Research Article   **Open Access**

Augustyn Markiewicz and Simo Puntanen*

Upper bounds for the Euclidean distances between the BLUPs

Abstract: In this article we consider the general linear model $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, where $\mathbf{y}$ is the observable random vector with expectation $\mathbf{X}\boldsymbol{\beta}$ and covariance matrix $\mathbf{V}$. Our interest is on predicting the unobservable random vector $\mathbf{y}_*$, which comes from $\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*$, where the expectation of $\mathbf{y}_*$ is $\mathbf{X}_*\boldsymbol{\beta}$ and the covariance matrix of $\mathbf{y}_*$ is known as well as the cross-covariance matrix between $\mathbf{y}_*$ and $\mathbf{y}$. The random vector $\mathbf{y}_*$ can be considered as a kind of unknown future value. We introduce upper bounds for the Euclidean distances between the BLUPs, the best linear unbiased predictors, when the prediction is based on the original model and when it is based on the transformed model $\{\mathbf{F}\mathbf{y}, \mathbf{F}\mathbf{X}\boldsymbol{\beta}, \mathbf{F}\mathbf{V}\mathbf{F}'\}$. We also show how the upper bounds are related to the concept of linear sufficiency, and we apply our results into the mixed linear model.

Keywords: Best linear unbiased estimator, best linear unbiased predictor, Euclidean norm, linear sufficiency, transformed linear model.

1 Introduction

Let us begin with some words about the notation. The symbol $\mathbb{R}^{m \times n}$ denotes the set of $m \times n$ real matrices, while $\mathbf{A}'$, $\mathbf{A}^-$, $\mathbf{A}^+$, $\mathcal{C}(\mathbf{A})$, and $\mathcal{C}(\mathbf{A})^\perp$, denote, respectively, the transpose, a generalized inverse, the (unique) Moore–Penrose inverse, the column space, and the orthogonal complement of the column space of the matrix $\mathbf{A}$. By $(\mathbf{A} : \mathbf{B})$ we denote the partitioned matrix with $\mathbf{A}_{a \times b}$ and $\mathbf{B}_{c \times d}$ as submatrices, where $a = c$. The symbol $\mathbf{A}^\perp$ stands for any matrix satisfying $\mathcal{C}(\mathbf{A}^\perp) = \mathcal{C}(\mathbf{A})^\perp$. 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 $\mathcal{C}(\mathbf{A})$, and $\mathbf{Q}_{\mathbf{A}} = \mathbf{I} - \mathbf{P}_{\mathbf{A}}$, where $\mathbf{I}$ refers to the identity matrix of conformable dimension. In particular, we use notation $\mathbf{M} = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}}$, where $\mathbf{X}_{n \times p}$ refers to the model matrix, see (1.1). One convenient choice for $\mathbf{X}^\perp$ is obviously $\mathbf{M}$; convenience follows from the symmetry and idempotence of $\mathbf{M}$.

We will consider the general linear model

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

where $\mathbf{X} \in \mathbb{R}^{n \times p}$ is a known model matrix, the vector $\mathbf{y}$ is an observable n-dimensional random vector (so-called response vector), $\boldsymbol{\beta} \in \mathbb{R}^p$ is vector of unknown parameters, and $\boldsymbol{\varepsilon}$ is an unobservable vector of random errors with expectation $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, and covariance matrix $\mathrm{cov}(\boldsymbol{\varepsilon}) = \mathbf{V}$. 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 our considerations σ^2 has no role and hence we omit it. The nonnegative definite matrix $\mathbf{V}$ is known and can be singular.

Let $\mathbf{y}_*$ denote a $q \times 1$ unobservable random vector containing new future observations. The new observations are assumed to be generated from

$$\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*, \tag{1.2}$$

Augustyn Markiewicz: Poznań University of Life Sciences, Poland, E-mail: amark@up.poznan.pl
*Corresponding Author: Simo Puntanen: University of Tampere, Finland, E-mail: simo.puntanen@uta.fi

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 $\mathcal{M}$, and $\boldsymbol{\varepsilon}_*$ is a q-dimensional random error vector with $\mathrm{E}(\boldsymbol{\varepsilon}_*) = \mathbf{0}$. We will also use the notation $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, $\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}$. 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

$$\mathrm{E}\begin{pmatrix} \mathbf{y} \\ \mathbf{y}_* \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}_* \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{X}_* \end{pmatrix}\boldsymbol{\beta}, \quad \mathrm{cov}\begin{pmatrix} \mathbf{y} \\ \mathbf{y}_* \end{pmatrix} = \mathrm{cov}\begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon}_* \end{pmatrix} = \begin{pmatrix} \mathbf{V} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} = \boldsymbol{\Gamma} \in \mathrm{NND}_{n+q}, \qquad (1.3)$$

where NND_{n+q} refers to the set of $(n + q) \times (n + q)$ nonnegative definite matrices. This setup can be denoted shortly as

$$\mathcal{M}_* = \left\{ \begin{pmatrix} \mathbf{y} \\ \mathbf{y}_* \end{pmatrix}, \begin{pmatrix} \mathbf{X} \\ \mathbf{X}_* \end{pmatrix}\boldsymbol{\beta}, \begin{pmatrix} \mathbf{V} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} \right\}. \qquad (1.4)$$

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 $\mathbf{By}$, where $\mathbf{B} \in \mathbb{R}^{q \times n}$, that would "conveniently" utilize the knowledge of (1.4). It is noteworthy that, literally taken, $\mathcal{M}_*$ in (1.4) is not a standard linear model—this is due to the fact that the $\mathbf{y}_*$-part is not a proper response variable which in usual linear model is assumed to be observable. To clarify the situation, we will call $\mathcal{M}_*$ "the linear model with new future observations".

One of the first articles to consider the setup $\mathcal{M}_*$ was Goldberger [9, 1962], who assumed $\boldsymbol{\Gamma}$ to be positive definite and $\mathbf{y}_*$ a scalar so that $y_* = \mathbf{x}_*'\boldsymbol{\beta} + \varepsilon_*$. Goldberger called $\mathbf{x}_*$ the vector "prediction regressors" and ε_* the "prediction disturbance".

Premultiplying the model $\mathcal{M}$ by an $f \times n$ matrix $\mathbf{F}$ yields the transformed model

$$\mathbf{Fy} = \mathbf{FX}\boldsymbol{\beta} + \mathbf{F}\boldsymbol{\varepsilon}, \quad \text{or shortly } \mathcal{M}_t = \{\mathbf{Fy}, \mathbf{FX}\boldsymbol{\beta}, \mathbf{FVF}'\}. \qquad (1.5)$$

Suppose we wish to do the prediction using the transformed model $\mathcal{M}_t$. Corresponding to $\mathcal{M}_*$, we then have the following transformed setup:

$$\mathcal{M}_{t*} = \left\{ \begin{pmatrix} \mathbf{Fy} \\ \mathbf{y}_* \end{pmatrix}, \begin{pmatrix} \mathbf{FX} \\ \mathbf{X}_* \end{pmatrix}\boldsymbol{\beta}, \begin{pmatrix} \mathbf{FVF}' & \mathbf{FV}_{12} \\ \mathbf{V}_{21}\mathbf{F}' & \mathbf{V}_{22} \end{pmatrix} \right\}. \qquad (1.6)$$

We shall concentrate on the linear unbiased estimators, LUEs, and predictors, LUPs, and hence we need the concept of estimability. For example, $\mathbf{X}_*\boldsymbol{\beta}$ is estimable under $\mathcal{M}$ if there exists a matrix $\mathbf{B}$ such that $\mathrm{E}(\mathbf{By}) = \mathbf{X}_*\boldsymbol{\beta}$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$. Such a matrix $\mathbf{B} \in \mathbb{R}^{q \times n}$ exists only when $\mathcal{C}(\mathbf{X}_*') \subset \mathcal{C}(\mathbf{X}')$. The LUE $\mathbf{By}$ is the best linear unbiased estimator, BLUE, of estimable $\mathbf{X}_*\boldsymbol{\beta}$ if $\mathbf{By}$ has the smallest covariance matrix in the Löwner sense among all linear unbiased estimators of $\mathbf{X}_*\boldsymbol{\beta}$:

$$\mathrm{cov}(\mathbf{By}) \leq_{\mathrm{L}} \mathrm{cov}(\mathbf{B}_\#\mathbf{y}) \quad \text{for all } \mathbf{B}_\# : \mathbf{B}_\#\mathbf{X} = \mathbf{X}_*, \qquad (1.7)$$

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

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

$$\mathrm{cov}(\mathbf{y}_* - \mathbf{Dy}) \leq_{\mathrm{L}} \mathrm{cov}(\mathbf{y}_* - \mathbf{D}_\#\mathbf{y}) \quad \text{for all } \mathbf{D}_\# : \mathbf{D}_\#\mathbf{X} = \mathbf{X}_*. \qquad (1.8)$$

The Lemma 1.1 below provides so-called fundamental BLUE- and BLUP-equations. For the BLUP, see, e.g., Christensen [6, p. 294], and Isotalo & Puntanen [18, p. 1015], and for the BLUE, Drygas [7, p. 55], Rao [28, p. 282], and Puntanen et al. [27, Th. 10]. For the reviews of the BLUP-properties, see, Robinson [31] and Haslett & Puntanen [14].

