LINEAR MODELS AND CONVEX PROGRAMS:
UNBIASED NONNEGATIVE ESTIMATION
IN VARIANCE COMPONENT MODELS

BY

FRIEDRICHPUKELSHEIM

TECHNICAL REPORT NO. 104
DECEMBER 1977

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT MPS74-21416

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
LINEAR MODELS AND CONVEX PROGRAMS:
UNBIASED NONNEGATIVE ESTIMATION
IN VARIANCE COMPONENT MODELS

BY

FRIEDRICH PUKELSHEIM

TECHNICAL REPORT NO. 104
DECEMBER 1977

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT MPS74-21416

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
LINEAR MODELS AND CONVEX PROGRAMS:
UNBIASED NONNEGATIVE ESTIMATION
IN VARIANCE COMPONENT MODELS

by Friedrich Pukelsheim
Albert-Ludwigs-Universität Freiburg im Breisgau
and Stanford University

Author's mailing address: Friedrich Pukelsheim
Department of Statistics
Stanford University
Stanford, California 94305

Abbreviated title: ESTIMATION OF VARIANCE COMPONENTS


Key words and phrases: Negative estimates of variances; unbiased
nonnegative estimability; best nonnegative estimator; negativity
eliminating projectors.
SUMMARY

Linear Models and Convex Programs: Unbiased Nonnegative Estimation in Variance Component Models

For a subclass of linear models, unbiased nonnegative-definite quadratic estimability of variance components is characterized in terms of the matrices that specify the linear decomposition of the model's dispersion matrix.

Assuming this kind of estimability, two representations of the "best nonnegative estimator," i.e., the minimum (Euclidean matrix) norm — unbiased — nonnegative-definite quadratic estimator, are derived. The first characterization is based on the "best defective estimator" (= MINQUE) by correcting it appropriately. The second characterization is built on "negativity eliminating projectors" which, by definition, generate a reduced model whose best defective estimator is equal to the best nonnegative estimator of the original model.

The above setting shifts the problem from linearity (linear space of symmetric matrices) to convexity (convex cone of nonnegative-definite matrices), and programming techniques provide the main tool to establish the present results.
1. Summary and introduction

Negative estimates of variances represent one of the major problems of linear model theory. For variance component models, this paper offers a twofold solution when at least one unbiased nonnegative-definite quadratic estimator exists: In Section 5, the best nonnegative estimator, i.e., the minimum norm — unbiased — nonnegative-definite quadratic estimator, is represented by appropriately adjusting C. R. Rao's MINQUE; in Section 6, the same estimator is obtained from a reduced model in which the problem generating negativity has been eliminated.

At the outset of Section 2, the problem is viewed at as a special case of the more general fundamental defect of linear model theory. Specializing to nonnegative estimation of variance components, the best nonnegative estimator is then characterized in Section 3 as the optimal solution of a convex minimization program. Section 4 investigates when this program has at least one feasible solution, or equivalently, in terms of the statistical model, when at least one unbiased nonnegative-definite quadratic estimator exists. In the concluding Section 7 we collect some indications what the explicit solution may actually look like.

Notations. A linear model is denoted by

\[(1.1) \quad \mathbf{y} \sim (\Sigma b_{\pi=\Pi} ; \Sigma t_{k=k})\]

and consists of an \(\mathbb{R}^n\)-valued observation vector \(\mathbf{y}\) assuming linear decompositions of both its mean vector \(\mathbf{e}_{\pi} = \Sigma b_{\pi=\Pi}\) and its dispersion
matrix $\mathbf{V} = \sum_{k=1}^{p} t_k \mathbf{V}_k$, where the $p$ decomposing $\mathbb{R}^n$-vectors $x_i$ and the $k$ decomposing symmetric $n \times n$ matrices $\mathbf{V}_k$ are known, while the coefficients $\mathbf{b} := (b_1, \ldots, b_p)' \in \mathbb{R}^p$ and $\mathbf{t} := (t_1, \ldots, t_k)' \in \mathbb{R}^k$ are to be estimated. If, in addition, the $k$ decomposing matrices $\mathbf{V}_k$ are assumed to be nonnegative-definite, and the components of $\mathbf{t}$ are restricted to be nonnegative, we call (1.1) a variance component model.

For a variance component model we could write $\mathbf{V}_k = \mathbf{U}_k \mathbf{U}_k'$, and $t_k = \sigma_k^2$, as is usually done with analysis of variance models. Our 'variance component model' excludes covariances, deviating from this name's use by Scheffé (1959, p. 221) and Rao (1973, pp. 258, 302).

As for further notations, we use $\mathbb{R}^{n \times n}$ to denote the space of all real $n \times n$ matrices, endowed with the Euclidean matrix inner product $\langle \mathbf{A}, \mathbf{B} \rangle = \text{trace} \mathbf{A} \mathbf{B}'$. A prime indicates transposition. $\mathfrak{R}$ stands for range (column space), $\mathfrak{H}$ for nullspace, $\text{sp}$ for linear span. Nonnegative-definite is abbreviated by NND. $\mathbf{A}$ is said to be NND if $\mathbf{y}' \mathbf{A} \mathbf{y} \geq 0$ for all $\mathbf{y}$ and $\mathbf{A}$ is symmetric. $\mathbf{A} \succcurlyeq \mathbf{B}$ is used if and only if $\mathbf{A} - \mathbf{B}$ is NND. We write $\text{Sym}(n)$ for the subspace of symmetric matrices of $\mathbb{R}^{n \times n}$, and NND(n) for the closed convex cone of NND matrices. Orthogonality is indicated by $\perp$; all our projectors project orthogonally. The nonnegative orthant of $\mathbb{R}^k$ is denoted by $\mathbb{R}_+^k$.

