Math. Operationsforsch. Statist., Ser. Statistics, Vol. 12 (1981) No. 2, 271–286

Linear Models and Convex Geometry: Aspects of Non-Negative Variance Estimation[1]

FRIEDRICH PUKELSHEIM[2]

Summary. Geometric aspects of linear model theory are surveyed as they bear on mean estimation, or variance covariance component estimation. It is outlined that notions associated with linear subspaces suffice for those of the customary procedures which are solely based on linear, or multilinear algebra. While conceptually simple, these methods do not always respect convexity constraints which naturally arise in variance component estimation.

Previous work on negative estimates of variance is reviewed, followed by a more detailed study of the non-negative definite analogue of the MINQUE procedure. Some characterizations are proposed which are based on convex duality theory. Optimal estimators now correspond to (non-linear) projections onto closed convex cones, they are easy to visualize, but hard to compute. No ultimate solution can be recommended, instead the paper concludes with a list of open problems.

Key words: Mean estimation, variance covariance component estimation, analysis of variance, negative estimates of variance, projections on convex cones, convex duality, non-negative MINQUE.

1. Introduction

A. C. AITKEN (1935), A. N. KOLMOGOROV (1946), H. SCHEFFÉ (1959) and many others since have emphasized ease and appeal of matrix notation and geometric visualization in linear model theory. Familiar pairs of notions derived from statistical concepts on the one hand and geometric concepts on the other comprise

- estimation space and range of the design matrix,
- least squares estimation and projection,
- degrees of freedom and dimension,
- dispersion matrix and inner product.

Section 2 (*Mean estimation*) highlights further geometrical aspects, as far as linear estimation of the mean is concerned.

Variance covariance component estimation originally developed as a field in its own right and separate from mean estimation. This separation was overcome

---

1 Extended version of a paper presented at the VI International Conference on Mathematical Statistics, Wisła, Poland, December 7–13, 1978.

2 Institut für Mathematische Stochastik der Universität, Hermann-Herder-Straße 10, D-7800 Freiburg im Breisgau, Federal Republic of Germany.

when J. SEELY (1970) showed that both mean estimation and dispersion estimation pose the same theoretical problem and only differ in their interpretation as parts of an applied setting. The common denominator of this unified theory is regression analysis in finite dimensional linear spaces with inner product; points may then be realized as usual column vectors for mean estimation, or as symmetric matrices for dispersion estimation. In this way pairs of concepts evolve referring to dispersion estimation on the one hand and mean estimation on the other:

- C. R. RAO's (1973, p. 303) MINQUE and least squares estimates,
- MIVQUE and BLUE,
- J. SEELY's (1971) quadratic subspaces of symmetric matrices and G. ZYSKIND's (1967) invariance of the estimation space.

In view of this correspondence the geometry of most of the customary theory of quadratic estimation of dispersion components coincides with the geometry of linear estimation of the mean, when the latter is reinterpreted in the space of symmetric matrices.

However, integration of variance covariance component estimation into (multi-) linear algebra encounters natural limits: There is no possibility to properly incorporate restrictions such as non-negativity of variance estimates, or non-negative definiteness of estimates of the dispersion matrix. Section 3 (*Negative estimates of variance*) is devoted to a short review of previous work concerned with this negativity defect.

An ultimate solution to non-negative variance estimation has not yet been obtained, and hence it seems reasonable to restrict further investigations to a conceptually simple and thus transparent procedure such as C. R. RAO's MINQUE. Nonnegative MINQUE and its characterization by convex programming methods is the prime interest of J. HARTUNG (1981), and F. PUKELSHEIM (1977a, 1977b, 1978a, 1978b), and the present paper is intended to survey these results and complement their analytical calculus by its geometric counterpart. Section 4 (*Minimum bias and non-negative definiteness*) shows that the minimum bias requirement of J. HARTUNG (1981) corresponds to a projection argument in the parameter space, and does not really lead beyond unbiased non-negative definite estimation as discussed in F. PUKELSHEIM (1981). The main results and various characterizations of the non-negative analogue of MINQUE are discussed and illustrated in Section 5 (*Non-negative definite* MINQUE). Roughly one may say that with a restriction to an appropriate sub-model (and the associated loss in degrees of freedom) the customary MINQUE procedure coincides with its non-negative analogue, but explicit determination of these sub-models may not be easy. The final Section 6 (*Open problems*) concludes with some questions that are raised by this discussion.

General references pertaining to the subject are the textbooks C. R. RAO (1973) and S. R. SEARLE (1971), the proceedings volumes J. N. SRIVASTAVA (1975) and L. D. VAN VLECK & S. R. SEARLE (1979), the survey paper J. KLEFFE (1977), and the bibliography H. SAHAI (1979). More specific references are mentioned throughout the paper.

As in F. PUKELSHEIM (1981) a linear model is represented by its moments, according to

$$Y \sim \left(X\beta; \sum_{j=1}^{l} \tau_j V_j\right),$$

with the tacit understanding that

- Y is a random R^n-vector of observations,
- the mean vector of Y is $X\beta$, where the $n \times k$ design matrix X is given and fixed while $\beta \in \mathrm{R}^k$ is the unknown mean parameter,
- and the dispersion matrix of Y is $\sum \tau_j V_j$, where the l real symmetric $n \times n$ matrices V_j are given and fixed while $\tau = (\tau_1, \ldots, \tau_l)'$ is the unknown dispersion parameter.

The natural parameter set for τ is the set $\bar{G}$ of those R^l-vectors $t = (t_1, \ldots, t_l)'$ for which $\sum t_j V_j$ is non-negative definite. Depending on the particular structure of V_j one may interprete τ_j as a variance component or as a covariance component.

Because of the correspondence between mean estimation and dispersion estimation as mentioned above we will have frequent opportunity to refer to the geometry of the space Sym (n) of real symmetric $n \times n$ matrices. Its Euclidean inner product is $\langle A, B \rangle = \text{trace } AB$, with associated norm $\|A\| = \sqrt{\text{trace } A^2}$. The subset NND$(n)$ of non-negative definite matrices in Sym(n) forms a closed convex cone whose interior consists of the set of all positive definite matrices. A prime denotes transposition.