Lemma 1.1. *Consider the linear model with new observations defined as $\mathcal{M}_*$ in (1.4), where $\mathcal{C}(\mathbf{X}_*') \subset \mathcal{C}(\mathbf{X}')$, i.e., $\mathbf{y}_*$ is predictable.*

(a) *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*

$$\mathbf{A}(\mathbf{X} : \mathbf{V}\mathbf{X}^\perp) = (\mathbf{X}_* : \mathbf{V}_{21}\mathbf{X}^\perp) \,. \tag{1.9}$$

(b) *The linear estimator* $\mathbf{By}$ *is the* BLUE *of* $\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}$ *if and only if* $\mathbf{B} \in \mathbb{R}^{q \times n}$ *satisfies the equation*

$$\mathbf{B}(\mathbf{X} : \mathbf{V}\mathbf{X}^\perp) = (\mathbf{X}_* : \mathbf{0}) \,. \tag{1.10}$$

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*

$$\mathbf{C}(\mathbf{X} : \mathbf{V}\mathbf{X}^\perp) = (\mathbf{X} : \mathbf{0}) \,. \tag{1.11}$$

(c) *The linear predictor* $\mathbf{Dy}$ *is the* BLUP *for* $\boldsymbol{\varepsilon}_*$ *if and only if* $\mathbf{D} \in \mathbb{R}^{q \times n}$ *satisfies the equation*

$$\mathbf{D}(\mathbf{X} : \mathbf{V}\mathbf{X}^\perp) = (\mathbf{0} : \mathbf{V}_{21}\mathbf{X}^\perp) \,. \tag{1.12}$$

We will use the following short notations:

$$\tilde{\mathbf{y}}_* = \mathrm{BLUP}(\mathbf{y}_* \mid \mathcal{M}_*) \,, \quad \tilde{\boldsymbol{\mu}}_* = \mathrm{BLUE}(\boldsymbol{\mu}_* \mid \mathcal{M}_*) \,, \quad \tilde{\boldsymbol{\varepsilon}}_* = \mathrm{BLUP}(\tilde{\boldsymbol{\varepsilon}} \mid \mathcal{M}_*) \,. \tag{1.13}$$

Notice that obviously $\mathrm{BLUE}(\boldsymbol{\mu}_* \mid \mathcal{M}_*) = \mathrm{BLUE}(\boldsymbol{\mu}_* \mid \mathcal{M})$.

Lemma 2.2.4 of Rao & Mitra [30] appears very useful for our considerations. It says that for nonnull matrices $\mathbf{A}$ and $\mathbf{C}$ the following holds:

$$\mathbf{AB}^-\mathbf{C} = \mathbf{AB}^+\mathbf{C} \iff \mathcal{C}(\mathbf{C}) \subset \mathcal{C}(\mathbf{B}) \ \& \ \mathcal{C}(\mathbf{A}') \subset \mathcal{C}(\mathbf{B}') \,. \tag{1.14}$$

In other words, (1.14) characerizes when the matrix product $\mathbf{AB}^-\mathbf{C}$ is invariant with respect to the choice of $\mathbf{B}^-$. In particular, we observe the following:

$$\mathbf{AB}^-\mathbf{B} = \mathbf{AB}^+\mathbf{B} \iff \mathcal{C}(\mathbf{A}') \subset \mathcal{C}(\mathbf{B}') \,. \tag{1.15}$$

One well-known solution for $\mathbf{C}$ in (1.11) is

$$\mathbf{P}_{\mathbf{X};\mathbf{W}^-} := \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^- \,, \tag{1.16}$$

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

$$\mathcal{W} = \left\{ \mathbf{W} \in \mathrm{NND}_n : \mathbf{W} = \mathbf{V} + \mathbf{X}\mathbf{U}\mathbf{U}'\mathbf{X}', \ \mathcal{C}(\mathbf{W}) = \mathcal{C}(\mathbf{X} : \mathbf{V}) \right\} \,. \tag{1.17}$$

In (1.17) the matrix $\mathbf{U}$ (having p rows) can be chosen arbitrarily subject to the condition $\mathcal{C}(\mathbf{W}) = \mathcal{C}(\mathbf{X} : \mathbf{V})$. One obvious choice is $\mathbf{I}_p$ and we can choose $\mathbf{U} = \mathbf{0}$ if $\mathcal{C}(\mathbf{X}) \subset \mathcal{C}(\mathbf{V})$. We could replace $\mathcal{W}$ with a set of matrices of the type $\mathbf{W} = \mathbf{V} + \mathbf{X}\mathbf{U}\mathbf{X}'$, where $\mathcal{C}(\mathbf{W}) = \mathcal{C}(\mathbf{X} : \mathbf{V})$, $\mathbf{U} \in \mathbb{R}^{p \times p}$, and thus $\mathbf{W}$ would not necessarily be nonnegative definite. However, to simplify our considerations, we will use (1.17) to define the set $\mathcal{W}$. In view of (1.14), 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 $\mathbf{P}_{\mathbf{X};\mathbf{W}^-}\mathbf{y}$ for $\mathbf{y} \in \mathcal{C}(\mathbf{X} : \mathbf{V})$. For a review of the properties of $\mathcal{W}$, see, e.g., Puntanen et al. [27, Sec. 12.3].

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

$$\mathbf{y} \in \mathcal{C}(\mathbf{X} : \mathbf{V}) = \mathcal{C}(\mathbf{X} : \mathbf{V}\mathbf{X}^\perp) = \mathcal{C}(\mathbf{X} : \mathbf{V}\mathbf{M}) = \mathcal{C}(\mathbf{X}) \oplus \mathcal{C}(\mathbf{V}\mathbf{M}) \,, \tag{1.18}$$

where $\oplus$ refers to the direct sum. For the equality $\mathcal{C}(\mathbf{X} : \mathbf{V}) = \mathcal{C}(\mathbf{X} : \mathbf{V}\mathbf{M})$, see, e.g., Rao [29, Lemma 2.1]. The corresponding consistency as in (1.18) is assumed in all models that we will deal with.

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

$$\mathbf{Ay} = \mathbf{By} \quad \text{for all } \mathbf{y} \in \mathcal{C}(\mathbf{X} : \mathbf{V}) \,, \tag{1.19}$$

which will be a crucial property in our considerations. For the equality of two estimators, see, e.g., Groß & Trenkler [11].

As for the structure of this article, our main goal is to introduce upper bounds for the Euclidean distances between the best linear unbiased predictors (BLUPs) of $\mathbf{y}_*$ when the prediction is based on the original model $\mathcal{M}_*$, defined in (1.4), and when it is based on the transformed model $\mathcal{M}_{t*}$, defined in (1.6). Corresponding considerations are made for the BLUPs of $\boldsymbol{\varepsilon}_*$. Our attempt has been to make the paper self-readable so that the necessary background concepts, starting from BLUP and BLUE, have been presented to a reasonable amount. In Section 2 we introduce, review and comment on some presentations for the BLUPs under the original and the transformed model. Because the concept of linear sufficiency is strongly connected with the transformed model, we devote Section 3 for this topic. In Section 5 we apply our results into the mixed linear model which is a special case of the model with new future observations. Our considerations are rather mathematical and for this paper, we have no practical statistical applications in mind.

2 Representations for the BLUPs under the original and the transformed model

If $\mathbf{B}_1\mathbf{y} = \text{BLUE}(\mathbf{X}_*\boldsymbol{\beta})$ and $\mathbf{B}_2\mathbf{y} = \text{BLUP}(\boldsymbol{\varepsilon}_*)$ under $\mathcal{M}_*$, then, in view of Lemma 1.1, we have

$$\begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{pmatrix} (\mathbf{X} : \mathbf{VM}) = \begin{pmatrix} \mathbf{X}_* & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{21}\mathbf{M} \end{pmatrix}. \tag{2.1}$$

Premultiplying (2.1) by $(\mathbf{I}_q : \mathbf{I}_q)$ leads to

$$(\mathbf{B}_1 + \mathbf{B}_2)(\mathbf{X} : \mathbf{VM}) = (\mathbf{X}_* : \mathbf{V}_{21}\mathbf{M}), \tag{2.2}$$

and thereby $(\mathbf{B}_1 + \mathbf{B}_2)\mathbf{y} = \text{BLUP}(\mathbf{y}_*)$, i.e.,

$$\text{BLUP}(\mathbf{y}_*) = \text{BLUE}(\mathbf{X}_*\boldsymbol{\beta}) + \text{BLUP}(\boldsymbol{\varepsilon}_*), \quad \text{or shortly}, \quad \tilde{\mathbf{y}}_* = \tilde{\boldsymbol{\mu}}_* + \tilde{\boldsymbol{\varepsilon}}_*. \tag{2.3}$$

Recall that equations of the type (2.3) hold "with probability 1", that is, they hold for all $\mathbf{y} \in \mathcal{C}(\mathbf{X} : \mathbf{V})$.

Now one solution for $\mathbf{B}_2$ satisfying $\mathbf{B}_2(\mathbf{X} : \mathbf{VM}) = (\mathbf{0} : \mathbf{V}_{21}\mathbf{M})$ is $\mathbf{B}_2 = \mathbf{AM}$, where $\mathbf{A}$ satisfies $\mathbf{AMVM} = \mathbf{V}_{21}\mathbf{M}$. Thus one expression for the BLUP of $\boldsymbol{\varepsilon}_*$ is

$$\text{BLUP}(\boldsymbol{\varepsilon}_*) = \mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^-\mathbf{M}\mathbf{y} = \mathbf{V}_{21}\dot{\mathbf{M}}\mathbf{y}, \tag{2.4}$$

where we have denoted

$$\dot{\mathbf{M}} = \mathbf{M}(\mathbf{MVM})^-\mathbf{M}. \tag{2.5}$$

The matrix $\dot{\mathbf{M}}$ appears to be very useful in this context. As noted by Isotalo et al. [17, p. 1439], the matrix $\dot{\mathbf{M}}$ is not necessarily unique with respect to the choice of the generalized inverse $(\mathbf{MVM})^-$. It is unique if and only if $\mathcal{C}(\mathbf{M}) \subset \mathcal{C}(\mathbf{MV})$, which further is equivalent to $\mathbb{R}^n = \mathcal{C}(\mathbf{X} : \mathbf{V})$. However, in view of (1.14) and the fact that $\mathbf{y} \in \mathcal{C}(\mathbf{X} : \mathbf{V})$, the expression $\mathbf{V}_{21}\dot{\mathbf{M}}\mathbf{y}$ is invariant with respect to the choice of $(\mathbf{MVM})^-$. For the Moore–Penrose inverse we have

$$\mathbf{M}(\mathbf{MVM})^+\mathbf{M} = (\mathbf{MVM})^+\mathbf{M} = \mathbf{M}(\mathbf{MVM})^+ = (\mathbf{MVM})^+. \tag{2.6}$$

Hence in light of (1.14) and (2.6),

$$\text{BLUP}(\boldsymbol{\varepsilon}_*) = \mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^-\mathbf{M}\mathbf{y} = \mathbf{V}_{21}(\mathbf{MVM})^+\mathbf{y} \quad \text{for all } \mathbf{y} \in \mathcal{C}(\mathbf{X} : \mathbf{V}). \tag{2.7}$$

For further properties of $\dot{\mathbf{M}}$, see Isotalo et al. [17] and Puntanen et al. [27, Ch. 17].

