The Annals of Mathematical Statistics
1971, Vol. 42, No. 2, 691–703

LINEAR SPACES AND MINIMUM VARIANCE UNBIASED ESTIMATION[1]

JUSTUS SEELY AND GEORGE ZYSKIND

Oregon State University and Iowa State University

Consideration is given to minimum variance unbiased estimation when the choice of estimators is restricted to a finite-dimensional linear space. The discussion gives generalizations and minor extensions of known results in linear model theory utilizing both the coordinate-free approach of Kruskal and the usual parametric representations. Included are (i) a restatement of a theorem on minimum variance unbiased estimation by Lehmann and Scheffé; (ii) a minor extension of a theorem by Zyskind on best linear unbiased estimation; (iii) a generalization of the covariance adjustment procedure described by Rao; (iv) a generalization of the normal equations; and (v) criteria for existence of minimum variance unbiased estimators by means of invariant subspaces. Illustrative examples are included.

1. Introduction and summary. Consideration is given to minimum variance unbiased estimation when the choice of estimators is restricted to a finite-dimensional linear space of rv's (random variables). The results form a continuation of an enquiry on unbiased estimation in Seely [10] and the terminology introduced in [10] is used throughout. Each section (except Section 2) is based on a minor extension or a generalization of a known result in linear model theory and examples are included for most results. It is hoped that the following will allow a more convenient and direct usage of linear model theory (e.g., Examples 2 and 8) when considering linear spaces of rv's other than linear combinations of a random vector, and that the examples on b.l.u. (best linear unbiased) estimation for the fixed and mixed linear models will be of interest.

In Section 2 the basic groundwork additional to that in [10] is discussed as well as the fixed and mixed linear model as interpreted in this paper. A theorem due to Lehmann and Scheffé [5] which states that an estimator is a uniformly minimum variance unbiased estimator if and only if the estimator has zero covariance with the unbiased estimators of zero is stated in the context of this paper in Section 3. This theorem is especially useful for verifying results in later sections and at times may be used directly to obtain minimum variance unbiased estimators.

For a linear model $Y = X\beta + e$ with covariance matrix $\sigma^2 V$, a theorem due to Zyskind [14] states that $a'Y$ is a b.l.u. estimator if and only if $Va \in \underline{R}(X)$. A minor extension of this theorem is given in Section 4 and two examples are included. One example extends the result of Zyskind to the mixed linear model and the other illustrates very concisely the notions of linear model theory as applied to the multivariate linear model.

Received July 25, 1969; revised August 26, 1970.

[1] This work was prepared partly in connection with work on Contract AF33(615)68C1343, Office of Aerospace Research, United States Air Force, Wright–Patterson Air Force Base, Ohio, and an NSF Traineeship.

Theorem 3 is the main result in Section 5. This theorem may be viewed as a generalization of the covariance adjustment procedure given by Rao [7] in the sense that Rao's procedure may be obtained as a special case of the necessity portion of the theorem. Two rather interesting examples concerning b.l.u. estimation in a fixed linear model are also given in Section 5. In Section 6 a generalization of the normal equations in linear model theory is given, and necessary and sufficient conditions are obtained for when these generalized equations give best estimators. Examples are given showing the generality of the normal type equations as applied to obtaining b.l.u. estimators in a fixed linear model situation for all linearly estimable parametric functions (Example 5) and for subsets of the linearly estimable parametric functions (Example 6). The last section characterizes, via invariant subspaces, when minimum variance unbiased estimators exist within an arbitrary linear space of rv's. The approach taken in the last section is similar to that used by Kruskal [4] and Zyskind [14] and more recently by Eaton [3]. Examples illustrating the invariant subspace notion as applied to the mixed linear model and as applied to a simple situation involving quadratic estimators are given.

2. Preliminary notions. The contents of Section 2 in [10] are assumed in the sequel. Thus, $\bar{\mathscr{A}}$ is a linear space of rv's of the form $\{(a, Y): a \in \mathscr{A}\}$ where $(\mathscr{A}, (-, -))$ is a real finite-dimensional inner product space and Y is a rv from a measurable space $(\mathscr{U}, S)$ into $\mathscr{A}$. Assumptions additional to those in [10] are (i) each $\bar{a} \in \bar{\mathscr{A}}$ is P-square integrable with respect to all $P \in \mathscr{P}$ and (ii) there is a function Var on $\bar{\mathscr{A}} \times \Omega$ by which the variance of an element in $\bar{\mathscr{A}}$ may be characterized over the class $\mathscr{P}$. Concerning additional notation, the covariance operator of Y with respect to $\theta \in \Omega$ is denoted by Σ_θ, i.e., Σ_θ denotes the unique linear operator on $\mathscr{A}$ such that

$$a, b \in \mathscr{A} \Rightarrow \text{Cov}\,[(a, Y), (b, Y) \mid \theta] = (a, \Sigma_\theta b).$$

Also, the notation $\Sigma \geqq 0$ means that Σ is a nonnegative linear operator and the set $\text{sp}\,\{\Sigma_\theta : \theta \in \Omega\}$ is denoted by $\mathscr{V}$

DEFINITION 1. An element $\bar{t} \in \bar{\mathscr{A}}$ is said to be $\bar{\mathscr{A}}$-best for a parametric function g if and only if $\bar{t} \in \bar{\mathscr{A}}_g$ and

$$\bar{h} \in \bar{\mathscr{A}}_g \Rightarrow \text{Var}\,[\bar{t} \mid \theta] \leqq \text{Var}\,[\bar{h} \mid \theta] \quad \text{for all} \quad \theta \in \Omega.$$

The statement that $\bar{t}$ is an $\bar{\mathscr{A}}$-best estimator implies that $\bar{t}$ is $\bar{\mathscr{A}}$-best for $g(\theta) = E[\bar{t} \mid \theta]$.