2. Mean estimation

In a model $Y \sim (X\beta; \sigma^2 V)$ familiar linear estimators for the mean vector $\mu = X\beta$ are

- the least squares estimate PY, $P = XX^+ = X(X'X)^- X'$, with residual vector MY, $M = I_n - P$,
- the covariance adjusted estimate CY, $C = P - PVM(MVM)^+ M$,
- and the Aitken estimate AY, $A = X(X'V^+X)^+ X'V^+$.

The matrix M also determines the set of all linear estimates which are unbiased for 0, namely, $z'Y$ has expectation 0 if and only if $z = Mz$. Thus the LEHMANN-SCHEFFÉ criterion reads: a linear estimate LY is a BLUE for μ if and only if $LVM = 0$.

A straightforward calculation shows that the covariance adjusted estimate CY is a BLUE for μ. Its second term $PVM(MVM)^+MY$ is most easily interpreted when X is taken to be in canonical form $\begin{bmatrix} I_k \\ 0 \end{bmatrix}$, and when $V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$ and $Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$ are partioned accordingly, since in this case CY equals $Y_1 - V_{12} V_{22}^+ Y_2$, see the original paper by C. R. RAO (1967), or A. ALBERT (1973), G. ZYSKIND (1975). Moreover, the second term vanishes if and only if $PVM = 0$, i.e., precisely when the least squares estimate is BLUE. The condition $PVM = 0$ means geometrically that the range of VM be contained in the range of M, so that the dispersion matrix V transforms the error space range M into itself, and the residual vector MY does not contribute any information concerning the estimation space range X. As an example, consider the transformed model $V^{\frac{1}{2}+} Y \sim (V^{\frac{1}{2}+} X\beta; \sigma^2 V V^+)$, here the least squares estimate is a BLUE for $V^{\frac{1}{2}+} \mu$.

Intuitively one would tend to believe that the transformation $y \to V^{\frac{1}{2}+} y$ does not leave out essential aspects of the original model provided all of the estimation space range X is accounted for under the transformation $V^{\frac{1}{2}+}$. In fact, if the range of V covers the range of X, then in the transformed model the least squares estimate for μ is $X(X'V^+X)^+X'V^+Y$, and thus coincides with the Aitken estimate of the original model where it is a BLUE for μ. More precisely, the two matrices A and C are equal if and only if the range of V covers the range of X. Geometrically this means that all of the estimation space is subject to random variation, algebraically an equivalent formulation is $VV^-X = X$. That the Aitken estimate needs an assumption of this kind is easy to see: Suppose there exists some $\beta_0 \in \mathrm{R}^k$ such that $a_0 = (I_n - VV^+) X\beta_0 \neq 0$, then although the normal distribution $\mathrm{N}_n(X\beta_0; V)$ is supported by the affine space $a_0 +$ range V the Aitken estimate will project the affine shift a_0 into 0. (The covariance adjusted estimate will do so if and only if range $V \supset$ range X). Further representations of the Aitken estimate are discussed by G. ZYSKIND (1975), J. K. BAKSALARY & R. KALA (1978), C. R. RAO (1978).

A more general model $Y \sim (X\beta; \sum \tau_j V_j)$ also calls for estimates of the dispersion components τ_j, or linear combinations thereof. In a coordinate-free framework J. SEELY (1970) showed the equivalence of quadratic estimation of a combination $q'\tau$, and linear estimation of a combination $c'\beta$. Alternatively, one may choose some fixed isomorphism to translate one problem into the other, see F. PUKELSHEIM (1974, 1976), K. G. BROWN (1977, 1978). However, these purely linear and multilinear methods do not respect the non-negativity constraint which is inherent in dispersion parameter estimation.

3. Negative estimates of variance

It is well known that analysis of variance estimates of variance components may yield negative estimates, see R. L. ANDERSON (1965), S. R. SEARLE (1971, pp. 406—408). Two kinds of reaction suggest themselves: either change the model assumptions, or change the class of estimators. In the first case, a model assumption of a finite underlying population does, in fact, lead to some explanation, see J. A. NELDER (1954), R. B. MCHUGH & P. W. MIELKE Jr. (1968), H. SAHAI (1974).

Changes in the class of estimators have been discussed by many authors. H. O. HARTLEY & J. N. K. RAO (1967), W. J. HEMMERLE & H. O. HARTLEY (1973) investigate algorithms for maximizing the normal likelihood under a non-negativity constraint, see also D. A. HARVILLE (1977). L. H. HERBACH (1959), H. DRY-

GAS (1972), H. DRYGAS & G. HUPET (1979) fairly completely study non-negativity of the estimators they discuss, but these results are restricted to their particular models. The estimator of W. T. FEDERER (1968) is non-negative but lacks other desirable properties. Unbiasedness and non-negative definiteness are frequently incompatible, see L. R. LAMOTTE (1973), F. PUKELSHEIM (1981). A collection of biased estimators may be found in W. A. FULLER & J. N. K. RAO (1978), S. D. HORN & R. A. HORN (1975), J. N. K. RAO (1973), P. S. R. S. RAO & Y. P. CHAUBEY (1978).

Negative variance estimates invalidate the estimation procedure, but do not necessarily contradict the model or question the data. L. R. VERDOOREN (1980) vividly demonstrates that analysis of variance estimates may be negative with positive probability, and provides tables of these probabilities for various choices of parameters for the 1-way classification, random effect model.

The approach to non-negative variance estimation presented below is based on convex analysis and convex programming. Programming methods have been applied to linear model theory by G. G. JUDGE & T. TAKAYAMA (1966), D. J. HUDSON (1969), and C. K. LIEW (1976) for constraint mean estimation, and D. K. RICH & K. G. BROWN (1979) for variance component estimation. Those authors restrict their study to computational algorithms and numerical aspects. Quite differently programming methods may also be employed to characterize, or derive properties of interest, of optimal non-negative variance estimates. The motivation for this approach originates from the theory of tests where best tests appear as optimal solutions of linear programs, as described in H. WITTING (1966, pp. 69–73), or O. KRAFFT (1970).