It is also of interest to substitute $\mathbf{y} = \mathbf{Xa} + \mathbf{Vb}$ (for some vectors $\mathbf{a}$ and $\mathbf{b}$) into (2.4) and obtain

$$\begin{aligned} \text{BLUP}(\boldsymbol{\varepsilon}_*) &= \mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^-\mathbf{M}(\mathbf{Xa} + \mathbf{Vb}) \\ &= \mathbf{V}_{21}\mathbf{V}^{+1/2}\mathbf{P}_{\mathbf{V}^{1/2}\mathbf{M}}\mathbf{V}^{+1/2}\mathbf{Vb} \\ &= \mathbf{V}_{21}\mathbf{V}^{+1/2}\mathbf{P}_{\mathbf{V}^{1/2}\mathbf{M}_0}\mathbf{V}^{+1/2}\mathbf{Vb} \end{aligned}$$

$$= \mathbf{V}_{21}\mathbf{M}_0(\mathbf{M}_0'\mathbf{V}\mathbf{M}_0)^{-}\mathbf{M}_0'\mathbf{y}, \tag{2.8}$$

where $\mathbf{M}_0$ is any matrix satisfying $\mathcal{C}(\mathbf{M}_0) = \mathcal{C}(\mathbf{M})$; here $\mathbf{V}^{+1/2}$ is the nonnegative definite square root of $\mathbf{V}^{+}$, and thereby $\mathbf{V}^{+1/2}\mathbf{V}^{1/2} = \mathbf{V}^{1/2}\mathbf{V}^{+1/2} = \mathbf{P_V}$. Observe that $\mathbf{V}_{21}\mathbf{P_V} = \mathbf{V}_{21}$ because $\mathcal{C}(\mathbf{V}_{12}) \subset \mathcal{C}(\mathbf{V})$ in light of the nonnegative definiteness of $\boldsymbol{\Gamma}$ in (1.3).

In view of the identity, see Haslett et al. [13, Sec. 2],

$$\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{+} = \mathbf{P_W} - \mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{P_W}, \tag{2.9}$$

the residual of the BLUE($\boldsymbol{\mu} \mid \mathcal{M}$) can be written as

$$\mathbf{y} - \text{BLUE}(\boldsymbol{\mu} \mid \mathcal{M}) = (\mathbf{I}_n - \mathbf{G})\mathbf{y} = \mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y} \quad \text{for all } \mathbf{y} \in \mathcal{C}(\mathbf{W}), \tag{2.10}$$

and the BLUP($\boldsymbol{\varepsilon}_*$) can be expressed, for example, as follows:

$$\text{BLUP}(\boldsymbol{\varepsilon}_*) = \mathbf{V}_{21}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y} = \mathbf{V}_{21}\mathbf{W}^{-}(\mathbf{I}_n - \mathbf{G})\mathbf{y} = \mathbf{V}_{21}\mathbf{V}^{-}(\mathbf{I}_n - \mathbf{G})\mathbf{y}, \tag{2.11}$$

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

For our purposes we assume that the parametric function $\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}$ is estimable under $\mathcal{M}$ as well as under $\mathcal{M}_t$, which happens if and only if $\mathcal{C}(\mathbf{X}_*') \subset \mathcal{C}(\mathbf{X}') \cap \mathcal{C}(\mathbf{X}'\mathbf{F}') = \mathcal{C}(\mathbf{X}'\mathbf{F}')$ so that

$$\mathbf{X}_* = \mathbf{L}\mathbf{F}\mathbf{X} \quad \text{for some matrix } \mathbf{L} \in \mathbb{R}^{q \times f}, \quad \boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta} = \mathbf{L}\mathbf{F}\mathbf{X}\boldsymbol{\beta} = \mathbf{L}\mathbf{F}\boldsymbol{\mu}. \tag{2.12}$$

In other words, estimability under $\mathcal{M}_t$ implies the estimability under $\mathcal{M}$. The parametric function $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ is of course always estimable under $\mathcal{M}$ while under $\mathcal{M}_t$ it is estimable whenever

$$\mathcal{C}(\mathbf{X}') = \mathcal{C}(\mathbf{X}'\mathbf{F}'), \quad \text{i.e.,} \quad \text{rank}(\mathbf{X}) = \text{rank}(\mathbf{F}\mathbf{X}). \tag{2.13}$$

In light of (2.12), one expression for BLUE($\mathbf{X}_*\boldsymbol{\beta}$) under $\mathcal{M}$ is

$$\text{BLUE}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathcal{M}) = \text{BLUE}(\mathbf{L}\mathbf{F}\mathbf{X}\boldsymbol{\beta} \mid \mathcal{M}) = \mathbf{L}\mathbf{F}\,\text{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathcal{M}) = \mathbf{L}\mathbf{F}\mathbf{G}\mathbf{y}, \tag{2.14}$$

where $\mathbf{G} = \mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-} = \mathbf{P}_{\mathbf{X};\mathbf{W}^{-}}$.

If $\mathbf{X}\boldsymbol{\beta}$ is estimable under $\mathcal{M}_t$, then, in view of part (b) of Lemma 1.1, the statistic $\mathbf{B}\mathbf{F}\mathbf{y}$ is the BLUE for $\mathbf{X}\boldsymbol{\beta}$ under $\mathcal{M}_t$ if and only if $\mathbf{B}$ satisfies

$$\mathbf{B}(\mathbf{F}\mathbf{X} : \mathbf{F}\mathbf{V}\mathbf{F}'\mathbf{Q_{FX}}) = (\mathbf{X} : \mathbf{0}). \tag{2.15}$$

Thus, see Kala et al. [19, Sec. 6] and Markiewicz & Puntanen [24, Sec. 3], the BLUE of $\mathbf{X}\boldsymbol{\beta}$ under $\mathcal{M}_t$ has, for example, the representation

$$\text{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathcal{M}_t) = \text{BLUE}(\boldsymbol{\mu} \mid \mathcal{M}_t) = \tilde{\boldsymbol{\mu}}_t = \mathbf{G}_t\mathbf{y}, \tag{2.16}$$

where

$$\mathbf{G}_t = \mathbf{X}[\mathbf{X}'\mathbf{F}'(\mathbf{F}\mathbf{W}\mathbf{F}')^{-}\mathbf{F}\mathbf{X}]^{-}\mathbf{X}'\mathbf{F}'(\mathbf{F}\mathbf{W}\mathbf{F}')^{-}\mathbf{F}. \tag{2.17}$$

Correspondingly,

$$\text{BLUE}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathcal{M}_t) = \text{BLUE}(\mathbf{L}\mathbf{F}\mathbf{X}\boldsymbol{\beta} \mid \mathcal{M}_t) = \mathbf{L}\mathbf{F}\mathbf{G}_t\mathbf{y} = \tilde{\boldsymbol{\mu}}_t. \tag{2.18}$$

The BLUP of $\boldsymbol{\varepsilon}_*$ under $\mathcal{M}_{t*}$ is $\mathbf{C}\mathbf{F}\mathbf{y}$ if and only if $\mathbf{C}$ satisfies

$$\mathbf{C}(\mathbf{F}\mathbf{X} : \mathbf{F}\mathbf{V}\mathbf{F}'\mathbf{Q_{FX}}) = (\mathbf{0} : \mathbf{V}_{21}\mathbf{F}'\mathbf{Q_{FX}}), \tag{2.19}$$

that is, $\mathbf{C} = \mathbf{A}\mathbf{Q_{FX}}$, where $\mathbf{A}$ satisfies $\mathbf{A}\mathbf{Q_{FX}}\mathbf{F}\mathbf{V}\mathbf{F}'\mathbf{Q_{FX}} = \mathbf{V}_{21}\mathbf{F}'\mathbf{Q_{FX}}$. Thus one expression for BLUP of $\boldsymbol{\varepsilon}_*$ under $\mathcal{M}_{t*}$ is

$$\text{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathcal{M}_{t*}) = \mathbf{V}_{21}\mathbf{F}'\mathbf{Q_{FX}}(\mathbf{Q_{FX}}\mathbf{F}\mathbf{V}\mathbf{F}'\mathbf{Q_{FX}})^{-}\mathbf{Q_{FX}}\mathbf{F}\mathbf{y}. \tag{2.20}$$

According to Markiewicz & Puntanen [24, Sec. 2], we have

$$\mathcal{C}(\mathbf{F}'\mathbf{Q_{FX}}) = \mathcal{C}(\mathbf{F}') \cap \mathcal{C}(\mathbf{M}), \quad \mathbf{M}\mathbf{F}'\mathbf{Q_{FX}} = \mathbf{F}'\mathbf{Q_{FX}}. \tag{2.21}$$

