbf{W}^+}\mathbf{y})'\mathbf{W}^+(\mathbf{y} - \mathbf{P}_{\mathbf{X};\mathbf{W}^+}\mathbf{y}) \le (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}^+(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \tag{4.10}$$

where the left-hand side can be expressed as $\mathbf{y}'\dot{\mathbf{M}}\mathbf{y}$; see Section 7. For the concept of generalized orthogonal projector with respect to nonnegative definite inner product matrix, see Mitra & Rao [49] and Puntanen et al. [56, §2.5]. It is noteworthy that $\mathbf{P}_{\mathbf{X};\mathbf{W}^-}\mathbf{y}$ can be expressed as $\mathbf{X}\tilde{\boldsymbol{\beta}}$, where $\tilde{\boldsymbol{\beta}}$ is any solution to

$$\mathbf{X}'\mathbf{W}^-\mathbf{X}\tilde{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{W}^-\mathbf{y}. \tag{4.11}$$

The following lemma comprises some useful properties of the class $\mathcal{W}_*$.

LEMMA 4.2. *Consider the model* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ *and let* $\mathcal{W}_*$ *be defined as in* (4.7). *Then the following statements concerning* $\mathbf{W}$ *belonging to* $\mathcal{W}_*$ *are equivalent:*

(a) $\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{W})$,

(b) $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{W})$,

(c) $\mathbf{X}'\mathbf{W}^{-}\mathbf{X}$ *is invariant for any choice of* $\mathbf{W}^{-}$*,*

(d) $\mathscr{C}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X}) = \mathscr{C}(\mathbf{X}')$ *for any choice of* $\mathbf{W}^{-}$*,*

(e) $\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-}\mathbf{X} = \mathbf{X}$ *for any choices of* $\mathbf{W}^{-}$ *and* $(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}$.

Moreover, each of these statements is equivalent also to $\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{W}')$*, and hence to the statements* (b)–(e) *by replacing* $\mathbf{W}$ *with* $\mathbf{W}'$*.*

Observe that obviously $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{W}')$ and that the invariance properties in (d) and (e) concern not only the the choice of the generalized inverse of $\mathbf{W}$ but also the choice of $\mathbf{W} \in \mathcal{W}_*$. For further properties of $\mathcal{W}_*$, see, e.g., Baksalary & Puntanen [8, Th. 1], Baksalary et al. [10, Th. 2], Baksalary & Mathew [7, Th. 2], and Puntanen et al. [56, §12.3].

5. General expressions for the BLUE

Using Lemma 4.2 and the equality

$$\mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MVM} = \mathbf{VM}\,, \tag{5.1}$$

which follows from part (b) of Lemma 2.1, it is easy to prove the following:

THEOREM 5.1. *The solution for* $\mathbf{G}$ *satisfying*

$$\mathbf{G}(\mathbf{X} : \mathbf{VM}) = (\mathbf{X} : \mathbf{0}) \tag{5.2}$$

can be expressed, for example, in the following ways:

(a) $\mathbf{G}_1 = \mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-}$, *where* $\mathbf{W} \in \mathcal{W}_*$*,*

(b) $\mathbf{G}_2 = \mathbf{I}_n - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{M}$,

(c) $\mathbf{G}_3 = \mathbf{H} - \mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{M}$,

and thus each $\mathbf{G}_i\mathbf{y} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})$ *under* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ *and*

$$\mathbf{G}_1\mathbf{y} = \mathbf{G}_2\mathbf{y} = \mathbf{G}_3\mathbf{y} \quad \textit{for all } \mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{W}). \tag{5.3}$$

It is important to observe that the multipliers $\mathbf{G}_i$ of $\mathbf{y}$ are not necessarily the same. Equation (5.2) has a unique solution if and only if $\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathbb{R}^n$. It is also worth noting that in light of part (b) of Theorem 5.1,

$$\tilde{\boldsymbol{\mu}} = \mathbf{y} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{My}, \tag{5.4}$$

and hence the BLUE's residual $\tilde{\varepsilon}$, say, and its covariance matrix are

$$\tilde{\varepsilon} = \mathbf{y} - \tilde{\boldsymbol{\mu}} = \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{My}, \quad \mathrm{cov}(\tilde{\varepsilon}) = \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MV}. \tag{5.5}$$

The linear model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ where

$$\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{V}), \tag{5.6}$$

is often called a *weakly singular* linear model or Zyskind–Martin model, see Zyskind & Martin [72]. When dealing with such a model we can choose $\mathbf{W} = \mathbf{V} \in \mathcal{W}_*$ and thus by Theorem 5.1 we have

$$\tilde{\boldsymbol{\mu}} = \mathbf{P}_{\mathbf{X};\mathbf{V}^-}\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^-\mathbf{y}, \quad \mathrm{cov}(\tilde{\boldsymbol{\mu}}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'. \tag{5.7}$$

All expressions in (5.7) are invariant with respect to the choice of generalized inverses involved. Now one might be curious to know whether

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^+\mathbf{X})^-\mathbf{X}'\mathbf{V}^+\mathbf{y}, \tag{5.8}$$

where $(\mathbf{X}'\mathbf{V}^+\mathbf{X})^-$ is some given generalized inverse of $\mathbf{X}'\mathbf{V}^+\mathbf{X}$, is the BLUE for $\mathbf{X}\boldsymbol{\beta}$, i.e., it satisfies the BLUE equation

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^+\mathbf{X})^-\mathbf{X}'\mathbf{V}^+(\mathbf{X} : \mathbf{VM}) = (\mathbf{X} : \mathbf{0}), \tag{5.9}$$

that is,

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^+\mathbf{X})^-\mathbf{X}'\mathbf{V}^+\mathbf{X} = \mathbf{X}, \tag{5.10a}$$

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^+\mathbf{X})^-\ \mathbf{X}'\mathbf{P}_\mathbf{V}\mathbf{M} = \mathbf{0}. \tag{5.10b}$$

In light of part (b) of Theorem 5.1, (5.10a) implies that $\mathscr{C}(\mathbf{X}') \subset \mathscr{C}(\mathbf{X}'\mathbf{V}^+\mathbf{X}) = \mathscr{C}(\mathbf{X}'\mathbf{V})$ and thus

$$\mathrm{rank}(\mathbf{X}') = \mathrm{rank}(\mathbf{X}'\mathbf{V}) = \mathrm{rank}(\mathbf{P}_\mathbf{V}\mathbf{X}). \tag{5.11}$$

Premultiplying (5.10b) by $\mathbf{X}'\mathbf{V}^+$ and using (b) of Theorem 5.1 yields $\mathbf{X}'\mathbf{P}_\mathbf{V}\mathbf{M} = \mathbf{0}$ and so

$$\mathscr{C}(\mathbf{P}_\mathbf{V}\mathbf{X}) \subset \mathscr{C}(\mathbf{X}). \tag{5.12}$$

Now (5.11) and (5.12) together imply (5.6). Thus we have proved that

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^+\mathbf{X})^-\mathbf{X}'\mathbf{V}^+\mathbf{y} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) \iff \mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{V}). \tag{5.13}$$

The following lemma gives some "mathematical" equalities which are related to "statistical" equalities in Theorem 5.1.; for further details, see, e.g., Isotalo et al. [29, pp. 1444–1446].

LEMMA 5.2. *Using the earlier notation and letting $\mathbf{W} \in \mathcal{W}_*$, the following matrix equalities hold:*

(a) $\mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MV} + \mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}' = \mathbf{W}$,

(b) $\mathbf{P_W}\dot{\mathbf{M}}\mathbf{P_W} = \mathbf{W}^{+} - \mathbf{W}^{+}\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{+}$,

(c)
$$\begin{aligned}\mathbf{P}_{\mathbf{X};\mathbf{W}^{+}} &= \mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{+}\\ &= \mathbf{H} - \mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{MP_W}\\ &= \mathbf{H} - \mathbf{HVM}(\mathbf{MVM})^{+}\mathbf{M}\\ &= \mathbf{P_W} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MP_W}\,,\end{aligned}$$

where $\dot{\mathbf{M}} = \mathbf{M}(\mathbf{MVM})^{-}\mathbf{M}$. *In particular, in the above representations, we use the Moore–Penrose inverse whenever the superscript* $^{+}$ *is used while the superscript* $^{-}$ *means that we can use any generalized inverse.*

Using Theorem 5.1 and Lemma 5.2 it is straightforward to introduce the following representations for the covariance matrix of the $\tilde{\boldsymbol{\mu}} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})$.

- General case:

$$\begin{aligned}\mathrm{cov}(\tilde{\boldsymbol{\mu}}) &= \mathbf{HVH} - \mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{MVH}\\ &= \mathbf{V} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MV}\\ &= \mathbf{V} - \mathbf{V}\dot{\mathbf{M}}\mathbf{V}\\ &= \mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}' - \mathbf{XTX}',\end{aligned} \tag{5.14}$$

  where $\dot{\mathbf{M}} = \mathbf{M}(\mathbf{MVM})^{-}\mathbf{M}$, and $\mathbf{W} = \mathbf{V} + \mathbf{XTX}' \in \mathcal{W}_{*}$.

- $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{V})$, i.e., the model is weakly singular:

$$\begin{aligned}\mathrm{cov}(\tilde{\boldsymbol{\mu}}) &= \mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}\mathbf{X}'\\ &= \mathbf{X}_b(\mathbf{X}_b'\mathbf{V}^{-}\mathbf{X}_b)^{-}\mathbf{X}_b'\\ &= \mathbf{H}(\mathbf{HV}^{-}\mathbf{H})^{-}\mathbf{H}\\ &= (\mathbf{HV}^{-}\mathbf{H})^{+},\end{aligned} \tag{5.15}$$

  where $\mathbf{X}_b$ is a matrix with a property $\mathscr{C}(\mathbf{X}_b) = \mathscr{C}(\mathbf{X})$.