DEFINITION 2. Let $\Sigma \geqq 0$ and let g be a parametric function, then t is said to be Σ-min for g if and only if $t \in \mathscr{A}_g$ and

$$\inf\,\{(a, \Sigma a): a \in \mathscr{A}_g\} = (t, \Sigma t).$$

The statement that t is a Σ-min element means that t is Σ-min for $g(\theta) = (t, \mu_\theta)$.

The notion of a Σ-min element is similar to the idea of a locally best estimator as used by Barankin [2] and Stein [12]. Moreover, (t, Y) is $\bar{\mathscr{A}}$-best for a parametric function g if and only if t is Σ_θ-min for g for all $\theta \in \Omega$. Thus, $\mathscr{A}$-best estimators may be investigated via Σ-min elements and since Σ-min elements are easy to work with much of the following is devoted to Σ-min elements.

The notation $Y \sim F_n(X\beta, \sum_{i=1}^m v_i V_i, \Omega)$ is used to indicate that Y is an $n \times 1$ random vector with expectation $X\beta$ and covariance matrix $\sum_{i=1}^m v_i V_i$ where X is a known $n \times p$ matrix, each V_i is a known $n \times n$ symmetric matrix, and $\Omega = \{(\beta, v)\}$ is a subset of $R^p \times R^m$ which describes the ranges and relationships of the parameter vectors $\beta = (\beta_1, \cdots, \beta_p)'$ and $v = (v_1, \cdots, v_m)'$. The notation

$$Y \sim F_n(X\beta, \textstyle\sum_{i=1}^m v_i V_i, \Omega)$$

describes what is often referred to as a mixed linear model and this terminology, i.e., mixed linear model, is employed in the following sections. When $m = 1$ the term fixed linear model is used and the covariance matrix for this case is denoted by $\sigma^2 V$. The term linear model is used to indicate either a fixed or a mixed linear model and for a linear model the sets Ω_1 and Ω_2 denote, respectively, the range of β and the range of v.

When $Y \sim F_n(X\beta, \sum_{i=1}^m v_i V_i, \Omega)$ and interest is in the linear space of rv's $\overline{\mathscr{L}} = \{a'Y : a \in R^n\}$, we assume that $\mathscr{L} = R^n$ with the usual inner product $(a, b) = a'b$. For this situation observe that $\mathscr{E} = \text{sp}\,\{X\beta : \beta \in \Omega_1\}$, $\mathscr{V} = \text{sp}\,\{\sum_{i=1}^m v_i V_i : v \in \Omega_2\}$, and that a natural $\mu_\theta = \mathbf{H}\xi_\theta$ representation is given by $\mathbf{H} = X$, $\xi_\theta = \beta$, $\mathscr{H} = R^p$ with the usual inner product, and $\Omega_H = \Omega_1$.

The main emphasis is on the fixed linear model with β restricted only through the covariance matrix. This statement is taken to imply that $Y \sim F_n(X\beta, \sigma^2 V, \Omega)$ where Ω is any parameter space consistent with

$$\Omega_1 = \{\beta : C'X\beta = C'Y, \underline{R}(C) = \underline{N}(V)\}, \tag{2.1}$$

and it is tacitly assumed $\Omega_2 \neq \{0\}$ so that $\mathscr{V} = \text{sp}\,\{V\}$, although V may be any positive semidefinite matrix. Concerning (2.1) suppose $Z \sim F_q(U\beta, \sigma^2 Q, \Omega)$ with Q nonsingular and with β restricted by k consistent equations $\Delta\beta = c$, then under the setup

$$Y = \begin{bmatrix} Z \\ c \end{bmatrix}, \quad X = \begin{bmatrix} U \\ \Delta \end{bmatrix}, \quad V = \begin{bmatrix} Q & 0 \\ 0 & 0 \end{bmatrix}, \quad n = q + k, \tag{2.2}$$

the expression for Ω_1 in (2.1) is equivalent to the restrictions $\Delta\beta = c$. In the usual sense, however, (2.1) does not constitute a parameter space, i.e., Ω_1 is dependent upon the observed outcome of the vector Y. Nevertheless, since $C'X\beta = C'Y$ with probability one, the parameter space for β is taken as in (2.1). Also note that with probability one the equations $C'X\beta = C'Y$ are consistent. In the following the repetitious "with probability one" is usually omitted and $C'Y$ is treated as a constant vector with the property that $C'Y \in \underline{R}(C'X)$.