Denoting

$$N = P_{F'Q_{FX}} = P_{\mathcal{C}(F') \cap \mathcal{C}(M)} \, , \tag{2.22}$$

and proceeding as in (2.8) we get the following expression:

$$\text{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathcal{M}_{t*}) = V_{21}N(NVN)^-Ny. \tag{2.23}$$

Thus, see also Isotalo et al. [16, Sec. 4], the BLUP(y_*) under $\mathcal{M}_*$ can be written as

$$\begin{aligned}
\text{BLUP}(y_* \mid \mathcal{M}_*) &= \text{BLUE}(\boldsymbol{\mu}_* \mid \mathcal{M}) + V_{21}V^-[y - \text{BLUE}(\boldsymbol{\mu} \mid \mathcal{M})] \\
&= LFGy + V_{21}V^-(I_n - G)y \\
&= LFGy + V_{21}M(MVM)^-My \\
&= \text{BLUE}(\boldsymbol{\mu}_* \mid \mathcal{M}) + \text{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathcal{M}_*) \, ,
\end{aligned} \tag{2.24}$$

or shortly, $\tilde{y}_* = \tilde{\boldsymbol{\mu}}_* + \tilde{\boldsymbol{\varepsilon}}_*$, and

$$\begin{aligned}
\text{BLUP}(y_* \mid \mathcal{M}_{t*}) &= \text{BLUE}(\boldsymbol{\mu}_* \mid \mathcal{M}_t) + V_{21}F'(FVF')^-F[y - \text{BLUE}(\boldsymbol{\mu} \mid \mathcal{M}_t)] \\
&= LFG_ty + V_{21}F'(FVF')^-F(I_n - G_t)y \\
&= LFG_ty + V_{21}N(NVN)^-Ny \\
&= \text{BLUE}(\boldsymbol{\mu}_* \mid \mathcal{M}_t) + \text{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathcal{M}_{t*}) \, ,
\end{aligned} \tag{2.25}$$

or shortly, $\tilde{y}_{t*} = \tilde{\boldsymbol{\mu}}_{t*} + \tilde{\boldsymbol{\varepsilon}}_{t*}$. Recall that $N = P_{F'Q_{FX}} = P_{\mathcal{C}(F') \cap \mathcal{C}(M)}$ and that N has properties

$$\mathcal{C}(N) = \mathcal{C}(F'Q_{FX}) = \mathcal{C}(F') \cap \mathcal{C}(M) \, , \quad N = MN = NM = N^2 \, . \tag{2.26}$$

Notice that the use of term BLUE($X\boldsymbol{\beta} \mid \mathcal{M}_t$), as in the first two expressions in (2.25), requires, of course, that $X\boldsymbol{\beta}$ is estimable under the transformed model $\mathcal{M}_t$. The use of other expressions in (2.25) does not require this assumption; the estimability of $X_*\boldsymbol{\beta}$ under $\mathcal{M}_t$ is only needed.

We observe that the random vectors $\tilde{\boldsymbol{\mu}}_*$ and $\tilde{\boldsymbol{\varepsilon}}_*$ are uncorrelated and the corresponding property holds also for $\tilde{\boldsymbol{\mu}}_{t*}$ and $\tilde{\boldsymbol{\varepsilon}}_{t*}$. Hence we have

$$\text{cov}(\tilde{y}_*) = \text{cov}(\tilde{\boldsymbol{\mu}}_*) + \text{cov}(\tilde{\boldsymbol{\varepsilon}}_*) \, , \quad \text{cov}(\tilde{y}_{t*}) = \text{cov}(\tilde{\boldsymbol{\mu}}_{t*}) + \text{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*}) \, . \tag{2.27}$$

Now we have $\tilde{\boldsymbol{\varepsilon}}_* = V_{21}M(MVM)^-My$, and $\tilde{\boldsymbol{\varepsilon}}_{t*} = V_{21}N(NVN)^-Ny$, with covariance matrices

$$\text{cov}(\tilde{\boldsymbol{\varepsilon}}_*) = V_{21}M(MVM)^-MV_{12} \, , \quad \text{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*}) = V_{21}N(NVN)^-NV_{12} \, . \tag{2.28}$$

Straightforward calculation shows that $\text{cov}(\tilde{\boldsymbol{\varepsilon}}_*, \, \tilde{\boldsymbol{\varepsilon}}_{t*}) = \text{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*})$, and

$$\text{cov}(\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}) = \text{cov}(\tilde{\boldsymbol{\varepsilon}}_*) - \text{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*}) \, , \tag{2.29}$$

and thereby we have the Löwner ordering $\text{cov}(\tilde{\boldsymbol{\varepsilon}}_*) \geq_L \text{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*})$. It is worth noting that for $\tilde{\boldsymbol{\mu}}_*$ and $\tilde{\boldsymbol{\mu}}_{t*}$ we have the reverse Löwner ordering $\text{cov}(\tilde{\boldsymbol{\mu}}_*) \leq_L \text{cov}(\tilde{\boldsymbol{\mu}}_{t*})$.

3 Conditions for linear sufficiency

Consider the model $\mathcal{M} = \{y, X\boldsymbol{\beta}, V\}$ and let F be an $f \times n$ matrix. Then Fy is called linearly sufficient (sometimes called BLUE-sufficient) for estimable $X_*\boldsymbol{\beta}$, where $X_* \in \mathbb{R}^{q \times p}$, if there exists a matrix $A_{q \times f}$ such that AFy is the BLUE for $X_*\boldsymbol{\beta}$. We use the notation $Fy \in \mathcal{S}(X_*\boldsymbol{\beta})$ do indicate that Fy is linearly sufficient for $X_*\boldsymbol{\beta}$.

Let y_* be predictable under the model $\mathcal{M}_*$, i.e., $\mathcal{C}(X'_*) \subset \mathcal{C}(X')$. Then Fy is called linearly (prediction) sufficient (BLUP-sufficient) for y_* if there exists a matrix $A_{q \times f}$ such that AFy is the BLUP for y_*; that is, there exists A such that

$$AF(X : VM) = (X_* : V_{21}M) \, . \tag{3.1}$$

If we want to emphasize that we are dealing with the prediction, we could talk about linear *prediction* sufficiency. We use the short notation $\mathbf{Fy} \in \mathcal{S}(\mathbf{y}_*)$.

The transformed model $\mathcal{M}_t$ 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 the linear sufficiency and correspondingly the equality of the BLUPs. For the following Lemma 3.1 collects some useful results. For (a)–(c), see, e.g., Baksalary & Kala [3, 4], Drygas [8], Tian & Puntanen [32, Th. 2.8], and Kala et al. [20, Th. 2], and for (d)–(f), see Isotalo & Puntanen [18], and Isotalo et al. [16].

Lemma 3.1. *Let $\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}$ be estimable under $\mathcal{M}_t$ (and thereby under $\mathcal{M}$). Then the following statements are equivalent:*
(a) $\mathbf{Fy}$ *is linearly sufficient for* $\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}$, *i.e.,* $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta})$,
(b) $\mathrm{BLUE}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathcal{M}_*) = \mathrm{BLUE}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathcal{M}_{t*})$, *or shortly,* $\tilde{\boldsymbol{\mu}}_* = \tilde{\boldsymbol{\mu}}_{t*}$ *with probability* 1,
(c) $\mathrm{cov}(\tilde{\boldsymbol{\mu}}_*) = \mathrm{cov}(\tilde{\boldsymbol{\mu}}_{t*})$.
     Moreover, the following statements are equivalent:
(d) $\mathbf{Fy}$ *is linearly sufficient for* $\mathbf{y}_*$, *i.e.,* $\mathbf{Fy} \in \mathcal{S}(\mathbf{y}_*)$,
(e) $\mathrm{BLUP}(\mathbf{y}_* \mid \mathcal{M}_*) = \mathrm{BLUP}(\mathbf{y}_* \mid \mathcal{M}_{t*})$, *or shortly,* $\tilde{\mathbf{y}}_* = \tilde{\mathbf{y}}_{t*}$ *with probability* 1,
(f) $\mathrm{cov}(\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}) = \mathbf{0}$.

The following lemma gives some BLUP-sufficiency properties of $\mathbf{Fy}$ for $\boldsymbol{\varepsilon}_*$; see Isotalo et al. [16].

Lemma 3.2. *The following statements are equivalent:*
(a) $\mathbf{Fy}$ *is linearly sufficient for* $\boldsymbol{\varepsilon}_*$, *i.e.,* $\mathbf{Fy} \in \mathcal{S}(\boldsymbol{\varepsilon}_*)$,
(b) $\mathcal{C}(\mathbf{MV}_{12}) \subset \mathcal{C}(\mathbf{MVF}'\mathbf{Q_{FX}})$,
(c) $\mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathcal{M}_*) = \mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathcal{M}_{t*})$, *or shortly,* $\tilde{\boldsymbol{\varepsilon}}_* = \tilde{\boldsymbol{\varepsilon}}_{t*}$ *with probability* 1,
(d) $\mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*) = \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*})$,
(e) $\mathcal{C}(\mathbf{V}_{12}) \subset \mathcal{C}(\mathbf{VN} : \mathbf{X}) = \mathcal{C}(\mathbf{VF}'\mathbf{Q_{FX}} : \mathbf{X})$, *where* $\mathbf{N} = \mathbf{P}_{\mathbf{F}'\mathbf{Q_{FX}}}$,
(f) $\mathbf{V}_{21}\mathbf{M} = \mathbf{V}_{21}\mathbf{N}(\mathbf{NVN})^-\mathbf{NVM}$.