- $\mathbf{V}$ is positive definite:

$$\begin{aligned}\mathrm{cov}(\tilde{\boldsymbol{\mu}}) &= \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\\ &= \mathbf{X}_b(\mathbf{X}_b'\mathbf{V}^{-1}\mathbf{X}_b)^{-1}\mathbf{X}_b'\\ &= \mathbf{H}(\mathbf{HV}^{-1}\mathbf{H})^{-}\mathbf{H}\\ &= (\mathbf{HV}^{-1}\mathbf{H})^{+}.\end{aligned} \tag{5.16}$$

In passing we may note that in view of (5.14) and (5.5) we have $\operatorname{cov}(\tilde{\boldsymbol{\mu}}) = \mathbf{V} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MV}$ and thereby

$$\operatorname{cov}(\mathbf{y}) = \operatorname{cov}(\tilde{\boldsymbol{\mu}}) + \operatorname{cov}(\tilde{\varepsilon}), \tag{5.17}$$

where $\tilde{\varepsilon} = \mathbf{y} - \tilde{\boldsymbol{\mu}} = \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{My}$ refers to the residual of the BLUE of $\mathbf{X}\boldsymbol{\beta}$.

There is one further special situation worth attention. This concerns the case when there are no unit canonical correlations between $\mathbf{Hy}$ and the vector of the OLS residuals $\mathbf{My}$. We recall, see, e.g., Anderson [2, §12.2] and Styan [68], that when

$$\operatorname{cov}\begin{pmatrix}\mathbf{Hy}\\ \mathbf{My}\end{pmatrix} = \begin{pmatrix}\mathbf{HVH} & \mathbf{HVM}\\ \mathbf{MVH} & \mathbf{MVM}\end{pmatrix} := \boldsymbol{\Sigma} = \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}, \tag{5.18}$$

then the nonzero canonical correlations between $\mathbf{Hy}$ and $\mathbf{My}$ are the nonzero eigenvalues of the matrix $\boldsymbol{\Psi}$, say, where

$$\boldsymbol{\Psi} = \boldsymbol{\Sigma}_{11}^{+}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{+}\boldsymbol{\Sigma}_{21} = (\mathbf{HVH})^{+}\mathbf{HVM}(\mathbf{MVM})^{+}\mathbf{MVH}. \tag{5.19}$$

The number of unit canonical correlations, say u, appears to be

$$\begin{aligned} u = \operatorname{rank}(\mathbf{HP_V M}) &= \dim \mathscr{C}(\mathbf{VH}) \cap \mathscr{C}(\mathbf{VM}) \\ &= \dim \mathscr{C}(\mathbf{V}^{1/2}\mathbf{H}) \cap \mathscr{C}(\mathbf{V}^{1/2}\mathbf{M}), \end{aligned} \tag{5.20}$$

see, e.g., Baksalary et al. [10, p. 289], Puntanen et al. [56, §15.10], and Puntanen & Scott [53, Th. 2.6]. Now the following can be shown.

- The situation when $\mathbf{HP_V M} = \mathbf{0}$, i.e., there are no unit canonical correlations between $\mathbf{Hy}$ and $\mathbf{My}$:

$$\begin{aligned} \operatorname{cov}(\tilde{\boldsymbol{\mu}}) &= \mathbf{X}_o(\mathbf{X}_o'\mathbf{V}^{+}\mathbf{X}_o)^{+}\mathbf{X}_o' \\ &= \mathbf{H}(\mathbf{HV}^{+}\mathbf{H})^{+}\mathbf{H} \\ &= \mathbf{P_V H}(\mathbf{HV}^{+}\mathbf{H})^{-}\mathbf{HP_V} \\ &= \mathbf{P_V X}_o(\mathbf{X}_o'\mathbf{V}^{+}\mathbf{X}_o)^{-}\mathbf{X}_o'\mathbf{P_V} \\ &= \mathbf{P_V X}(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{-}\mathbf{X}'\mathbf{P_V}, \end{aligned} \tag{5.21}$$

  where $\mathbf{X}_o$ is a matrix whose columns form an orthonormal basis for $\mathscr{C}(\mathbf{X})$.

It is interesting to observe that the covariance matrix of $\tilde{\boldsymbol{\mu}}$ is a special Schur complement: it is the Schur complement of $\mathbf{MVM}$ in $\boldsymbol{\Sigma}$ in (5.18):

$$\operatorname{cov}(\tilde{\boldsymbol{\mu}}) = \mathbf{HVH} - \mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{MVH} := \boldsymbol{\Sigma}_{11\cdot 2}. \tag{5.22}$$

Since the rank is additive on the Schur complement, see, e.g., Puntanen & Styan [55, §6.3.3], that is,

$$\operatorname{rank}(\boldsymbol{\Sigma}) = \operatorname{rank}(\boldsymbol{\Sigma}_{22}) + \operatorname{rank}(\boldsymbol{\Sigma}_{11\cdot 2}), \tag{5.23}$$

we have

$$\operatorname{rank}(\mathbf{\Sigma}) = \operatorname{rank}(\mathbf{V}) = \operatorname{rank}(\mathbf{MVM}) + \operatorname{rank}[\operatorname{cov}(\tilde{\boldsymbol{\mu}})]\,, \tag{5.24}$$

and so

$$\operatorname{rank}[\operatorname{cov}(\tilde{\boldsymbol{\mu}})] = \operatorname{rank}(\mathbf{V}) - \operatorname{rank}(\mathbf{VM}) = \dim \mathscr{C}(\mathbf{X}) \cap \mathscr{C}(\mathbf{V})\,, \tag{5.25}$$

where we have used the rank rule of Marsaglia & Styan [40, Cor. 6.2], which gives $\operatorname{rank}(\mathbf{VM}) = \operatorname{rank}(\mathbf{V}) - \dim \mathscr{C}(\mathbf{X}) \cap \mathscr{C}(\mathbf{V})$.

6. The relative efficiency of OLSE

Consider the linear model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, where $\mathbf{X}$ has full column rank and $\mathbf{V}$ is positive definite. Then the covariance matrices of OLSE and BLUE of $\boldsymbol{\beta}$ are

$$\operatorname{cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \quad \operatorname{cov}(\tilde{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \tag{6.1}$$

By Lemma 3.1,

$$\begin{aligned}\operatorname{cov}(\tilde{\boldsymbol{\beta}}) &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}[\mathbf{X}'\mathbf{V}\mathbf{X} - \mathbf{X}'\mathbf{V}\mathbf{M}(\mathbf{MVM})^{-}\mathbf{MVX}](\mathbf{X}'\mathbf{X})^{-1} \\ &= \operatorname{cov}(\hat{\boldsymbol{\beta}}) - \mathbf{X}^{+}\mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MV}(\mathbf{X}^{+})',\end{aligned} \tag{6.2}$$

and hence

$$\operatorname{cov}(\hat{\boldsymbol{\beta}}) - \operatorname{cov}(\tilde{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MVX}(\mathbf{X}'\mathbf{X})^{-1}. \tag{6.3}$$

The relative efficiency, so-called Watson efficiency, see [70, p. 330], of OLSE vs. BLUE is defined as the ratio determinants of the covariance matrices:

$$\operatorname{eff}(\hat{\boldsymbol{\beta}}) = \frac{|\operatorname{cov}(\tilde{\boldsymbol{\beta}})|}{|\operatorname{cov}(\hat{\boldsymbol{\beta}})|}\,. \tag{6.4}$$

We have $0 < \operatorname{eff}(\hat{\boldsymbol{\beta}}) \le 1$, with $\operatorname{eff}(\hat{\boldsymbol{\beta}}) = 1$ if and only if $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}$. Moreover, the efficiency can be expressed as

$$\begin{aligned}\operatorname{eff}(\hat{\boldsymbol{\beta}}) = \frac{|\operatorname{cov}(\tilde{\boldsymbol{\beta}})|}{|\operatorname{cov}(\hat{\boldsymbol{\beta}})|} &= \frac{|\mathbf{X}'\mathbf{X}|^2}{|\mathbf{X}'\mathbf{VX}| \cdot |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|} \\ &= \frac{|\mathbf{X}'\mathbf{VX} - \mathbf{X}'\mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MVX}|}{|\mathbf{X}'\mathbf{VX}|} \\ &= |\mathbf{I}_p - \mathbf{X}'\mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MVX}(\mathbf{X}'\mathbf{VX})^{-1}| \\ &= (1-\kappa_1^2)\cdots(1-\kappa_p^2) \\ &= \theta_1^2 \cdots \theta_p^2\,,\end{aligned} \tag{6.5}$$

where κ_i and the θ_i are the canonical correlations between $\mathbf{X}'\mathbf{y}$ and $\mathbf{My}$, and $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$, respectively. Notice that

$$\operatorname{cov}\begin{pmatrix}\mathbf{X}'\mathbf{y}\\ \mathbf{My}\end{pmatrix} = \begin{pmatrix}\mathbf{X}'\mathbf{VX} & \mathbf{X}'\mathbf{VM}\\ \mathbf{MVX} & \mathbf{MVM}\end{pmatrix}, \tag{6.6a}$$

$$\begin{aligned}\operatorname{cov}\begin{pmatrix}\hat{\boldsymbol{\beta}}\\ \tilde{\boldsymbol{\beta}}\end{pmatrix} &= \begin{pmatrix}\operatorname{cov}(\hat{\boldsymbol{\beta}}) & \operatorname{cov}(\tilde{\boldsymbol{\beta}})\\ \operatorname{cov}(\tilde{\boldsymbol{\beta}}) & \operatorname{cov}(\tilde{\boldsymbol{\beta}})\end{pmatrix}\\ &= \begin{pmatrix}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{VX}(\mathbf{X}'\mathbf{X})^{-1} & (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\\ (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} & (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\end{pmatrix},\end{aligned} \tag{6.6b}$$

and thus

$$\{\theta_1^2, \dots, \theta_p^2\} = \operatorname{ch}\big[[\operatorname{cov}(\hat{\boldsymbol{\beta}})]^{-1}\operatorname{cov}(\tilde{\boldsymbol{\beta}})\big], \tag{6.7}$$

where $\operatorname{ch}(\cdot)$ denotes the set of the eigenvalues of the matrix argument. On account of (6.3), it can be shown that indeed

$$\{\theta_1^2, \dots, \theta_p^2\} = \operatorname{ch}[\mathbf{I}_p - (\mathbf{X}'\mathbf{VX})^{-1}\mathbf{X}'\mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MVX}]. \tag{6.8}$$

The efficiency formula (6.5) in terms of κ_i's and θ_i's was first introduced by Bartmann & Bloomfield [11]. It is interesting to observe that in view of (6.7), the squared canonical correlations θ_i^2's are the roots of the equation

$$|\operatorname{cov}(\tilde{\boldsymbol{\beta}}) - \theta^2\operatorname{cov}(\hat{\boldsymbol{\beta}})| = 0, \tag{6.9}$$

and thereby they are solutions to

$$\operatorname{cov}(\tilde{\boldsymbol{\beta}})\mathbf{w} = \theta^2\operatorname{cov}(\hat{\boldsymbol{\beta}})\mathbf{w}, \quad \mathbf{w} \neq \mathbf{0}. \tag{6.10}$$

Here θ^2 is an eigenvalue and $\mathbf{w}$ the corresponding eigenvector of $\operatorname{cov}(\tilde{\boldsymbol{\beta}})$ with respect to $\operatorname{cov}(\hat{\boldsymbol{\beta}})$; see (8.6) in Section 8.

