Applied and Computational Matrix Analysis : MAT-TRIAD, Coimbra, Portugal, September 2015 Selected, Revised Contributions. Cham, Switzerland: Springer. Springer proceedings in mathematics & statistics vol. 192. ISBN: 978-3-319-49982-6. ISSN: 2194-1009.
http://dx.doi.org/10.1007/978-3-319-49984-0_19

Chapter 1
Some further remarks on the linear sufficiency in the linear model

Radosław Kala, Augustyn Markiewicz and Simo Puntanen

Prepared for *Applied and Computational Matrix Analysis: MatTriad, Coimbra, Portugal, September 2015, Selected, Revised Contributions* (Natália Bebiano, editor), Springer Proceedings in Mathematics & Statistics, vol 192, 275–294.

Abstract In this article we consider the linear sufficiency of statistic $\mathbf{Fy}$ when estimating the estimable parametric function of β under the linear model $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}$. We review some properties that have not been received much attention in the literature and provide some new results and insight into the meaning of the linear sufficiency. In particular, we consider the best linear unbiased estimation (BLUE) under the transformed model $\mathscr{A}_t = \{\mathbf{Fy}, \mathbf{FX}\beta, \mathbf{FVF}'\}$ and study the possibilities to measure the relative linear sufficiency of $\mathbf{Fy}$ by comparing the BLUEs under $\mathscr{A}$ and $\mathscr{A}_t$. We also consider some new properties of the Euclidean norm of the distance of the BLUEs under $\mathscr{A}$ and $\mathscr{A}_t$. The concept of linear sufficiency was essentially introduced in early 1980s by Baksalary, Kala and Drygas, but to our knowledge the concept of relative linear sufficiency nor the Euclidean norm of the difference between the BLUEs under $\mathscr{A}$ and $\mathscr{A}_t$ have not appeared in the literature. To make the article more self-readable we go through some basic concepts related to linear sufficiency. We also provide a rather extensive list of relevant references.

1.1 Introduction

In this paper we consider the linear model defined by

Radosław Kala
Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Wojska Polskiego 28, PL-60637 Poznań, Poland, e-mail: kalar@up.poznan.pl

Augustyn Markiewicz
Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Wojska Polskiego 28, PL-60637 Poznań, Poland, e-mail: amark@up.poznan.pl

Simo Puntanen
School of Information Sciences, FI-33014 University of Tampere, Finland, e-mail: simo.puntanen@uta.fi

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon\,, \quad \text{or shortly notated} \quad \mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}\,, \tag{1.1}$$

where $\mathbf{y}$ is an n-dimensional observable response variable, $\mathbf{X}$ is a known $n \times p$ matrix, i.e., $\mathbf{X} \in \mathbb{R}^{n\times p}$, $\beta \in \mathbb{R}^p$ is a vector of fixed (but unknown) parameters, and ε is an unobservable random error with a known covariance matrix $\operatorname{cov}(\varepsilon) = \mathbf{V} = \operatorname{cov}(\mathbf{y})$ and expectation $\mathrm{E}(\varepsilon) = \mathbf{0}$.

Under the model $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}$, the statistic $\mathbf{Gy}$, where $\mathbf{G} \in \mathbb{R}^{n\times n}$, is the best linear unbiased estimator, BLUE, of $\mathbf{X}\beta$ whenever $\mathbf{Gy}$ is unbiased, i.e., $\mathbf{GX} = \mathbf{X}$, and it has the minimal covariance matrix in the Löwner sense among all unbiased linear estimators of $\mathbf{X}\beta$. The BLUE of an estimable parametric function $\mathbf{K}\beta$, where $\mathbf{K} \in \mathbb{R}^{k\times p}$, is defined in the corresponding way. Recall that $\mathbf{K}\beta$ is said to be estimable under $\mathscr{A}$ if it has a linear unbiased estimator $\mathbf{Ly}$, say, so that $\mathrm{E}(\mathbf{Ly}) = \mathbf{LX}\beta = \mathbf{K}\beta$ for all $\beta \in \mathbb{R}^p$, which happens if and only if

$$\mathscr{C}(\mathbf{K}') \subset \mathscr{C}(\mathbf{X}')\,, \tag{1.2}$$

where $\mathscr{C}(\cdot)$ stands for the column space (range) of the matrix argument.

In what follows, we frequently refer to the following lemma; see, e.g., Drygas (1970, p. 55), Rao (1973, p. 282), and Baksalary (2004).

Lemma 1. *Consider the general linear model $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}$. Then the statistic* $\mathbf{Gy}$ *is the* BLUE *for* $\mathbf{X}\beta$ *if and only if* $\mathbf{G}$ *satisfies the equation*

$$\mathbf{G}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{X} : \mathbf{0})\,. \tag{1.3}$$

The corresponding condition for $\mathbf{By}$ *to be the* BLUE *of an estimable parametric function* $\mathbf{K}\beta$ *is*

$$\mathbf{B}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{K} : \mathbf{0})\,. \tag{1.4}$$

The notation $(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp})$ refers to a columnwise partitioned matrix by juxtaposing matrices $\mathbf{X}$ and $\mathbf{V}\mathbf{X}^{\perp}$. The matrix $\mathbf{X}^{\perp}$ refers to a matrix spanning the orthocomplement of the column space $\mathscr{C}(\mathbf{X})$. One convenient choice for $\mathbf{X}^{\perp}$ is $\mathbf{M} := \mathbf{I}_n - \mathbf{P}_{\mathbf{X}} = \mathbf{I}_n - \mathbf{H}$, with $\mathbf{P}_{\mathbf{X}} = \mathbf{X}\mathbf{X}^{+} =: \mathbf{H}$ denoting the orthogonal projector onto $\mathscr{C}(\mathbf{X})$ and $\mathbf{X}^{+}$ referring to the Moore–Penrose inverse of $\mathbf{X}$. Of course, $\mathscr{C}(\mathbf{X}^{\perp}) = \mathscr{C}(\mathbf{M}) = \mathscr{N}(\mathbf{X}')$, where $\mathscr{N}(\cdot)$ stands for the nullspace.

The solution $\mathbf{G}$ for (1.3) always exists but is unique if and only if $\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathbb{R}^n$. However, the numerical observed value of $\mathbf{Gy}$ is unique (with probability 1) once the random vector $\mathbf{y}$ has realized its value in the space

$$\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{X}) \oplus \mathscr{C}(\mathbf{VM})\,. \tag{1.5}$$

In (1.5) the symbol $\oplus$ stands for the direct sum. Two estimators $\mathbf{G}_1\mathbf{y}$ and $\mathbf{G}_2\mathbf{y}$ are said to be equal (with probability 1) whenever $\mathbf{G}_1\mathbf{y} = \mathbf{G}_2\mathbf{y}$ for all $\mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V})$. When talking about the equality of estimators we sometimes may drop the phrase "with probability 1". The consistency of the model $\mathscr{A}$ means that the observed $\mathbf{y}$ lies in $\mathscr{C}(\mathbf{X} : \mathbf{V})$ which is assumed to hold whatever model we have. For the consistency concept, see, e.g., Baksalary et al. (1992).

In this paper we use the notation $\mathscr{W}$ for the set of nonnegative definite matrices defined as

$$\mathscr{W} = \left\{\mathbf{W} \in \mathbb{R}^{n\times n} : \mathbf{W} = \mathbf{V} + \mathbf{X}\mathbf{U}\mathbf{U}'\mathbf{X}',\ \mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})\right\}. \tag{1.6}$$

In (1.6) $\mathbf{U}$ can be any $p \times m$ matrix as long as $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$ is satisfied. One obvious choice is of course $\mathbf{U} = \mathbf{I}_p$. In particular, if $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{V})$, we can choose $\mathbf{U} = \mathbf{0}$. The set $\mathscr{W}$ appears to be a very useful class of matrices and it has numerous applications related to linear models. For example, it is easy to confirm the following lemma.

Lemma 2. *Let* $\mathbf{W} \in \mathscr{W}$. *Then* $\mathbf{Gy}$ *is the* BLUE *for* $\mathbf{X}\beta$ *under* $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}$ *if and only if* $\mathbf{Gy}$ *is the* BLUE *for* $\mathbf{X}\beta$ *under* $\mathscr{A}_{\mathbf{W}} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{W}\}$.

We will later consider some interesting properties of $\mathscr{W}$ and the corresponding extended set

$$\mathscr{W}_* = \left\{\mathbf{W} \in \mathbb{R}^{n\times n} : \mathbf{W} = \mathbf{V} + \mathbf{X}\mathbf{T}\mathbf{X}',\ \mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})\right\}. \tag{1.7}$$

Notice that $\mathbf{W}$ that belongs to $\mathscr{W}_*$ is not necessarily nonnegative definite and it can be nonsymmetric. For example, the following statements concerning $\mathbf{W}$ belonging to $\mathscr{W}_*$ are equivalent:

$$\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{W}), \tag{1.8a}$$

$$\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{W}), \tag{1.8b}$$

$$\mathbf{X}'\mathbf{W}^-\mathbf{X} \text{ is invariant for any choice of } \mathbf{W}^-, \tag{1.8c}$$

$$\mathscr{C}(\mathbf{X}'\mathbf{W}^-\mathbf{X}) = \mathscr{C}(\mathbf{X}') \text{ for any choice of } \mathbf{W}^-, \tag{1.8d}$$

$$\mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^-\mathbf{X} = \mathbf{X} \text{ for any choices of } \mathbf{W}^- \text{ and } (\mathbf{X}'\mathbf{W}^-\mathbf{X})^-. \tag{1.8e}$$

Moreover, each of these statements is equivalent also to $\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{W}')$, and hence to the statements (1.8b)–(1.8e) by replacing $\mathbf{W}$ with $\mathbf{W}'$. Notice that obviously $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{W}')$ and that the invariance properties in (1.8d) and (1.8e) concern also the choice of $\mathbf{W} \in \mathscr{W}_*$. For further properties of $\mathscr{W}_*$, see, e.g., Baksalary & Puntanen (1989, Th. 1), Baksalary et al. (1990, Th. 2), Baksalary & Mathew (1990, Th. 2), and Puntanen et al. (2011, §12.3).