Details of Lemma 3.2 are proved in Markiewicz & Puntanen [25] but let us take a brief look at the claim (f), which will be needed later on. To confirm (f), we can start from (c):

$$\mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathcal{M}_*) = \mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathcal{M}_{t*}) \quad \text{with probability } 1, \tag{3.2}$$

i.e.,

$$\mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^-\mathbf{My} = \mathbf{V}_{21}\mathbf{N}(\mathbf{NVN})^-\mathbf{Ny} \quad \text{for all } \mathbf{y} \in \mathcal{C}(\mathbf{X} : \mathbf{VM}), \tag{3.3}$$

where $\mathbf{N} = \mathbf{P}_{\mathbf{F}'\mathbf{Q_{FX}}}$ and $\mathbf{N}$ has properties like in (2.26). Choosing $\mathbf{y} \in \mathcal{C}(\mathbf{X})$ yields zeros on both sides of (3.3). For $\mathbf{y} \in \mathcal{C}(\mathbf{VM})$ the left-hand side of (3.3) becomes

$$\mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^-\mathbf{MVM} = \mathbf{V}_{21}\mathbf{M}, \tag{3.4}$$

where we have used (1.15). Hence (3.3) can be expressed as

$$\mathbf{V}_{21}\mathbf{M} = \mathbf{V}_{21}\mathbf{N}(\mathbf{NVN})^-\mathbf{NVM} = \mathbf{V}_{21}\mathbf{MN}(\mathbf{NVN})^-\mathbf{NVM} := \mathbf{V}_{21}\mathbf{ME}, \tag{3.5}$$

where $\mathbf{E} = \mathbf{N}(\mathbf{NVN})^-\mathbf{NVM} \in \mathbb{R}^{n \times n}$.

4 Some upper bounds for the Euclidean distance between the BLUPs

In this section we provide new results giving upper bounds for the Euclidean norms of differences

$$\mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathcal{M}_*) - \mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathcal{M}_{t*}) \quad \text{and} \quad \mathrm{BLUP}(\mathbf{y}_* \mid \mathcal{M}_*) - \mathrm{BLUP}(\mathbf{y}_* \mid \mathcal{M}_{t*}). \tag{4.1}$$

For this purpose, let $\mathrm{ch}_1(\cdot)$ denote the largest eigenvalue of the matrix argument and let the matrix norm be defined as $\|\mathbf{A}\|_2 = \sqrt{\mathrm{ch}_1(\mathbf{A}\mathbf{A}')}$. In particular, $\|\mathbf{a}\|_2^2 = \mathbf{a}'\mathbf{a}$, where $\mathbf{a} \in \mathbb{R}^n$.

Denoting $\tilde{\boldsymbol{\varepsilon}}_* = \mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathcal{M}_*)$, $\tilde{\boldsymbol{\varepsilon}}_{t*} = \mathrm{BLUP}(\boldsymbol{\varepsilon}_{t*} \mid \mathcal{M}_{t*})$, and $\mathbf{N} = \mathbf{P}_{\mathbf{F}'\mathbf{Q}_{\mathbf{FX}}} = \mathbf{MN}$, we have

$$\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*} = \mathbf{V}_{21}[\mathbf{M}(\mathbf{MVM})^-\mathbf{M} - \mathbf{N}(\mathbf{NVN})^-\mathbf{N}]\mathbf{y} = \mathbf{S}_0\mathbf{y}, \tag{4.2}$$

where

$$\mathbf{S}_0 = \mathbf{V}_{21}[\mathbf{M}(\mathbf{MVM})^-\mathbf{M} - \mathbf{N}(\mathbf{NVN})^-\mathbf{N}]. \tag{4.3}$$

Putting $\mathbf{y} = \mathbf{Xa} + \mathbf{VMb}$ and denoting $\mathbf{E} = \mathbf{N}(\mathbf{NVN})^-\mathbf{NVM}$, gives, on account of (3.5),

$$\begin{aligned}
\|\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}\|_2^2 &= \|\mathbf{S}_0(\mathbf{Xa} + \mathbf{VMb})\|_2^2 \\
&= \|\mathbf{V}_{21}\mathbf{M}[\mathbf{I}_n - \mathbf{N}(\mathbf{NVN})^-\mathbf{NVM}]\mathbf{Mb}\|_2^2 \\
&= \|\mathbf{V}_{21}\mathbf{M}(\mathbf{I}_n - \mathbf{E})\mathbf{Mb}\|_2^2 \\
&:= \|\mathbf{SMb}\|_2^2 \\
&\le \|\mathbf{S}\|_2^2 \|\mathbf{Mb}\|_2^2 \\
&= \mathrm{ch}_1(\mathbf{SS}')\,\mathbf{b}'\mathbf{Mb}, \tag{4.4}
\end{aligned}$$

where

$$\mathbf{S} = \mathbf{S}_0\mathbf{VM} = \mathbf{V}_{21}\mathbf{M}(\mathbf{I}_n - \mathbf{E}). \tag{4.5}$$

The inequality in (4.4) follows from the consistency and multiplicativity of the matrix norm $\|\cdot\|_2$; see, for example, Ben-Israel & Greville [5, pp. 19–20]. In view of part (f) of Lemma 3.2, we know that $\mathbf{S} = \mathbf{0}$ if and only if $\mathbf{Fy} \in \mathcal{S}(\boldsymbol{\varepsilon}_*)$. The scalar $\mathbf{b}'\mathbf{Mb}$ is obviously zero if and only if $\mathbf{b} \in \mathcal{C}(\mathbf{X})$. Thus we have proved the following theorem.

Theorem 4.1. *Consider the model* $\mathcal{M}_*$. *Then for all* $\mathbf{y} = \mathbf{Xa} + \mathbf{VMb}$,

$$\begin{aligned}
\|\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}\|_2^2 &= \|\mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathcal{M}_*) - \mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathcal{M}_{t*})\|_2^2 \\
&\le \mathrm{ch}_1(\mathbf{SS}')\,\mathbf{b}'\mathbf{Mb} := \alpha_1, \tag{4.6}
\end{aligned}$$

where

$$\mathbf{S} = \mathbf{V}_{21}\mathbf{M}(\mathbf{I}_n - \mathbf{E}) \in \mathbb{R}^{q \times n}, \quad \mathbf{E} = \mathbf{N}(\mathbf{NVN})^-\mathbf{NVM} \in \mathbb{R}^{n \times n}. \tag{4.7}$$

If $\mathbf{b} \notin \mathcal{C}(\mathbf{X})$, *then the upper bound* α_1 *in* (4.6) *is equal to zero if and only if* $\mathbf{S} = \mathbf{0}$, *i.e.,* $\mathbf{Fy}$ *is linearly sufficient for* $\boldsymbol{\varepsilon}_*$.

An alternative upper bound can be found out as follows:

$$\begin{aligned}
\|\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}\|_2^2 &= \|\mathbf{S}_0\mathbf{My}\|_2^2 \le \|\mathbf{S}_0\|_2^2\,\mathbf{y}'\mathbf{My} \\
&= \mathrm{ch}_1(\mathbf{S}_0\mathbf{S}_0')\,\mathbf{y}'\mathbf{My} = \mathrm{ch}_1(\mathbf{S}_0\mathbf{S}_0')\,\mathbf{b}'\mathbf{MVMVMb} := \alpha_2, \tag{4.8}
\end{aligned}$$

where $\mathbf{S}_0 = \mathbf{V}_{21}[\mathbf{M}(\mathbf{MVM})^-\mathbf{M} - \mathbf{N}(\mathbf{NVN})^-\mathbf{N}]$. It is easy to conclude that $\mathbf{y}'\mathbf{My} = 0$ for all $\mathbf{y} \in \mathcal{C}(\mathbf{X} : \mathbf{V})$ if and only if $\mathbf{VM} = \mathbf{0}$, or, equivalently, $\mathcal{C}(\mathbf{V}) \subset \mathcal{C}(\mathbf{X})$. Groß [10, p. 317] calls a model with property $\mathbf{VM} = \mathbf{0}$ a *degenerated* model.

Remark 4.1. As one of the referees pointed out, the upper bounds in (4.6) and (4.8) depend on vector $\mathbf{y}$, which is an arbitrary vector in $\mathbb{R}^n$ belonging to the column space $\mathcal{C}(\mathbf{X} : \mathbf{V})$. If $\mathbf{VM} = \mathbf{0}$, then $\alpha_1 = \alpha_2 = 0$, but of course the situation is somewhat pathological as $\mathbf{y}'\mathbf{My} = 0$ for all $\mathbf{y} \in \mathcal{C}(\mathbf{X} : \mathbf{V})$. $\qquad\square$

If $\mathcal{M}_*$ is not a degenerated model, then the upper bound α_2 in (4.8) is equal to zero if and only if

$$\mathbf{S}_0 = \mathbf{V}_{21}[\mathbf{M}(\mathbf{MVM})^-\mathbf{M} - \mathbf{N}(\mathbf{NVN})^-\mathbf{N}] = \mathbf{0}, \tag{4.9}$$

or, equivalently,

$$\mathbf{S}_0\mathbf{S}_0' = \mathbf{V}_{21}[\mathbf{M}(\mathbf{MVM})^-\mathbf{M} - \mathbf{N}(\mathbf{NVN})^-\mathbf{N}]^2\mathbf{V}_{12} = \mathbf{0}. \tag{4.10}$$

Of course, (4.9) implies

$$\mathbf{V}_{21}[\mathbf{M}(\mathbf{MVM})^-\mathbf{M} - \mathbf{N}(\mathbf{NVN})^-\mathbf{N}]\mathbf{V}_{12} = \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*) - \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*}) = \mathbf{0}\,, \tag{4.11}$$

as well as

$$\mathbf{S}_0\mathbf{VM} = \mathbf{V}_{21}[\mathbf{M}(\mathbf{MVM})^-\mathbf{M} - \mathbf{N}(\mathbf{NVN})^-\mathbf{N}]\mathbf{VM} = \mathbf{S} = \mathbf{0}\,, \tag{4.12}$$

which both are necessary and sufficient conditions for $\mathbf{Fy}$ being linearly sufficient for $\boldsymbol{\varepsilon}_*$. Interestingly, but somewhat unwishfully, the linear sufficiency, i.e., (4.12), does not imply that $\alpha_2 = 0$; exception for this is the case when $\mathcal{C}(\mathbf{X} : \mathbf{V}) = \mathbb{R}^n$. It remains an open question which upper bound α_1 or α_2 is sharper.