It can be shown that the nonzero canonical correlations between $\mathbf{X}'\mathbf{y}$ and $\mathbf{My}$ are the same as those between $\mathbf{Hy}$ and $\mathbf{My}$. For the further references regarding the relative efficiency and canonical correlations, see Chu et al. [14, 15] and Drury et al. [17].

In this context we may also mention the following Löwner inequality:

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{VX}(\mathbf{X}'\mathbf{X})^{-1} \leq_{\mathrm{L}} \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}, \tag{6.11a}$$

$$\operatorname{cov}(\tilde{\boldsymbol{\beta}}) \leq_{\mathrm{L}} \operatorname{cov}(\hat{\boldsymbol{\beta}}) \leq_{\mathrm{L}} \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}\operatorname{cov}(\tilde{\boldsymbol{\beta}}). \tag{6.11b}$$

Further generalizations of the matrix inequalities of the type (6.11) appear in Baksalary & Puntanen [9], Pecaric et al. [51], and Drury et al. [17].

As regards the lower bound of the OLSE's efficiency, we may mention that [12] and [36] proved the following inequality:

$$\operatorname{eff}(\hat{\boldsymbol{\beta}}) \geq \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2} \cdot \frac{4\lambda_2\lambda_{n-1}}{(\lambda_2+\lambda_{n-1})^2} \cdots \frac{4\lambda_p\lambda_{n-p+1}}{(\lambda_p+\lambda_{n-p+1})^2} = \tau_1^2\tau_2^2\cdots\tau_p^2\,, \tag{6.12}$$

i.e.,

$$\min_{\mathbf{X}} \operatorname{eff}(\hat{\boldsymbol{\beta}}) = \prod_{i=1}^{p} \frac{4\lambda_i\lambda_{n-i+1}}{(\lambda_i+\lambda_{n-i+1})^2} = \prod_{i=1}^{p} \tau_i^2, \tag{6.13}$$

where $\lambda_i = \operatorname{ch}_i(\mathbf{V})$, and $\tau_i = i$th antieigenvalue of $\mathbf{V}$. The concept of antieigenvalue was introduced by Gustafson [23]. For further papers in the antieigenvalues, see, e.g., Gustafson [24, 25], and Rao [62, 63].

We conclude this section by commenting on the equality of OLSE and BLUE which happens precisely when they have identical covariance matrices. In view of of (5.14), we have

$$\begin{aligned}\operatorname{cov}(\tilde{\boldsymbol{\mu}}) &= \mathbf{HVH} - \mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{MVH}\\ &= \operatorname{cov}(\hat{\boldsymbol{\mu}}) - \mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{MVH}\,,\end{aligned} \tag{6.14}$$

which means that $\hat{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}$ (with probability 1) if and only if

$$\mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{MVH} = \mathbf{0}\,. \tag{6.15}$$

It is easy to conclude that (6.15) holds if and only if $\mathbf{HVM} = \mathbf{0}$. In Theorem 6.1 we collect some characterizations for the OLSE and the BLUE to be equal. For the proofs, see, e.g., Rao [57] and Zyskind [71], and for a detailed review, see Puntanen & Styan [54].

THEOREM 6.1. *Consider the general linear model* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$. *Then* $\operatorname{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \operatorname{BLUE}(\mathbf{X}\boldsymbol{\beta})$ *if and only if any one of the following six equivalent conditions holds:*

(a) $\mathbf{HV} = \mathbf{VH}$, (b) $\mathbf{HVM} = \mathbf{0}$, (c) $\mathscr{C}(\mathbf{VX}) \subset \mathscr{C}(\mathbf{X})$,

(d) $\mathscr{C}(\mathbf{X})$ *has a basis comprising a set of* $r = \operatorname{rank}(\mathbf{X})$ *orthonormal eigenvectors of* $\mathbf{V}$,

(e) $\mathbf{V} = \mathbf{HN}_1\mathbf{H} + \mathbf{MN}_2\mathbf{M}$ *for some* $\mathbf{N}_1, \mathbf{N}_2 \in \operatorname{NND}_n$,

(f) $\mathbf{V} = \alpha\mathbf{I}_n + \mathbf{HN}_3\mathbf{H} + \mathbf{MN}_4\mathbf{M}$ *for some* $\alpha \in \mathbb{R}$, *and* $\mathbf{N}_3$ *and* $\mathbf{N}_4$ *are symmetric.*

7. Weighted sum of squares of errors

The ordinary, unweighted sum of squares of errors SSE is defined as

$$\operatorname{SSE}(I) = \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 = \mathbf{y}'\mathbf{M}\mathbf{y}\,, \tag{7.1}$$

while the weighted SSE, when $\mathbf{V}$ is positive definite, is

$$\begin{aligned}\mathrm{SSE}(V) &= \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2_{\mathbf{V}^{-1}} = \|\mathbf{y} - \mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}}\mathbf{y}\|^2_{\mathbf{V}^{-1}} \\ &= \mathbf{y}'[\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}]\mathbf{y} \\ &= \mathbf{y}'\mathbf{M}(\mathbf{MVM})^{-}\mathbf{M}\mathbf{y} = \mathbf{y}'\dot{\mathbf{M}}\mathbf{y}\,. \end{aligned} \tag{7.2}$$

In the general case, the weighted SSE can be defined as

$$\mathrm{SSE}(W) = (\mathbf{y} - \tilde{\boldsymbol{\mu}})'\mathbf{W}^{-}(\mathbf{y} - \tilde{\boldsymbol{\mu}})\,, \tag{7.3}$$

where $\mathbf{W} \in \mathcal{W}_*$. Then, recalling that by (5.5), the BLUE's residual is

$$\tilde{\varepsilon} = \mathbf{y} - \tilde{\boldsymbol{\mu}} = \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{M}\mathbf{y} = \mathbf{V}\dot{\mathbf{M}}\mathbf{y}, \tag{7.4}$$

it is straightforward to confirm the following:

$$\begin{aligned}\mathrm{SSE}(W) &= (\mathbf{y} - \tilde{\boldsymbol{\mu}})'\mathbf{W}^{-}(\mathbf{y} - \tilde{\boldsymbol{\mu}}) \\ &= \tilde{\varepsilon}'\mathbf{W}^{-}\tilde{\varepsilon} = \mathbf{y}'\dot{\mathbf{M}}\mathbf{V}\mathbf{W}^{-}\mathbf{V}\dot{\mathbf{M}}\mathbf{y} \\ &= \mathbf{y}'\dot{\mathbf{M}}\mathbf{W}\mathbf{W}^{-}\mathbf{W}\dot{\mathbf{M}}\mathbf{y} = \mathbf{y}'\dot{\mathbf{M}}\mathbf{W}\dot{\mathbf{M}}\mathbf{y} \\ &= \mathbf{y}'\dot{\mathbf{M}}\mathbf{V}\dot{\mathbf{M}}\mathbf{y} = \mathbf{y}'\dot{\mathbf{M}}\mathbf{V}\mathbf{V}^{-}\mathbf{V}\dot{\mathbf{M}}\mathbf{y} \\ &= \tilde{\varepsilon}'\mathbf{V}^{-}\tilde{\varepsilon} = \mathbf{y}'\dot{\mathbf{M}}\mathbf{y}\,. \end{aligned} \tag{7.5}$$

Note that $\mathrm{SSE}(W)$ is invariant with respect to the choice of $\mathbf{W}^{-}$.