The mixed linear model with β restricted only through the covariance matrix is also used for illustration purposes. By such a linear model we mean that

$$Y \sim F_n(X\beta, \textstyle\sum_{i=1}^m v_i V_i, \Omega)$$

where Ω is any parameter space consistent with

$$\Omega_1 = \{\beta : C'X\beta = C'Y,\ \underline{R}(C) = \textstyle\bigcap_{T \in \mathscr{V}} \underline{N}(T)\}. \tag{2.3}$$

The comments concerning Ω_1 in the previous paragraph apply here also and in addition we assume

$$\mathscr{V} = \text{sp}\,\{\textstyle\sum_i v_i V_i : v \in \Omega_2\} = \text{sp}\,\{V_1, \cdots, V_m\}. \tag{2.4}$$

By assuming (2.4) most statements concerning $\mathscr{V}$ may be formulated in terms of $V_1, V_2, \cdots, V_m$; for example, in (2.3) the condition on C becomes $\underline{R}(C) = \bigcap_{i=1}^m \underline{N}(V_i)$. The assumption in (2.4) is satisfied in most commonly used linear models; however, if (2.4) is not satisfied any spanning set for $\mathscr{V}$ may be substituted for $\{V_1, \cdots, V_m\}$ in the sequel.

3. Lehmann–Scheffé theorem. A useful starting point for characterizing $\mathscr{A}$-best estimators is Theorem 5.3 in Lehmann and Scheffé [5]. In the present context, this theorem is essentially Corollary 1.1 below except that Corollary 1.1 is stated with an arbitrary spanning set for $\mathscr{V}$ instead of the particular spanning set $\{\Sigma_\theta : \theta \in \Omega\}$ used in [5]. A proof for the following theorem from which Corollaries 1.1, 1.2, and 1.3 follow immediately may be constructed from Theorem 5.3 in [5].

THEOREM 1. *Let* $\Sigma \geqq 0$, *then* $t \in \mathscr{A}$ *is a* Σ-min *element if and only if* $(t, \Sigma z) = 0$ *for all* $z \in \mathscr{A}_0$.

COROLLARY 1.1. *Let* $\mathscr{V}_1$ *be any spanning set for* $\mathscr{V}$. *For* $t \in \mathscr{A}$ *the* rv (t, Y) *is an* $\mathscr{A}$*-best estimator if and only if* $(t, \Gamma z) = 0$ *for all* $z \in \mathscr{A}_0$ *and all* $\Gamma \in \mathscr{V}_1$.

COROLLARY 1.2. *Let* $\Sigma \geqq 0$, *then* Σ*-min elements are unique if and only if* $\underline{N}(\Sigma) \cap \mathscr{A}_0 = \{0\}$.

COROLLARY 1.3. *An* $\mathscr{A}$*-best estimator is unique if and only if* $t \in \text{N}(\Sigma_\theta) \cap \mathscr{A}_0$ *for all* $\theta \in \Omega$ *implies that* (t, Y) *is the zero function.*

Suppose $Y \sim F_n(X\beta, \sigma^2 V, \Omega)$ and that (2.1) describes Ω_1, then the subspace $\mathscr{L}_0$ corresponding to the zero estimators in $\mathscr{L}$ is

$$\mathscr{L}_0 = \{X\beta : \beta \in \Omega_1\}^\perp = \{X\beta : \beta \in \beta_0 + \text{N}(C'X)\}^\perp,$$

where $\underline{R}(C) = \underline{N}(V)$ and β_0 is such that $C'X\beta_0 = C'Y$. If $a \in \underline{N}(V) \cap \mathscr{L}_0$, then $a = C\rho$ for some ρ and $a'X\beta_0 = 0$ so that

$$(a, Y) = \rho' C' Y = \rho' C' X\beta_0 = a'X\beta_0 = 0;$$

and hence Corollary 1.3 implies b.l.u. estimators are unique. Generally this uniqueness is only with probability one; however, when $C'Y$ is actually a constant vector (e.g., the setup in (2.2) with Q invertible) the phrase with probability one may be omitted. Thus, for the fixed linear model (also for the mixed linear model) with β restricted only through the covariance matrix any uniqueness statement concerning b.l.u. estimators should be interpreted according to the behavior of the vector $C'Y$.

For a nonnegative linear operator Σ on $\mathscr{A}$, let $\mathscr{K}(\Sigma)$ denote the set of Σ-min elements. Clearly, from Theorem 1

$$\mathscr{K}(\Sigma) = \Sigma[\mathscr{A}_0]^{\perp}, \tag{3.1}$$

and thus finite linear combinations of Σ-min elements are Σ-min elements. Moreover, finite linear combinations of $\bar{\mathscr{A}}$-best estimators are also $\bar{\mathscr{A}}$-best estimators since the set $\mathscr{K}$ of elements t such that (t, Y) is an $\bar{\mathscr{A}}$-best estimator is the intersection over Ω of all subspaces $\mathscr{K}(\Sigma_\theta)$. Also (3.1) implies that $\mathscr{K}(\Sigma)+\mathscr{A}_0 = \mathscr{A}$, and this observation with Theorem 1 in [10] establishes the existence of a Σ-min element for each $\bar{\mathscr{A}}$-estimable parametric function.