4. Minimum bias and non-negative definiteness

Consider a model $Y \sim (X\beta; \sum \tau_j V_j)$ in which $q'\tau$ is to be estimated, $q \in \mathrm{R}^l$. Define Unb (q) to be the set of all $A \in$ Sym (n) such that $Y'AY$ is unbiased for $q'\tau$. The non-negative definite analogue of the MINQUE procedure leads to the following problem:

(P) Minimize $\|A\|^2$ subject to $A \in \mathrm{Unb}\ (q) \cap \mathrm{NND}(n)$.

If $q'\tau$ is *non-negatively estimable* , i.e., if there exists a matrix $A \in \mathrm{NND}(n)$ such that $Y'AY$ is unbiased for $q'\tau$, then problem (P) has a unique solution, and this solution will be denoted by A^*.

Every non-negative definite estimate $Y'AY$ which is unbiased for $q'\tau$ is location-invariant (M. ATIQULLAH 1962), i.e., the matrix A satisfies $A = MAM$, where $M = I_n - XX^+$ as above. This connects (P) with the MINQUE problem:

Minimize $\|A\|^2$ subject to $A \in \mathrm{Unb}\ (q)$, $A = MAM$.

The unique optimal solution of the latter problem will be denoted by $\hat{A}$, determining the estimate "MINQUE given I_n" (J. KLEFFE 1977) and leading to minimum variance estimates when the true dispersion matrix is I_n.

Let $\overline{G_M}$ be the set of all those $\mathbb{R}^l$-vectors $t=(t_1, \ldots, t_l)'$ such that $\sum t_j M V_j M$ is non-negative definite, and let $\mathfrak{Q}$ be the set of those coefficient vectors $q \in \mathbb{R}^l$ such that $q'\tau$ is non-negatively estimable. For the remainder of this paper we adopt the following

Assumption. *There is at least one choice of t such that* $\sum t_j V_j$ *is positive definite.*

Then it is shown in ([37], Theorem 1) that $\mathfrak{Q}$ and $\overline{G_M}$ are closed convex cones which are dual to each other, see Figure 1. The following example demonstrates that some assumption is needed in order to ascertain closedness of $\mathfrak{Q}$.

Example 1 (R. BELLMAN & K. FAN, 1963). In the model $Y \sim \left(0;\, \tau_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \tau_2 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right)$ the convex cone $\mathfrak{Q}$ is the union of the open right half-plane and the origin, hence $\mathfrak{Q}$ is not closed.

For a form $p'\tau$ which is *not* non-negatively estimable, $p \in \mathbb{R}^l$, J. HARTUNG (1981) proposes a minimum bias procedure: First find all location-invariant non-

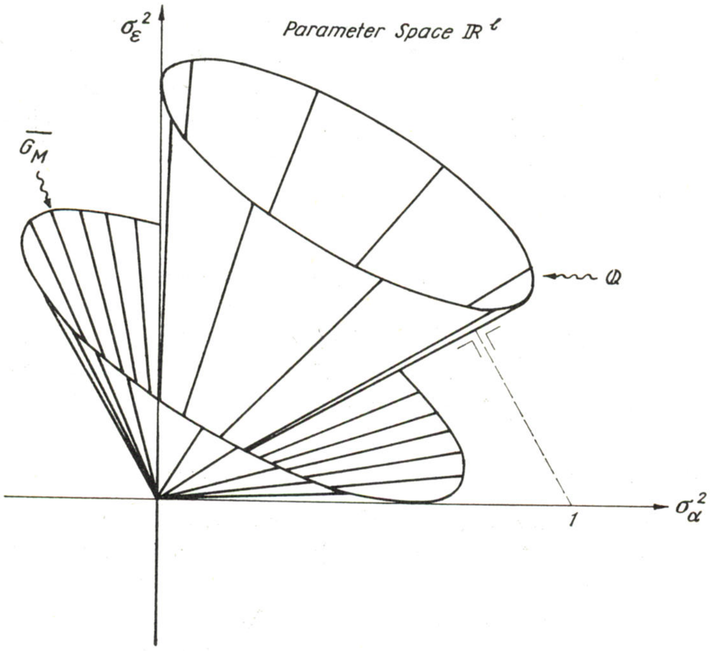

Fig. 1. The set of all coefficient vectors $q \in \mathbb{R}^l$ for which $q'\tau$ is non-negatively estimable forms a closed convex cone $\mathfrak{Q}$. Its dual is the cone $\overline{G_M}$ of those parameter values $t=(t_1, \ldots, t_e)'$ for which $\sum t_j M V_j M$ is non-negative definite. Unbiased non-negative definite estimation for $q'\tau$ coincides with minimum bias non-negative definite estimation for all those forms $p'\tau$ for which q is the projection of p onto $\mathfrak{Q}$. In the 2-dimensional version of the figure $p=\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is projected onto $\mathfrak{Q}$ as required in Example 2.

negative definite estimates $Y'AY$ which minimize the bias function $\sum (p_j - \text{trace } AMV_jM)^2$ over NND (n), and secondly find the minimum norm element among all these solutions. However, the minimum of $\sum (p_j - \text{trace } AMV_jM)^2$ subject to $A \in \text{NND}(n)$ is the same as the minimum of $\|p-q\|^2$ subject to $q \in \mathfrak{Q}$, since the sets $\mathfrak{Q}$ and $\{(\text{trace } AMV_1M, \ldots, \text{trace } AMV_lM)' \mid A \in \text{NND}(n)\}$ are equal. Hence we have:

Theorem 1. *Minimum bias non-negative definite estimation for $p'\tau$ coincides with unbiased non-negative definite estimation for $q'\tau$, where q is the projection of p onto the closed convex cone $\mathfrak{Q}$.*

This is the exact analogue of the interplay between unbiasedness and minimum biasedness as in the unconstrained case, see J. Kleffe (1977, p. 217). Note that while in mean estimation the question of identifiability and estimability lead to the same answer, this is no longer the case in dispersion component estimation, see ([36], Theorem 1). The following example illustrates the use of Theorem 1, for an alternative discussion see J. Hartung (1981).