In light of part (b) of Lemma 5.2, the following holds:

$$\mathbf{P}_{\mathbf{W}}\dot{\mathbf{M}}\mathbf{P}_{\mathbf{W}} = \mathbf{W}^{+} - \mathbf{W}^{+}\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{+}. \tag{7.6}$$

From (7.6) it follows that for every $\mathbf{y} \in \mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$,

$$\mathbf{y}'\mathbf{P}_{\mathbf{W}}\dot{\mathbf{M}}\mathbf{P}_{\mathbf{W}}\mathbf{y} = \mathbf{y}'[\mathbf{W}^{+} - \mathbf{W}^{+}\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{+}]\mathbf{y}, \tag{7.7}$$

i.e.,

$$\mathrm{SSE}(W) = \mathbf{y}'\dot{\mathbf{M}}\mathbf{y} = \mathbf{y}'[\mathbf{W}^{-} - \mathbf{W}^{-}\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-}]\mathbf{y}. \tag{7.8}$$

It can be further shown that $\mathrm{SSE}(W)$ provides an unbiased estimator of σ^2:

$$\mathrm{E}(\mathbf{y}'\dot{\mathbf{M}}\mathbf{y}/f) = \sigma^2, \quad \text{where } f = \mathrm{rank}(\mathbf{VM})\,. \tag{7.9}$$

The weighted SSE has an essential role in testing linear hypothesis. Consider the model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V}\}$, where $\mathrm{rank}(\mathbf{X}) = r$, $\mathbf{V}$ is positive definite, $\mathbf{y}$ follows normal distribution with parameters $\mathbf{X}\boldsymbol{\beta}$ and $\sigma^2\mathbf{V}$, i.e., $\mathbf{y} \sim \mathrm{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V})$, and F is the F-statistic for testing linear hypothesis H: $\mathbf{K}\boldsymbol{\beta} = \mathbf{d}$, where $\mathbf{d} \in \mathbb{R}^q$, $\boldsymbol{\eta} = \mathbf{K}\boldsymbol{\beta}$ is estimable and $\mathrm{rank}(\mathbf{K}_{q\times p}) = q$. Denoting $\tilde{\boldsymbol{\eta}} = \mathrm{BLUE}(\mathbf{K}\boldsymbol{\beta})$, we have, for example, the following:

(a) $F = \dfrac{Q/q}{\mathrm{SSE}(V)/(n-r)} \sim \mathrm{F}(q, n-r, \delta)$, (b) $\mathrm{SSE}(V)/\sigma^2 \sim \chi^2(n-r)$,

(c) $\mathrm{cov}(\tilde{\boldsymbol{\eta}}) = \sigma^2 \mathbf{K}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{K}'$,

(d) $Q = (\tilde{\boldsymbol{\eta}} - \mathbf{d})'[\mathrm{cov}(\tilde{\boldsymbol{\eta}})]^{-1}(\tilde{\boldsymbol{\eta}} - \mathbf{d})\sigma^2$, (e) $Q/\sigma^2 \sim \chi^2(q, \delta)$,

(f) $\delta = (\mathbf{K}\boldsymbol{\beta} - \mathbf{d})'[\mathrm{cov}(\tilde{\boldsymbol{\eta}})]^{-1}(\mathbf{K}\boldsymbol{\beta} - \mathbf{d})/\sigma^2$,

where $\mathrm{F}(\cdot,\cdot,\cdot)$ and $\chi^2(\cdot)$ refer to F- and χ^2-distributions, respectively.

8. BLUE's covariance matrix and the proper eigenvalues

Let us begin with a simple example when the linear model is $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, where $\mathbf{X}$ has full column rank and $\mathbf{V}$ is positive definite and we have the following determinant equation

$$\det(\mathbf{V} - \lambda\mathbf{H}) = 0, \text{ i.e, } \quad \det[\mathbf{V} - \lambda\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{X}'] = 0, \tag{8.1}$$

Let our task be to solve the scalar λ from (8.1). Premultiplying (8.1) by $\mathbf{V}^{-1}$ and assuming that $\lambda \neq 0$ yields

$$\det[\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{X}' - \tfrac{1}{\lambda}\mathbf{I}_n] = 0\,. \tag{8.2}$$

Hence $\frac{1}{\lambda}$ is a nonzero eigenvalue of $\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{X}'$. Because the matrix products $\mathbf{AB}$ and $\mathbf{BA}$ have the same the nonzero eigenvalues, we observe that $\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{X}'$ and $(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}$ have the same nonzero eigenvalues:

$$\tfrac{1}{\lambda} \in \mathrm{nzch}[(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}] \tag{8.3}$$

and thus

$$\begin{aligned} \lambda &\in \mathrm{nzch}[(\mathbf{X}'\mathbf{X})^{1/2}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})^{1/2}] \\ &= \mathrm{nzch}[\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}] \\ &= \mathrm{nzch}[\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'] \\ &= \mathrm{nzch}[\mathrm{cov}(\tilde{\boldsymbol{\mu}})]\,, \end{aligned} \tag{8.4}$$

where $\tilde{\boldsymbol{\mu}} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})$ and $\mathrm{nzch}(\cdot)$ denotes the set of nonzero eigenvalues. Putting into words: the nonzero roots of (8.1) are the nonzero eigenvalues of the covariance matrix of the BLUE of $\mathbf{X}\boldsymbol{\beta}$. This, somewhat surprising result can be generalized (done below), but before that we need to recall the concept of the proper eigenvalue and eigenvector in the spirit of Rao & Mitra [64, §6.3]; see also Mitra & Rao [48], as well as de Leeuw [16], Mitra & Moore [42, Appendix], Scott & Styan [66], Isotalo et al. [30, §2], and Hauke et al. [28].

For this purpose, let $\mathbf{A}$ and $\mathbf{B}$ be two symmetric $n \times n$ matrices of which $\mathbf{B}$ is nonnegative definite and thus possibly singular. Let $\lambda \in \mathbb{R}$ be a scalar and $\mathbf{w}$ a vector such that

$$\mathbf{A}\mathbf{w} = \lambda \mathbf{B}\mathbf{w}\,, \quad \mathbf{B}\mathbf{w} \neq \mathbf{0}\,. \tag{8.5}$$

In (8.5), we call λ a proper eigenvalue and $\mathbf{w}$ a proper eigenvector of $\mathbf{A}$ with respect to $\mathbf{B}$, or shortly, $(\lambda, \mathbf{w})$ is a proper eigenpair for $(\mathbf{A}, \mathbf{B})$. The set of all proper eigenvalues of pair $(\mathbf{A}, \mathbf{B})$ is denoted as $\mathrm{ch}(\mathbf{A}, \mathbf{B})$. For a positive definite $\mathbf{B}$ we obviously have

$$\mathrm{ch}(\mathbf{A}, \mathbf{B}) = \mathrm{ch}(\mathbf{B}^{-1}\mathbf{A}, \mathbf{I}_n) = \mathrm{ch}(\mathbf{B}^{-1}\mathbf{A}) = \mathrm{ch}(\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2})\,, \tag{8.6}$$

where $\mathrm{ch}(\cdot)$ refers to the set of the eigenvalues (including multiplicities) of the matrix argument. If $\mathbf{B}$ is singular, we might wonder whether, for example, the following might be true:

$$\mathrm{ch}(\mathbf{A}, \mathbf{B}) = \mathrm{ch}(\mathbf{B}^{+}\mathbf{A})\,? \tag{8.7}$$

Statement (8.7) does not always hold, but if $\mathscr{C}(\mathbf{A}) \subset \mathscr{C}(\mathbf{B})$, then indeed (8.7) holds for nonzero proper eigenvalues; see Lemma 8.1 below.

For completeness, we state the following two lemmas appearing in Rao & Mitra [64, §6.3]; see also Mitra & Rao [48]. Using the notation $\mathrm{nzch}(\cdot)$ for the set of the nonzero eigenvalues and $\mathrm{nzch}(\mathbf{A}, \mathbf{B})$ for the set of the nonzero proper eigenvalues of $\mathbf{A}$ with respect to $\mathbf{B}$ we have the following lemma.

LEMMA 8.1. *Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$ nonnegative definite matrices and let $\mathbf{N} \in \{\mathbf{B}^{\perp}\}$. Then the nonzero proper eigenvalues of $\mathbf{A}$ with respect to $\mathbf{B}$ are the same as the nonzero eigenvalues of $[\mathbf{A} - \mathbf{A}\mathbf{N}(\mathbf{N}'\mathbf{A}\mathbf{N})^{-}\mathbf{N}'\mathbf{A}]\mathbf{B}^{-}$ and vice versa for any generalized inverses involved; i.e.,*

$$\mathrm{nzch}(\mathbf{A}, \mathbf{B}) = \mathrm{nzch}\big([\mathbf{A} - \mathbf{A}\mathbf{N}(\mathbf{N}'\mathbf{A}\mathbf{N})^{-}\mathbf{N}'\mathbf{A}]\mathbf{B}^{-}\big)\,. \tag{8.8}$$

In particular,

$$\mathscr{C}(\mathbf{A}) \subset \mathscr{C}(\mathbf{B}) \implies \mathrm{nzch}(\mathbf{A}, \mathbf{B}) = \mathrm{nzch}(\mathbf{A}\mathbf{B}^{-})\,. \tag{8.9}$$

THEOREM 8.2. *Consider the linear model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$. The nonzero proper eigenvalues of $\mathbf{V}$ with respect to $\mathbf{H}$ are the same as the nonzero eigenvalues of the covariance matrix of the* BLUE *of $\mathbf{X}\boldsymbol{\beta}$.*

Proof. Following Puntanen et al. [56, p. 376], consider the equation

$$\mathbf{V}\mathbf{w} = \lambda \mathbf{H}\mathbf{w}, \quad \mathbf{H}\mathbf{w} \neq \mathbf{0}. \tag{8.10}$$

Lemma 8.1 immediately implies that the nonzero proper eigenvalues of $\mathbf{V}$ with respect to $\mathbf{H}$ are the nonzero eigenvalues of

$$[\mathbf{V} - \mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}]\mathbf{H}^{-}, \tag{8.11}$$

which are precisely the same as the nonzero eigenvalues of

$$\mathbf{H}[\mathbf{V} - \mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}]\mathbf{H}\,, \tag{8.12}$$

which is the covariance matrix of the BLUE of $\mathbf{X}\boldsymbol{\beta}$. □

An alternative proof of Theorem 8.2 appears in Isotalo et al. [30, Th. 2.3].

9. BLUE's covariance matrix as a shorted matrix

Following Isotalo et al. [30, §4], let us consider a simple linear model $\{\mathbf{y},\, \mathbf{1}\beta,\, \mathbf{V}\}$, where $\mathbf{V}$ is positive definite. Let our task be to find a nonnegative definite matrix $\mathbf{S}$ which belongs to the set

$$\mathcal{U} = \{\, \mathbf{U} : \mathbf{0} \leq_{\mathrm{L}} \mathbf{U} \leq_{\mathrm{L}} \mathbf{V},\ \mathscr{C}(\mathbf{U}) \subset \mathscr{C}(\mathbf{1})\,\}\,, \tag{9.1}$$

and which is maximal in the Löwner sense; that is, a nonnegative definite matrix which is "as close to $\mathbf{V}$ as possible" in the Löwner partial ordering, but whose column space is in that of $\mathbf{1}$. This matrix $\mathbf{S}$ is called the *shorted matrix* of $\mathbf{V}$ with respect to $\mathbf{1}$, and denoted as $\mathrm{Sh}(\mathbf{V} \mid \mathbf{1})$.