Previous work. Most comments on negative estimates of variance components content themselves with rather ad hoc reasoning what to do
if a particular estimate turns out to be negative; see the discussion in Searle (1971, pp. 406-408). Two kinds of reaction suggest themselves: change either the model assumptions or the class of estimators. The first course of action is taken by Nelder (1954) who interprets negative variances by assuming randomized blocks, and by McHugh & Mielke (1968) who argue with a finite population. The second possibility is chosen by Thompson (1962) who discusses an algorithm for maximizing the likelihood under a nonnegativity constraint — see also the recent review by Harville (1977) —, and by Federer (1968) whose estimator is nonnegative but lacks other desirable properties. Herbach (1959) and Drygas (1972) investigate fairly completely nonnegativity of the estimators they discuss, but these results are restricted to their particular models.

In general, MINQUE is consistent, cf. Brown (1976), and hence asymptotically nonnegative. But this offers a bad comfort to the experimenter if no replicates are available, and in any case it is not satisfying if an otherwise finite theory is patched up by an asymptotic argument.

Programming methods have been applied to linear model theory by Judge & Takayama (1966), Hudson (1969), and Liew (1976) who deal with inequality constrained linear estimation of the mean; Brown (1977) indicates an extension to nonnegative estimation of variance components; Harville (1977) outlines the use of programming methods in maximum likelihood approaches. These authors, however, restrict their investigations to computational algorithms which, when given a
particular sample, single out a reasonable value in the parameter space. Quite differently, we shall use those programming techniques to find best estimators, i.e., functions from the sample space into the parameter space. The motivation for our approach originates from the theory of tests where best tests appear as optimal solutions of linear programs, as described by Witting (1966, pp. 69-73) or reviewed by Krafft (1970).
2. The fundamental defect of linear model theory

Definition (1.1) of a linear model lacks an exact statement concerning the sets $\Theta_\mathbb{P} \subset \mathbb{R}^p$ and $\Theta_\mathbb{K} \subset \mathbb{R}^k$ of parameter values $\underline{\theta}$ and $\underline{t}$ that can be attained, or more precisely, for which $\underline{b} \in \mathbb{R}^p$ and $\underline{t} \in \mathbb{R}^k$ there exists a probability measure $\mathcal{P} \in \mathcal{P}$ such that $\varepsilon_{\mathcal{P} \underline{y}} = \Sigma \underline{b} \underline{x}_{\underline{x_\pi}}$, and $\varphi_{\mathcal{P} \underline{y}} = \Sigma \underline{t} \underline{V}_{\underline{K} = \underline{K}}$, when $\mathcal{P}$ is the underlying class of distributions.

Implicitly, linear model theory assumes that

$$
(2.1) \quad \text{sp}\{\varepsilon_{\mathcal{P} \underline{y}}/\mathcal{P} \in \mathcal{P}\} = \text{sp}\{\underline{x}_1, \ldots, \underline{x}_p\}, \quad \text{sp}\{\varphi_{\mathcal{P} \underline{y}}/\mathcal{P} \in \mathcal{P}\} = \text{sp}\{\underline{v}_1, \ldots, \underline{v}_k\}.
$$

The prevailing theory of linear estimation does not make any effort to maintain the difference between the exact set of parameter values and its linear span; this deficiency we call the fundamental defect of linear model theory.

As a first example, consider the estimation of the expected value $\varepsilon_{\underline{y}}$: the resulting estimates always lie in $\text{sp}\{\underline{x}_1, \ldots, \underline{x}_p\}$, but need not be in $\{\varepsilon_{\mathcal{P} \underline{y}}/\mathcal{P} \in \mathcal{P}\}$. As a second example, consider the estimation of the mean parameter $\underline{b}$: The resulting estimates always lie in $\mathbb{R}^p$ ($= \text{sp}_{\mathcal{P}} \Theta_\mathbb{P}$ under assumption (2.1)), but need not be in the set $\Theta_\mathbb{P}$ of attainable parameter values.

In mean estimation, the fundamental defect may be passed over by admitting every $\underline{b} \in \mathbb{R}^p$ to be attainable, i.e., $\Theta_\mathbb{P} = \mathbb{R}^p$. As for variance covariance component estimation, however, $\Sigma t \underline{V}_{\underline{K} = \underline{K}}$ has to be NND, and $\Theta_\mathbb{K}$ cannot possibly be all of $\mathbb{R}^k$. Thus the fundamental defect cannot be denied in general; with variance covariance components, it even appears in two distinct forms.
Firstly, in any linear model, \( \mathbf{t} \) and any estimate \( \hat{\mathbf{t}} \) thereof is restricted to yield NND sums \( \sum_{k=K}^K \mathbf{t}_k \mathbf{V}_k \) and \( \sum_{k=K}^K \hat{\mathbf{t}}_k \mathbf{V}_k \). A sufficient condition for MINQUE to have this property is given in Pukelsheim & Styan (1977).