Example 2 (J. Hartung 1981). Consider a 1-way classification, random model, with balanced data $Y \sim (1_n\mu;\ \sigma_\alpha^2 I_a \otimes J_n + \sigma_\varepsilon^2 I_a \otimes I_n)$, where $a,\ n > 1$, and 1_n is the $\mathbb{R}^n$-vector consisting of one's, and $J_n = 1_n 1_n'$. Here $\mathfrak{Q}$ contains those vectors $q = \begin{pmatrix} q_\alpha \\ q_\varepsilon \end{pmatrix}$ for which $q_\varepsilon \geqq q_\alpha / n \geqq 0$, see L. R. LaMotte (1973). The variance component σ_α^2 is not non-negatively estimable. The projection of its associated coefficient vector $p = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ onto $\mathfrak{Q}$ is $q = n^2 (n^2+1)^{-1} \begin{pmatrix} 1 \\ 1/n \end{pmatrix}$. Thus minimum bias non-negative definite estimation for σ_α^2 is the same as unbiased non-negative definite estimation for $n^2 (n^2+1)^{-1} \sigma_\alpha^2 + n (n^2+1)^{-1} \sigma_\varepsilon^2$ for which the optimal solution of (P) is the analysis of variance estimate, see ([37], Theorem 2).

5. Non-negative definite MINQUE

Suppose the form $q'\tau$ is non-negatively estimable.

Both A^* and $\hat{A}$ are defined as smallest norm elements and thus are closest to 0 in their respective sets. Another description of this type assures that A^* is closest to $\hat{A}$ among all $A \in \text{Unb}(q) \cap \text{NND}(n)$, this following from the fact $\|\hat{A} - A\|^2 = \|\hat{A} - A^*\|^2 + \|A\|^2 - \|A^*\|^2 \geqq \|\hat{A} - A^*\|^2$. In other words: The optimal solution A^* of problem (P) is the projection of the MINQUE matrix $\hat{A}$ onto the closed convex cone NND (n) along the affine subspace Unb (q) ([33], Lemma 5.2). Although geometrically appealing, this does not, in general, facilitate computation of A^*.

A second characterization is due to A. Kozek (1980). A major theorem of convex analysis (Rockafellar 1970, Theorem 27.4) states that A^* is the optimal solution of (P) if and only if $-2A^*$ is normal to Unb $(q) \cap$ NND (n) at A^*. This means that for every matrix N such that $Y'NY$ is a location-invariant estimate which is unbiased for 0 and such that A^*+N is non-negative definite one has $\langle A^*, N\rangle \geqq 0$, i. e., that $Y'A^*Y$ and $Y'NY$ have non-negative correlation under true dispersion matrix I_n. Applicability of this result again amounts to effectively handle the set Unb $(q) \cap$ NND (n) which is hard to achieve.

A third approach is proposed by J. HARTUNG (1981): He defines a sequence of functions f_m which attain their minimum over the unconstrained space Sym (n) at a unique matrix A_m such that the sequence A_m so obtained converges to the optimal solution A^* of problem (P). Here the auxiliary tool consists of an infinity of additional problems without constraints, whereas in duality theory it is one additional problem with constraints.

The auxiliary problem of duality theory is called *the dual of* (P), and standard theory determines it to be

(D) Maximize $g(B)$ subject to $B \in$ NND (n) ,

where $g(B) = \min \{\|A\|^2 - \langle A, B\rangle \mid A \in \text{Unb}(q),\ A = MAM\}$ ([32], [33], [35]). For an explicit representation of g define the operator $N_M(A) = MAM - \sum (S_M^+)_{ij} \times \times \langle A, MV_iM\rangle MV_jM$, with $S_M = ((\text{trace } MV_iMV_j))$. Then $Y'AY$ is a location-invariant quadratic estimator which is unbiased for 0 if and only if $A = N_M(A)$, and with an arbitrary unbiased location-invariant quadratic estimator A_q for $q'\tau$, or the particular choice $\hat{A}$, one obtains

$$g(B) = \|A_q\|^2 - \langle A_q, B\rangle - \left\| N_M\left(A_q - \frac{1}{2}B\right)\right\|^2$$
$$= \|\hat{A}\|^2 - \langle \hat{A}, B\rangle - (1/4)\,\|N_M(B)\|^2 .$$

It is immediate that $A \in$ Unb $(q) \cap$ NND (n) and $B \in$ NND (n) satisfy the *weak duality* relation $\|A\|^2 \geqq g(B)$. Hence a sufficient condition for optimality of A^* in (P) is that there exists a matrix $B^* \in$ NND (n) such that $\|A^*\|^2 = g(B^*)$. In fact, more can be said on the interplay between A^* and such B^*.

Theorem 2. *For every $B^* \in$ NND (n) the following statements are equivalent:*
(a) *B^* is an optimal solution of the dual problem* (D).
(b) $\hat{A} + \frac{1}{2} N_M(B^*) \in$ NND (n), *and* $\left\langle \hat{A} + \frac{1}{2} N_M(B^*), B^*\right\rangle = 0$.
(c) $A^* = \hat{A} + \frac{1}{2} N_M(B^*)$, *and* $\langle A^*, B^*\rangle = 0$.

Proof ([33], Theorem 5.1). Employ the subgradient theorem in ROCKAFELLAR (1970, p. 270).

It may happen, however, that the dual problem (D) does not admit an optimal solution B^*. Note in the following example that $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ is a direction of recession common to g and NND (2), compare Theorem 27.3 in ROCKAFELLAR (1970).

Example 3. Consider a model $Y \sim \left(0; \tau_1 \begin{pmatrix}1 & 0\\0 & 0\end{pmatrix} + \tau_2 \begin{pmatrix}1 & 1\\1 & 1\end{pmatrix}\right)$. Here $Y'AY$ is unbiased for τ_2 if and only if $A = \begin{pmatrix}0 & -\alpha\\ -\alpha & 2\alpha+1\end{pmatrix}$ for some real α. Its unique non-negative definite member is $A^* = \begin{pmatrix}0 & 0\\0 & 1\end{pmatrix}$. For $B \in \mathrm{NND}\,(2)$ the condition $\langle A^*, B\rangle = 0$ necessitates $B = \begin{pmatrix}\beta & 0\\0 & 0\end{pmatrix}$. On the other hand define $B_k = \begin{pmatrix}4k & -2\\ -2 & 1/k\end{pmatrix}$. Then $g\begin{pmatrix}\beta & 0\\0 & 0\end{pmatrix} = 1/3 < \lim g(B_k) = 1 = \|A^*\|^2$, whence B_k is an optimizing sequence for (D), but no optimal B^* exists.