Because $\mathbf{S}$ is nonnegative definite, we must have $\mathbf{S} = \mathbf{L}\mathbf{L}'$ for some $\mathbf{L}$ of full column rank. Further, the condition $\mathscr{C}(\mathbf{S}) \subset \mathscr{C}(\mathbf{1})$ implies that $\mathbf{L} = \alpha\mathbf{1}$ for some nonzero scalar α and hence $\mathbf{S} = \alpha^2\mathbf{1}\mathbf{1}'$. Our objective is to find a scalar α so that $\alpha^2\mathbf{1}\mathbf{1}'$ is maximal in the Löwner sense, which means that α^2 must be maximal. The choice of α^2 must be made under the condition

$$\alpha^2\mathbf{1}\mathbf{1}' \leq_{\mathrm{L}} \mathbf{V}. \tag{9.2}$$

We show that the maximal value for α^2 is

$$\alpha^2 = (\mathbf{1}'\mathbf{V}^{-1}\mathbf{1})^{-1}. \tag{9.3}$$

It is well known that for two symmetric nonnegative definite matrices $\mathbf{A}$ and $\mathbf{B}$ the following holds see, for example, Liski & Puntanen [38]:

$$\mathbf{A} \leq_{\mathrm{L}} \mathbf{B} \iff \text{(i) } \mathscr{C}(\mathbf{A}) \subset \mathscr{C}(\mathbf{B}) \text{ and (ii) } \mathrm{ch}_1(\mathbf{A}\mathbf{B}^{+}) \leq 1. \tag{9.4}$$

Using (9.4) we observe that (9.2) is equivalent to $\alpha^2 \leq (\mathbf{1}'\mathbf{V}^{-1}\mathbf{1})^{-1}$. Hence the shorted matrix is

$$\mathrm{Sh}(\mathbf{V} \mid \mathbf{1}) = (\mathbf{1}'\mathbf{V}^{-1}\mathbf{1})^{-1}\mathbf{1}\mathbf{1}' = \mathbf{1}(\mathbf{1}'\mathbf{V}^{-1}\mathbf{1})^{-1}\mathbf{1}' \tag{9.5}$$

which is precisely the covariance matrix of $\mathrm{BLUE}(\mathbf{1}\beta)$ under $\{\mathbf{y},\, \mathbf{1}\beta,\, \mathbf{V}\}$. This result can be also generalized as shown below.

Consider now a general case of $\mathcal{U}$:

$$\mathcal{U} = \{\, \mathbf{U} : \mathbf{0} \leq_{\mathrm{L}} \mathbf{U} \leq_{\mathrm{L}} \mathbf{V},\ \mathscr{C}(\mathbf{U}) \subset \mathscr{C}(\mathbf{X})\,\}\,. \tag{9.6}$$

The maximal element $\mathbf{U}$ in $\mathcal{U}$ is the shorted matrix of $\mathbf{V}$ with respect to $\mathbf{X}$, and denoted as $\mathrm{Sh}(\mathbf{V} \mid \mathbf{X})$. The concept of shorted matrix (or operator) was first introduced by Krein [37], and later rediscovered by Anderson [3], who introduced the term "shorted operator". As shown by Anderson [3], and Anderson & Trapp [4], the set $\mathcal{U}$ in (9.6) indeed has a maximal element and it, the shorted matrix, is unique. Mitra & Puri [46, 47] were apparently the first to consider statistical applications of the shorted matrix and the shorted operator.

Mitra & Puntanen [43] proved the following.

THEOREM 9.1. *Consider the general linear model* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$. *Then*

$$\mathrm{cov}[\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})] = \mathrm{Sh}(\mathbf{V} \mid \mathbf{X}). \tag{9.7}$$

Proof. Let us go through the proof which is rather easy, while the result (9.7) itself is somewhat unexpected. To prove (9.7), let $\mathbf{G}_{n\times n}$ be such a matrix that $\mathbf{G}\mathbf{y} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})$ and so

$$\mathrm{cov}(\mathbf{G}\mathbf{y}) = \mathbf{G}\mathbf{V}\mathbf{G}' \leq_{\mathrm{L}} \mathbf{V} = \mathrm{cov}(\mathbf{y}), \tag{9.8}$$

because $\mathbf{y}$ is an unbiased estimator of $\mathbf{X}\boldsymbol{\beta}$. Let $\mathbf{U}$ be an arbitrary member of $\mathcal{U}$, which implies that $\mathbf{U} = \mathbf{H}\mathbf{A}\mathbf{A}'\mathbf{H}$ for some matrix $\mathbf{A}$ [here $\mathbf{H} = \mathbf{P}_{\mathbf{X}}$] and

$$\mathbf{U} = \mathbf{H}\mathbf{A}\mathbf{A}'\mathbf{H} \leq_{\mathrm{L}} \mathbf{V}. \tag{9.9}$$

Premultiplying (9.9) by $\mathbf{G}$ and postmultiplying it by $\mathbf{G}'$ yields

$$\mathbf{G}\mathbf{U}\mathbf{G}' = \mathbf{G}\mathbf{H}\mathbf{A}\mathbf{A}'\mathbf{H}\mathbf{G}' = \mathbf{H}\mathbf{A}\mathbf{A}'\mathbf{H} = \mathbf{U} \leq_{\mathrm{L}} \mathbf{G}\mathbf{V}\mathbf{G}', \tag{9.10}$$

where we have used the fact that $\mathbf{G}\mathbf{H} = \mathbf{H}$. Now (9.10) confirms that $\mathbf{G}\mathbf{V}\mathbf{G}'$ is the maximal element in the class $\mathcal{U}$, i.e., (9.7) holds. □

Definition of the shorted matrix via (9.6) is applicable only for nonnegative definite matrix $\mathbf{V}$. The generalization to rectangular matrices was made by Mitra & Puri [47]. We do not go into this definition in details but we mention that if it is applied into the case of nonnegative definite $\mathbf{V}$, then, according to this generalized definition, we consider the following set of matrices $\mathbf{T}$:

$$\mathcal{T} = \{\mathbf{T}_{n\times n} : \mathscr{C}(\mathbf{T}) \subset \mathscr{C}(\mathbf{X}),\, \mathscr{C}(\mathbf{T}') \subset \mathscr{C}(\mathbf{X})\}. \tag{9.11}$$

Then the shorted matrix of $\mathbf{V}$ with respect to $\mathbf{X}$ is the matrix $\mathbf{S}$ satisfying the property

$$\mathrm{rank}(\mathbf{V} - \mathbf{S}) \leq \mathrm{rank}(\mathbf{V} - \mathbf{T}) \quad \text{for all } \mathbf{T} \in \mathcal{T}. \tag{9.12}$$

This definition yields the same shorted matrix as done by maximizing the matrix $\mathbf{U} \in \mathcal{U}$ in (9.6). The generalized definition is related to the concept of minus (or rank-subtractivity) partial ordering for $\mathbf{A}_{n\times m}$ and $\mathbf{B}_{n\times m}$ defined as

$$\mathbf{A} \leq^{-} \mathbf{B} \iff \mathrm{rank}(\mathbf{B} - \mathbf{A}) = \mathrm{rank}(\mathbf{B}) - \mathrm{rank}(\mathbf{A}). \tag{9.13}$$

For the following lemma, see Mitra et al. [44].

LEMMA 9.2. *The following statements are equivalent when considering the linear model* $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ *with* $\mathbf{X}^{\sim}$ *being a generalized inverse of* $\mathbf{X}$.

(a) $\mathbf{X}\mathbf{X}^{\sim}\mathbf{V}(\mathbf{X}^{\sim})'\mathbf{X}' \leq_{\mathrm{L}} \mathbf{V}$,

(b) $\mathbf{X}\mathbf{X}^{\sim}\mathbf{y}$ *is the* BLUE *for* $\mathbf{X}\boldsymbol{\beta}$,

(c) $\mathbf{X}\mathbf{X}^{\sim}\mathbf{V}(\mathbf{X}^{\sim})'\mathbf{X}' \leq^{-} \mathbf{V}$,

(d) $\mathbf{X}\mathbf{X}^{\sim}\mathbf{V}(\mathbf{X}^{\sim})'\mathbf{X}' = \mathrm{Sh}(\mathbf{V} \mid \mathbf{X})$.

For a review of shorted matrices and their applications in statistics, see Mitra et al. [44] and Mitra et al. [41], and for relations of shorted matrices and matrix partial orderings, see, e.g., Mitra & Prasad [45], Eagambaram et al. [20] and Prasad et al. [52].

10. Best linear unbiased predictor, BLUP

We can extend the model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ by considering a $q \times 1$ random vector $\mathbf{y}_*$, which is an unobservable random vector containing new future observations. These new observations are assumed to be generated from

$$\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_* \,, \tag{10.1}$$

where $\mathbf{X}_*$ is a known $q \times p$ matrix, $\boldsymbol{\beta} \in \mathbb{R}^p$ is the same vector of fixed but unknown parameters as in $\mathscr{M}$, and $\boldsymbol{\varepsilon}_*$ is a q-dimensional random error vector with $\mathrm{E}(\boldsymbol{\varepsilon}_*) = \mathbf{0}$. We will also use the notations $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ and $\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}$. The covariance matrix of $\mathbf{y}_*$ and the cross-covariance matrix between $\mathbf{y}$ and $\mathbf{y}_*$ are assumed to be known and thus we have

$$\mathrm{E}\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_*\end{pmatrix} = \begin{pmatrix}\boldsymbol{\mu}\\ \boldsymbol{\mu}_*\end{pmatrix} = \begin{pmatrix}\mathbf{X}\\ \mathbf{X}_*\end{pmatrix}\boldsymbol{\beta}\,, \quad \mathrm{cov}\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_*\end{pmatrix} = \begin{pmatrix}\mathbf{V} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{pmatrix} = \boldsymbol{\Gamma}\,, \tag{10.2}$$

where the $(n+q) \times (n+q)$ covariance matrix $\boldsymbol{\Gamma}$ is known. This setup can be denoted shortly as

$$\mathscr{M}_* = \left\{\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_*\end{pmatrix}, \begin{pmatrix}\mathbf{X}\\ \mathbf{X}_*\end{pmatrix}\boldsymbol{\beta}, \begin{pmatrix}\mathbf{V} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{pmatrix}\right\}. \tag{10.3}$$

We are particularly interested in predicting the unobservable $\mathbf{y}_*$ on the basis of the observable $\mathbf{y}$. While doing this, we look for linear predictors of the type $\mathbf{A}\mathbf{y}$, where $\mathbf{A} \in \mathbb{R}^{q\times n}$.