The usefulness of $\mathscr{W}_*$ appears, e.g., from the following well-known representation of the BLUE of $\mathbf{X}\beta$:

$$\mathrm{BLUE}(\mathbf{X}\beta \mid \mathscr{A}) = \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^-\mathbf{y} =: \mathbf{Cy}, \tag{1.9}$$

where $\mathbf{W} \in \mathscr{W}_*$. The *general* representation for the BLUE can be written as $\mathbf{Ay}$, where

$$\mathbf{A} = \mathbf{C} + \mathbf{N}(\mathbf{I}_n - \mathbf{P}_{\mathbf{W}}), \tag{1.10}$$

with $\mathbf{N} \in \mathbb{R}^{n\times n}$ being free to vary. In this context we might mention also the following expression:

$$\mathrm{BLUE}(\mathbf{X}\beta \mid \mathscr{A}) = [\mathbf{I}_n - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{M}]\mathbf{y}\,. \tag{1.11}$$

For further expressions, see, e.g., Puntanen et al. (2011, §10.4).

Recall that the multipliers of the random vector $\mathbf{y}$ in (1.9) and (1.11) are not necessarily the same but the following holds:

$$\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-}\mathbf{y} = [\mathbf{I}_n - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{M}]\mathbf{y} \quad \text{for all } \mathbf{y} \in \mathscr{C}(\mathbf{W})\,. \tag{1.12}$$

One more property requiring attention before proceeding into the concept of linear sufficiency is the invariance of the matrix product $\mathbf{AB}^{-}\mathbf{C}$. According to Rao & Mitra (1971, Lemma 2.2.4), for any nonnull $\mathbf{A}$ and $\mathbf{C}$ the following holds:

$$\mathbf{AB}^{-}\mathbf{C} = \mathbf{AB}^{+}\mathbf{C} \text{ for all } \mathbf{B}^{-} \iff \mathscr{C}(\mathbf{C}) \subset \mathscr{C}(\mathbf{B}) \text{ and } \mathscr{C}(\mathbf{A}') \subset \mathscr{C}(\mathbf{B}')\,. \tag{1.13}$$

We shall frequently need the invariance property (1.13). For example, we immediately see that for $\mathbf{W} \in \mathscr{W}_*$, the matrices $\mathbf{X}'\mathbf{W}^{-}\mathbf{X}$ and $\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'$ are invariant for any choice of $\mathbf{W}^{-}$. Similarly in (1.12) we can use any generalized inverses involved.

1.2 Definition of the linear sufficiency

Now we can formally define the concept of linear sufficiency as done by Baksalary & Kala (1981). Actually they talked about "linear transformations preserving best linear unbiased estimators" and it was Drygas (1983) who adopted the term "linear sufficiency".

Definition 1. A linear statistic $\mathbf{Fy}$, where $\mathbf{F} \in \mathbb{R}^{f\times n}$, is called linearly sufficient for $\mathbf{X}\beta$ under the model $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}$, if there exists a matrix $\mathbf{A} \in \mathbb{R}^{n\times f}$ such that $\mathbf{AFy}$ is the BLUE for $\mathbf{X}\beta$. Correspondingly, $\mathbf{Fy}$ is linearly sufficient for estimable $\mathbf{K}\beta$, where $\mathbf{K} \in \mathbb{R}^{k\times p}$, if there exists a matrix $\mathbf{A} \in \mathbb{R}^{k\times f}$ such that $\mathbf{AFy}$ is the BLUE for $\mathbf{K}\beta$.

Sometimes we may use the short notations

$$\mathbf{Fy} \in \mathrm{S}(\mathbf{X}\beta), \quad \mathbf{Fy} \in \mathrm{S}(\mathbf{K}\beta) \tag{1.14}$$

to indicate that $\mathbf{Fy}$ is linearly sufficient for $\mathbf{X}\beta$ or for $\mathbf{K}\beta$, respectively.

By definition, $\mathbf{Fy}$ is linearly sufficient for $\mathbf{X}\beta$ if and only if the equation

$$\mathbf{AF}(\mathbf{X} : \mathbf{VM}) = (\mathbf{X} : \mathbf{0}) \tag{1.15}$$

has a solution for $\mathbf{A}$, which happens if and only if

$$\mathscr{C}\begin{pmatrix}\mathbf{X}'\\ \mathbf{0}\end{pmatrix} \subset \mathscr{C}\begin{pmatrix}\mathbf{X}'\mathbf{F}'\\ \mathbf{MVF}'\end{pmatrix}. \tag{1.16}$$

The concept of linear minimal sufficiency, introduced by Drygas (1983), is defined as follows.

Definition 2. A linear statistic $\mathbf{Fy}$ is called linearly minimal sufficient if for any other linearly sufficient statistics $\mathbf{Sy}$, there exists a matrix $\mathbf{A}$ such that $\mathbf{Fy} = \mathbf{ASy}$ almost surely.

In Lemma 3 we collect some well-known equivalent conditions for $\mathbf{Fy}$ being linearly sufficient for $\mathbf{X}\beta$. For the proofs of parts (c) and (d), see Baksalary & Kala (1981); part (e), see Baksalary & Kala (1986, Cor. 2); and part (f), Müller (1987, Prop. 3.1a). For further related references, see Drygas (1983), Baksalary & Mathew (1986), Baksalary & Drygas (1992), Groß (1998), Isotalo & Puntanen (2006a,b), Kornacki (2007), and Kala & Pordzik (2009).

Lemma 3. *The statistic* $\mathbf{Fy}$ *is linearly sufficient for* $\mathbf{X}\beta$ *under the linear model* $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}$ *if and only if any of the following equivalent statements holds:*

(a) $\mathscr{C}\begin{pmatrix}\mathbf{X}'\\ \mathbf{0}\end{pmatrix} \subset \mathscr{C}\begin{pmatrix}\mathbf{X}'\mathbf{F}'\\ \mathbf{MVF}'\end{pmatrix}$,
(b) $\mathscr{N}(\mathbf{FX} : \mathbf{FVX}^{\perp}) \subset \mathscr{N}(\mathbf{X} : \mathbf{0})$,
(c) $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WF}')$, *where* $\mathbf{W} \in \mathscr{W}$,
(d) $\operatorname{rank}(\mathbf{X} : \mathbf{VF}') = \operatorname{rank}(\mathbf{WF}')$, *where* $\mathbf{W} \in \mathscr{W}$,
(e) $\mathscr{C}(\mathbf{X}'\mathbf{F}') = \mathscr{C}(\mathbf{X}')$ *and* $\mathscr{C}(\mathbf{FX}) \cap \mathscr{C}(\mathbf{FVX}^{\perp}) = \{\mathbf{0}\}$,
(f) $\mathscr{N}(\mathbf{F}) \cap \mathscr{C}(\mathbf{X} : \mathbf{V}) \subset \mathscr{C}(\mathbf{VX}^{\perp})$,
(g) *there exists a matrix* $\mathbf{A}$ *such that* $\mathbf{AF}(\mathbf{X} : \mathbf{VX}^{\perp}) = (\mathbf{X} : \mathbf{0})$.

Moreover, $\mathbf{Fy}$ *is linearly minimal sufficient for* $\mathbf{X}\beta$ *if and only if* $\mathscr{C}(\mathbf{X}) = \mathscr{C}(\mathbf{WF}')$, *or equivalently, the equality holds in (a), (b) or (f).*

Baksalary & Kala (1986) proved the following:

Lemma 4. *Let* $\mathbf{K}\beta$ *be an estimable parametric function under* $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}$, *i.e.,* $\mathscr{C}(\mathbf{K}') \subset \mathscr{C}(\mathbf{X}')$. *Then* $\mathbf{Fy}$ *is linearly sufficient for* $\mathbf{K}\beta$ *under* $\mathscr{A}$ *if and only if any of the following equivalent statements holds:*

(a) $\mathscr{C}\begin{pmatrix}\mathbf{K}'\\ \mathbf{0}\end{pmatrix} \subset \mathscr{C}\begin{pmatrix}\mathbf{X}'\mathbf{F}'\\ \mathbf{MVF}'\end{pmatrix}$,
(b) $\mathscr{N}(\mathbf{FX} : \mathbf{FVX}^{\perp}) \subset \mathscr{N}(\mathbf{K} : \mathbf{0})$,
(c) $\mathscr{C}[\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{K}'] \subset \mathscr{C}(\mathbf{WF}')$, *where* $\mathbf{W} \in \mathscr{W}$,
(d) *there exists a matrix* $\mathbf{A}$ *such that* $\mathbf{AF}(\mathbf{X} : \mathbf{VX}^{\perp}) = (\mathbf{K} : \mathbf{0})$.