Let us take a look at the Euclidean distance between the BLUEs of $\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}$ in the original and the transformed model. As in Kala et al. [19, Sec. 6], we can observe that $\mathbf{G}_t\mathbf{G} = \mathbf{G}$ and hence

$$(\mathbf{G}_t - \mathbf{G})\mathbf{y} = (\mathbf{G}_t - \mathbf{G}_t\mathbf{G})\mathbf{y} = \mathbf{G}_t(\mathbf{I}_n - \mathbf{G})\mathbf{y} = \mathbf{G}_t\mathbf{VM}(\mathbf{MVM})^-\mathbf{My} \quad \text{for all } \mathbf{y} \in \mathcal{C}(\mathbf{W})\,, \tag{4.13}$$

where we have used (2.10). Then, for all $\mathbf{y} \in \mathcal{C}(\mathbf{W})$, and $\boldsymbol{\mu}_* = \mathbf{LFX}\boldsymbol{\beta}$, we have

$$
\begin{aligned}
\|\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}\|_2^2 &= \|\mathbf{LF}(\mathbf{G}_t - \mathbf{G})\mathbf{y}\|_2^2 \\
&= \|\mathbf{LFG}_t\mathbf{VM}(\mathbf{MVM})^-\mathbf{My}\|_2^2 \\
&\le \|\mathbf{LFG}_t\mathbf{VM}\|_2^2\,\|(\mathbf{MVM})^+\|_2^2\,\|\mathbf{My}\|_2^2 \\
&= \|\mathbf{R}\|_2^2\,\|(\mathbf{MVM})^+\|_2^2\,\|\mathbf{My}\|_2^2 \\
&= \frac{a}{b^2}\,\mathbf{y}'\mathbf{My} := \gamma_2\,,
\end{aligned}
\tag{4.14}
$$

where $\mathbf{R} = \mathbf{LFG}_t\mathbf{VM}$, the scalar a is the largest eigenvalue of $\mathbf{RR}'$, and b is the smallest nonzero eigenvalue of $\mathbf{MVM}$. Moreover, if $\mathcal{M}$ is not a degenerated model then γ_2 is zero if and only if $\mathbf{Fy}$ is linearly sufficient for $\mathbf{X}_*\boldsymbol{\beta}$.

An alternative upper bound for $\|\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}\|_2^2$ can be obtained by substituting $\mathbf{y} = \mathbf{Xa} + \mathbf{VMb}$ into (4.14). This yields

$$
\begin{aligned}
\|\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}\|_2^2 &= \|\mathbf{LFG}_t\mathbf{VM}(\mathbf{MVM})^-\mathbf{MVMb}\|_2^2 \\
&= \|\mathbf{LFG}_t\mathbf{VMb}\|_2^2 \\
&\le \|\mathbf{LFG}_t\mathbf{VM}\|_2^2\,\mathbf{b}'\mathbf{Mb} \\
&= \|\mathbf{R}\|_2^2\,\mathbf{b}'\mathbf{Mb} = a\,\mathbf{b}'\mathbf{Mb} := \gamma_1\,,
\end{aligned}
\tag{4.15}
$$

where $a = \mathrm{ch}_1(\mathbf{RR}')$. Similarly as for α_1 and α_2 in (4.6) and (4.8), the question whether γ_1 is a sharper upper bound than γ_2 remains open.

The BLUPs of $\mathbf{y}_*$ in the original and the transformed model, respectively, are

$$\mathrm{BLUP}(\mathbf{y}_* \mid \mathcal{M}_*) = \mathbf{LFGy} + \mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^-\mathbf{My}\,, \quad \text{or shortly,} \quad \tilde{\mathbf{y}}_* = \tilde{\boldsymbol{\mu}}_* + \tilde{\boldsymbol{\varepsilon}}_*\,, \tag{4.16a}$$

$$\mathrm{BLUP}(\mathbf{y}_* \mid \mathcal{M}_{t*}) = \mathbf{LFG}_t\mathbf{y} + \mathbf{V}_{21}\mathbf{N}(\mathbf{NVN})^-\mathbf{Ny}\,, \quad \text{or shortly,} \quad \tilde{\mathbf{y}}_{t*} = \tilde{\boldsymbol{\mu}}_{t*} + \tilde{\boldsymbol{\varepsilon}}_{t*}\,. \tag{4.16b}$$

Putting $\mathbf{y} = \mathbf{Xa} + \mathbf{VMb}$ and using earlier notation, gives, in light of the triangle inequality,

$$
\begin{aligned}
\|\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}\|_2 &= \|\mathrm{BLUP}(\mathbf{y}_* \mid \mathcal{M}_*) - \mathrm{BLUP}(\mathbf{y}_* \mid \mathcal{M}_{t*})\|_2 \\
&= \|(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}) + (\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*})\|_2 \\
&\le \|\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}\|_2 + \|\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}\|_2 \\
&= \|\mathbf{RMb}\|_2 + \|\mathbf{SMb}\|_2 \\
&= \sqrt{\mathrm{ch}_1(\mathbf{RR}')\,\mathbf{b}'\mathbf{Mb}} + \sqrt{\mathrm{ch}_1(\mathbf{SS}')\,\mathbf{b}'\mathbf{Mb}} \\
&= \gamma_1 + \alpha_1 := \alpha\,.
\end{aligned}
\tag{4.17}
$$

We can now write the following theorem.

Theorem 4.2. *Consider the model* $\mathcal{M}_*$ *where* $\boldsymbol{\mu}_* = \mathbf{LFX}\boldsymbol{\beta}$. *Then for all* $\mathbf{y} = \mathbf{Xa} + \mathbf{VMb}$,

$$\|\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}\|_2 \leq \sqrt{\mathrm{ch}_1(\mathbf{RR}')\,\mathbf{b}'\mathbf{Mb}} + \sqrt{\mathrm{ch}_1(\mathbf{SS}')\,\mathbf{b}'\mathbf{Mb}}$$

$$= \gamma_1 + \alpha_1 = \alpha\,, \tag{4.18}$$

where $\mathbf{R} = \mathbf{LFG}_t\mathbf{VM} \in \mathbb{R}^{q\times n}$ *and*

$$\mathbf{S} = \mathbf{V}_{21}\mathbf{M}(\mathbf{I}_n - \mathbf{E}) \in \mathbb{R}^{q\times n}, \quad \mathbf{E} = \mathbf{N}(\mathbf{NVN})^-\mathbf{NVM} \in \mathbb{R}^{n\times n}. \tag{4.19}$$

If $\mathbf{b} \notin \mathcal{C}(\mathbf{X})$, *then the upper bound* α *in* (4.18) *is equal to zero if and only if* $\mathbf{S} = \mathbf{R} = \mathbf{0}$, *i.e.,* $\mathbf{Fy}$ *is linearly sufficient for* $\boldsymbol{\mu}_*$ *and for* $\boldsymbol{\varepsilon}_*$.

Another formulation for the upper bound $\|\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}\|_2$ can be formulated using (4.6) and (4.14).

It is interesting to observe that $\mathbf{S} = \mathbf{R} = \mathbf{0}$ is sufficient but not necessary for the equality

$$\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*} = \mathbf{0}\,, \tag{4.20}$$

which holds (with probability 1) if and only if $\mathbf{Fy} \in \mathcal{S}(\mathbf{y}_*)$. However, $\mathbf{Fy} \in \mathcal{S}(\mathbf{y}_*)$ does not necessarily imply that $\mathbf{Fy} \in \mathcal{S}(\boldsymbol{\mu}_*) \cap \mathcal{S}(\boldsymbol{\varepsilon}_*)$, i.e., $\mathbf{S} = \mathbf{R} = \mathbf{0}$. For further discussion in this matter, see Markiewicz & Puntanen [25, Sec. 5].

5 BLUPs under mixed linear models

Consider the mixed linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{e}\,, \quad \text{denoted as} \quad \mathcal{L} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu}, \boldsymbol{\Delta}, \boldsymbol{\Phi}, \boldsymbol{\Psi}\}\,, \tag{5.1}$$

where $\mathbf{X}_{n\times p}$ and $\mathbf{Z}_{n\times q}$ are known matrices, $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of unknown fixed effects, $\mathbf{u}$ is an unobservable vector (q elements) of *random effects* with $\mathrm{E}(\mathbf{u}) = \mathbf{0}_q$, $\mathrm{cov}(\mathbf{u}) = \boldsymbol{\Delta}_{q\times q}$, $\mathbf{e}$ is a random error vector (n elements) with $\mathrm{E}(\mathbf{e}) = \mathbf{0}_n$, $\mathrm{cov}(\mathbf{e}) = \boldsymbol{\Phi}_{n\times n}$, and $\mathrm{cov}(\mathbf{e}, \mathbf{u}) = \boldsymbol{\Psi}_{n\times q}$. Denoting $\mathbf{g} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu}$, we have

$$\mathrm{cov}(\mathbf{y}) = \mathrm{cov}(\mathbf{Zu} + \mathbf{e}) = \mathbf{Z}\boldsymbol{\Delta}\mathbf{Z}' + \boldsymbol{\Phi} + \mathbf{Z}\boldsymbol{\Psi}' + \boldsymbol{\Psi}\mathbf{Z}' = \boldsymbol{\Sigma}\,, \tag{5.2a}$$

$$\mathrm{cov}\begin{pmatrix}\mathbf{y}\\\mathbf{g}\end{pmatrix} = \mathrm{cov}\begin{pmatrix}\mathbf{y}\\\mathbf{Zu}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\Sigma} & (\mathbf{Z}\boldsymbol{\Delta} + \boldsymbol{\Psi})\mathbf{Z}'\\\mathbf{Z}(\mathbf{Z}\boldsymbol{\Delta} + \boldsymbol{\Psi})' & \mathbf{Z}\boldsymbol{\Delta}\mathbf{Z}'\end{pmatrix} = \begin{pmatrix}\boldsymbol{\Sigma} & \boldsymbol{\Sigma}_{12}\\\boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix} := \boldsymbol{\Omega} \in \mathrm{NND}_{2n}\,. \tag{5.2b}$$