The random vector $\mathbf{y}_*$ is called predictable under $\mathscr{M}_*$ if there exists a matrix $\mathbf{A}$ such that the expected prediction error is zero, i.e., $\mathrm{E}(\mathbf{y}_* - \mathbf{A}\mathbf{y}) = \mathbf{0}$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$. Then $\mathbf{A}\mathbf{y}$ is a linear unbiased predictor (LUP) of $\mathbf{y}_*$. Such a matrix $\mathbf{A} \in \mathbb{R}^{q\times n}$ exists if and only if $\mathscr{C}(\mathbf{X}_*') \subset \mathscr{C}(\mathbf{X}')$, that is, $\mathbf{X}_*\boldsymbol{\beta}$ is estimable under $\mathscr{M}$. Thus $\mathbf{y}_*$ is predictable under $\mathscr{M}_*$ if and only if $\mathbf{X}_*\boldsymbol{\beta}$ is estimable. Now a LUP $\mathbf{Ay}$ is the best linear unbiased predictor, BLUP, for $\mathbf{y}_*$, if we have the Löwner ordering

$$\operatorname{cov}(\mathbf{y}_* - \mathbf{Ay}) \leq_{\mathrm{L}} \operatorname{cov}(\mathbf{y}_* - \mathbf{A}_{\#}\mathbf{y}) \quad \text{for all } \mathbf{A}_{\#} : \mathbf{A}_{\#}\mathbf{X} = \mathbf{X}_* \,. \tag{10.4}$$

Theorem 10.1 below provides so-called fundamental BLUP equations, see, e.g., Christensen [13, p. 294], and Isotalo & Puntanen [32, p. 1015], For the reviews of the BLUP-properties, see, Robinson [65] and Haslett & Puntanen [27].

THEOREM 10.1. *Consider the linear model with new observations defined as $\mathscr{M}_*$ in* (10.3)*, where $\mathscr{C}(\mathbf{X}_*') \subset \mathscr{C}(\mathbf{X}')$, i.e., $\mathbf{y}_*$ is predictable. Then the linear predictor $\mathbf{Ay}$ is the* BLUP *for $\mathbf{y}_*$ if and only if $\mathbf{A} \in \mathbb{R}^{q\times n}$ satisfies the equation*

$$\mathbf{A}(\mathbf{X} : \mathbf{VX}^{\perp}) = (\mathbf{X}_* : \mathbf{V}_{21}\mathbf{X}^{\perp}) \,. \tag{10.5}$$

Moreover, the linear predictor $\mathbf{By}$ is the BLUP *for $\boldsymbol{\varepsilon}_*$ if and only if $\mathbf{B} \in \mathbb{R}^{q\times n}$ satisfies the equation*

$$\mathbf{B}(\mathbf{X} : \mathbf{VX}^{\perp}) = (\mathbf{0} : \mathbf{V}_{21}\mathbf{X}^{\perp}) \,. \tag{10.6}$$

We will use the following short notations:

$$\tilde{\mathbf{y}}_* = \mathrm{BLUP}(\mathbf{y}_*)\,, \quad \tilde{\boldsymbol{\mu}}_* = \mathrm{BLUE}(\boldsymbol{\mu}_*)\,, \quad \tilde{\varepsilon}_* = \mathrm{BLUP}(\boldsymbol{\varepsilon}_*)\,. \tag{10.7}$$

Suppose that the parametric function $\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}$ is estimable under $\mathscr{M}_*$ which happens if and only if $\mathscr{C}(\mathbf{X}_*') \subset \mathscr{C}(\mathbf{X}')$ so that

$$\mathbf{X}_* = \mathbf{LX} \quad \text{for some matrix } \mathbf{L} \in \mathbb{R}^{q\times f}, \quad \boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta} = \mathbf{LX}\boldsymbol{\beta} = \mathbf{L}\boldsymbol{\mu} \,. \tag{10.8}$$

Now the $\mathrm{BLUP}(\mathbf{y}_*)$ under $\mathscr{M}_*$, see, e.g., Isotalo et al. [31, Sec. 4], can be written as

$$\tilde{\mathbf{y}}_* = \tilde{\boldsymbol{\mu}}_* + \tilde{\varepsilon}_* \,, \tag{10.9}$$

and further as

$$\begin{aligned} \mathrm{BLUP}(\mathbf{y}_*) &= \mathrm{BLUE}(\boldsymbol{\mu}_*) + \mathrm{BLUP}(\boldsymbol{\varepsilon}_*) \\ &= \mathbf{LGy} + \mathbf{V}_{21}\mathbf{V}^{-}(\mathbf{I}_n - \mathbf{G})\mathbf{y} \\ &= \mathbf{LGy} + \mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^{-}\mathbf{My} \,, \end{aligned} \tag{10.10}$$

where $\mathbf{G} = \mathbf{P}_{\mathbf{X};\mathbf{W}^-} = \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^-$, and $\mathbf{W} \in \mathcal{W}_*$.

What about the covariance matrix of $\mathrm{BLUP}(\mathbf{y}_*)$? We observe that the random vectors $\tilde{\boldsymbol{\mu}}_*$ and $\tilde{\varepsilon}_*$ are uncorrelated and so

$$\operatorname{cov}(\tilde{\mathbf{y}}_*) = \operatorname{cov}(\tilde{\boldsymbol{\mu}}_*) + \operatorname{cov}(\tilde{\varepsilon}_*) \,. \tag{10.11}$$

Now we have

$$\operatorname{cov}(\tilde{\boldsymbol{\mu}}_*) = \mathbf{L}\operatorname{cov}(\tilde{\boldsymbol{\mu}})\mathbf{L}', \quad \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_*) = \mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^{-}\mathbf{MV}_{12}\,. \tag{10.12}$$

For calculating the covariance matrix of the prediction error $\mathbf{y}_* - \tilde{\mathbf{y}}_*$, it is convenient to express the prediction error as

$$\mathbf{y}_* - \tilde{\mathbf{y}}_* = (\mathbf{y}_* - \mathbf{V}_{21}\mathbf{V}^{-}\mathbf{y}) + (\mathbf{V}_{21}\mathbf{V}^{-}\mathbf{y} - \tilde{\mathbf{y}}_*)\,, \tag{10.13}$$

see Sengupta & Jammalamadaka [67, p. 292]. In view of (10.10), we get

$$\begin{aligned}\mathbf{y}_* - \tilde{\mathbf{y}}_* &= (\mathbf{y}_* - \mathbf{V}_{21}\mathbf{V}^{-}\mathbf{y}) + (\mathbf{V}_{21}\mathbf{V}^{-} - \mathbf{L})\mathbf{Gy} \\ &= (\mathbf{y}_* - \mathbf{V}_{21}\mathbf{V}^{-}\mathbf{y}) + \mathbf{N}\tilde{\boldsymbol{\mu}}\,,\end{aligned} \tag{10.14}$$

where $\mathbf{N} = \mathbf{V}_{21}\mathbf{V}^{-} - \mathbf{L}$. The random vectors $\mathbf{y}_* - \mathbf{V}_{21}\mathbf{V}^{-}\mathbf{y}$ and $\mathbf{N}\tilde{\boldsymbol{\mu}}$ are uncorrelated and hence

$$\begin{aligned}\operatorname{cov}(\mathbf{y}_* - \tilde{\mathbf{y}}_*) &= \operatorname{cov}(\mathbf{y}_* - \mathbf{V}_{21}\mathbf{V}^{-}\mathbf{y}) + \operatorname{cov}(\mathbf{N}\tilde{\boldsymbol{\mu}}) \\ &= \mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{V}^{-}\mathbf{V}_{12} + \mathbf{N}\operatorname{cov}(\tilde{\boldsymbol{\mu}})\mathbf{N}'.\end{aligned} \tag{10.15}$$

The first term $\boldsymbol{\Gamma}_{22\cdot 1} := \mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{V}^{-}\mathbf{V}_{12}$ in (10.15) is the Schur complement of $\mathbf{V}$ in

$$\boldsymbol{\Gamma} = \begin{pmatrix} \mathbf{V} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}, \tag{10.16}$$

and as Sengupta & Jammalamadaka [67, p. 293] point out, $\boldsymbol{\Gamma}_{22\cdot 1}$ is the covariance matrix of the prediction error associated with the best linear predictor, BLP, (supposing that $\mathbf{X}\boldsymbol{\beta}$ were known) while the second term represents the increase in the covariance matrix of the prediction error due to estimation of $\mathbf{X}\boldsymbol{\beta}$.

We may complete our paper by briefly touching the concept of the best linear predictor, BLP. Notice first that the word "unbiased" is missing in this concept. The following lemma is essential when dealing with the best linear prediction; see, e.g., Puntanen et al. [56, Ch. 9].

LEMMA 10.2.