Secondly, in a variance component model, each component \( \mathbf{t}_K \) of \( \mathbf{t} \) is nonnegative and the same should hold for its estimates \( \hat{\mathbf{t}}_K \). Slightly more general, consider a linear form \( \mathbf{g}' \mathbf{t} \) where \( \mathbf{g} \in \mathbb{R}^k \) is fixed. Since \( \mathbf{t} \) is restricted to vary over \( \mathbb{R}^k_+ \), the linear form \( \mathbf{g}' \mathbf{t} \) is nonnegative if and only if \( \mathbf{g} \in \mathbb{R}^k_+ \), and exactly these nonnegative linear forms call for nonnegative estimates. It is this case that now will be pursued further.
3. The best nonnegative estimator

Assume a variance component model (1.1) in which a nonnegative linear form \( q'x \) is to be estimated, \( q \in \mathbb{R}_+^k \). A widely approved estimate is C. R. Rao's (1973, pp. 303-305) MINQUE \( X' \hat{\Lambda}X \), where \( \hat{\Lambda} \) is defined to solve the following minimization program (R):

\[
\text{(R) } \quad \text{Minimize } \| \hat{\Lambda} \|_2^2 \text{ subject to } \hat{\Lambda} \in \text{Unb}(q), \hat{\Lambda} = \frac{MM'}{MM'}.
\]

\( \text{Unb}(q) \subset \text{Sym}(n) \) denotes the affine space of all unbiased quadratic estimators for \( q'x \), and the second constraint ensures translation-invariance with respect to the group \( \{ x + y + \sum b \frac{x_n}{b} \in \mathbb{R}^p \} \) of all 'mean translations', \( M \) being the projector onto the orthogonal complement of \( \text{sp}\{ x_1, \ldots, x_p \} \) in \( \mathbb{R}^n \).

MINQUE does not take into account any nonnegativity, hence the optimal solution \( \hat{\Lambda} \) of program (R) may happen to be NND; but this need not be the case.

To solve the problem of nonnegativity, one would like to minimize the norm among all unbiased invariant quadratic estimators which, in addition, are NND. To this end we define the minimization program (P):

\[
\text{(P) } \quad \text{Minimize } \| \Lambda \|_2^2 \text{ subject to } \Lambda \in \text{Unb}(q), \Lambda \in \text{NND}(n).
\]

Its optimal solution \( \Lambda^* \) we shall call the best nonnegative estimator for \( q'x \), in contrast to the best defective estimator \( \text{MINQUE} \hat{\Lambda} \).

In order to relate programs (R) and (P) to each other, and for further reference, we state the following matrix lemma whose easy proof is omitted.
Lemma 3.1. (a) For any two matrices $A$ and $B$ with the same number of rows

\[ R_A \perp R_B \quad \text{if and only if} \quad R_A' R_B = 0. \]

(b) For any symmetric matrix $A$ and for any projector $Q$

\[ R_A \subseteq R_Q \quad \text{if and only if} \quad A = Q A Q. \]

(c) For any two NND matrices $A$ and $B$

\[ R_A + R_B = R(A + B), \quad \cap A \cap B = \cap (A + B). \]

(d) For any two NND matrices $A$ and $B$

\[ R_A \perp R_B \quad \text{if and only if} \quad A \perp B. \]

(e) $A \in \text{Sym}(n)$ is NND if and only if $<A, B> \geq 0$ for all $B \in \text{NND}(n)$.

According to part (d), for $A, B \in \text{NND}(n)$, orthogonality of their ranges (as subspaces of $\mathbb{R}^n$) is equivalent to orthogonality of $A$ and $B$ (as points in the Hilbert spaces $\text{Sym}(n)$ or $\mathbb{R}^{n \times n}$). A first example that this statement may be helpful is the proof of the next lemma.

Lemma 3.2. Any unbiased NND quadratic estimator is translation-invariant.

Proof. In a linear model (1.1), the expected value of a quadratic form $Y' A Y$ is

\[ E[Y' A Y] = <A, V + E[Z]Z>' = E_k <A, V_k>' + E[Z] [X] A X = 0, \]

where $X = [x_1: \ldots : x_p] \in \mathbb{R}^{n \times p}$. Unbiasedness for $q = t$ implies $X'AX = 0$, hence $<A, X X>' = 0$, and $R_A \perp R_X$ by Lemma 3.1(d).
From this follows $\mathcal{R} \bar{A} \subset \mathcal{R} X = \mathcal{R} \bar{M}, \bar{A} = \frac{M\bar{M}}{\lambda}$, by Lemma 3.1(b), and invariance. □

Lemma 3.2 implies that (i) the best nonnegative estimator is 'best' irrespective of whether invariance is required or not, (ii) any feasible solution of (P) is feasible in (R) as well, and (iii) the optimal solutions $\hat{\bar{A}}$ of (R) and $\bar{A}^*$ of (P) satisfy $\|\hat{\bar{A}}\|^2 \leq \|\bar{A}^*\|^2$.