Although (D) may fail to have optimal solutions there always exists an optimizing sequence B_k for (D) such that $\|A^*\|^2 = \lim g(B_k)$, due to the following Fenchel-type duality theorem. Also a sufficient condition for the existence of optimal B^* may be given.

Theorem 3. *One has*

$$\min_{A \in \mathrm{Unb}(q) \cap \mathrm{NND}(n)} \|A\|^2 = \sup_{B \in \mathrm{NND}(n)} g(B) .$$

For the right supremum to be attained at some $B^* \in \mathrm{NND}\,(n)$ a sufficient condition is that there exists an unbiased non-negative definite estimate $Y'AY$ for $q'\tau$ such that A has the same rank as M. In general, the matrices A^* and B^* are optimal solutions of problems (P) *and* (D), *respectively, if and only if*

(1) $A^* \in \mathrm{Unb}\,(q) \cap \mathrm{NND}\,(n)$,

(2) $B^* \in \mathrm{NND}\,(n)$,

(3) $\langle A^*, B^*\rangle = 0$,

(4) $2A^* - MB^*M = \sum u_j M V_j M,$ *for some* $u_1, \ldots, u_l \in \mathrm{R}$.

Proof ([33], Theorem 5.2). Apply Theorem 31.4 in Rockafellar (1970). For the existence statement use a decomposition $M = UU'$ with $U'U = I_r$, $r = \mathrm{rank}\ M$, to rewrite problems (P) and (D) with the variables $C = U'AU$, and $D = U'BU$. The hypothesis of the theorem ensures that there exists a positive definite matrix C such that $(UY)'C(UY)$ is unbiased for $q'\tau$, and Theorems 31.4 (a) or 27.3 in Rockafellar (1970) yield the assertion.

Condition (4) closely relates to the unconstrained MINQUE procedure. To this end a projection matrix $Q = Q^2 = Q'$ will be called a *negativity eliminating projector for estimating* $q'\tau$ if the range of Q is contained in the range of M and the optimal solution A^* of problem (P) appears as the MINQUE matrix for $q'\tau$ in the Q-reduced model $QY \sim (0; \sum \tau_j Q V_j Q)$, i.e., $A^* = \sum u_j Q V_j Q$ with coefficients $u_1, \ldots, u_l$ determined from unbiasedness for $q'\tau$.

Theorem 4. *Suppose $B^* \in \mathrm{NND}\,(n)$ is an optimal solution of problem* (D). *Define Q_* to be the projector onto the range of A^*, and Q^* to be the projector onto the nullspace of $XX' + B^*$. Then the range of Q_* is contained in the range of Q^*, and every pro-*

jector Q whose range lies between range Q_* *and* range Q^* *is a negativity eliminating projector for estimating $q'\tau$.*

Proof ([33], Theorem 6.1). Utilizing $A^*=QA^*Q$, $MQ=Q$, $B^*Q=0$, the assertion follows from condition (4).

This points to an interesting connection between nonnegative and unconstrained variance estimation: First reduce by a negativity eliminating projector — which results in a loss of degrees of freedom —, then employ the customary methods. Another reduction of this type is given in Lemma 3 in ([37]). Geometrically a negativity eliminating projector splits the sample space $\mathbb{R}^n$ into three mutually orthogonal subspaces, see Fig. 2:

- the mean estimation space range X,
- the space range Q for estimating $q'\tau$,
- and the remaining waste space range $(M-Q)$.

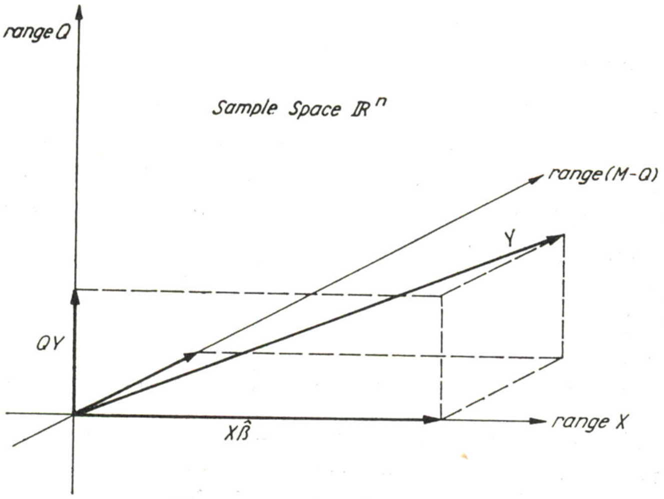

Fig. 2. Unbiased non-negative definite estimation of a form $q'\tau$ splits the sample space into three mutually orthogonal subspaces:
(x) the mean estimation space range X,
(y) the variance estimation space range Q where Q is a negativity eliminating projector for estimating $q'\tau$, and
(z) the waste space range $(M-Q)$.
The last two components may vary with different negativity eliminating projectors Q, for details see Theorem 4.

Even when the unconstrained procedure is carried through in the original model the resulting MINQUE matrix $\hat{A}$ contains interesting information on problems (P) and (D).

Theorem 5. *Suppose* $B\in\mathrm{NND}\,(n)$. *If* $\langle\hat{A}, B\rangle\geqq 0$, *then* $g(B)\leqq g(0)$. *If* $\langle\hat{A}, B\rangle<0$, *then* $N_M(B)\neq 0$ *and the maximum of g along the ray* $\{\alpha B \mid \alpha>0\}$ *is*

$$g\left(-2\,\langle\hat{A}, B\rangle\,\|N_M(B)\|^{-2}\,B\right)=\|\hat{A}\|^2+\langle\hat{A}, B\rangle^2\,\|N_M(B)\|^{-2},$$

in fact, if the matrix $\hat{A} - \langle \hat{A}, B \rangle \|N_M(B)\|^{-2} N_M(B)$ *is nonnegative definite then it actually coincides with* A^*.

Proof ([33], Lemmas 7.1, 7.2). Use Theorem 2(b).