4. The condition $\Sigma t \in \mathscr{E}$. For a fixed linear model with β restricted only through the covariance matrix, Zyskind [14] shows that $a'Y$ is a b.l.u. estimator if and only if $Va \in \underset{\sim}{R}(X)$. (A slightly stronger statement is actually given in [14] in the sense that only a certain structure need be inferred about V.) The following theorem states this result under the present formulation and shows (Example 1) that $\underset{\sim}{R}(X)$ in Zyskind's statement could possibly be replaced (depending upon V) by other sets in R^n. As a corollary to the theorem a characterization for $\bar{\mathscr{A}}$-best estimators is given and the corollary is applied in Example 1 to extend the result of Zyskind to a mixed linear model.

THEOREM 2. *If $\Sigma \geqq 0$ and $\mathscr{M}$ is any subset of $\mathscr{A}$ such that $\mathscr{M}\cap \underset{\sim}{R}(\Sigma) = \mathscr{E}\cap \underset{\sim}{R}(\Sigma)$, then a necessary and sufficient condition for t to be a Σ-min element is that $\Sigma t \in \mathscr{M}$.*

PROOF. It follows from Theorem 1 that t is a Σ-min element if and only if $\Sigma t \in \mathscr{E}$. The result follows by observing

$$\Sigma t \in \mathscr{E} \Leftrightarrow \Sigma t \in \underset{\sim}{R}(\Sigma)\cap\mathscr{E} \Leftrightarrow \Sigma t \in \underset{\sim}{R}(\Sigma)\cap\mathscr{M} \Leftrightarrow \Sigma t \in \mathscr{M}.$$

COROLLARY 2.1. *Let $\mathscr{V}_1$ be a spanning set for $\mathscr{V}$ and let $\mathscr{M}$ be such that $\mathbf{R}(\Gamma)\cap\mathscr{M} = \underset{\sim}{R}(\Gamma)\cap\mathscr{E}$ for each $\Gamma \in \mathscr{V}_1$. A necessary and sufficient condition for (t, Y) to be an $\bar{\mathscr{A}}$-best estimator is that $\Sigma t \in \mathscr{M}$ for all $\Sigma \in \mathscr{V}_1$.*

EXAMPLE 1. Consider a mixed linear model with β restricted only through the covariance matrix and let C be such that $\underset{\sim}{R}(C) = \bigcap_{i=1}^{m} \underset{\sim}{N}(V_i)$. Suppose $X\rho \in \underset{\sim}{R}(V_i)$, then $\rho \in \underset{\sim}{N}(C'X)$ which implies $X\rho \in \mathscr{E}$ and hence $\underset{\sim}{R}(X)\cap\underset{\sim}{R}(V_i)\subset\mathscr{E}\cap\underset{\sim}{R}(V_i)$. Thus, $\underset{\sim}{R}(X)\cap\underset{\sim}{R}(V_i) = \mathscr{E}\cap\underset{\sim}{R}(V_i)$ for $i = 1, 2, \cdots, m$. Let $\mathscr{V}_1 = \{V_1, V_2, \cdots, V_m\}$ and let $\mathscr{M}$ be such that $\underset{\sim}{R}(V_i)\cap\mathscr{E}\subset\mathscr{M}\subset\underset{\sim}{R}(X)$ for $i = 1, 2, \cdots, m$. From Corollary 2.1 it follows that $a'Y$ is a b.l.u. estimator if and only if $V_i a \in \mathscr{M}$ for $i = 1, 2, \cdots, m$. By selecting $\mathscr{M} = \underset{\sim}{R}(X)$ when $m = 1$ the result of Zyskind is obtained.

Eaton [3] and Rao [6] consider the multivariate linear model very conveniently from the ideas and notions of fixed linear model theory. The following example further exemplifies this correspondence and illustrates an application of Corollary 2.1.

EXAMPLE 2. Let $(\mathscr{A}, (-,-))$ denote the inner product space of all $n \times k$ real matrices with the trace inner product, i.e., $(A, B) = \operatorname{tr}(A'B)$ for all $A, B \in \mathscr{A}$. Suppose Y is an $n \times k$ random matrix composed of rows of independent random vectors with a common unknown covariance matrix Φ and with expectations such that $E(Y) = X\boldsymbol{\beta}$ where X is a known $n \times p$ matrix and $\boldsymbol{\beta}$ is an unknown $p \times k$ matrix. Let $\bar{\mathscr{A}} = \{(A, Y): A \in \mathscr{A}\}$. Then for $\theta = (\boldsymbol{\beta}, \Phi)$ it is clear that

(a) $\mu_{\boldsymbol{\beta}} = X\boldsymbol{\beta}$ where $\mu_\theta = \mu_{\boldsymbol{\beta}}$,

(b) $A \in \mathscr{A} \Rightarrow \boldsymbol{\Sigma}_\Phi A = A\Phi$ where $\boldsymbol{\Sigma}_\Phi = \boldsymbol{\Sigma}_\theta$,

(c) $\mathscr{E} = \{XG : G$ an arbitrary $p \times k$ matrix$\}$, and

(d) $\mathscr{V} = \operatorname{sp}\{\boldsymbol{\Sigma}_\Phi : \Phi \geqq 0\}$.