Moreover, $\mathbf{Fy}$ *is linearly minimal sufficient for* $\mathbf{K}\beta$ *if and only if equality (instead of subspace inclusion) holds in (a), (b) or equivalently (c).*

Suppose that $\mathbf{Fy}$ is linearly sufficient for $\mathbf{X}\beta$ under the model $\mathscr{A}$, and $\mathbf{F}_1$ is some arbitrary matrix with n columns. Then it is interesting to observe that the extended statistic

$$\mathbf{F}_0\mathbf{y} := \begin{pmatrix}\mathbf{F}\\ \mathbf{F}_1\end{pmatrix}\mathbf{y} \tag{1.17}$$

is also linearly sufficient for $\mathbf{X}\beta$. This is so because

$$\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{W}\mathbf{F}') \subset \mathscr{C}[\mathbf{W}(\mathbf{F}' : \mathbf{F}_1')] = \mathscr{C}(\mathbf{W}\mathbf{F}_0') \,. \tag{1.18}$$

Similarly

$$\mathbf{F}\mathbf{y} \in \mathrm{S}(\mathbf{X}\beta) \implies \mathbf{F}_* \mathbf{y} \in \mathrm{S}(\mathbf{X}\beta)\,, \quad \text{if } \mathscr{C}(\mathbf{F}') = \mathscr{C}(\mathbf{F}_*') \,. \tag{1.19}$$

Thus if $\operatorname{rank}(\mathbf{F}) = r$ we can replace $\mathbf{F} \in \mathbb{R}^{f\times n}$ with $\mathbf{F}_* \in \mathbb{R}^{r\times n}$, where $r \le f$, i.e., the columns of $\mathbf{F}_*'$ provide a spanning basis for $\mathscr{C}(\mathbf{F}')$.

Notice also that the linear sufficiency condition $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{W}\mathbf{F}')$ implies that we necessarily must have

$$\operatorname{rank}(\mathbf{X}_{n\times p}) \le p \le \operatorname{rank}(\mathbf{F}_{f\times n}) \le f \,. \tag{1.20}$$

In passing we note that $\mathbf{X}'\mathbf{W}^-\mathbf{y}$ is linearly minimal sufficient for $\mathbf{X}\beta$ under the model $\mathscr{A}$; this follows from $\mathscr{C}(\mathbf{X}) = \mathscr{C}[\mathbf{W}(\mathbf{W}^-)'\mathbf{X}]$.

1.3 The transformed model $\mathscr{A}_t$

Consider the model $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}$ and let $\mathbf{F} \in \mathbb{R}^{f\times n}$ be such a matrix that $\mathbf{F}\mathbf{y}$ is linearly sufficient for $\mathbf{X}\beta$. Then the transformation $\mathbf{F}$ applied to $\mathbf{y}$ induces the transformed model

$$\mathscr{A}_t = \{\mathbf{F}\mathbf{y}, \mathbf{F}\mathbf{X}\beta, \mathbf{F}\mathbf{V}\mathbf{F}'\} \,. \tag{1.21}$$

Now, as the statistic $\mathbf{F}\mathbf{y}$ is linearly sufficient for $\mathbf{X}\beta$, it sounds intuitively believable that both models provide the same starting point for obtaining the BLUE of $\mathbf{X}\beta$. Indeed this appears to be true as proved by Baksalary & Kala (1981, 1986). Moreover, Tian & Puntanen (2009, Th. 2.8) and Kala et al. (2015, Th. 2) showed the following:

Lemma 5. *Consider the model $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}$ and its transformed version*

$$\mathscr{A}_t = \{\mathbf{F}\mathbf{y}, \mathbf{F}\mathbf{X}\beta, \mathbf{F}\mathbf{V}\mathbf{F}'\} \,, \tag{1.22}$$

and let $\mathbf{K}\beta$ be estimable under $\mathscr{A}$. Then the following statements are equivalent:

(a) $\mathbf{F}\mathbf{y}$ *is linearly sufficient for* $\mathbf{K}\beta$.
(b) $\mathrm{BLUE}(\mathbf{K}\beta \mid \mathscr{A}) = \mathrm{BLUE}(\mathbf{K}\beta \mid \mathscr{A}_t)$ *with probability* 1.
(c) *There exists at least one representation of* BLUE *of* $\mathbf{K}\beta$ *under* $\mathscr{A}$ *which is the* BLUE *also under the transformed model* $\mathscr{A}_t$.

It is noteworthy that if $\mathbf{F}\mathbf{y}$ is linearly sufficient for $\mathbf{X}\beta$, then, in view of (1.16), we have

$$\mathscr{C}(\mathbf{X}') = \mathscr{C}(\mathbf{X}'\mathbf{F}'), \text{ i.e., } \operatorname{rank}(\mathbf{F}\mathbf{X}) = \operatorname{rank}(\mathbf{X}) \,. \tag{1.23}$$

On the other hand, on account of (1.2), $\mathbf{X}\beta$ is estimable under the transformed model $\mathscr{A}_t = \{\mathbf{F}\mathbf{y}, \mathbf{F}\mathbf{X}\beta, \mathbf{F}\mathbf{V}\mathbf{F}'\}$ if and only if

$$\mathscr{C}(\mathbf{X}') \subset \mathscr{C}(\mathbf{X}'\mathbf{F}'), \tag{1.24}$$

i.e., $\mathscr{C}(\mathbf{X}') = \mathscr{C}(\mathbf{X}'\mathbf{F}')$, which is (1.23). This confirms the following:

$$\mathbf{Fy} \in \mathrm{S}(\mathbf{X}\beta) \implies \mathbf{X}\beta \text{ is estimable under } \mathscr{A}_t\,. \tag{1.25}$$

However, the reverse relation in (1.25) does not hold. In view of part (e) of Lemma 3, we need the following *two* conditions for $\mathbf{Fy} \in \mathrm{S}(\mathbf{X}\beta)$:

$$\mathscr{C}(\mathbf{X}'\mathbf{F}') = \mathscr{C}(\mathbf{X}') \quad \text{and} \quad \mathscr{C}(\mathbf{FX}) \cap \mathscr{C}(\mathbf{FVX}^{\perp}) = \{\mathbf{0}\}, \tag{1.26}$$

which can be expressed equivalently as

$$\mathbf{X}\beta \text{ is estimable under } \mathscr{A}_t \quad \text{and} \quad \mathscr{C}(\mathbf{FX}) \cap \mathscr{C}(\mathbf{FVX}^{\perp}) = \{\mathbf{0}\}. \tag{1.27}$$

Let us consider some special choices of $\mathbf{F}$. For example, if $\mathbf{F}$ has the property $\mathscr{C}(\mathbf{F}') = \mathbb{R}^n$ (implying that the number of the rows in $\mathbf{F} \in \mathbb{R}^{f \times n}$ is at least n), then

$$\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{WF}'), \tag{1.28}$$

and thereby $\mathbf{Fy}$ is linearly sufficient for $\mathbf{X}\beta$. In particular, for a nonsingular $\mathbf{F} \in \mathbb{R}^{n \times n}$, the statistic $\mathbf{Fy}$ is linearly sufficient. For a positive definite $\mathbf{V}$ the linear sufficiency condition becomes simply

$$\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{VF}'). \tag{1.29}$$

Supposing that $\mathbf{V}^{1/2}$ is the positive definite square root of $\mathbf{V}$ we observe that $\mathbf{V}^{-1/2}\mathbf{y}$ is linearly sufficient and thus the BLUE of $\mathbf{X}\beta$ under the transformed model

$$\mathscr{A}_t = \{\mathbf{V}^{-1/2}\mathbf{y},\, \mathbf{V}^{-1/2}\mathbf{X}\beta,\, \mathbf{I}_n\} \tag{1.30}$$

is the same as in the original model $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}$, i.e., the BLUE($\mathbf{X}\beta$) under $\mathscr{A}$ equals the ordinary least squares estimator of $\mathbf{X}\beta$, OLSE($\mathbf{X}\beta$), under $\mathscr{A}_t$:

$$\mathrm{BLUE}(\mathbf{X}\beta \mid \mathscr{A}) = \mathrm{OLSE}(\mathbf{X}\beta \mid \mathscr{A}_t). \tag{1.31}$$

This technique, sometimes referred to as the Aitken-approach, see Aitken (1935), is well known in statistical textbooks. However, usually these textbooks do not mention anything about linear sufficiency feature of this transformation.

Consider then a more general case. By Lemma 2 we know that the BLUEs under $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}$ and $\mathscr{A}_{\mathbf{W}} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{W}\}$ are equal. Suppose that $\mathrm{rank}(\mathbf{W}) = w$ and that $\mathbf{W}$ has the eigenvalue decomposition $\mathbf{W} = \mathbf{Z}\Lambda\mathbf{Z}'$, where the columns of $\mathbf{Z} \in \mathbb{R}^{n \times w}$ are orthonormal eigenvectors of $\mathbf{W}$ with respect to nonzero eigenvalues $\lambda_1 \geq \cdots \geq \lambda_w > 0$ of $\mathbf{W}$, and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_w)$. Choosing

$$\mathbf{F} = \Lambda^{-1/2}\mathbf{Z}' \in \mathbb{R}^{w \times n}, \tag{1.32}$$

we observe that

$$\mathscr{C}(\mathbf{W}\mathbf{F}') = \mathscr{C}(\mathbf{W}\mathbf{Z}\Lambda^{-1/2}) = \mathscr{C}(\mathbf{W}) \tag{1.33}$$

and hence $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{W}\mathbf{F}')$ and thereby $\mathbf{Fy}$ is linearly sufficient in $\mathscr{A}_{\mathbf{W}}$. Thus the BLUE of $\mathbf{X}\beta$ under the original model $\mathscr{A}$ is the same as under $\mathscr{A}_{\mathbf{W}}$ and further the same as under the transformed model

$$\mathscr{A}_t = \{\, \Lambda^{-1/2}\mathbf{Z}'\mathbf{y}, \Lambda^{-1/2}\mathbf{Z}'\mathbf{X}\beta, \mathbf{I}_w \} \,. \tag{1.34}$$

Because $\mathbf{F}'\mathbf{F} = \mathbf{Z}\Lambda^{-1}\mathbf{Z}' = \mathbf{W}^{+}$, we have

$$\begin{aligned} \mathrm{BLUE}(\mathbf{X}\beta \mid \mathscr{A}) &= \mathrm{BLUE}(\mathbf{X}\beta \mid \mathscr{A}_t) = \mathrm{OLSE}(\mathbf{X}\beta \mid \mathscr{A}_t) \\ &= \mathbf{X}(\mathbf{X}'\mathbf{W}^{+}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{+}\mathbf{y} \,, \end{aligned} \tag{1.35}$$

where we actually can use any generalized inverses involved.