As the text above foreshadows, program (P) has, if at all, a unique optimal solution.

Lemma 3.3. If program (P) has at least one feasible solution,

then it has a unique optimal solution.

Proof. Existence follows from Theorem 27.3 in Rockafellar (1970), uniqueness from the norm's strict convexity. □
4. Unbiased NND quadratic estimability

In terms of the statistical model (1.1), the hypothesis of Lemma 3.3 requires \( g' \bar{t} \) to be unbiasedly NND quadratically estimable. For a single variance component \( t_\kappa \), a tractable criterion is provided by the following theorem which first appears, with a different proof, in Pukelsheim (1977).

Theorem 4.1. Assume a variance component model (1.1). A single variance component \( t_\kappa \) is unbiasedly NND quadratically estimable if and only if

\[
(4.1) \quad \eta \overset{M}{=}_{\lambda \kappa} \overset{\Sigma}{=}_{\lambda} \overset{V}{=}_{\kappa} \overset{M}{=}, \quad \not\subset \quad \eta \overset{M}{=}_{\kappa} \overset{V}{=}_{\kappa} \overset{M}{=}.
\]

Proof. Following Lemma 3.2 and its proof, \( t_\kappa \) is unbiasedly NND quadratically estimable if and only if there exists a NND matrix \( \Lambda \) with \( \langle A, \overset{M}{=}_{\lambda} \overset{M}{=} \rangle = \delta_{\kappa \lambda} \), for \( \lambda = 1, \ldots, k \). Lemma 3.1(d) yields that, firstly, \( \Omega \overset{A}{=}_{\lambda \kappa} \overset{M}{=}_{\lambda} \overset{\Sigma}{=}_{\lambda} \overset{V}{=}_{\kappa} \overset{M}{=} \eta \overset{M}{=}_{\lambda \kappa} \overset{\Sigma}{=}_{\lambda} \overset{V}{=}_{\kappa} \overset{M}{=} \), and secondly, \( \Omega \overset{A}{=} \not\subset \overset{M}{=}_{\kappa} \overset{V}{=}_{\kappa} \overset{M}{=} \) is not orthogonal to \( \Omega \overset{M}{=} \overset{V}{=} \overset{M}{=} \); this implies (4.1). If, conversely, (4.1) holds, then there is some \( \xi \in \mathbb{R}^n \) for which \( \overset{M}{=}_{\lambda \kappa} \overset{V}{=}_{\lambda} \overset{M}{=} = 0 \not\subset \overset{M}{=} \overset{V}{=} \overset{M}{=} \), and \( \Lambda := \| \overset{V}{=} \|^{-2} \overset{M}{=} \overset{V}{=} \overset{M}{=} \) is an unbiased NND quadratic estimator for \( t_\kappa \).

Setting \( \bar{V} := \overset{V}{=} \overset{M}{=} \), condition (4.1) is easily seen to be equivalent to each of the following three statements:

\[
(4.2) \quad \Omega \overset{M}{=} \overset{V}{=} \overset{M}{=} \not\subset \quad \Omega \overset{M}{=} \overset{\Sigma}{=}_{\lambda \kappa} \overset{V}{=}_{\kappa} \overset{M}{=} ;
\]
\[
(4.3) \quad \mathcal{O} = \sum_{\lambda \perp \kappa = \lambda} M \quad \dagger \quad \mathcal{O} = \sum_{\lambda \perp \kappa} M \nu M \] 

\[
(4.4) \quad \text{rank } M = \sum_{\lambda \perp \kappa = \lambda} M < \text{rank } M \nu M .
\]

The range of \( M \nu M \) may be interpreted as the 'error space' associated with the model (1.1); thus \( t_{\kappa} \) is nonnegatively estimable if and only if \( \nu_{\kappa} \) properly contributes to covering the model's error space.

LaMotte (1973) was the first to deal with nonnegative estimability, and his results easily follow from Theorem 4.1:

**Corollary 4.1.1.** (L. R. LaMotte) (a) If \( \sum_{\lambda \perp \kappa = \lambda} M \nu \) is positive-definite, then \( t_{\kappa} \) is not unbiasedly NND quadratically estimable.

(b) If \( \nu_{\kappa} \) is positive-definite, then none of the other variance components \( t_{\lambda}, \lambda \perp \kappa \), is unbiasedly NND quadratically estimable.

In a variance component model, if \( \hat{t}_{\kappa} \) is a nonnegative estimate of \( t_{\kappa} \), then \( \sum_{\kappa = \kappa} \hat{t}_{\kappa} \nu_{\kappa} \) is NND; Theorem 4.1 indicates when the converse is true, too:

**Corollary 4.1.2.** Suppose all \( k \) variance components \( t_{\kappa} \) to be unbiasedly NND quadratically estimable. Then, for any estimate \( \hat{t}_{\kappa} \in \mathbb{R}^k \), the matrix \( \sum_{\kappa = \kappa} \hat{t}_{\kappa} \nu_{\kappa} \) is NND if and only if \( \hat{t}_{\kappa} \in \mathbb{R}^k_+ \).