Now the mixed linear model can be expressed as a version of the model with "new observations", the new observations, corresponding $\mathbf{y}_*$ in (1.2), being in $\mathbf{g} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu}$:

$$\mathcal{L}_* = \left\{\begin{pmatrix}\mathbf{y}\\\mathbf{g}\end{pmatrix}, \begin{pmatrix}\mathbf{X}\\\mathbf{X}\end{pmatrix}\boldsymbol{\beta}, \begin{pmatrix}\boldsymbol{\Sigma} & \boldsymbol{\Sigma}_{12}\\\boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}\right\}\,. \tag{5.3}$$

Notice that $\boldsymbol{\Omega}$ in (5.2b) corresponds to $\boldsymbol{\Gamma}$ in (1.3).

Transforming the mixed model $\mathcal{L}$ by premultiplying it by $\mathbf{F} \in \mathbb{R}^{f\times n}$ gives

$$\mathbf{Fy} = \mathbf{FX}\boldsymbol{\beta} + \mathbf{FZu} + \mathbf{Fe}\,, \quad \text{denoted as} \quad \mathcal{L}_t = \{\mathbf{Fy}, \mathbf{FX}\boldsymbol{\beta} + \mathbf{FZu}, \boldsymbol{\Delta}, \mathbf{F}\boldsymbol{\Phi}\mathbf{F}', \mathbf{F}\boldsymbol{\Psi}\}\,. \tag{5.4}$$

Our aim is to do the prediction of $\mathbf{g} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu}$ using this transformed model $\mathcal{L}_t$. Corresponding to $\mathcal{L}_*$, the resulting transformed setup is

$$\mathcal{L}_{t*} = \left\{\begin{pmatrix}\mathbf{Fy}\\\mathbf{g}\end{pmatrix}, \begin{pmatrix}\mathbf{FX}\\\mathbf{X}\end{pmatrix}\boldsymbol{\beta}, \begin{pmatrix}\mathbf{F}\boldsymbol{\Sigma}\mathbf{F}' & \mathbf{F}\boldsymbol{\Sigma}_{12}\\\boldsymbol{\Sigma}_{21}\mathbf{F}' & \boldsymbol{\Sigma}_{22}\end{pmatrix}\right\}\,. \tag{5.5}$$

Remark 5.1. It is worth noting that $\mathcal{L}$ and $\mathcal{L}_*$ refer to the same mixed model. The difference is that when using $\mathcal{L}_*$ we wish to emphasize that the "new observation" (corresponding to $\mathbf{y}_*$ in $\mathcal{M}_*$) is $\mathbf{g}$. It is clear that

BLUP($\mathbf{g} \mid \mathcal{L}_*$) means precisely the same as BLUP($\mathbf{g} \mid \mathcal{L}$) and thus we can drop the subscript $*$ from $\mathcal{L}_*$ and $\mathcal{L}_{t*}$. □

Corresponding to $\mathcal{M}_*$, we can express the BLUP for $\mathbf{g} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$ as follows:

$$
\begin{aligned}
\text{BLUP}(\mathbf{g} \mid \mathcal{L}) = \tilde{\mathbf{g}} &= \text{BLUE}(\boldsymbol{\mu} \mid \mathcal{L}) + \mathbf{Z}\,\text{BLUP}(\mathbf{u} \mid \mathcal{L}) \\
&= \mathbf{Ty} + \mathbf{Z}(\mathbf{Z}\boldsymbol{\Delta} + \boldsymbol{\Psi})'\boldsymbol{\Sigma}^-(\mathbf{I}_n - \mathbf{T})\mathbf{y} \\
&= \mathbf{Ty} + \boldsymbol{\Sigma}_{21}\mathbf{M}(\mathbf{M}\boldsymbol{\Sigma}\mathbf{M})^-\mathbf{My} \\
&= \tilde{\boldsymbol{\mu}} + \mathbf{Z}\tilde{\mathbf{u}} \, ,
\end{aligned}
\tag{5.6}
$$

where $\mathbf{T} = \mathbf{X}(\mathbf{X}'\mathbf{W}_{\boldsymbol{\Sigma}}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}_{\boldsymbol{\Sigma}}^-$ and $\mathbf{W}_{\boldsymbol{\Sigma}} \in \mathcal{W}_{\boldsymbol{\Sigma}}$,

$$
\mathcal{W}_{\boldsymbol{\Sigma}} = \left\{ \mathbf{W}_{\boldsymbol{\Sigma}} \in \text{NND}_n : \mathbf{W}_{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \mathbf{X}\mathbf{U}\mathbf{U}'\mathbf{X}', \ \mathcal{C}(\mathbf{W}_{\boldsymbol{\Sigma}}) = \mathcal{C}(\mathbf{X} : \boldsymbol{\Sigma}) \right\} .
\tag{5.7}
$$

In (5.7) the matrix $\mathbf{U}$ is free to vary subject to condition $\mathcal{C}(\mathbf{W}_{\boldsymbol{\Sigma}}) = \mathcal{C}(\mathbf{X} : \boldsymbol{\Sigma})$. The BLUP of $\mathbf{g} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$ under the transformed model $\mathcal{L}_t$ can be expressed as

$$
\begin{aligned}
\text{BLUP}(\mathbf{g} \mid \mathcal{L}_t) = \tilde{\mathbf{g}}_t &= \text{BLUE}(\boldsymbol{\mu} \mid \mathcal{L}_t) + \mathbf{Z}\,\text{BLUP}(\mathbf{u} \mid \mathcal{L}_t) \\
&= \mathbf{T}_t\mathbf{y} + \mathbf{Z}(\mathbf{Z}\boldsymbol{\Delta} + \boldsymbol{\Psi})'\mathbf{F}'\boldsymbol{\Sigma}^-(\mathbf{I}_n - \mathbf{T}_t)\mathbf{y} \\
&= \mathbf{T}_t\mathbf{y} + \boldsymbol{\Sigma}_{21}\mathbf{N}(\mathbf{N}\boldsymbol{\Sigma}\mathbf{N})^-\mathbf{Ny} \\
&= \tilde{\boldsymbol{\mu}}_t + \mathbf{Z}\tilde{\mathbf{u}}_t \, ,
\end{aligned}
\tag{5.8}
$$

where $\mathbf{N} = \mathbf{P}_{\mathbf{F}'\mathbf{Q}_{\mathbf{FX}}}$ and

$$
\mathbf{T}_t = \mathbf{X}[\mathbf{X}'\mathbf{F}'(\mathbf{F}\mathbf{W}_{\boldsymbol{\Sigma}}\mathbf{F}')^-\mathbf{F}\mathbf{X}]^-\mathbf{X}'\mathbf{F}'(\mathbf{F}\mathbf{W}_{\boldsymbol{\Sigma}}\mathbf{F}')^-\mathbf{F}.
\tag{5.9}
$$

Theorem 4.2 gives immediately Corollary 5.1.

Corollary 5.1. *Consider the mixed linear model $\mathcal{L}$ where $\mathcal{C}(\mathbf{X}') = \mathcal{C}(\mathbf{F}'\mathbf{X}')$. Then for all $\mathbf{y} = \mathbf{Xa} + \boldsymbol{\Sigma}\mathbf{Mb}$,*

$$
\begin{aligned}
\|\tilde{\mathbf{g}} - \tilde{\mathbf{g}}_t\|_2 &= \|\text{BLUP}(\mathbf{g} \mid \mathcal{L}) - \text{BLUP}(\mathbf{g} \mid \mathcal{L}_t)\|_2 \\
&= \|(\tilde{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}_t) + \mathbf{Z}(\tilde{\mathbf{u}} - \tilde{\mathbf{u}}_t)\|_2 \\
&= \|\mathbf{R}_m\mathbf{Mb}\|_2 + \|\mathbf{S}_m\mathbf{Mb}\|_2 \\
&\leq \sqrt{\text{ch}_1(\mathbf{R}_m\mathbf{R}_m')\,\mathbf{b}'\mathbf{Mb}} + \sqrt{\text{ch}_1(\mathbf{S}_m\mathbf{S}_m')\,\mathbf{b}'\mathbf{Mb}} \, ,
\end{aligned}
\tag{5.10}
$$

where $\mathbf{R}_m = \mathbf{T}_t\boldsymbol{\Sigma}\mathbf{M} \in \mathbb{R}^{n \times n}$ and

$$
\mathbf{S}_m = \boldsymbol{\Sigma}_{21}\mathbf{M}(\mathbf{I}_n - \mathbf{E}_m) \in \mathbb{R}^{n \times n}, \quad \mathbf{E}_m = \mathbf{N}(\mathbf{N}\boldsymbol{\Sigma}\mathbf{N})^-\mathbf{N}\boldsymbol{\Sigma}\mathbf{M} \in \mathbb{R}^{n \times n}.
\tag{5.11}
$$

If $\mathbf{b} \notin \mathcal{C}(\mathbf{X})$, then the upper bound in (5.10) is equal to zero if and only if $\mathbf{R}_m = \mathbf{S}_m = \mathbf{0}$, i.e., $\mathbf{Fy}$ is linearly sufficient for $\boldsymbol{\mu}$ and for $\mathbf{Zu}$.