We may note that Christensen (2011, p. 239) uses the transformation matrix $\Lambda^{-1/2}\mathbf{Z}'$ when considering the so-called weakly singular linear model, i.e., when $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{V})$, and Hauke et al. (2012, §4) while comparing the BLUEs under two linear models with different covariance matrices.

We complete this section by considering a partitioned linear model

$$\mathscr{A}_{12} = \{\mathbf{y}, \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2, \mathbf{V}\} \,. \tag{1.36}$$

Le us assume that $\mathscr{C}(\mathbf{X}_1) \cap \mathscr{C}(\mathbf{X}_2) = \{\mathbf{0}\}$ implying that $\mathbf{X}_1\beta_1$ is estimable. Premultiplying the model $\mathscr{A}_{12}$ by $\mathbf{M}_2 = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}_2}$ yields the reduced model

$$\mathscr{A}_{12\cdot 2} = \{\mathbf{M}_2\mathbf{y},\ \mathbf{M}_2\mathbf{X}_1\beta_1,\ \mathbf{M}_2\mathbf{V}\mathbf{M}_2\} \,. \tag{1.37}$$

Now the well-known Frisch–Waugh–Lovell theorem, see, e.g., Groß & Puntanen (2000, 2005), and Arendacká & Puntanen (2015, Th. 1), states that the BLUEs of $\mathbf{X}_1\beta_1$ under $\mathscr{A}_{12}$ and $\mathscr{A}_{12\cdot 2}$ coincide. Hence, in view of Lemma 5, the statistic $\mathbf{M}_2\mathbf{y}$ is linearly sufficient for $\mathbf{X}_1\beta_1$. One expression for the BLUE of $\mathbf{X}_1\beta_1$, obtainable from the reduced model $\mathscr{A}_{12\cdot 2}$, is

$$\mathbf{Ay} := \mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_2\mathbf{X}_1)^{-}\mathbf{X}_1'\dot{\mathbf{M}}_2\mathbf{y} \,, \tag{1.38}$$

where $\dot{\mathbf{M}}_2 = \mathbf{M}_2(\mathbf{M}_2\mathbf{W}_1\mathbf{M}_2)^{-}\mathbf{M}_2$ and $\mathbf{W}_1 = \mathbf{V} + \mathbf{X}_1\mathbf{U}_1\mathbf{U}_1'\mathbf{X}_1'$ is such that $\mathscr{C}(\mathbf{W}_1) = \mathscr{C}(\mathbf{X}_1 : \mathbf{V})$. Notice that of course the BLUE of $\mathbf{X}_1\beta_1$ can be written also as

$$\mathbf{By} := (\mathbf{X}_1 : \mathbf{0})(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-}\mathbf{y} = \mathbf{K}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-}\mathbf{y} \,, \tag{1.39}$$

where $\mathbf{K} = (\mathbf{X}_1 : \mathbf{0}) \in \mathbb{R}^{n\times p}$ and $\mathbf{W} \in \mathscr{W}$. The equality $\mathbf{AW} = \mathbf{BW}$ implies

$$\mathbf{W}\dot{\mathbf{M}}_2\mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_2\mathbf{X}_1)^{-}\mathbf{X}_1' = \mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{K}', \tag{1.40}$$

and it is easy to confirm that $\mathscr{C}[\mathbf{W}\dot{\mathbf{M}}_2\mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_2\mathbf{X}_1)^{-}\mathbf{X}_1'] = \mathscr{C}(\mathbf{W}\dot{\mathbf{M}}_2\mathbf{X}_1)$. Thus, in view of part (c) of Lemma 4, the statistic $\mathbf{Fy}$ is linearly sufficient for $\mathbf{X}_1\beta_1$ if and only if

$$\mathscr{C}(\mathbf{W}\dot{\mathbf{M}}_2\mathbf{X}_1) \subset \mathscr{C}(\mathbf{W}\mathbf{F}')\,. \tag{1.41}$$

From (1.41) we immediately see that $\mathbf{X}_1'\dot{\mathbf{M}}_2\mathbf{y}$ is linearly minimal sufficient for $\mathbf{X}_1\beta_1$, as observed by Isotalo & Puntanen (2006a, Th. 2).

1.4 Properties of $\mathscr{C}(\mathbf{W}\mathbf{F}')$

Consider the linear sufficiency condition

$$\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{W}\mathbf{F}')\,, \quad \text{where } \mathbf{W} \in \mathscr{W}\,. \tag{1.42}$$

One question: is the column space $\mathscr{C}(\mathbf{W}\mathbf{F}')$ unique, i.e., does it remain invariant for any choice of $\mathbf{W} \in \mathscr{W}$? In statistical literature, the invariance of $\mathscr{C}(\mathbf{W}\mathbf{F}')$ is not discussed. It might be somewhat tempting to conjecture that for a given $\mathbf{F}$, the column space $\mathscr{C}(\mathbf{W}\mathbf{F}')$ would be invariant. However, our counterexample below shows that this is not the case. In any event, it is of interest to study the mathematical properties of the possible invariance.

Before our counterexample, we will take a quick look at the rank of $\mathbf{W}\mathbf{F}'$ by allowing $\mathbf{W}$ to belong to set $\mathscr{W}_*$, defined as in (1.7),

$$\mathscr{W}_* = \left\{\mathbf{W} \in \mathbb{R}^{n\times n} : \mathbf{W} = \mathbf{V} + \mathbf{X}\mathbf{T}\mathbf{X}',\ \mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})\right\}. \tag{1.43}$$

Now, on account of (1.5) and the equality $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{W}') = \mathscr{C}(\mathbf{X} : \mathbf{V})$, we have $\mathscr{C}(\mathbf{F}\mathbf{W}') = \mathscr{C}(\mathbf{F}\mathbf{W}) = \mathscr{C}[\mathbf{F}(\mathbf{X} : \mathbf{V}\mathbf{M})]$. Using the rank rule for the partitioned matrix: $\operatorname{rank}(\mathbf{A} : \mathbf{B}) = \operatorname{rank}(\mathbf{A}) + \operatorname{rank}[(\mathbf{I} - \mathbf{P}_{\mathbf{A}})\mathbf{B}]$, see, e.g., Marsaglia & Styan (1974, Th. 19), we get

$$\operatorname{rank}(\mathbf{W}\mathbf{F}') = \operatorname{rank}(\mathbf{F}\mathbf{W}') = \operatorname{rank}(\mathbf{F}\mathbf{W}) = \operatorname{rank}(\mathbf{F}\mathbf{X}) + \operatorname{rank}(\mathbf{Q}_{\mathbf{F}\mathbf{X}}\mathbf{F}\mathbf{V}\mathbf{M})\,, \tag{1.44}$$

where $\mathbf{Q}_{\mathbf{F}\mathbf{X}} = \mathbf{I} - \mathbf{P}_{\mathbf{F}\mathbf{X}}$. Now (1.44) means that $\operatorname{rank}(\mathbf{W}\mathbf{F}')$ is invariant with respect to $\mathbf{W} \in \mathscr{W}_*$. In particular, if $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{W}\mathbf{F}')$, we obtain

$$\begin{aligned} \operatorname{rank}(\mathbf{W}\mathbf{F}') &= \operatorname{rank}(\mathbf{X} : \mathbf{W}\mathbf{F}') = \operatorname{rank}(\mathbf{X}) + \operatorname{rank}(\mathbf{M}\mathbf{W}\mathbf{F}') \\ &= \operatorname{rank}(\mathbf{X}) + \operatorname{rank}(\mathbf{M}\mathbf{V}\mathbf{F}') \\ &= \operatorname{rank}(\mathbf{X} : \mathbf{V}\mathbf{F}')\,. \end{aligned} \tag{1.45}$$

We can summarise our observations as follows:

Theorem 1. *Consider the linear model $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}$. Then:*

(a) *The rank of $\mathbf{W}\mathbf{F}'$ is invariant for any $\mathbf{W} \in \mathscr{W}_*$ and it can be expressed as*

$$\operatorname{rank}(\mathbf{W}\mathbf{F}') = \operatorname{rank}(\mathbf{F}\mathbf{X}) + \operatorname{rank}(\mathbf{Q}_{\mathbf{F}\mathbf{X}}\mathbf{F}\mathbf{V}\mathbf{M})\,. \tag{1.46}$$

(b) *For any $\mathbf{W} \in \mathscr{W}_*$, the inclusion $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{W}\mathbf{F}')$ holds if and only if*

$$\operatorname{rank}(\mathbf{W}\mathbf{F}') = \operatorname{rank}(\mathbf{X}) + \operatorname{rank}(\mathbf{F}\mathbf{V}\mathbf{M}) = \operatorname{rank}(\mathbf{X} : \mathbf{V}\mathbf{F}') \,. \tag{1.47}$$

(c) *For any* $\mathbf{W} \in \mathscr{W}_*$*, we have* $\operatorname{rank}(\mathbf{W}'\mathbf{F}') = \operatorname{rank}(\mathbf{W}\mathbf{F}')$.