Let $(\mathscr{H}, <-, ->)$ denote the inner product space of real $p \times k$ matrices with the trace inner product and define $\mathbf{H}$ from $\mathscr{H}$ into $\mathscr{A}$ by $\mathbf{H}\Lambda = X\Lambda$ for all $\Lambda \in \mathscr{H}$. For $\xi_{\boldsymbol{\beta}} = \boldsymbol{\beta}$ we have a $\mu_{\boldsymbol{\beta}} = \mathbf{H}\xi_{\boldsymbol{\beta}}$ representation. Observing that $\mathbf{H}^*A = X'A$ and applying Theorem 2 in [10] we obtain the following:

The parametric function $\operatorname{tr}(\Lambda'\boldsymbol{\beta})$ is $\bar{\mathscr{A}}$-estimable if and only if there exists a C such that $X'C = \Lambda$.

From Corollary 2.1 it is clear that (A, Y) is an $\bar{\mathscr{A}}$-best estimator if and only if

$$\Phi \geqq 0 \Rightarrow \boldsymbol{\Sigma}_\Phi A = A\Phi \in \mathscr{E}.$$

Thus, the following conclusion may be drawn:

The rv (A, Y) is an $\bar{\mathscr{A}}$-best estimator if and only if there exists a G such that $A = XG$.

If $\boldsymbol{\beta}$ is restricted by a consistent set of matrix equations $\Delta'\boldsymbol{\beta} = \Gamma$ and if $\bar{\mathscr{A}}$ is the linear space spanned by the constant function and all rv's of the form (A, Y), then Corollary 2.1 establishes the following statement:

The rv (A, Y) is an $\bar{\mathscr{A}}$-best estimator if and only if there exists a G such that $A = XG$ and $\Delta'G = 0$.

From this statement it is easily verified that the random matrix $X\tilde{\beta} = QY + (I-Q)X\boldsymbol{\beta}_0$ (Q is the symmetric idempotent matrix such that $\underset{\sim}{R}(Q) = X[(\underset{\sim}{N}\Delta')]$ and $\boldsymbol{\beta}_0$ is such that $\Delta'\boldsymbol{\beta}_0 = \Gamma$) is such that $(A, X\tilde{\beta})$ is the $\bar{\mathscr{A}}$-best estimator for the parametric function $(A, X\boldsymbol{\beta})$ for arbitrary $A \in \mathscr{A}$. Other analogies with the fixed linear model may be made in a straightforward manner. One final comment is perhaps worth noting. Assuming a multivariate normal distribution for the rows of Y the pertinent matrices regarding a hypothesis of the form $\Delta'\boldsymbol{\beta} = \Gamma$ may be set up directly without any assumption of estimability, i.e., without assuming $\underset{\sim}{R}(\Delta) \subset \underset{\sim}{R}(X')$. To see this let P denote the symmetric idempotent matrix such that $\underset{\sim}{R}(P) = \underset{\sim}{R}(X)$, then

$$R_1 = (Y - X\boldsymbol{\beta}_0)'(P-Q)(Y - X\boldsymbol{\beta}_0) \quad \text{and} \quad R_0 = Y'(I-P)Y$$

have independent Wishart distributions and under the null hypothesis both have central Wishart distributions so that the usual tests (e.g., see Rao [6]) may be used. When $\underline{R}(\Delta) \subset \underline{R}(X')$ it may be noted that a matrix A exists such that $X'A = \Delta$ and $\underline{R}(A) = \underline{R}(P-Q)$ so that R_1 may be written in the more familiar form

$$R_1 = (\Delta'\hat{\beta}-\Gamma)'(A'A)^-(\Delta'\hat{\beta}-\Gamma),$$

where $\hat{\beta}$ is such that $X'X\hat{\beta} = X'Y$.

5. Σ-min elements via $\mathscr{A}_0$. For $\Sigma \geqq 0$ the expression in (3.1) implies that t is a Σ-min element if and only if t is orthogonal to $\mathscr{A}_0$ with respect to the quasi-inner product $(a, b)_\Sigma = (a, \Sigma b)$. Thus, by modifying a procedure such as a Gram–Schmidt process, an element $t \in \mathscr{A}_g$ may be adjusted to a new element $h \in \mathscr{A}_g$ such that h is a Σ-min element. The following lemma (a proof is given in [8]) is useful in the next example for illustrating this adjustment procedure.

LEMMA 1. *Let $\mathscr{M}$ be a subspace of R^n with the usual inner product, let T be a real $n \times n$ positive semidefinite matrix and let $k = \dim T[\mathscr{M}]$. There exists a matrix Z such that*

$$Z'TZ = I_k \quad \textit{and} \quad \mathscr{M} = \underline{R}(Z) \oplus \underline{N}(T) \cap \mathscr{M}.$$

Moreover, for such a Z the matrix $ZZ'T$ is the projection on $\underline{R}(Z)$ along $\underline{N}(Z'T) = T[\mathscr{M}]^\perp = T^{-1}[\mathscr{M}^\perp]$.