Remark 5.2. One of the referees suggested that in addition to predicting $\mathbf{y}_* = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$, one could also consider predicting the vector of new observations of the type

$$
\boldsymbol{\eta} = \mathbf{X}_*\boldsymbol{\beta} + \mathbf{Z}_*\mathbf{u} + \mathbf{e}_* \, ,
\tag{5.12}
$$

where $\mathbf{X}_*$ and $\mathbf{Z}_*$ are given matrices, $\mathbf{e}_*$ is a random error vector with $\text{E}(\mathbf{e}_*) = \mathbf{0}$, $\text{cov}(\mathbf{e}_*) = \boldsymbol{\Phi}_*$, $\text{cov}(\mathbf{e}, \mathbf{e}_*) = \mathbf{0}$ and $\text{cov}(\mathbf{u}, \mathbf{e}_*) = \mathbf{0}$. Now we have

$$
\text{cov}\begin{pmatrix} \mathbf{y} \\ \boldsymbol{\eta} \end{pmatrix} = \text{cov}\begin{pmatrix} \mathbf{y} \\ \mathbf{Z}_*\mathbf{u} + \mathbf{e}_* \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma} & (\mathbf{Z}\boldsymbol{\Delta} + \boldsymbol{\Psi})\mathbf{Z}_*' \\ \mathbf{Z}_*(\mathbf{Z}\boldsymbol{\Delta} + \boldsymbol{\Psi})' & \mathbf{Z}_*\boldsymbol{\Delta}\mathbf{Z}_*' + \boldsymbol{\Phi}_* \end{pmatrix} ,
\tag{5.13}
$$

and thus the prediction of $\boldsymbol{\eta}$ can be based on the model

$$
\mathcal{K} = \left\{ \begin{pmatrix} \mathbf{y} \\ \boldsymbol{\eta} \end{pmatrix}, \begin{pmatrix} \mathbf{X} \\ \mathbf{X}_* \end{pmatrix}\boldsymbol{\beta}, \begin{pmatrix} \boldsymbol{\Sigma} & (\mathbf{Z}\boldsymbol{\Delta} + \boldsymbol{\Psi})\mathbf{Z}_*' \\ \mathbf{Z}_*(\mathbf{Z}\boldsymbol{\Delta} + \boldsymbol{\Psi})' & \mathbf{Z}_*\boldsymbol{\Delta}\mathbf{Z}_*' + \boldsymbol{\Phi}_* \end{pmatrix} \right\} .
\tag{5.14}
$$

The setup (5.14) offers interesting further problems but hey go beyond our main focus and are thus left for further research. □

6 Concluding remarks

In this article we have introduced upper bounds for the Euclidean distances between the best linear unbiased predictors (BLUPs) of $\mathbf{y}_*$ when the prediction is based on the original model $\mathcal{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ and when it is based on the transformed model $\mathcal{M}_t = \{\mathbf{F}\mathbf{y}, \mathbf{F}\mathbf{X}\boldsymbol{\beta}, \mathbf{F}\mathbf{V}\mathbf{F}'\}$. The unobservable "new future" random vector $\mathbf{y}_*$ is generated from $\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*$. Corresponding considerations are made for the BLUPs of $\boldsymbol{\varepsilon}_*$. The original setup and the transformed setup for a linear model with new observations can be described, respectively, as

$$\mathcal{M}_* = \left\{ \begin{pmatrix} \mathbf{y} \\ \mathbf{y}_* \end{pmatrix}, \begin{pmatrix} \mathbf{X} \\ \mathbf{X}_* \end{pmatrix} \boldsymbol{\beta}, \begin{pmatrix} \mathbf{V} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} \right\}, \quad \mathcal{M}_{t*} = \left\{ \begin{pmatrix} \mathbf{F}\mathbf{y} \\ \mathbf{y}_* \end{pmatrix}, \begin{pmatrix} \mathbf{F}\mathbf{X} \\ \mathbf{X}_* \end{pmatrix} \boldsymbol{\beta}, \begin{pmatrix} \mathbf{F}\mathbf{V}\mathbf{F}' & \mathbf{F}\mathbf{V}_{12} \\ \mathbf{V}_{21}\mathbf{F}' & \mathbf{V}_{22} \end{pmatrix} \right\}. \tag{6.1}$$

We also consider the mixed linear model which is a special case of $\mathcal{M}_*$. We show how the upper bounds are related to the concept of linear sufficiency. The concept of linear sufficiency is strongly connected to the transformed model $\mathcal{M}_t$, because for example, if $\mathbf{F}\mathbf{y}$ is linearly sufficient for $\mathbf{y}_*$, then every representation of the BLUP for $\mathbf{y}_*$ based on the transformed model is BLUP also under the original model.

Our attempt has been to make the paper self-readable so that the necessary background tools have been presented to a reasonable amount. Considerations are pretty mathematical and for this paper, we have no practical statistical applications in mind.

The Euclidean distances between estimator/predictors in linear models have not been studied very much in literature. However, we wish to mention some related articles. Baksalary & Kala [2, p. 680] provided an upper bound for the Euclidean distance of the difference of the ordinary least squares estimator (OLSE) and the BLUE of of $\mathbf{X}\boldsymbol{\beta}$, i.e., $\|\mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}) - \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})\|_2$ under $\mathcal{M}$; see also Haberman [12], Baksalary & Kala [1], Mäkinen [21, 22], and Trenkler [33, p. 261]. The corresponding considerations for the BLUEs under two linear models $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}_1\}$ and $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}_2\}$, have been made by Hauke et al. [15], and for

$$\|\mathrm{BLUP}(\mathbf{y}_*) - \mathrm{BLUE}(\mathbf{X}_*\boldsymbol{\beta})\|_2 \quad \text{and} \quad \|\mathrm{BLUP}(\mathbf{y}_*) - \mathrm{OLSE}(\mathbf{X}_*\boldsymbol{\beta})\|_2 \tag{6.2}$$

under $\mathcal{M}_*$ by Haslett et al. [13, Sec. 3]. See also Kala et al. [19, Sec. 6], and Pordzik [26].

References

[1] Baksalary, J.K. & Kala, R. (1978). A bound for the Euclidean norm of the difference between the least squares and the best linear unbiased estimators. *Ann. Statist.*, 6, 1390–1393. DOI

[2] Baksalary, J.K. & Kala, R. (1980). A new bound for the Euclidean norm of the difference between the least squares and the best linear unbiased estimators. *Ann. Statist.*, 8, 679–681. DOI

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

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

[5] Ben-Israel, A. & Greville, T.N.E. (2003). *Generalized Inverses: Theory and Applications*, 2nd Edition. Springer, New York. DOI

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

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

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

[9] Goldberger, A.S. (1962). Best linear unbiased prediction in the generalized linear regression model. *J. Amer. Statist. Assoc.*, 58, 369–375. DOI

[10] Groß, J. (2004). The general Gauss–Markov model with possibly singular dispersion matrix. *Stat. Pap.*, 45, 311–336. DOI

[11] 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. [Proceedings of the Second International Triennial Calcutta Symposium on Probability and Statistics, Calcutta, India, 30 December 1994 – 2 January 1995.]

[12] Haberman, S.J. (1975). How much do Gauss–Markov and least square estimates differ? A coordinate-free approach. *Ann. Statist.*, 3, 982–990. DOI

[13] 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. DOI

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

[15] Hauke, J., Markiewicz, A. & Puntanen, S. (2012). Comparing the BLUEs under two linear models. *Comm. Statist. Theory Methods*, 41, 2405–2418. DOI

[16] Isotalo, J., Markiewicz, A. & Puntanen, S. (2018). Some properties of linear prediction sufficiency in the linear model. *Trends and Perspectives in Linear Statistical Inference: LinStat, Istanbul, August 2016.* (International Conference on Trends and Perspectives in Linear Statistical Inference, Istanbul, Turkey, 22–25 August 2016). (M. Tez & D. von Rosen, eds.) Springer, pages 111–129. DOI.

[17] 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. DOI

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

[19] 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, editor), Springer Proceedings in Mathematics & Statistics, vol. 192, 275–294. DOI

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

[21] Mäkinen, J. (2000). Bounds for the difference between a linear unbiased estimate and the best linear unbiased estimate. *Physics and Chemistry of the Earth - Part A: Solid Earth and Geodesy*, 25, 693–698. DOI

[22] Mäkinen, J. (2002). A bound for the Euclidean norm of the difference between the best linear unbiased estimator and a linear unbiased estimator. *Journal of Geodesy*, 76, 317–322. DOI

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

[24] Markiewicz, A. & Puntanen, S. (2017). Further properties of the linear sufficiency in the partitioned linear model. *Matrices, Statistics and Big Data* (Proceedings of the 25th International Workshop on Matrices and Statistics, IWMS-2016, held in Funchal, Madeira. Portugal, 6–9 June 2016). (S.E. Ahmed, F. Carvalho & S. Puntanen, eds.) Springer, in press.

[25] Markiewicz, A. & Puntanen, S. (2018). Further properties of linear prediction sufficiency and the BLUPs in the linear model with new observations. *Afrika Statistika*, in press.

[26] Pordzik, P.R. (2012). A bound for the Euclidean distance between restricted and unrestricted estimators of parametric functions in the general linear model. *Statist. Papers*, 53, 299–304. DOI

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

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

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

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

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

[32] 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. DOI

[33] Trenkler, G. (1994). Characterizations of oblique and orthogonal projectors. *Proceedings of the International Conference on Linear Statistical Inference LINSTAT '93 (Poznań, 1993)*. (T. Caliński and R. Kala, eds.) Kluwer, Dordrecht, 255–270.