Example 1. Our purpose is to confirm that the following statement is not correct:
Let $\mathbf{W}_1, \mathbf{W}_2 \in \mathscr{W}$. Then for any matrix $\mathbf{F}$,

$$\mathscr{C}(\mathbf{W}_1\mathbf{F}') = \mathscr{C}(\mathbf{W}_2\mathbf{F}') \,. \tag{1.48}$$

Consider the model where

$$\mathbf{V} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{F}' = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \tag{1.49}$$

and let $\mathbf{U}_1\mathbf{U}_1' = \mathbf{I}_2$, $\mathbf{U}_2\mathbf{U}_2' = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. Denoting $\mathbf{W}_i = \mathbf{V} + \mathbf{X}\mathbf{U}_i\mathbf{U}_i'\mathbf{X}'$, we have

$$\mathscr{C}(\mathbf{W}_1\mathbf{F}') = \mathscr{C}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \neq \mathscr{C}(\mathbf{W}_2\mathbf{F}') = \mathscr{C}\begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}, \tag{1.50}$$

and hence the statement (1.48) is not correct. □

It is interesting to observe that in the above Example 1 the linear sufficiency condition $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{W}\mathbf{F}')$ does not hold. Actually $\mathbf{X}\beta$ is not even estimable under the transformed model $\mathscr{A}_t$ since $\operatorname{rank}(\mathbf{X}'\mathbf{F}') \neq \operatorname{rank}(\mathbf{X})$. For $\mathbf{F}\mathbf{y}$ to be linearly sufficient it is necessary that $\operatorname{rank}(\mathbf{X}) \leq \operatorname{rank}(\mathbf{F})$, which in this case would mean $\operatorname{rank}(\mathbf{F}) \geq 2$. Consider the Example 1 by extending the matrix $\mathbf{F}'$ by one column:

$$\mathbf{F}' = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{X}. \tag{1.51}$$

Then we immediately observe that $\mathscr{C}(\mathbf{W}_1\mathbf{F}') = \mathscr{C}(\mathbf{W}_2\mathbf{F}')$. Actually,

$$\mathscr{C}(\mathbf{X}) = \mathscr{C}(\mathbf{W}_i\mathbf{F}') = \mathscr{C}(\mathbf{W}_i\mathbf{X}) \,, \; i = 1, 2, \tag{1.52}$$

implying that in this situation $\mathbf{F}\mathbf{y} = \mathbf{X}'\mathbf{y}$ is linearly minimal sufficient for $\mathbf{X}\beta$. This provokes the following questions:

(A) When is $\mathbf{X}'\mathbf{y}$ linearly sufficient for $\mathbf{X}\beta$?
(B) What can be said about $\mathscr{C}(\mathbf{W}\mathbf{F}')$ in such a case when $\operatorname{rank}(\mathbf{X}'\mathbf{F}') = \operatorname{rank}(\mathbf{X})$, i.e., $\mathbf{X}\beta$ is estimable under $\mathscr{A}_t$?
(C) Is $\mathscr{C}(\mathbf{W}\mathbf{F}')$ invariant for any choice of $\mathbf{W}$ if $\mathbf{F}\mathbf{y} \in \mathrm{S}(\mathbf{X}\beta)$?

Let us first take a look at the problem (A). Now $\mathbf{X}'\mathbf{y}$ is linearly sufficient for $\mathbf{X}\beta$ if and only if $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{W}\mathbf{X})$, which, in light of $\operatorname{rank}(\mathbf{W}\mathbf{X}) = \operatorname{rank}(\mathbf{X})$, becomes equality

$$\mathscr{C}(\mathbf{X}) = \mathscr{C}(\mathbf{WX}). \tag{1.53}$$

The column space equality (1.53) holds if and only if

$$\mathbf{HWX} = \mathbf{WX}, \tag{1.54}$$

where $\mathbf{H} = \mathbf{P}_{\mathbf{X}}$. Now (1.54) can be equivalently expressed as

$$\mathbf{HV} = \mathbf{VH}, \tag{1.55}$$

which is the well-known condition for the equality of the $\text{OLSE}(\mathbf{X}\beta) = \mathbf{Hy}$ and $\text{BLUE}(\mathbf{X}\beta)$ under the model $\mathscr{A}$; see, e.g., Puntanen & Styan (1989) and Puntanen et al. (2011, Ch. 10). We can express our conclusion as follows:

Theorem 2. *The statistic* $\mathbf{X}'\mathbf{y}$ *is linearly sufficient for* $\mathbf{X}\beta$ *under the model* $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}$ *if and only if*

$$\text{OLSE}(\mathbf{X}\beta) = \text{BLUE}(\mathbf{X}\beta). \tag{1.56}$$

In this situation $\mathbf{X}'\mathbf{y}$ *is linearly minimal sufficient.*

The corresponding result as in Theorem 2, for a positive definite $\mathbf{V}$, appears also in Baksalary & Kala (1981, p. 913). We recall that expression (1.56) is supposed to hold with probability 1, just like any other equality between estimators.

Example 2. As a reply to question (B) above, let us consider the situation where

$$\mathbf{V} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{F}' = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{1.57}$$

In this situation the estimability condition $\text{rank}(\mathbf{FX}) = \text{rank}(\mathbf{X})$ holds but $\mathbf{Fy}$ is not linearly sufficient for $\mathbf{X}\beta$. Choosing $\mathbf{U}_1\mathbf{U}_1' = \mathbf{I}_2$, $\mathbf{U}_2\mathbf{U}_2' = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, and denoting $\mathbf{W}_i = \mathbf{V} + \mathbf{X}\mathbf{U}_i\mathbf{U}_i'\mathbf{X}'$, we have

$$\mathscr{C}(\mathbf{W}_1\mathbf{F}') = \mathscr{C}\begin{pmatrix} 2 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \neq \mathscr{C}(\mathbf{W}_2\mathbf{F}') = \mathscr{C}\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{1.58}$$

Thus the estimability condition is not enough for the invariance of $\mathscr{C}(\mathbf{WF}')$. □

The following theorem is a reply to question (C) above. However, we formulate it in a more general setup by using the set $\mathscr{W}_*$ of $\mathbf{W}$-matrices defined by (1.7) instead of $\mathscr{W}$.

Theorem 3. *Consider the linear model* $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}$*, let* $\mathbf{W} \in \mathscr{W}_*$ *and suppose that* $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WF}')$*. Then the column space* $\mathscr{C}(\mathbf{WF}')$ *is invariant for any choice of* $\mathbf{W} \in \mathscr{W}_*$ *and*

$$\mathscr{C}(\mathbf{WF}') = \mathscr{C}(\mathbf{X}) \oplus \mathscr{C}(\mathbf{MVF}') = \mathscr{C}(\mathbf{W}'\mathbf{F}'). \tag{1.59}$$

Proof. Suppose that $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WF}')$. Then

$$\mathscr{C}(\mathbf{WF}') = \mathscr{C}(\mathbf{X} : \mathbf{WF}') = \mathscr{C}(\mathbf{X}) \oplus \mathscr{C}(\mathbf{MVF}')\,, \tag{1.60}$$

and the proof is completed. □

Next we present the following extended version of Lemma 3:

Theorem 4. *Let $\mathbf{W} \in \mathscr{W}_*$. Then the statistic $\mathbf{Fy}$ is linearly sufficient for $\mathbf{X}\beta$ under the linear model $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}$ if and only if*

$$\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WF}')\,, \tag{1.61}$$

or, equivalently,

$$\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{W}'\mathbf{F}')\,. \tag{1.62}$$

Proof. The proof is parallel to that of Baksalary & Kala (1981, p. 914) who utilize the fact that $\mathbf{By}$ is a BLUE of estimable $\mathbf{K}\beta$ if and only if

$$\mathbf{BW} = \mathbf{K}(\mathbf{X}'\mathbf{W}^+\mathbf{X})^+\mathbf{X}', \quad \text{where } \mathbf{W} \in \mathscr{W}\,. \tag{1.63}$$

However, it is easy to confirm, using (1.8a)–(1.8e), that in this condition the set $\mathscr{W}$ can be replaced with $\mathscr{W}_*$. Moreover, if $\mathbf{W} \in \mathscr{W}_*$, then also $\mathbf{W}' \in \mathscr{W}_*$ and (1.63) can be replaced with

$$\mathbf{BW}' = \mathbf{K}[\mathbf{X}'(\mathbf{W}')^+\mathbf{X}]^+\mathbf{X}'\,. \tag{1.64}$$

Proceeding then along the same lines as Baksalary & Kala (1981), we observe that $\mathbf{AFy}$ is the BLUE for $\mathbf{X}\beta$ under $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}$ if and only if

$$\mathbf{AFW}' = \mathbf{X}[\mathbf{X}'(\mathbf{W}')^+\mathbf{X}]^+\mathbf{X}'\,. \tag{1.65}$$

Now (1.65) has a solution for $\mathbf{A}$, i.e., $\mathbf{Fy}$ is linearly sufficient for $\mathbf{X}\beta$, if and only if

$$\mathscr{C}[\mathbf{X}(\mathbf{X}'\mathbf{W}^+\mathbf{X})^+\mathbf{X}'] \subset \mathscr{C}(\mathbf{WF}')\,. \tag{1.66}$$

Using (1.8a)–(1.8e), we observe that $\mathscr{C}[\mathbf{X}(\mathbf{X}'\mathbf{W}^+\mathbf{X})^+\mathbf{X}'] = \mathscr{C}(\mathbf{X})$ and so we have obtained (1.61). Notice also that in light of Theorem 3, the statements (1.61) and (1.62) are equivalent. □

According to our knowledge, in all linear sufficiency considerations appearing in literature, it is assumed that $\mathbf{W}$ is nonnegative definite. However, this is not necessary, and $\mathbf{W}$ can also be nonsymmetric. Of course, sometimes it is simpler to have $\mathbf{W}$ from set $\mathscr{W}$.