Theorem 5 has to do with the following iterative procedure: (i) project $\hat{A}$ onto NND (n), and (ii) force this projection into the direction prescribed by Unb (q). (i) For any $A \in \text{Sym}\,(n)$ the projection onto NND (n) is called its *positive part* A_+, it is obtained from a spectral decomposition of A by deleting all negative eigenvalues and their associated projectors. The difference $A_+ - A$ then gives the *negative part* A_- by which A deviates from NND (n). Unless $\hat{A}$ itself is nonnegative definite and hence coincides with A^*, its negative part $\hat{A}_-$ does not vanish and is a viable candidate for the matrix B in Theorem 5. Let H be the hyperplane orthogonal to $\hat{A}_-$ when the latter is not 0. Then H separates $\hat{A}$ and NND (n), and supports NND (n) in the point $\hat{A}_+$, see Fig. 3. When going from $\hat{A}$ in the direction of $N_M(\hat{A}_-)$ towards NND (n) one can at least proceed until

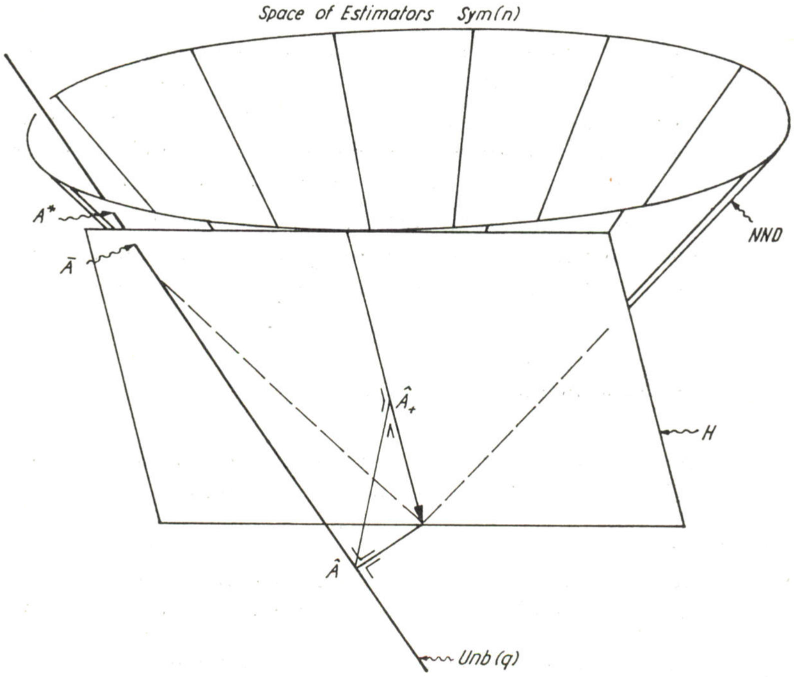

Fig. 3. The non-negative analogue of the MINQUE procedure yields the matrix A^* which is the projection of the ordinary MINQUE matrix $\hat{A}$ onto the cone NND of non-negative definite matrices along the affine subspace Unb (q) of all unbiased location-invariant quadratic estimators for $q'\tau$. A reasonable approximation of A^* is the matrix $\bar{A} = \hat{A} + \|\hat{A}_-\|^2 \|N_M(\hat{A}_-)\|^{-2} N_M(\hat{A}_-)$ which is obtained by projecting the negative part $\hat{A}_- = \hat{A}_+ - \hat{A}$ of $\hat{A}$ onto Unb (q) and then proceeding along this direction until the separating hyperplane $H = \{\hat{A}_-\}^\perp$ is met.

the separating hyperplane H is met, this point being the matrix

$$\hat{A}+\|\hat{A}_-\|^2\|N_M(\hat{A}_-)\|^{-2}\,N_M(\hat{A}_-),$$

and corresponding to the choice $B=\hat{A}_-$ in Theorem 5. If this matrix is non-negative definite then $B^*=2\,\|\hat{A}_-\|^2\|N_M(\hat{A}_-)\|^{-2}\,\hat{A}_-$ is an optimal solution of problem (D). Example 3 shows that this construction cannot always yield optimal solutions (see also [34], Counterexample 3.9), it is more surprising that it does work in certain cases.

Example 4. Consider a model $Y\sim(0;\,\sigma_1^2\,\mathrm{Diag}\,[I_a\!:\!I_b\!:\!0]+\sigma_2^2\,\mathrm{Diag}\,[0\!:\!I_b\!:\!I_c])$, $a, b, c>1$, in which σ_1^2 is to be estimated. The MINQUE matrix is $\hat{A}==d^{-1}\,\mathrm{Diag}\,[(b+c)\,I_a\!:\!cI_b\!:\!-bI_c]$, $d=ab+bc+ca$, and leads to a non-negative definite matrix $\mathrm{Diag}\,[a^{-1}I_a\!:\!0\!:\!0]=A^*$, as described above. Moreover, since $B^*==(2/a)\,\mathrm{Diag}\,[0\!:\!0\!:\!I_c]$ is an optimal solution of (D) the sufficient condition in Theorem 3 is not, in general, necessary. The negativity eliminating projectors of Theorem 4 are $Q_*=\mathrm{Diag}\,[I_a\!:\!0\!:\!0]$, and $Q^*=\mathrm{Diag}\,[I_a\!:\!I_b\!:\!0]$.

The model with a common mean and heteroscedastic variances behaves similarly, in that the ranges of MV_iM and MV_jM, $i\neq j$, overlap, without being contained in each other. Details for 2 heteroscedastic variances are given in [34], Lemma 3.8), for more components the computational difficulties are prohibitive since the spectral decomposition is needed of a matrix of size $n\times n$. Nevertheless the calculations may be carried through, as far as an individual component σ_i^2 is concerned rather than a combination $\sum q_j\sigma_j^2$ ([34], Lemma 3.4, 3.7), here the optimal solution of (P) is the sample variance of the i-th group, see ([37]).

6. Open problems

1. How do the matrices $M, V_1, \ldots, V_l$ determine those situations when $\hat{A}+\|\hat{A}_-\|^2\,\|N_M(\hat{A}_-)\|^{-2}\,N_M(\hat{A}_-)$ is non-negative definite?