EXAMPLE 3. Consider a fixed linear model with β restricted only through the covariance matrix and let $Z = (z_1, z_2, \cdots, z_k)$ be a matrix as described in Lemma 1 when $\mathscr{M} = \mathscr{L}_0$, $V = T$, and $k = \dim V[\mathscr{L}_0]$. For $t \in R^n$ it follows from Theorem 1 that

$$t - \textstyle\sum_{i=1}^k (t, Vz_i)z_i = (I-ZZ'V)t$$

is a V-min element, and since $\mathscr{V} = \operatorname{sp}\{V\}$ it is clear that $t'(I-VZZ')Y$ is a b.l.u. estimator for $t'X\beta$. In addition the matrix $(I-ZZ'V)$ projects onto

$$\underline{N}(Z'V) = V[\mathscr{L}_0]^\perp = V[\underline{N}(X')]^\perp = V^{-1}[\underline{R}(X)];$$

and so Example 1 also verifies that $t'(I-VZZ')Y$ is a b.l.u. estimator. Note that the matrix Z may be obtained using $\mathscr{M} = \underline{N}(X')$. This follows since a matrix Z satisfies $\mathscr{L}_0 = \underline{R}(Z) \oplus \underline{N}(V) \cap \mathscr{L}_0$ if and only if Z satisfies $\underline{N}(X') = \underline{R}(Z) \oplus \underline{N}(V) \cap \underline{N}(X')$.

The following theorem is the main result in this section. The essential part of the theorem is similar to a covariance adjustment procedure (Rao [7]) or a Gram–Schmidt process in that an element $t \in \mathscr{A}_g$ may be adjusted to a new element in $\mathscr{A}_g$ in such a way that the adjusted element is a Σ-min element.

THEOREM 3. *Let $\mathbf{A}$ and $\mathbf{N}$ be linear operators from finite-dimensional inner product spaces into $\mathscr{A}$ such that $\underline{R}(\mathbf{N}) \subset \mathscr{A}_0$ and $\underline{R}(\mathbf{A}^*\mathbf{N}) = \underline{R}(\mathbf{A}^*)$ and let $\Sigma \geqq 0$. A necessary and sufficient condition for $\Sigma[\mathscr{A}_0] \subset \underline{R}(\mathbf{A})$ is that $t - \mathbf{N}\rho$ be a Σ-min element whenever t and ρ are such that $\mathbf{A}^*\mathbf{N}\rho = \mathbf{A}^*t$.*

PROOF. Note that $\Sigma[\mathscr{A}_0] \subset \underset{\sim}{R}(\mathbf{A})$ is equivalent to $\underset{\sim}{N}(\mathbf{A}^*) \subset \mathscr{K}(\Sigma)$. If $N(\mathbf{A}^*) \subset \mathscr{K}(\Sigma)$ and $\mathbf{A}^*\mathbf{N}\rho = \mathbf{A}^*t$, then $t-\mathbf{N}\rho \in \underset{\sim}{N}(\mathbf{A}^*) \subset \mathscr{K}(\Sigma)$. Conversely, if $t \in \underset{\sim}{N}(\mathbf{A}^*)$, then $t \in \mathscr{K}(\Sigma)$ since $\rho = 0$ satisfies $\mathbf{A}^*\mathbf{N}\rho = \mathbf{A}^*t = 0$. Thus, the proof is complete.

COROLLARY 3.1. *Let* $\mathbf{A}$ *and* $\mathbf{N}$ *be such that* $\underset{\sim}{R}(\mathbf{N}) \subset \mathscr{A}_0$ *and* $\Gamma[\mathscr{A}_0] \subset \underset{\sim}{R}(\mathbf{A})$ *for all* Γ *in some spanning set for* $\mathscr{V}$. *For any* t *and* ρ *such that* $\mathbf{A}^*\mathbf{N}\rho = \mathbf{A}^*t$, *the* rv $(t-\mathbf{N}\rho, Y)$ *is an* $\bar{\mathscr{A}}$*-best estimator for the parametric function* $g(\theta) = (t, \mu_\theta)$. *Furthermore, if* $\underset{\sim}{R}(\mathbf{A}^*\mathbf{N}) = \underset{\sim}{R}(\mathbf{A}^*)$ *then to any* $t \in \mathscr{A}_g$ *for an* $\bar{\mathscr{A}}$*-estimable* g *there exists a* ρ *such that* $\mathbf{A}^*\mathbf{N}\rho = \mathbf{A}^*t$.