Remark 1. There is one feature in the paper of Baksalary & Kala (1981) that is worth special attention. Namely in their considerations they need the "$\mathbf{W}$-matrix" in the transformed model $\mathscr{A}_t = \{\mathbf{Fy}, \mathbf{FX}\beta, \mathbf{FVF}'\}$. The appropriate set is the following:

$$\mathscr{W}_t = \{\mathbf{W}_t : \mathbf{W}_t = \mathbf{F}(\mathbf{V} + \mathbf{XSX}')\mathbf{F}',\ \mathscr{C}(\mathbf{W}_t) = \mathscr{C}[\mathbf{F}(\mathbf{X} : \mathbf{V})]\}. \tag{1.67}$$

Let $\mathbf{W} = \mathbf{V} + \mathbf{XSX}'$ be some matrix from $\mathscr{W}_*$, and so $\mathbf{W}$ may not be nonnegative definite. We then have

$$\mathscr{C}(\mathbf{W}_t) = \mathscr{C}(\mathbf{FWF}') \subset \mathscr{C}(\mathbf{FW}) = \mathscr{C}[\mathbf{F}(\mathbf{X} : \mathbf{V})]\,. \tag{1.68}$$

If $\mathbf{W}$ is nonnegative definite, as Baksalary & Kala (1981) have, then we have equality in (1.68). However, if $\mathbf{W}$ belongs to $\mathscr{W}_*$ and is not nonnegative definite, then we must add the condition

$$\operatorname{rank}(\mathbf{FWF}') = \operatorname{rank}(\mathbf{FW}) \tag{1.69}$$

if we want to have $\mathbf{FWF}' \in \mathscr{W}_t$. Thus one representation for the BLUE of $\mathbf{FX}\beta$ under $\mathscr{A}_t$ is

$$\mathbf{FX}[\mathbf{X}'\mathbf{F}'(\mathbf{FWF}')^-\mathbf{FX}]^-\mathbf{X}'\mathbf{F}'(\mathbf{FWF}')^-\mathbf{Fy}\,, \tag{1.70}$$

where $\mathbf{W} \in \mathscr{W}_*$ and $\mathbf{W}$ satisfies (1.69). □

1.5 Comments on the relative linear sufficiency

When studying the relative efficiency of OLSE vs BLUE of β we are dealing with two linear models

$$\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}, \quad \mathscr{A}_{\mathbf{I}} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{I}_n\}, \tag{1.71}$$

where the corresponding BLUEs are

$$\tilde{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \quad \hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \tag{1.72}$$

Then it is assumed that model $\{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}$ is correct and then the relative goodness of $\hat{\beta}$ with respect to $\tilde{\beta}$ is measured by various means. The most common measure is the Watson efficiency, see Watson (1955) and Bloomfield & Watson (1975),

$$\phi = \frac{|\operatorname{cov}(\tilde{\beta})|}{|\operatorname{cov}(\hat{\beta})|} = \frac{|\mathbf{X}'\mathbf{X}|^2}{|\mathbf{X}'\mathbf{VX}| \cdot |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|}\,, \tag{1.73}$$

where $|\cdot|$ refers to the determinant. Obviously $0 < \phi \leq 1$ and the upper bound is attained when $\tilde{\beta} = \hat{\beta}$.

Let us consider the models

$$\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}\,, \quad \mathscr{A}_t = \{\mathbf{Fy}, \mathbf{FX}\beta, \mathbf{FVF}'\}\,, \tag{1.74}$$

and try to do something similar with

$$\operatorname{BLUE}(\beta \mid \mathscr{A}) = \tilde{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\,, \tag{1.75}$$

$$\operatorname{BLUE}(\beta \mid \mathscr{A}_t) = \tilde{\beta}_t = [\mathbf{X}'\mathbf{F}'(\mathbf{FVF}')^-\mathbf{FX}]^{-1}\mathbf{X}'\mathbf{F}'(\mathbf{FVF}')^-\mathbf{Fy}\,. \tag{1.76}$$

Above we have some rank problems. To simplify the considerations, we have assumed that $\mathbf{V}$ is positive definite. The model matrix $\mathbf{X}$ has to have full column rank so that β would be estimable under $\mathscr{A}$. Similarly, $\mathbf{FX}$ has to have full column rank for β to be estimable under $\mathscr{A}_t$; using the rank rule of Marsaglia & Styan (1974, Cor. 6.2) for the matrix product, we must have

$$p = \operatorname{rank}(\mathbf{X}) = \operatorname{rank}(\mathbf{FX}) = \operatorname{rank}(\mathbf{X}) - \dim \mathscr{C}(\mathbf{X}) \cap \mathscr{C}(\mathbf{F}')^{\perp}, \tag{1.77}$$

so that

$$\mathscr{C}(\mathbf{X}) \cap \mathscr{C}(\mathbf{F}')^{\perp} = \{\mathbf{0}\}. \tag{1.78}$$

It is noteworthy that in view of $\mathscr{C}(\mathbf{FX}) \subset \mathscr{C}(\mathbf{FVF}') = \mathscr{C}(\mathbf{F})$ the model $\mathscr{A}_t = \{\mathbf{Fy}, \mathbf{FX}\beta, \mathbf{FVF}'\}$ is so-called weakly singular linear, or Zyskind–Martin model, see Zyskind & Martin (1969), and hence the representation (1.76) indeed is valid for any $(\mathbf{FVF}')^{-}$. Moreover, it is easy to confirm that $\mathbf{X}'\mathbf{F}'(\mathbf{FVF}')^{-}\mathbf{FX}$ is positive definite.

Notice that $\mathrm{E}(\tilde{\beta}) = \mathrm{E}(\tilde{\beta}_t) = \beta$ and

$$\operatorname{cov}(\tilde{\beta}_t) = [\mathbf{X}'\mathbf{F}'(\mathbf{FVF}')^{-}\mathbf{FX}]^{-1}, \quad \operatorname{cov}(\tilde{\beta}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \tag{1.79}$$

Remark 2. The following Löwner ordering obviously holds:

$$\operatorname{cov}(\tilde{\beta}) \leq_{\mathrm{L}} \operatorname{cov}(\tilde{\beta}_t), \tag{1.80}$$

i.e.,

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \leq_{\mathrm{L}} [\mathbf{X}'\mathbf{F}'(\mathbf{FVF}')^{-}\mathbf{FX}]^{-1}. \tag{1.81}$$

We can rewrite (1.81) as

$$\mathbf{X}'\mathbf{V}^{-1/2}\mathbf{P}_{\mathbf{V}^{1/2}\mathbf{F}'}\mathbf{V}^{-1/2}\mathbf{X} \leq_{\mathrm{L}} \mathbf{X}'\mathbf{V}^{-1/2}\mathbf{V}^{-1/2}\mathbf{X}, \tag{1.82}$$

where the equality is obtained if and only if $\mathscr{C}(\mathbf{V}^{-1/2}\mathbf{X}) \subset \mathscr{C}(\mathbf{V}^{1/2}\mathbf{F}')$, i.e., $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{VF}')$, which is precisely the condition for linear sufficiency (when $\mathbf{V}$ is positive definite). □

Corresponding to Watson efficiency, we could consider the ratio

$$\begin{aligned} \gamma = \frac{|\operatorname{cov}(\tilde{\beta})|}{|\operatorname{cov}(\tilde{\beta}_t)|} &= \frac{|(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}|}{|[\mathbf{X}'\mathbf{F}'(\mathbf{FVF}')^{-}\mathbf{FX}]^{-1}|} \\ &= \frac{|\mathbf{X}'\mathbf{F}'(\mathbf{FVF}')^{-}\mathbf{FX}|}{|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|} \\ &= \frac{|\mathbf{X}'\mathbf{V}^{-1/2}\mathbf{P}_{\mathbf{V}^{1/2}\mathbf{F}'}\mathbf{V}^{-1/2}\mathbf{X}|}{|\mathbf{X}'\mathbf{V}^{-1/2}\mathbf{V}^{-1/2}\mathbf{X}|}. \end{aligned} \tag{1.83}$$

Clearly

$$0 < \gamma \leq 1, \tag{1.84}$$

where the upper bound is attained if and only if $\mathbf{Fy}$ is linearly sufficient for β. What might be the lower bound? Here we now keep $\mathbf{X}$ and $\mathbf{V}$ given and try to figure out which $\mathbf{F}$ yields the minimum of γ subject to the condition $\operatorname{rank}(\mathbf{X}) = \operatorname{rank}(\mathbf{FX})$. The lower bound for the Watson efficiency was found by Bloomfield & Watson (1975) [actually it appeared already in Watson (1955) but there was a flaw in the proof]. However, it seems to be nontrivial to find the lower bound for γ. The (attainable) lower bound zero does not make sense, of course.