Let $\mathbf{u}$ *and* $\mathbf{v}$ *be* u*- and* v*-dimensional random vectors, respectively, and let* $\mathbf{z} = \left(\begin{smallmatrix}\mathbf{u}\\ \mathbf{v}\end{smallmatrix}\right)$ *be a partitioned random vector with covariance matrix*

$$\operatorname{cov}(\mathbf{z}) = \operatorname{cov}\begin{pmatrix}\mathbf{u}\\ \mathbf{v}\end{pmatrix} = \begin{pmatrix} \boldsymbol{\Omega}_{\mathbf{uu}} & \boldsymbol{\Omega}_{\mathbf{uv}} \\ \boldsymbol{\Omega}_{\mathbf{vu}} & \boldsymbol{\Omega}_{\mathbf{vv}} \end{pmatrix} = \boldsymbol{\Omega}\,. \tag{10.17}$$

Then

$$\operatorname{cov}(\mathbf{v} - \mathbf{Fu}) \geq_{\mathrm{L}} \operatorname{cov}(\mathbf{v} - \boldsymbol{\Omega}_{\mathbf{vu}}\boldsymbol{\Omega}_{\mathbf{uu}}^{-}\mathbf{u}) \quad \textit{for all } \mathbf{F} \in \mathbb{R}^{v\times u}, \tag{10.18}$$

and the minimal covariance matrix is

$$\operatorname{cov}(\mathbf{v} - \mathbf{\Omega}_{\mathbf{vu}}\mathbf{\Omega}_{\mathbf{uu}}^{-}\mathbf{u}) = \mathbf{\Omega}_{\mathbf{vv}} - \mathbf{\Omega}_{\mathbf{vu}}\mathbf{\Omega}_{\mathbf{uu}}^{-}\mathbf{\Omega}_{\mathbf{uv}} = \mathbf{\Omega}_{\mathbf{vv}\cdot\mathbf{u}}\,, \tag{10.19}$$

the Schur complement of $\mathbf{\Omega}_{\mathbf{uu}}$ *in* $\mathbf{\Omega}$.

A linear predictor $\mathbf{Cu} + \mathbf{c}$, where $\mathbf{C} \in \mathbb{R}^{v\times u}$ and $\mathbf{c} \in \mathbb{R}^{v}$, is the best linear predictor, BLP, of the random vector $\mathbf{v}$ on the basis of $\mathbf{u}$ if it minimizes, in the Löwner sense, the mean squared error matrix

$$\mathrm{E}[\mathbf{v} - (\mathbf{Cu}+\mathbf{c})][\mathbf{v} - (\mathbf{Cu}+\mathbf{c})]' = \operatorname{cov}(\mathbf{v} - \mathbf{Cu}) + \|\boldsymbol{\mu}_{\mathbf{v}} - (\mathbf{C}\boldsymbol{\mu}_{\mathbf{u}} + \mathbf{c})\|^2, \tag{10.20}$$

where $\boldsymbol{\mu}_{\mathbf{u}} = \mathrm{E}(\mathbf{u})$ and $\boldsymbol{\mu}_{\mathbf{v}} = \mathrm{E}(\mathbf{v})$. The random vector $\tilde{\mathbf{v}} = \boldsymbol{\mu}_{\mathbf{v}} + \mathbf{\Sigma}_{\mathbf{vu}}\mathbf{\Sigma}_{\mathbf{uu}}^{-}(\mathbf{u} - \boldsymbol{\mu}_{\mathbf{u}})$ appears to be the BLP of $\mathbf{v}$ on the basis of $\mathbf{u}$, and the covariance matrix of the prediction error $\mathbf{v} - \tilde{\mathbf{v}}$ is $\mathbf{\Omega}_{\mathbf{vv}\cdot\mathbf{u}}$.

In Lemma 10.2 our attempt is to find a matrix $\mathbf{F}$ which minimizes the covariance matrix of the difference $\mathbf{v} - \mathbf{Fu}$. This is strikingly close to the task of finding the BLUP for $\mathbf{y}_*$, where we are minimizing the covariance matrix of the prediction error $\mathbf{y}_* - \mathbf{Ay}$. However, the major difference between these two tasks is that in (10.18) we have no restrictions to $\mathbf{F}$ while in (10.4) we assume that $\mathbf{AX} = \mathbf{X}_*$.

Acknowledgements

The authors thank the support of so and so project/funding. Part of this article was presented by Simo Puntanen in *The Fourth DAE-BRNS Theme Meeting on Generation and use of Covariance Matrices in the Applications of Nuclear Data*, Manipal University, Manipal, Karnataka, India, 9–13 December 2017. Thanks go to Professor K. Manjunatha Prasad and his team for excellent hospitality. Due to the topic of this conference, it might be appropriate to cite a few words from the Appendix of article [58], published by Professor C. Radhakrishna Rao in 1971. The title of the Appendix was "The Atom Bomb and Generalized Inverse". Below the first and last paragraph of the Appendix are quoted.

> "The author was first led to the definition of a pseudo-inverse (now called generalized inverse or g-inverse) of a singular matrix in 1945–1955 when he undertook to carry out multivariate analysis of anthropometric data obtained on families of Hiroshima and Nagasaki to study the effects of radiation due atom bomb explosions, on request from Dr. W.J. Schull of the University of Michigan. The computation and use of a pseudo-inverse are given in a statistical report prepared by the author, which is incorporated in Publication No. 461 of the National Academy of Sciences, U.S.A., by Neel and Schull (1956), [50]. It may be of interest to the audience to know the circumstances under which the pseudo-inverse had to be introduced."
>
> "It is hard to believe that scientist have found in what has been described as the greatest tragedy a source for providing material and simulation for research in many directions."

REFERENCES

[1] Aitken A.C. (1936). On least squares and linear combination of observations. *Proc. Roy. Soc. Edinburgh Sect. A*, 55, 42–48.

[2] Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd Ed. Wiley, New York.

[3] Anderson W.N. Jr. (1971). Shorted operators. *SIAM J. Appl. Math.*, 20, 520–525.

[4] Anderson, W.N. Jr. & Trapp G.E. (1975). Shorted operators, II. *SIAM J. Appl. Math.*, 28, 60–71.

[5] Baksalary J.K. & Kala R. (1981). Linear transformations preserving best linear unbiased estimators in a general Gauss–Markoff model. *Ann. Statist.*, 9, 913–916.

[6] Baksalary J.K. & Kala R. (1986). Linear sufficiency with respect to a given vector of parametric functions. *J. Statist. Plann. Inference*, 14, 331–338.

[7] Baksalary J.K. & Mathew T. (1990). Rank invariance criterion and its application to the unified theory of least squares. *Linear Algebra Appl.*, 127, 393–401.

[8] Baksalary J.K. & Puntanen S. (1989). Weighted-least-squares estimation in the general Gauss–Markov model. *Statistical Data Analysis and Inference*. (Y. Dodge, ed.) Elsevier Science Publishers B.V., Amsterdam, 355–368.

[9] Baksalary J.K. & Puntanen S. (1991). Generalized matrix versions of the Cauchy–Schwarz and Kantorovich inequalities. *Aequationes Mathematicae*, 41, 103–110.

[10] Baksalary J.K., Puntanen S. & Styan G.P.H. (1990). A property of the dispersion matrix of the best linear unbiased estimator in the general Gauss–Markov model. *Sankhyā Ser. A*, 52, 279–296.

[11] Bartmann F.C. & Bloomfield P. (1981). Inefficiency and correlation. *Biometrika*, 68, 67–71.

[12] Bloomfield P. & Watson G.S. (1975). The inefficiency of least squares. *Biometrika*, 62, 121–128.

[13] Christensen R. (2011). *Plane Answers to Complex Questions: the Theory of Linear Models*, 4th Ed. Springer, New York.

[14] Chu K.L., Isotalo J., Puntanen S. & Styan G.P.H. (2004). On decomposing the Watson efficiency of ordinary least squares in a partitioned weakly singular linear model. *Sankhyā*, 66, 634–651.

[15] Chu K.L., Isotalo J., Puntanen S. & Styan G.P.H. (2005). Some further results concerning the decomposition of the Watson efficiency in partitioned linear models. *Sankhyā*, 67, 74–89.

[16] de Leeuw J. (1982). Generalized eigenvalue problems with positive semidefinite matrices. *Psychometrika*, 1, 87–93.

[17] Drury S.W., Liu S., Lu C.-Y., Puntanen S. & Styan G.P.H. (2002). Some comments on several matrix inequalities with applications to canonical correlations: historical background and recent developments. *Sankhyā Ser. A*, 64, 453–507.

[18] Drygas H. (1970). *The Coordinate-Free Approach to Gauss–Markov Estimation*. Springer, Berlin.

[19] Drygas H. (1983). Sufficiency and completeness in the general Gauss–Markov model. *Sankhyā Ser. A*, 45, 88–98.

[20] Eagambaram M., Prasad K.M. & Mohana K.S. (2013). Column space decomposition and partial order on matrices. *Electron. J. Linear Algebra*, 26, 795–815.

[21] Groß J. & Puntanen S. (2000). Estimation under a general partitioned linear model. *Linear Algebra Appl.* 321,131–144.

[22] Groß J. & Trenkler G. (1998). On the equality of linear statistics in general Gauss–Markov Model. *Frontiers in Probability and Statistics*. (S.P. Mukherjee, S.K. Basu & B.K. Sinha, eds.) Narosa Publishing House, New Delhi, pp.189–194. [Proc. of the Second International Triennial Calcutta Symposium on Probability and Statistics, Calcutta, 30 December 1994 – 2 January 1995.]

[23] Gustafson K. (1972). Antieigenvalue inequalities in operator theory. *Inequalities, III (Proc. Third Sympos., Univ. California, Los Angeles, 1969; Dedicated to the Memory of Theodore S.Motzkin)*. (O. Shisha, ed.) Academic Press, New York, pp. 115–119.

[24] Gustafson K. (2006). The trigonometry of matrix statistics. *Internat. Statist. Rev.*, 74, 187–202.

[25] Gustafson K. (2012). *Antieigenvalue Analysis.* World Scientific, Singapore.

[26] Haslett S.J., Isotalo J., Liu Y. & Puntanen S. (2014). Equalities between OLSE, BLUE and BLUP in the linear model. *Statist. Papers*, 55, 543–561.

[27] Haslett S.J. & Puntanen S. (2017). Best linear unbiased prediction (BLUP). *Wiley StatsRef: Statistics Reference Online.*

[28] Hauke J., Markiewicz A. & Puntanen S. (2013). Revisiting the BLUE in a linear model via proper eigenvectors. *Combinatorial Matrix Theory and Generalized Inverses of Matrices*. (R.B. Bapat, S. Kirkland, K. M. Prasad & S. Puntanen, eds.) Springer, pp. 73–83.