Theorem 3 is stated as an if and only if condition. The main interest, however, centers upon two linear operators $\mathbf{A}$ and $\mathbf{N}$ satisfying $\underset{\sim}{R}(\mathbf{N}) \subset \mathscr{A}_0$, $\underset{\sim}{R}(\mathbf{A}^*\mathbf{N}) = \underset{\sim}{R}(\mathbf{A}^*)$, and $\Sigma[\mathscr{A}_0] \subset \underset{\sim}{R}(\mathbf{A})$. A convenient way to select $\mathbf{A}$ is to choose $\mathbf{A} = \Sigma\mathbf{N}$. If $\mathbf{A}$ is selected in this fashion, it is easily seen that the above three conditions on $\mathbf{A}$ and $\mathbf{N}$ may be replaced by the condition that $\mathbf{N}$ satisfy

$$\underset{\sim}{R}(\mathbf{N}) \subset \mathscr{A}_0 \subset \underset{\sim}{R}(\mathbf{N})+\underset{\sim}{N}(\Sigma). \tag{5.1}$$

To illustrate this condition and Theorem 3 we consider the following example.

EXAMPLE 4. Consider the situation and notation in Example 3. Let N be selected to satisfy (5.1), i.e., choose N such that

$$\underset{\sim}{R}(N) \subset \mathscr{L}_0 \subset \underset{\sim}{R}(N)+\underset{\sim}{N}(V).$$

Two particular choices worth noting for the matrix N are $N = Z$ and choosing N such that $\underset{\sim}{R}(N) = \underset{\sim}{N}(X')$. That $\underset{\sim}{R}(N) = \underset{\sim}{N}(X')$ will suffice follows from the observation $\mathscr{L}_0 = \underset{\sim}{N}(X')+(X\beta_0)^\perp \cap \underset{\sim}{N}(V)$. The matrices N and $A = VN$ satisfy the conditions in Corollary 3.1; thus, we obtain the following:

For $t \in R^n$ there exists a ρ such that $N'VN\rho = N'Vt$, and for any such t and ρ the rv $(t-N\rho)'Y$ is a b.l.u. estimator for the parametric function $t'X\beta$.

Further, if B is such that $N'VNB = N'V$, then the random vector $X\hat{\beta} = (I-NB)'Y$ has the property for arbitrary t that $t'X\hat{\beta}$ is a b.l.u. estimator for $t'X\beta$. Other interesting results may be obtained in a straightforward manner from these "error normal equations."

6. The normal equation approach. For a fixed linear model with β restricted only through the covariance matrix, various equations which depend upon V may be solved to obtain b.l.u. estimators for parametric functions of the form $\lambda'\beta$. When $V = I$ the normal equations $X'X\beta = X'Y$ may be used; when V is nonsingular the equations $X'V^{-1}X\beta = X'V^{-1}Y$ given by Aitken [1] may be used; and for arbitrary V the equations $X'V^{-}X\beta = X'V^{-}Y$ given by Zyskind and

Martin [16, 17] where V^- belongs to a particular subset of the g-inverses of V may be used. In each of these results the following characteristics are common:

(6.1)
(a) There exists a $\hat{\beta}$ satisfying the equations.
(b) If $\hat{\beta}_1$ and $\hat{\beta}_2$ satisfy the equations, then $X\hat{\beta}_1 = X\hat{\beta}_2$.
(c) If $\lambda \in \underset{\sim}{R}(X')$ and $\hat{\beta}$ satisfies the equations, then $\lambda'\hat{\beta}$ is a b.l.u. estimator for $\lambda'\beta$.

Using these properties as a basis, a generalization of the normal equations is given in the following corollary in terms of $\mathbf{\Sigma}$-best estimators, i.e., estimators of the form (t, Y) where t is a $\mathbf{\Sigma}$-min element. The theorem is given in terms of two linear operators $\mathbf{W}$ and $\mathbf{H}$ and the condition $\underset{\sim}{R}(\mathbf{W}) \oplus \underset{\sim}{N}(\mathbf{H}^*) = \mathscr{A}$ is based upon (6.1a) and (6.1b), i.e., this condition is equivalent to the statements that $\underset{\sim}{R}(\mathbf{W}^*\mathbf{H}) = \underset{\sim}{R}(\mathbf{W}^*)$ and that $\mathbf{W}^*\mathbf{H}\rho = \mathbf{W}^*\mathbf{H}\delta$ implies $\mathbf{H}\rho = \mathbf{H}\delta$.

THEOREM 4. *Let $\mathbf{\Sigma} \geqq 0$ and let $\mathbf{W}$ and $\mathbf{H}$ be linear operators from finite-dimensional inner product spaces into $\mathscr{A}$ such that $\underset{\sim}{R}(\mathbf{W}) \oplus N(\mathbf{H}^*) = \mathscr{A}$. The conditions*

$$\underset{\sim}{R}(\mathbf{W}) \subset \mathscr{K}(\mathbf{\Sigma}) \quad \textit{and} \quad \mathscr{E} \subset \underset{\sim}{R}(\mathbf{H}) \tag{6.2}$$

are satisfied if and only if to each t the rv $(t, \mathbf{H}\hat{\xi})$ *is $\mathbf{\Sigma}$-best for $g(\theta) = (t, \mu_\theta)$ whenever the* rv $\hat{\xi}$ *from $\mathscr{U}$ into the domain of H is such that* $\mathbf{W}^*\mathbf{H}\hat{\xi} = \mathbf{W}^*Y$.