Remark 3. Consider matrices $\mathbf{F}_1$ and $\mathbf{F}_2$ and the corresponding transformed models

$$\mathscr{A}_{ti} = \{\mathbf{F}_i\mathbf{y}, \mathbf{F}_i\mathbf{X}\beta, \mathbf{F}_i\mathbf{V}\mathbf{F}_i'\}, \; i = 1,2, \tag{1.85}$$

and suppose that $\operatorname{rank}(\mathbf{F}_1\mathbf{X}) = \operatorname{rank}(\mathbf{F}_2\mathbf{X}) = \operatorname{rank}(\mathbf{X}) = p$, so that β is estimable under both models. We observe that the Löwner ordering

$$\operatorname{cov}(\tilde{\beta}_{t1}) \leq_{\mathrm{L}} \operatorname{cov}(\tilde{\beta}_{t2}) \tag{1.86}$$

holds if and only if

$$\mathbf{X}'\mathbf{V}^{-1/2}\mathbf{P}_{\mathbf{V}^{1/2}\mathbf{F}_2'}\mathbf{V}^{-1/2}\mathbf{X} \leq_{\mathrm{L}} \mathbf{X}'\mathbf{V}^{-1/2}\mathbf{P}_{\mathbf{V}^{1/2}\mathbf{F}_1'}\mathbf{V}^{-1/2}\mathbf{X}, \tag{1.87}$$

i.e.,

$$\mathbf{X}'\mathbf{V}^{-1/2}(\mathbf{P}_{\mathbf{V}^{1/2}\mathbf{F}_1'} - \mathbf{P}_{\mathbf{V}^{1/2}\mathbf{F}_2'})\mathbf{V}^{-1/2}\mathbf{X} \geq_{\mathrm{L}} \mathbf{0}. \tag{1.88}$$

The matrix $\mathbf{P}_{\mathbf{V}^{1/2}\mathbf{F}_1'} - \mathbf{P}_{\mathbf{V}^{1/2}\mathbf{F}_2'}$ is nonnegative definite if and only if

$$\mathscr{C}(\mathbf{F}_2') \subset \mathscr{C}(\mathbf{F}_1'). \tag{1.89}$$

Hence we can conclude that (1.86) holds if $\mathscr{C}(\mathbf{F}_2') \subset \mathscr{C}(\mathbf{F}_1')$. In this case we can say that in a sense $\mathbf{F}_1\mathbf{y}$ is "more than or equally linearly sufficient" than $\mathbf{F}_2\mathbf{y}$ even though neither of them need to be "fully linearly sufficient". Notice that if $\mathscr{C}(\mathbf{F}_1') = \mathbb{R}^n$, i.e., $\mathbf{F}_1$ is a nonsingular $n \times n$ matrix, then $\operatorname{cov}(\tilde{\beta}_{t1})$ is the smallest in the Löwner sense in the set of $\operatorname{cov}(\tilde{\beta}_t)$: it is $\operatorname{cov}(\tilde{\beta})$.

However, it may well be that there is no Löwner ordering between the covariance matrices $\operatorname{cov}(\tilde{\beta}_{t1})$ and $\operatorname{cov}(\tilde{\beta}_{t2})$. Then some other criteria should be used to compare the "linear sufficiency" of $\mathbf{F}_1\mathbf{y}$ and $\mathbf{F}_2\mathbf{y}$. □

Bloomfield & Watson (1975) introduced also another measure of efficiency of the OLSE, based on the Frobenius norm of the commutator $\mathbf{HV} - \mathbf{VH}$:

$$\delta = \tfrac{1}{2}\,\|\mathbf{HV} - \mathbf{VH}\|_F^2 = \|\mathbf{HVM}\|_F^2 = \operatorname{tr}(\mathbf{HVMVH}), \tag{1.90}$$

where $\operatorname{tr}(\cdot)$ refers to the trace. They showed that the maximum of δ is attained in the same situation as the minimum of the Watson efficiency ϕ. Of course, $\delta = 0$ if and only if $\operatorname{OLSE}(\mathbf{X}\beta)$ equals $\operatorname{BLUE}(\mathbf{X}\beta)$.

We can now try to develop something similar as the commutator criterion for the linear linear sufficiency condition $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WF}')$ which is equivalent to

$$\mathbf{P}_{\mathbf{WF}'}\mathbf{X} = \mathbf{X}. \tag{1.91}$$

Hence one can wonder how "badly" (1.42) is satisfied by considering the difference

$$\mathbf{D} := \mathbf{X} - \mathbf{P}_{\mathbf{WF}'}\mathbf{X}. \tag{1.92}$$

The "size" of $\mathbf{D}$ could be measured by the Frobenius norm as

$$\|\mathbf{D}\|_F^2 = \operatorname{tr}(\mathbf{D}'\mathbf{D}) = \operatorname{tr}(\mathbf{X}'\mathbf{X}) - \operatorname{tr}(\mathbf{X}'\mathbf{P}_{\mathbf{WF}'}\mathbf{X}). \tag{1.93}$$

Hence the relative linear sufficiency of $\mathbf{Fy}$ could be defined as

$$\psi = \frac{\operatorname{tr}(\mathbf{X}'\mathbf{P}_{\mathbf{WF}'}\mathbf{X})}{\operatorname{tr}(\mathbf{X}'\mathbf{X})}. \tag{1.94}$$

Now

$$0 \le \psi \le 1, \tag{1.95}$$

where the lower bound is attained when $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WF}')^{\perp}$ and the upper bound is attained when $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WF}')$, i.e., when $\mathbf{Fy}$ is linearly sufficient for $\mathbf{X}\beta$.

1.6 Euclidean norm of the difference between the BLUEs under $\mathscr{A}$ and $\mathscr{A}_t$

In this section we will study the properties of the Euclidean norm of the difference between the BLUEs of $\mu := \mathbf{X}\beta$ under the models $\mathscr{A}$ and $\mathscr{A}_t$. We can denote shortly

$$\operatorname{BLUE}(\mathbf{X}\beta \mid \mathscr{A}) = \tilde{\mu}, \quad \text{and} \quad \operatorname{BLUE}(\mathbf{X}\beta \mid \mathscr{A}_t) = \tilde{\mu}_t. \tag{1.96}$$

The corresponding considerations for $\operatorname{OLSE}(\mathbf{X}\beta) - \operatorname{BLUE}(\mathbf{X}\beta)$ have been made by Baksalary & Kala (1978, 1980) and for the BLUEs under two models by Hauke et al. (2012); see also Mäkinen (2000, 2002), Pordzik (2012), Baksalary et al. (2013), and Haslett et al. (2014).

Suppose that $\mathbf{W} \in \mathscr{W}$. Then the BLUE under the original model $\mathscr{A}$ can be expressed as $\mathbf{Gy}$ where

$$\mathbf{G} = \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^-. \tag{1.97}$$

Moreover, assuming that $\mathbf{X}\beta$ is estimable under the transformed model $\mathscr{A}_t$, the estimator $\mathbf{BFy}$ is the BLUE for $\mathbf{X}\beta$ under $\mathscr{A}_t$ if and only if $\mathbf{B}$ satisfies

$$\mathbf{B}[\mathbf{FX} : \mathbf{FVF}'(\mathbf{FX})^{\perp}] = (\mathbf{X} : \mathbf{0}). \tag{1.98}$$

One choice for $\mathbf{B}$ is $\mathbf{X}[\mathbf{X}'\mathbf{F}'(\mathbf{FWF}')^-\mathbf{FX}]^-\mathbf{X}'\mathbf{F}'(\mathbf{FWF}')^-$ and so the BLUE of $\mathbf{X}\beta$ under $\mathscr{A}_t$ has representation $\mathbf{G}_t\mathbf{y}$, where

$$\mathbf{G}_t = \mathbf{X}[\mathbf{X}'\mathbf{F}'(\mathbf{FWF}')^-\mathbf{FX}]^-\mathbf{X}'\mathbf{F}'(\mathbf{FWF}')^-\mathbf{F}. \tag{1.99}$$

We observe that $\mathbf{G}_t\mathbf{G} = \mathbf{G}$ and hence for all $\mathbf{y} \in \mathscr{C}(\mathbf{W})$ we have

$$\begin{aligned}(\mathbf{G}_t - \mathbf{G})\mathbf{y} &= (\mathbf{G}_t - \mathbf{G}_t\mathbf{G})\mathbf{y} \\ &= \mathbf{G}_t(\mathbf{I}_n - \mathbf{G})\mathbf{y} \\ &= \mathbf{G}_t\mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y}\,, \end{aligned} \tag{1.100}$$

where we have used (1.12), i.e.,

$$(\mathbf{I}_n - \mathbf{G})\mathbf{y} = \mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y} \quad \text{for all } \mathbf{y} \in \mathscr{C}(\mathbf{W})\,. \tag{1.101}$$

Notice that in view of (1.13), the expression $\mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y}$ is invariant for the choice of $(\mathbf{M}\mathbf{V}\mathbf{M})^{-}$ for all $\mathbf{y} \in \mathscr{C}(\mathbf{W})$.

The Euclidean norm of vector $\mathbf{a}$ is of course $\|\mathbf{a}\|_2 = \sqrt{\mathbf{a}'\mathbf{a}}$ and the corresponding matrix norm (spectral norm) $\|\mathbf{A}\|_2$ is defined as the square root of the largest eigenvalue of $\mathbf{A}'\mathbf{A}$. Then, for all $\mathbf{y} \in \mathscr{C}(\mathbf{W})$, we have

$$\begin{aligned}\|\mathbf{G}_t\mathbf{y} - \mathbf{G}\mathbf{y}\|_2^2 &= \|\mathbf{G}_t\mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y}\|_2^2 \\ &\leq \|\mathbf{G}_t\mathbf{V}\mathbf{M}\|_2^2\ \|(\mathbf{M}\mathbf{V}\mathbf{M})^{+}\|_2^2\ \|\mathbf{M}\mathbf{y}\|_2^2\,. \end{aligned} \tag{1.102}$$

The inequality in (1.102) follows from the consistency and multiplicativity of the matrix norm $\|\mathbf{A}\|_2$; see, e.g., Ben-Israel & Greville (2003, pp. 19–20).