Obviously, all nonnegative linear forms \( q^{'t}, q \in \mathbb{R}^k_+ \), are nonnegatively estimable if and only if this is true for the \( k \) variance components \( t_{\kappa} \). The so determined subclass of variance component models is easily recognized by Theorem 4.1. It does not include analysis of variance models since the overall variance \( \sigma^2 \) leads to
$\sigma^2 \ I_{mn}$ and thus prevents all other variance components from being nonnegatively estimable.
5. Duality results

Since (P) is a convex minimization program, vital information on its optimal solution $A^*_n$ can be expected from primal-dual relationships of convex analysis.

Let $N_\theta^M$ be the orthogonal projector onto that subspace in $\text{Sym}(n)$ that is made up by all unbiased invariant quadratic estimators of zero, cf., (6.3) below.

Lemma 5.1. The following program, utilizing the best defective estimator $\hat{A}^\theta$, is a dual of (P):

\begin{align*}
(D) \quad \text{Maximize } g(B) : = & \|\hat{A}\|^2 - \langle \hat{A}, B \rangle - \frac{1}{4} \|N_M(B)\|^2 \text{ subject to } B \in \text{NND}(n).
\end{align*}

Proof. The cone $\text{NND}(n)$ is selfdual by Lemma 3.1(e), i.e., its polar cone is $-\text{NND}(n)$. Lemma 3.2 implies for program (P) that the feasible solutions do not change if we restrict $\hat{A}$ to the set $\hat{A} + \mathcal{R} N_M$ of all unbiased invariant quadratic estimators for $q^\theta$, instead of $\text{Unb}(q)$. With $\hat{A}$ varying over $\hat{A} + \mathcal{R} N_M$, the infimum of $\|\hat{A}\|^2 + \langle \hat{A}, \zeta \rangle$ is $g(-\zeta)$. Following Rockafellar (1974, p. 26), a dual of (P) is thus given by

\begin{align*}
(-D) \quad \text{Maximize } g(-\zeta) \text{ subject to } \zeta \in -\text{NND}(n).
\end{align*}

A change of variables proves the assertion. \(\square\)

We shortly digress and attempt a statistical interpretation of (D). Accepting a reduction by unbiasedness, i.e., $\hat{A} \in \text{Unb}(q)$, we would like to minimize the risk $r_{-1}(A) : = \|A\|^2$ under the restriction
\( \mathbb{A} \geq \mathbb{O} \), i.e., \( \mathbb{A} \in \text{NND}(n) \). The latter restriction has the equivalent parametrized version: \( <\mathbb{A}, \mathbb{B}> \geq 0 \) for all \( \mathbb{B} \geq \mathbb{O} \), by Lemma 3.1(e). This motivates to introduce a second risk \( r_2(\mathbb{A}, \mathbb{B}) := -<\mathbb{A}, \mathbb{B}> \) that reflects the constraint \( \mathbb{A} \geq \mathbb{O} \) in an antitonic manner: \( \mathbb{A} \geq \tilde{\mathbb{A}} \) if and only if, for all \( \mathbb{B} \geq \mathbb{O} \), \( r_2(\mathbb{A}, \mathbb{B}) \leq r_2(\tilde{\mathbb{A}}, \mathbb{B}) \). Combining both risk functions yields \( r(\mathbb{A}, \mathbb{B}) := r_1(\mathbb{A}) + r_2(\mathbb{A}, \mathbb{B}) \). Given a restriction \( \mathbb{B} \geq \mathbb{O} \), the value \( g(\mathbb{B}) = \inf_{\mathbb{A} \in \text{AC}(\tilde{\mathbb{A}}) + \mathbb{N}(\mathbb{M})} r(\mathbb{A}, \mathbb{B}) \) represents that amount of risk that has at least to be faced whatever unbiased invariant quadratic estimator \( \tilde{\mathbb{A}} \) is chosen. Hence an optimal solution \( \mathbb{B}^* \) of (D) is a least favorable restriction in the sense that it maximizes \( g(\mathbb{B}) \).

The following theorem characterizes the best nonnegative estimator \( \tilde{\mathbb{A}}^* \) in terms of an optimal solution \( \mathbb{B}^* \) of (D).

Theorem 5.1. For every \( \mathbb{B}^* \in \text{NND}(n) \) the following statements are equivalent:

(a) \( \mathbb{B}^* \) is an optimal solution of (D).

(b) \( \tilde{\mathbb{A}} + \frac{1}{2} \mathbb{M}(\mathbb{B}^*) \) is NND, and \( <\tilde{\mathbb{A}} + \frac{1}{2} \mathbb{M}(\mathbb{B}^*), \mathbb{B}^*> = 0 \).

(c) \( \tilde{\mathbb{A}}^* = \tilde{\mathbb{A}} + \frac{1}{2} \mathbb{M}(\mathbb{B}^*) \), and \( <\tilde{\mathbb{A}}^*, \mathbb{B}^*> = 0 \).