2. Let L be the line $\{\hat{A}+\alpha N_M(\hat{A}_-)\mid \alpha\in\mathrm{R}\}$. Is it possible that L by-passes NND (n)? Otherwise there is a smallest α^* which makes $\hat{A}+\alpha^* N_M(\hat{A}_-)$ non-negative definite; how does the matrix so obtained compare with A^*?

3. In Example 3 there exists a negativity eliminating projector although the hypothesis of Theorem 4 is not satisfied. Is it possible to generalize Theorem 4 in this direction?

4. It is shown in [37], Theorem 2, that if $MV_1M, \ldots, MV_lM$ span an l-dimensional quadratic subspace of symmetric matrices then the cone $\mathfrak{Q}$ is the image of $\overline{G_M}$ under the transformation $S_M=((\mathrm{trace}\;MV_iMV_j))$. Is the quadratic subspace condition also necessary, besides being sufficient?

5. Partial results on non-negative definiteness of the estimated dispersion matrix also need a quadratic subspace condition, see F. Pukelsheim & G. P. H. Styan (1979), or [37], Theorem 2. Would here a programming approach prove useful?

6. Programming results other than from duality theory are developed, e.g., by J. M. Borwein & H. Wolkowicz (1979). How do they relate to non-negative variance estimation?

7. The correspondence between mean estimation and dispersion estimation is not quite lost through non-negative variance estimation, as it may seem at first glance. A parallel might be constructed as follows: Suppose in a model $Y\sim(X\beta;\sigma^2V)$ that all components of Y are non-negative, or more generally, that Y takes its values in a closed convex cone K. Let $\overline{H}=\{\beta\in\mathrm{R}^k \mid X\beta\in K\}$ be the natural parameter set for β. Again one may discuss non-negative linear forms $c'\beta$, and non-negative estimates $a'Y$ which are optimal for $c'\beta$.

Acknowledgement

I would like to thank Professor H. Witting for valuable discussions and continued support during the course of this work. Parts of this work were carried out while I was a post-doctoral fellow of the Deutsche Forschungsgemeinschaft, Bonn, with the Department of Statistics, Stanford University.

References

[1] Aitken, A. C. (1935). On least squares and linear combination of observations. *Proc. Roy. Soc. Edinburgh* **55**, 42–48.

[2] Albert, A. (1973). The Gauss-Markov theorem for regression models with possibly singular covariances. *SIAM J. Appl. Math.* **24**, 182–187.

[3] Anderson, R. L. (1965). Query 9. Negative variance estimates. *Technometrics* **7**, 75–76.

[4] Atiqullah, M. (1962). The estimation of residual variance in quadratically balanced least-squares problems and the robustness of the F-test. *Biometrika* **49**, 83–91.

[5] Baksalary, J. K. & Kala, R. (1978). Relationships between some representations of the best linear unbiased estimator in the general Gauss-Markoff model. *SIAM J. Appl. Math.* **35**, 515–520.

[6] Bellman, R. & Fan, K. (1963). On systems of linear inequalities in Hermitian matrix variables. pp. 1–11 in: V. L. Klee (ed.). *Convexity. Proceedings of Symposia in Pure Mathematics, Vol. VII.* Providence, R. I.: American Mathematical Society.

[7] Borwein, J. M. & Wolkowicz, H. (1979). Characterizations of optimality for the abstract convex program. Typescript. Dalhousie University, Halifax, Nova Scotia.

[8] Brown, K. G. (1977). On estimation of diagonal covariance matrices by MINQUE. *Comm. Statist. A – Theory Methods* **6**, 471–484.

[9] Brown, K. G. (1978). Estimation of variance components using residuals. *J. Amer. Statist. Assoc.* **73**, 141–146.

[10] Drygas, H. (1972). The estimation of residual variance in regression analysis. *Math. Operationsforsch. Statist.* **3**, 373–388.

[11] Drygas, H. & Hupet, G. (1979). Non-negativeness of the best quadratic unbiased estimator of the variance in the linear regression model. Preprint Nr. 18/1979. Gesamthochschule Kassel.

[12] Federer, W. T. (1968). Non-negative estimators for components of variance. *J. Roy. Statist. Soc.* C **17**, 171–174.

[13] Fuller, W. A. & Rao, J. N. K. (1978). Estimation for a linear regression model with unknown diagonal covariance matrix. *Ann. Statist.* **6**, 1149–1158.

[14] Judge, G. G. & Takayama, T. (1966). Inequality restrictions in regression analysis. *J. Amer. Statist. Assoc.* **61**, 166–181.

[15] Hartley, H. O. & Rao, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* **54**, 93–108.

[16] Hartung, J. (1981). Non-negative minimum biased invariant estimation in variance component models. Ann. Statist. **9**, No. 2.

[17] Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.* **72**, 320–340.

[18] Hemmerle, W. J. & Hartley, H. O. (1973). Computing maximum likelihood estimates for the mixed A. O. V. model using the W transformation. *Technometrics* **15**, 819–831.

[19] Herbach, L. H. (1959). Properties of model II – type analysis of variance tests, A: Optimum nature of the F-test for model II in the balanced case. *Ann. Math. Statist.* **30**, 939–959.

[20] Horn, S. D. & Horn, R. A. (1975). Comparison of estimators of heteroscedastic variances in linear models. *J. Amer. Statist. Assoc.* **70**, 872–879.

[21] Hudson, D. J. (1969). Least-squares fitting of a polynomial constrained to be either non-negative, non-decreasing or convex. *J. Roy. Statist. Soc.* B **31**, 113–118.

[22] Kleffe, J. (1977). Optimal estimation of variance components – a survey. *Sankhyā* B **39**, 211–244.

[23] Kolmogorov, A. N. (1946). On the proof of the method of least squares. *Uspehi Mat. Nauk.* 1 (11), No. 1, 57–70. (In Russian.)

[24] Kozek, A. (1980). Personal communication.

[25] Krafft, O. (1970). Programming methods in statistics and probability theory. pp. 425–446 in: Rosen, J. B., Mangasarian, O. L. & Ritter, K. (eds.). *Nonlinear Programming.* Academic Press, New York.