[29] Isotalo J., Puntanen S. & Styan G.P.H. (2008). A useful matrix decomposition and its statistical applications in linear regression. *Comm. Statist. Theory Methods*, 37, 1436–1457.

[30] Isotalo J., Puntanen S. & Styan G.P.H. (2008). The BLUE's covariance matrix revisited: a review. *J. Statist. Plann. Inference*, 138/9, 2722–2737.

[31] Isotalo J., Markiewicz A. & Puntanen S. (2017). Some properties of linear prediction sufficiency in the linear model. *Proceedings of the LINSTAT-2016* (International Conference on Trends and Perspectives in Linear Statistical Inference, Istanbul, Turkey, 22–25 August 2016). (M. Tez & D. von Rosen, eds.) Springer, in press.

[32] Isotalo J. & Puntanen S. (2006). Linear prediction sufficiency for new observations in the general Gauss–Markov model. *Comm. Statist. Theory Methods*, 35, 1011–1023.

[33] Kala R., Markiewicz A. & Puntanen S. (2017). Some further remarks on the linear sufficiency in the linear model. *Applied and Computational Matrix Analysis: MatTriad, Coimbra, Portugal, September 2015, Selected, Revised Contributions.* (N. Bebiano, ed.) Springer Proceedings in Mathematics & Statistics, vol. 192, 275–294.

[34] Kala R., Puntanen S. & Tian Y. (2017). Some notes on linear sufficiency. *Statist. Papers*, 58, 1–17.

[35] Khatri, C.G. (1966). A note on a MANOVA model applied to problems in growth curves. *Ann. Inst. Stat. Math.*, 18, 75–86.

[36] Knott M. (1975). On the minimum efficiency of least squares. *Biometrika*, 62, 129–132.

[37] Krein M.G. (1947). The theory of self-adjoint extensions of semi-bounded Hermitian transformations and its applications, I, II (in Russian). *Rec. Math. [Mat. Sbornik] N.S.*, 20(62), 431–495; 21(63), 365–404.

[38] Liski E.P. & Puntanen S. (1989). A further note on a theorem on the difference of the generalized inverses of two nonnegative definite matrices. *Comm. Statist. Theory Methods*, 18, 1747–1751.

[39] Markiewicz A. & Puntanen S. (2015). All about the $\perp$ with its applications in the linear statistical models. *Open Mathematics*, 13, 33–50.

[40] Marsaglia G. & Styan G.P.H. (1974). Equalities and inequalities for ranks of matrices. *Linear Multilinear Algebra*, 2, 269–292.

[41] Mitra S.K., Bhimasankaram P. & Malik S.B. (2010). *Matrix Partial Orders, Shorted Operators and Applications*. World Scientific, Singapore.

[42] Mitra S.K. & Moore B.J. (1973). Gauss–Markov estimation with an incorrect dispersion matrix. *Sankhyā Ser. A*, 35, 139–152.

[43] Mitra S.K. & Puntanen S. (1991). The shorted operator statistically interpreted. *Calcutta Statist. Assoc. Bull.*, 40, 97–102.

[44] Mitra S.K., Puntanen S. & Styan G.P.H. (1995). Shorted matrices and their applications in linear statistical models: a review. *Multivariate Statistics and Matrices in Statistics: Proceedings of the 5th Tartu Conference, Tartu–Pühajärve, Estonia, 23–28 May 1994.* (E.-M. Tiit, T. Kollo & H. Niemi, eds.) New Trends in Probability and Statistics, vol. 3, VSP, Utrecht, Netherlands & TEV, Vilnius, Lithuania, pp. 289–311.

[45] Mitra S.K. & Prasad K.M. (1996). The nonunique shorted matrix. *Linear Algebra Appl.*, 237/238, 41–70.

[46] Mitra S.K. & Puri M.L. (1979). Shorted operators and generalized inverses of matrices. *Linear Algebra Appl.*, 25, 45–56.

[47] Mitra, S.K. & Puri M.L. (1982). Shorted matrices – an extended concept and some applications. *Linear Algebra Appl.*, 42, 57–79.

[48] Mitra S.K. & Rao C.R. (1968). Simultaneous reduction of a pair of quadratic forms. *Sankhyā Ser. A*, 30, 313–322.

[49] Mitra S.K. & Rao C.R. (1974). Projections under seminorms and generalized Moore–Penrose inverses. *Linear Algebra and Appl.*, 9, 155–167.

[50] Neel J.V. & Schull W.J. (1956). The effect of exposure to the atomic bomb on pregnancy termination in Hiroshima and Nagasaki. National Academy of Sciences National Research Council, Washington, D.C., Publ. No. 461.

[51] Pecaric J.E., Puntanen S. & Styan G.P.H. (1996). Some further matrix extensions of the Cauchy–Schwarz and Kantorovich inequalities, with some statistical applications. *Linear Algebra Appl.*, 237/238, 455–477.

[52] Prasad K.M, Mohana K.S. & Sheela Y.S. (2013). Matrix partial orders associated with space preorder. *Combinatorial Matrix Theory and Generalized Inverses of Matrices*. (R.B. Bapat, S. Kirkland, K.M. Prasad & S. Puntanen, eds.) Springer, pp. 195–226.

[53] Puntanen S. & Scott A.J. (1996). Some further remarks on the singular linear model. *Linear Algebra Appl.*, 237/238, 313–327.

[54] Puntanen S. & Styan G.P.H. (1989). The equality of the ordinary least squares estimator and the best linear unbiased estimator [with comments by O. Kempthorne & by S.R. Searle and with "Reply" by the authors]. *Amer. Statist.*, 43, 153–164.

[55] Puntanen S. & Styan G.P.H. (2005). Schur complements in statistics and probability. Chapter 6 In *The Schur Complement and Its Applications*. (F. Zhang, ed.) Springer, New York, pp. 163–226.

[56] Puntanen S., Styan G.P.H. & Isotalo J. (2011). *Matrix Tricks for Linear Statistical Models: Our Personal Top Twenty.* Springer, Heidelberg.

[57] Rao C.R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Berkeley, California, 1965/1966*, vol. 1. (L.M. Le Cam & J. Neyman, eds.) University of California Press, Berkeley, pp. 355–372.

[58] Rao, C.R. (1971). Unified theory of linear estimation. *Sankhyā Ser. A*, 33, 371–394. [Corrigendum (1972), 34, p. 194 and p. 477.]

[59] Rao C.R. (1973a). *Linear Statistical Inference and its Applications*, 2nd Ed. Wiley, New York.

[60] Rao C.R. (1973b). Representations of best linear estimators in the Gauss–Markoff model with a singular dispersion matrix. *J. Multivariate Anal.*, 3, 276–292.

[61] Rao C.R. (1974). Projectors, generalized inverses and the BLUE's. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 36, 442–448.

[62] Rao C.R. (2005). Antieigenvalues and antisingularvalues of a matrix and applications to problems in statistics. *Res. Lett. Inf. Math. Sci.*, 8, 53–76.

[63] Rao C.R. (2007). Antieigenvalues and antisingularvalues of a matrix and applications to problems in statistics. *Math. Inequal. Appl.*, 10, 471–489.

[64] Rao C.R. & Mitra S.K. (1971). *Generalized Inverse of Matrices and Its Applications.* Wiley, New York.

[65] Robinson G.K. (1991). That BLUP is a good thing: the estimation of random effects (discussion: pp. 32–51). *Stat. Sci.*, 6, 15–51.

[66] Scott A.J. & Styan G.P.H. (1985). On a separation theorem for generalized eigenvalues and a problem in the analysis of sample surveys. *Linear Algebra Appl.*, 70, 209–224.

[67] Sengupta D. & Jammalamadaka S.R. (2003) *Linear Models: An Integrated Approach*. World Scientific, River Edge, NJ.

[68] Styan G.P.H. (1985). Schur complements and linear statistical models. *Proceedings of the First International Tampere Seminar on Linear Statistical Models and their Applications: Tampere, Finland, August–September 1983.* (T. Pukkila & S. Puntanen, eds.) Dept. of Mathematical Sciences/Statistics, University of Tampere, Tampere, Finland, pp. 37–75.

[69] Tian Y. & Puntanen S. (2009). On the equivalence of estimations under a general linear model and its transformed models. *Linear Algebra Appl.*, 430, 2622–2641.

[70] Watson G.S. (1955). Serial correlation in regression analysis, I. *Biometrika*, 42, 327–341.

[71] Zyskind G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *Ann. Math. Stat.*, 38, 1092–1109.

[72] Zyskind G. & Martin F.B. (1969). On best linear estimation and general Gauss–Markov theorem in linear models with arbitrary nonnegative covariance structure. *SIAM J. Appl. Math.*, 17, 1190–1202. ,

Correction

Display

$$\mathrm{E}[\mathbf{v} - (\mathbf{Cu} + \mathbf{c})][\mathbf{v} - (\mathbf{Cu} + \mathbf{c})]' = \mathrm{cov}(\mathbf{v} - \mathbf{Cu}) + \|\boldsymbol{\mu}_{\mathbf{v}} - (\mathbf{C}\boldsymbol{\mu}_{\mathbf{u}} + \mathbf{c})\|^2, \quad (10.20)$$

should be replaced with

$$\begin{aligned}\mathrm{E}[\mathbf{v} - (\mathbf{Cu} + \mathbf{c})][\mathbf{v} - (\mathbf{Cu} + \mathbf{c})]' &= \mathrm{cov}(\mathbf{v} - \mathbf{Cu}) \\ &\quad + [\boldsymbol{\mu}_{\mathbf{v}} - (\mathbf{C}\boldsymbol{\mu}_{\mathbf{u}} + \mathbf{c})][\boldsymbol{\mu}_{\mathbf{v}} - (\mathbf{C}\boldsymbol{\mu}_{\mathbf{u}} + \mathbf{c})]'.\end{aligned}$$