Proof. (a) \( \Rightarrow \) (b). Theorem 27.4 in Rockafellar (1970) yields that \( \mathbb{B}^* \) solves (D) if and only if the gradient of \( g \) at \( \mathbb{B}^* \) is normal to \( \text{NND}(n) \) at \( \mathbb{B}^* \), i.e., for all \( \mathbb{B} \in \text{NND}(n) \), \( <\mathbb{B} - \mathbb{B}^*, -\tilde{\mathbb{A}} - \frac{1}{2} \mathbb{M}(\mathbb{B}^*)> \leq 0 \). Hence \( <\tilde{\mathbb{A}} + \frac{1}{2} \mathbb{M}(\mathbb{B}^*), \mathbb{B} > > <\tilde{\mathbb{A}} + \frac{1}{2} \mathbb{M}(\mathbb{B}^*), \mathbb{B}^* > \), for all \( \mathbb{B} \in \text{NND}(n) \). Since zero is the only possible lower bound when a linear functional is restricted to a cone, we obtain (b).
(b) $\Rightarrow$ (c). $\hat{A} + \frac{1}{2} N_M(B^*)$ is a feasible solution of (P), its squared norm being $\|\hat{A}\|^2 + \frac{1}{4} \|N_M(B^*)\|^2$. The second assumption yields the same value for $g(B^*)$, whence follows (c). In the same manner, (c) $\Rightarrow$ (a). $\square$

Conversely, an optimal solution $\hat{B}^*_M$ of (D) is characterized by its properties with respect to (P). To this end, let $\phi(U)$ be the optimal value of the $U$-pertubated version $(P;U)$ of (P):

$$
(P;U) \quad \text{Minimize} \quad \|A\|^2 \quad \text{subject to} \quad A \in \text{Unb}(q), \ A \gg U.
$$

The optimal values of (P), (D), and (−D) are the same, by Theorem 30.4(i) in Rockafellar (1970). Theorem 16 in Rockafellar (1974) states then that $\hat{C}^*_M$ is an optimal solution of (−D) if and only if $−\hat{C}^*_M$ is a subgradient of $\phi$ at zero. Equivalently, $\hat{B}^*_M$ is an optimal solution of (D) if and only if

$$
(5.1) \quad \phi(U) - \phi(0) > \langle U, \hat{B}^*_M \rangle \quad \text{for all } \ U \in \text{Sym}(n).
$$

The pertubated program $(P;U)$ either has optimal value $+\infty$, or there exists a unique optimal solution $\hat{A}^*_M(U)$. The following lemma states that $\hat{A}^*_M(U)$ may be obtained alternatively by projecting $\hat{A}$ parallel to $\cap N_M$ onto the convex set $U + \text{NND}(n)$.

Lemma 5.2. $\hat{A}^*_M(U)$ is closest to $\hat{A}$ among all $A \in \text{Unb}(q)$, $A \gg U$.

Proof. Since $\hat{A}$ is orthogonal to $\hat{A}^*_M(U) - \hat{A}$, the assertion follows from $\|\hat{A} - \hat{A}^*_M(U)\|^2 = \|\hat{A} - \hat{A}^*_M(U)\|^2 - \|\hat{A}^*_M(U)\|^2 + \|\hat{A}\|^2 > \|\hat{A} - \hat{A}^*_M(U)\|^2$. $\square$
Since $\hat{A}^*(\underline{a}) = \underline{A}^*$, and $\phi(\underline{a}) - \phi(\underline{a}) = \|\hat{A} - \underline{A}^*(\underline{u})\|^2 - \|\hat{A} - \underline{A}^*\|^2$, property (5.1) has the appealing geometric interpretation that $B^*_\underline{u}$ has to be such that for every possible perturbation $\underline{u}$ the difference of the squared distances of $\hat{A}_{\underline{u}}$ to $\text{Unb}(\underline{g}) \cap (\underline{u} + \text{NND}(\underline{n}))$ and to $\text{Unb}(\underline{g}) \cap \text{NND}(\underline{n})$ is at least $<\underline{u}, B^*_\underline{u}>$.

As a last characterization, we present the following Fenchel-type duality theorem, cf., Theorem 31.4 in Rockafellar (1970).

**Theorem 5.2.** $A^*$ and $B^*_\underline{n}$ are optimal solutions of (P) and (D), respectively, if and only if (1) $\underline{A}^* \in \text{Unb}(\underline{g})$, (2) $\underline{A}^* \in \text{NND}(\underline{n})$, (3) $<\underline{A}^*, B^*_\underline{n}>=0$, (4) $\exists \underline{u} \in \mathbb{R}^\underline{k}$: $2\underline{A}^* - M B^*_\underline{n} M = \Sigma_{\underline{k}} M_{\underline{k}} M_{\underline{k}}$, (5) $B^*_\underline{n} \in \text{NND}(\underline{n})$.