[26] LaMotte, L. R. (1973). On non-negative quadratic unbiased estimation of variance components. *J. Amer. Statist. Assoc.* **68**, 728–730.

[27] Liew, C. K. (1976). Inequality constrained least-squares estimation. *J. Amer. Statist. Assoc.* **71**, 746–751.

[28] McHugh, R. B. & Mielke Jr., P. W. (1968). Negative variance estimates and statistical dependence in nested sampling. *J. Amer. Statist. Assoc.* **63**, 1000–1003.

[29] Nelder, J. A. (1954). The interpretation of negative components of variance. *Biometrika* **41**, 544–548.

[30] Pukelsheim, F. (1974). Schätzen von Mittelwert und Streuungsmatrix in Gauss-Markoff-Modellen. Diplomarbeit. Institut für Mathematische Stochastik. Universität Freiburg im Breisgau.

[31] Pukelsheim, F. (1976). Estimating variance components in linear models. *J. Multivariate Anal.* **6**, 626–629.

[32] Pukelsheim, F. (1977a). Punktschätzungen in linearen statistischen Modellen. Dissertation. Institut für Mathematische Stochastik. Universität Freiburg im Breisgau.

[33] Pukelsheim, F. (1977b). Linear models and convex programs: unbiased nonnegative

estimation in variance component models. Technical Report No. 104. Department of Statistics. Stanford University.

[34] PUKELSHEIM, F. (1978a). Examples for unbiased nonnegative estimation in variance component models. Technical Report No. 113. Department of Statistics. Stanford University.

[35] PUKELSHEIM, F. (1978b). A quick introduction to mathematical programming with applications to most powerful tests, nonnegative variance estimation, and optimal design theory. Technical Report No. 128. Department of Statistics. Stanford University.

[36] PUKELSHEIM, F. (1979). Classes of linear models. pp. 69–83 in [53].

[37] PUKELSHEIM, F. (1981). On the existence of unbiased non-negative estimates of variance covariance components. Ann. Statist. **9**, No. 2.

[38] PUKELSHEIM, F. & STYAN, G. P. H. (1979). Nonnegative definiteness of the estimated dispersion matrix in a multivariate linear model. *Bull. Acad. Polon. Sci. Sér. Sci. Math.* **27**, 327–330.

[39] RAO, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1**, 355–372.

[40] RAO, C. R. (1973). *Linear Statistical Inference and Its Applications.* Second Edition. John Wiley & Sons, New York.

[41] RAO, C. R. (1978). Choice of best linear estimators in the Gauss-Markoff model with a singular dispersion matrix. *Comm. Statist. A – Theory Methods* **7**, 1199–1208.

[42] RAO, J. N. K. (1973). On the estimation of heteroscedastic variances. *Biometrics* **29**, 11–24.

[43] RAO, P. S. R. S. & CHAUBEY, Y. P. (1978). Three modifications of the principle of the MINQUE. *Comm. Statist. A – Theory Methods* **7**, 767–778.

[44] RICH, D. K. & BROWN, K. G. (1979). Estimation of variance components using residuals: some empirical evidence. pp. 139–154 in [53].

[45] ROCKAFELLAR, R. T. (1970). *Convex Analysis.* Princeton University Press, Princeton, N. J.

[46] SAHAI, H. (1974). On negative estimates of variance components under finite population models. *South African Statist. J.* **8**, 157–166.

[47] SAHAI, H. (1979). A bibliography on variance components. *Internat. Statist. Rev.* **47**, 177–222.

[48] SCHEFFÉ, H. (1959). *The Analysis of Variance.* John Wiley & Sons, New York.

[49] SEARLE, S. R. (1971). *Linear Models.* John Wiley & Sons, New York.

[50] SEELY, J. (1970). Linear spaces and unbiased estimation – application to the mixed model. *Ann. Math. Statist.* **41**, 1735–1748.

[51] SEELY, J. (1971). Quadratic subspaces and completeness. *Ann. Math. Statist.* **42**, 710–721.

[52] SRIVASTAVA, J. N. (ed.) (1975). *A Survey of Statistical Design and Linear Models.* North Holland Publishing Company, Amsterdam.

[53] VAN VLECK, L. D. & SEARLE, S. R. (eds.) (1979). *Variance Components and Animal Breeding. Proceedings of a Conference in Honor of C. R. Henderson.* Cornell University, Ithaca, N. Y.

[54] VERDOOREN, L. R. (1980). On estimation of variance components. Statist. Neerlandica **34**, 83–106.

[55] WITTING, H. (1966). *Mathematische Statistik.* B. G. Teubner, Stuttgart.

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

[57] ZYSKIND, G. (1975). Error structures, projections and conditional inverses in linear model theory. pp. 647–663 in [52].

Résumé

Ce travail présente quelques characterisations des estimateurs du type de MINQUE non-negatif modifiés pour l'estimation des composants de la variance dans le modèle linéaire générale. Les résultats sont obtenus par l'application de la théorie de dualité d'analyse convexe. En outre, on présente une revue générale des aspects geométriques des problèmes d'estimation dans ces modèles.

Zusammenfassung

Es werden geometrische Ansätze zur Mittelwert- und Streuungsschätzung in linearen Modellen diskutiert. Für diejenigen herkömmlichen Verfahren, deren Herleitung ausschließlich auf linearer oder multilinearer Algebra beruht, reichen die mit einem linearen Unterraum verbundenen Begriffe. Dabei müssen allerdings Nebenbedingungen wie die Nicht-Negativität von Varianzschätzern unberücksichtigt bleiben. Nach einer kurzen Literaturübersicht zu negativen Varianzschätzungen wird auf die nicht-negativ definite Abart des MINQUE-Verfahrens genauer eingegangen. Einige Charakterisierungen ergeben sich aus der konvexen Dualitätstheorie. Optimale Schätzer entsprechen dann (nicht-linearen) Projektionen auf abgeschlossene konvexe Kegel, ihre geometrische Darstellung ist einfach, ihre explizite Berechnung schwieriger. Für den Praktiker können aus diesen Teilergebnissen noch keine Empfehlungen hergeleitet werden, eine abschließende Zusammenstellung ungelöster Probleme betont, daß erst weitere theoretische Untersuchungen erforderlich sind.

Received February 1980.