The special situation when $\mathbf{V}\mathbf{M} = \mathbf{0}$, i.e., $\mathscr{C}(\mathbf{V}) \subset \mathscr{C}(\mathbf{X})$, deserves some attention. Notice also, as pointed out by Haslett et al. (2014, p. 554), that $\mathbf{y}'\mathbf{M}\mathbf{y} = 0$ for all $\mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V})$ holds if and only if $\mathbf{V}\mathbf{M} = \mathbf{0}$. Groß (2004, p. 317) calls a model with property $\mathbf{V}\mathbf{M} = \mathbf{0}$ a *degenerated* model. If $\mathscr{A}$ is not a degenerated model then the right-hand side of (1.102) is zero if and only if

$$\mathbf{G}_t\mathbf{V}\mathbf{M} = \mathbf{0}\,. \tag{1.103}$$

Noticing that obviously $\mathbf{G}_t$ satisfies $\mathbf{G}_t\mathbf{X} = \mathbf{X}$, we can conclude that (1.103) means that $\mathbf{G}_t\mathbf{y}$ is a BLUE also under the original model $\mathscr{A}$. Thus, in light of Lemma 5, (1.103) means also that $\mathbf{F}\mathbf{y}$ is linearly sufficient.

Thus we have proved the following:

Theorem 5. *Suppose that $\mu = \mathbf{X}\beta$ is estimable under the transformed model $\mathscr{A}_t$. Then, using the above notation,*

$$\begin{aligned}\|\tilde{\mu}_t - \tilde{\mu}\|_2^2 &\leq \|\mathbf{G}_t\mathbf{V}\mathbf{M}\|_2^2\ \|(\mathbf{M}\mathbf{V}\mathbf{M})^{+}\|_2^2\ \mathbf{y}'\mathbf{M}\mathbf{y} \\ &= \frac{a}{\alpha^2}\,\mathbf{y}'\mathbf{M}\mathbf{y}\,, \end{aligned} \tag{1.104}$$

where α is the smallest nonzero eigenvalue of $\mathbf{M}\mathbf{V}\mathbf{M}$, and a is the largest eigenvalue of $\mathbf{G}_t\mathbf{V}\mathbf{M}\mathbf{V}\mathbf{G}_t'$. Moreover, if $\mathscr{A}$ is not a degenerated model then the right-hand side of (1.104) *is zero if and only if $\mathbf{F}\mathbf{y}$ is linearly sufficient for $\mathbf{X}\beta$.*

1.7 Conclusions

The origins of the idea of transforming $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}$ by a matrix $\mathbf{F}$ of order $f \times n$ follow from a desire of reduction of the initial information delivered by an observed value of a random vector variable $\mathbf{y}$ in such a way that it is still possible to obtain the BLUE of $\mathbf{X}\beta$ from the transformed model $\mathscr{A}_t = \{\mathbf{Fy}, \mathbf{FX}\beta, \mathbf{FVF}'\}$. Hence the concept of the linear sufficiency has an essential role when studying the connection between $\mathscr{A}$ and its transformed version $\mathscr{A}_t$.

In the theory of linear models the classes of matrices

$$\mathscr{W} = \left\{\mathbf{W} \in \mathbb{R}^{n\times n} : \mathbf{W} = \mathbf{V} + \mathbf{XUU}'\mathbf{X}',\ \mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})\right\}, \tag{1.105a}$$

$$\mathscr{W}_* = \left\{\mathbf{W} \in \mathbb{R}^{n\times n} : \mathbf{W} = \mathbf{V} + \mathbf{XTX}',\ \mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})\right\}, \tag{1.105b}$$

have important roles. In our paper we study in details the properties of these **W**-matrices related to the concept of linear sufficiency. As far as we know, in all linear sufficiency considerations appearing in literature, it is assumed that $\mathbf{W}$ is nonnegative definite, i.e., $\mathbf{W}$ belongs to set $\mathscr{W}$. We have shown that this is not necessary: it is enough if $\mathbf{W}$ belongs to set $\mathscr{W}_*$.

If $\mathbf{Fy}$ is linearly sufficient then the BLUEs of $\mathbf{X}\beta$ under $\mathscr{A}$ and under $\mathscr{A}_t$ are equal (with probability 1). Hence it might be of interest to describe the relative linear sufficiency of $\mathbf{Fy}$ by comparing the BLUEs under $\mathscr{A}$ and under $\mathscr{A}_t$ by some means. Some suggestions on this matter are made in Section 1.5. The applicability of these measures is left for further research.

Acknowledgements Part of this research was done during the meetings of an International Research Group on Multivariate Models in the Mathematical Research and Conference Center, Będlewo, Poland, March 2015 and October 2015, supported by the Stefan Banach International Mathematical Center.

References

Aitken, A. C. (1935): On least squares and linear combination of observations. Proc. Roy. Soc. Edinburgh Sect. A **55**, 42–49.

Arendacká, B. & Puntanen, S. (2015): Further remarks on the connection between fixed linear model and mixed linear model. Stat. Pap. **56**, 1235–1247.

Baksalary, J. K. (2004): An elementary development of the equation characterizing best linear unbiased estimators. Linear Algebra Appl. **388**, 3–6.

Baksalary, J. K., Drygas, H. (1992): A note on the concepts of sufficiency in the general Gauss–Markov model: a coordinate-free approach. Forschungsbericht 92/2, Universität Dortmund, Fachbereich Statistik.

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. Stat. **6**, 1390–1393.

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. Stat. **8**, 679–681.

Baksalary, J. K., Kala, R. (1981): Linear transformations preserving best linear unbiased estimators in a general Gauss–Markoff model. Ann. Stat. **9**, 913–916.
Baksalary, J. K., Kala, R. (1986): Linear sufficiency with respect to a given vector of parametric functions. J. Stat. Plan. Inf. **14**, 331–338.
Baksalary, J. K., Mathew, T, (1986): Linear sufficiency and completeness in an incorrectly specified general Gauss–Markov model. Sankhyā Ser. A **48**, 169–180.
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.
Baksalary, J. K., Puntanen, S. (1989): Weighted-least-squares estimation in the general Gauss–Markov model. In: Statistical Data Analysis and Inference (Y. Dodge, ed.). Elsevier Science Publishers B.V., Amsterdam, 355–368.
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.
Baksalary, J. K., Rao, C. R., Markiewicz, A. (1992): A study of the influence of the "natural restrictions" on estimation problems in the singular Gauss–Markov model. J. Stat. Plan. Inf. **31**, 335–351.
Baksalary, O. M., Trenkler, G., Liski, E. (2013): Let us do the twist again. Stat. Pap. **54**, 1109–1119.
Ben-Israel, A., Greville, T. N. E. (2003): Generalized Inverses: Theory and Applications, 2nd Edition. Springer, New York.
Bloomfield, P., Watson, G. S. (1975): The inefficiency of least squares. Biometrika **62**, 121–128.
Christensen, R. (2011): Plane Answers to Complex Questions: the Theory of Linear Models, 4th Edition. Springer, New York.
Drygas, H. (1970): The Coordinate-Free Approach to Gauss–Markov Estimation. Springer, Berlin
Drygas, H. (1983): Sufficiency and completeness in the general Gauss–Markov model. Sankhyā Ser. A **45**, 88–98.
Groß, J. (1998): A note on the concepts of linear and quadratic sufficiency. J. Stat. Plan. Inf. **70**, 88–98.
Groß, J. (2004): The general Gauss–Markov model with possibly singular dispersion matrix. Stat. Pap. **45**, 311–336.
Groß, J., Puntanen, S. (2000): Estimation under a general partitioned linear model. Linear Algebra Appl. **321**,131–144.
Groß, J., Puntanen, S. (2005): Extensions of the Frisch–Waugh–Lovell Theorem. Discussiones Mathematicae – Probab. Stat. **25**, 39–49.
Haslett, S. J., Isotalo, J., Liu, Y., Puntanen, S. (2014): Equalities between OLSE, BLUE and BLUP in the linear model. Stat. Pap. **55**, 543–561.
Hauke, J., Markiewicz, A., Puntanen, S. (2012): Comparing the BLUEs under two linear models. Commun. Stat. Theory Methods **41**, 2405–2418.
Isotalo, J., Puntanen, S. (2006a): Linear sufficiency and completeness in the partitioned linear model. Acta Comment Univ. Tartu Math. **10**, 53–67.
Isotalo, J., Puntanen, S. (2006b): Linear prediction sufficiency for new observations in the general Gauss–Markov model. Commun. Stat. Theory Methods **35**, 1011–1023.
Kala, R., Pordzik, P. R. (2009): Estimation in singular partitioned, reduced or transformed linear models. Stat. Pap. **50**, 633–638.
Kala, R., Puntanen, S., Tian, Y. (2015): Some notes on linear sufficiency. Stat. Pap., available online. doi: 10.1007/s00362-015-0682-2
Kornacki, A. (2007): Different kinds of sufficiency in the general Gauss–Markov model. Math. Slovaca **57**, 389–392.
Marsaglia, G., Styan, G. P. H. (1974): Equalities and inequalities for ranks of matrices. Linear and Multilinear Algebra **2**, 269–292.
Müller, J. (1987): Sufficiency and completeness in the linear model. J. Multivar. Anal. **21**, 312–323.
Mäkinen, J. (2000): Bounds for the difference between a linear unbiased estimate and the best linear unbiased estimate. Phys. Chem. Earth Pt A. **25**, 693–698.

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

Pordzik, P. R. (2012): A bound for the Euclidean distance between restricted and unrestricted estimators of parametric functions in the general linear model. Stat. Pap. **53**, 299–304.

Puntanen, S., Styan, G. P. H. (1989): The equality of the ordinary least squares estimator and the best linear unbiased estimator (with discussion). Am. Stat. **43**, 151–161 [Commented by O. Kempthorne on pp. 161–162 and by S. R. Searle on pp. 162–163, Reply by the authors on p. 164.]

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

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.

Rao, C. R., Mitra, S. K. (1971): Generalized Inverse of Matrices and Its Applications. Wiley, New York.

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.

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

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.