**Proof.** If $A^*$ and $B^*_\underline{n}$ are optimal, they are in particular feasible and fulfill (1), (2), and (5). According to Theorem 5.1(c), they are also orthogonal, and $2\underline{A}^* - M B^*_\underline{n} M = 2\hat{A} - M B^*_\underline{n} M + N_{\underline{n}}(B^*_\underline{n})$ yields (4), see below (6.3). On the other hand, (1) and (4) force $\underline{A}^*$ to be equal to $\hat{A} + \frac{1}{2} N_{\underline{n}}(B^*_\underline{n})$; and optimality of $\underline{A}^*$ and $B^*_\underline{n}$ follows by again using Theorem 5.1. □

$A^*$ in its representation $\hat{A} + \frac{1}{2} N_{\underline{n}}(B^*_\underline{n})$ may be interpreted as taking the best defective estimator $\hat{A}$ and correcting its negativity by adding the term $\frac{1}{2} N_{\underline{n}}(B^*_\underline{n})$. We next provide a second approach that offers a more informative interpretation in the model (1.1).
6. **Negativity eliminating projectors**

Negativity eliminating projectors will be defined as to reduce to a model that is small enough to eliminate possible negativity, but not too small in order to maintain unbiasedness and equality with \( \hat{A}^* \). We first introduce some notations that are inspired by the dispersion-mean-correspondence, see Pukelsheim (1976).

Use the isometry \( \text{vec} \) that maps a matrix into a vector by ordering its entries lexicographically and define

\[
(6.1) \quad D := [\text{vec} V_1; \ldots; \text{vec} V_K], \quad \text{and} \quad D_M := M \otimes M \cdot D = [\text{vec} MV_1 M; \ldots; \text{vec} MV_K M].
\]

Recall that the best defective estimator \( \hat{A}_M \) obeys

\[
(6.2) \quad \hat{A}_M = \Sigma_{K} M V M \quad \text{whenever} \quad D'_M D_M u = q,
\]

see, e.g., Rao (1973, p. 304). Incidentally, the projector \( N_M \) is determined by

\[
(6.3) \quad \text{vec} \circ N_M = N_M \circ \text{vec}, \quad N_M := M \otimes M - D'_M (D'_M D_M)^{-1} D'_M.
\]

**Definition.** Consider a variance component model (1.1) and a projector \( Q \) in \( \mathbb{R}^n \) whose range is contained in the range of \( M, \) i.e., \( Q \preceq M. \)

(a) The model \( OY \sim (Q; \Sigma_K QVQ) \) that is generated by \( OY \) is called the \( Q \)-reduced model of the original model (1.1).

(b) \( Q \) is called a **negativity eliminating projector for estimating** \( q'_M, \) \( q \in \mathbb{R}^k, \) if the best defective estimator for \( q'_M \) in the \( Q \)-reduced model is equal to the best nonnegative estimator in the original model, i.e., if
(6.4) \[ A^*_\alpha = \Sigma u_{\kappa} Q_{V} Q \] whenever \[ D_{\alpha}^{-1} D_{\alpha} u = q \; ; \; D_{\alpha} : = Q \otimes Q \cdot D \].

The existence of negativity eliminating projectors is guaranteed by the following theorem.

**Theorem 6.1.** Consider a variance component model (1.1) and a nonnegative linear form \( a^t \), \( a \in \mathbb{R}^k \), that is unbiasedly NND quadratically estimable. Let \( Q_{\alpha}^* \) be the projector onto the range of \( A^*_\alpha \), and \( Q_{\alpha}^* \) be the projector onto the nullspace of \( XX' + B^* \), where \( B^* \) is an optimal solution of (D).

Then \( Q_{\alpha}^* \leq Q_{\alpha}^* \), and every projector \( Q \) in the interval \([Q_{\alpha}^*, Q_{\alpha}^*]\) is a negativity eliminating projector for estimating \( a^t \).

**Proof.** Note that for projectors the orderings \( \mathcal{R} Q_{\alpha}^* \subset \mathcal{R} Q_{\alpha}^* \), and \( Q_{\alpha}^* \leq Q_{\alpha}^* \) coincide. The following argument extensively uses Lemmata 3.1 and 3.2, and properties (1)-(5) in Theorem 5.2. From (2), (3), and (5) follows \( \mathcal{R} A^*_\alpha \subset \mathcal{R} B^*_\alpha = \mathcal{R} B^*_\alpha \), hence \( \mathcal{R} Q_{\alpha}^* = \mathcal{R} A^*_\alpha \subset \mathcal{R} B^*_\alpha \) + \( XX' = \mathcal{R} Q_{\alpha}^* \). Now fix a projector \( Q \in [Q_{\alpha}^*, Q_{\alpha}^*] \). Then \( Q_{\alpha}^* \leq Q \) implies \( \mathcal{R} A^*_\alpha \subset \mathcal{R} Q \), and \( A^*_\alpha = Q_{\alpha}^* Q \). And \( Q \leq Q_{\alpha}^* \) implies \( \mathcal{R} Q \subset \mathcal{R} B^* + XX' = \mathcal{R} B^* \cap \mathcal{R} M \); hence, firstly, \( \mathcal{R} Q \subset \mathcal{R} M \), and \( Q = Q \cdot M \), and, secondly, \( \mathcal{R} Q \subset \mathcal{R} M \), and \( B^* Q = Q \). Using (4) we obtain \( A^*_\alpha = 2MB^*_M + \Sigma u_{\kappa} M V_{\kappa} M \) = \( QA^*_\alpha Q = 0 + \Sigma u_{\kappa} Q V_{\kappa} Q \), and (1) determines \( u \) to be an arbitrary solution of \( D_{\alpha}^{-1} D_{\alpha} u = q \). □

Obviously \( Q_{\alpha}^* \) is the smallest negativity eliminating projector and does exist, along with \( A^*_\alpha \). It informs about the maximal model reduction, but cannot, in general, be expected to contribute to
computing $A^*$. To this extent, $Q^*$ seems to be more useful, its existence depending on the existence of a $B^*_m$. Whether program (D) does have an optimal solution, and whether $Q^*$ is the largest negativity eliminating projector remain to be open questions.
7. Dummy example

The dual program (D) is to some extent easier to handle than the primal program (P), hence we append some results that indicate where to drive at in order to determine $B^*$. 

Lemma 7.1. If (P) has a feasible solution, then $\langle \hat{A}, B \rangle < 0$ implies $\|N^s(M)(B)\|^2 > 0$, for every $B \in NND(n)$. 

Proof. If $\|N^s(M)(B)\|^2 = 0$, then as $\varepsilon$ tends to $\infty$ so does $g(\varepsilon B) = \|\hat{A}\|^2 + \varepsilon|\langle \hat{A}, B \rangle|$, and the optimal value of (D) is $\infty$. This contradicts the fact that the existing feasible solution of (P) induces an upper bound for the optimal value of (D). $\square$

The following lemma will easily be proved from elementary calculus.

Lemma 7.2. Let (P) have a feasible solution, and assume $B \in NND(n)$ to fulfill $\langle \hat{A}, B \rangle < 0$. Then the maximum of the dual objective function $g$ along the ray $\{\varepsilon B | \varepsilon > 0\}$ is attained at $-2\langle \hat{A}, B \rangle \|N^s(M)(B)\|^{-2} B$ and equal to $\|\hat{A}\|^2 + \langle \hat{A}, B \rangle^2 \|N^s(M)(B)\|^{-2}$.

We wish to conveniently transcribe the condition $\langle \hat{A}, B \rangle < 0$. To this end we choose the spectral decomposition $\Sigma_{\lambda \in \text{spec}(A)} \lambda E(\lambda)$ of a symmetric $A$, where $E(\lambda)$ is the projector onto the eigenspace associated with $\lambda$, and define the positive part $A^+$ and negative part $A^-$ of $A$ by

$$A^+ := \Sigma_{\lambda \in \text{spec}(A)} \lambda E(\lambda), \quad A^- := \Sigma_{\lambda \in \text{spec}(A)} \lambda E(\lambda).$$

21
cf., Riesz & Nagy (1955, p. 277). As usual, \( \lambda_+ = \max\{0, \lambda\} \), and \( \lambda_- = \max\{0, -\lambda\} \). The following properties are obvious:

(7.2) \( \hat{A} = \hat{A}_+ - \hat{A}_- ; \quad \hat{A}_+, \hat{A}_- \in \text{NN}(n) \);

(7.3) \( \hat{A}_+ \perp \hat{A}_- ; \quad \hat{A}_+ \hat{A}_- = 0 \);

(7.4) \( ||\hat{A}_-||^2 = ||\hat{A} - \hat{A}_+||^2 = \inf\{||A - B||^2/B \in \text{NN}(n)\} \);

(7.5) \( \hat{A} \notin \text{NN}(n) \Rightarrow \hat{A}_- \neq 0 \Rightarrow \exists B \in \text{NN}(n): \langle \hat{A}, B \rangle < 0 \).

Pursuing the solely interesting case when \( \hat{A} \) is not \( \text{NN} \), Lemma 7.2 applies to \( \hat{B} = \hat{A}_- \), and so the optimal value of (D) is at least

(7.6) \( g(2s^*_{\hat{A}_-}) = ||\hat{A}||^2 + s^*||\hat{A}_-||^2, \quad s^* = ||\hat{A}_-||^2/||N(\hat{A}_-)||^2 \).

A case that admits explicit solutions is provided by the dummy example \( \mathcal{Y} \sim (0; \sigma_1^2 \text{Diag}[I_{a_a}; I_{a_b}; I_{a_c}]) + \sigma_2^2 \text{Diag}[0; I_{b_a}; I_{c_a}] \); here \( B^* = 2s^*_{\hat{A}_-}, \quad A^* = \hat{A} + s^* N(\hat{A}_-), \) and estimating \( \sigma_1^2 \) one obtains \( Q^*_{a_a} = \text{Diag}[I_{a_a}; 0; 0], \) and \( Q^*_{a_b} = \text{Diag}[I_{a_b}; I_{b_a}; 0]. \) Less tractable, but more interesting is the model with heteroscedastic variances, see the companion paper [16].

When \( \hat{A}_- \) determines the optimal solutions, we are in a situation that, although the MINQUE \( \hat{A} \) may be defective, it contains all vital information on \( \hat{A}^* \), and leads to it in an intuitively rather appealing manner. In this case the general theory as developed above may be considerably shortened, as we elaborate in [17].
Acknowledgement

The author would like to sincerely thank Professor H. Witting for helpful discussions and encouraging guidance throughout the course of this work.

References