PROOF. Assume (6.2) and suppose $t \in \mathscr{A}$ and that $\hat{\xi}$ satisfies $\mathbf{W}^*\mathbf{H}\hat{\xi} = \mathbf{W}^*Y$. Let $t = \mathbf{W}\rho + f$ where $f \in \underset{\sim}{N}(\mathbf{H}^*)$, then $(t, \mathbf{H}\hat{\xi}) = (\mathbf{W}\rho, Y)$ and $(\mathbf{W}\rho, \mu_\theta) = (\mathbf{W}\rho + f, \mu_\theta)$ imply the desired result. Conversely, let $\hat{\xi}$ satisfy $\mathbf{W}^*\mathbf{H}\hat{\xi} = \mathbf{W}^*Y$. Since $(\mathbf{W}\rho, \mathbf{H}\hat{\xi}) = (\mathbf{W}\rho, Y)$ must be $\mathbf{\Sigma}$-best for $(\mathbf{W}\rho, \mu_\theta)$, it is clear that $\underset{\sim}{R}(\mathbf{W}) \subset \mathscr{K}(\mathbf{\Sigma})$. If $f \in \underset{\sim}{N}(\mathbf{H}^*)$ then $(f, \mathbf{H}\hat{\xi})$ is $\mathbf{\Sigma}$-best for (f, μ_θ); however, $(f, \mathbf{H}\hat{\xi})$ is the zero function so that $(f, \mu_\theta) = 0$ for all $\theta \in \Omega$, which implies that $f \in \mathscr{A}_0$. Thus, the proof is complete.

In Theorem 4 if $\mathbf{\Sigma}$ is such that $\underset{\sim}{N}(\mathbf{\Sigma}) \cap \mathscr{A}_0 = \{0\}$, then the hypothesis $\underset{\sim}{R}(\mathbf{W}) \oplus \underset{\sim}{N}(\mathbf{H}^*) = \mathscr{A}$ and the conditions in (6.2) are equivalent to

$$\underset{\sim}{R}(\mathbf{W}) = \mathscr{K}(\mathbf{\Sigma}) \quad \text{and} \quad \underset{\sim}{R}(\mathbf{H}) = \mathscr{E}.$$

A particular case for which this situation applies is when $\mathbf{\Sigma}$ is nonsingular so that $\mathbf{W}$ must satisfy $\underset{\sim}{R}(\mathbf{W}) = \underset{\sim}{R}(\mathbf{\Sigma}^{-1}\mathbf{H})$. Thus, an obvious choice is $\mathbf{W} = \mathbf{\Sigma}^{-1}\mathbf{H}$ so that $\mathbf{W}^*\mathbf{H} = \mathbf{H}^*\mathbf{\Sigma}^{-1}\mathbf{H}$ which is an expression analogous to the equations given by Aitken.

COROLLARY 4.1. *Consider a $\mu_\theta = \mathbf{H}\xi_\theta$ representation and let $\mathbf{\Sigma} \geqq 0$. If $\mathbf{W}$ is a linear operator into $\mathscr{A}$ such that*

$$\underset{\sim}{R}(\mathbf{W}) \oplus \underset{\sim}{N}(\mathbf{H}^*) = \mathscr{A} \quad \textit{and} \quad \underset{\sim}{R}(\mathbf{W}) \subset \mathscr{K}(\mathbf{\Sigma}), \tag{6.3}$$

then to any rv $\hat{\xi}$ *such that* $\mathbf{W}^*\mathbf{H}\hat{\xi} = \mathbf{W}^*Y$ *and to any $\lambda \in \underset{\sim}{R}(\mathbf{H}^*)$ it follows that* $\langle\lambda, \hat{\xi}\rangle$ is *$\mathbf{\Sigma}$-best for the parametric function $\langle\lambda, \xi_\theta\rangle = g(\theta)$.*

In Theorem 4 and Corollary 4.1 the condition $\underset{\sim}{R}(\mathbf{W}^*\mathbf{H}) = \underset{\sim}{R}(\mathbf{W}^*)$ is imposed to assure the existence of a $\hat{\xi}$ satisfying $\mathbf{W}^*\mathbf{H}\hat{\xi} = \mathbf{W}^*Y$. However, such a rv will exist provided the range of $\mathbf{W}^* Y$ is a subset of $\underset{\sim}{R}(\mathbf{W}^*\mathbf{H})$. Thus, Theorem 4 or Corollary 4.1 may be restated in a slightly more general setting. For example, in (6.3) replace $\underset{\sim}{R}(\mathbf{W}) \oplus \underset{\sim}{N}(\mathbf{H}^*) = \mathscr{A}$ by

$$\underset{\sim}{R}(\mathbf{W}) + \underset{\sim}{N}(\mathbf{H}^*) = \mathscr{A} \quad \text{and} \quad \{\mathbf{W}^* Y(u) : u \in \mathscr{U}\} \subset \underset{\sim}{R}(\mathbf{W}^*\mathbf{H}), \tag{6.4}$$

and the corollary still remains true.

EXAMPLE 5. Consider a fixed linear model with β restricted only through the covariance matrix. It is easily verified that a matrix W satisfies $\underset{\sim}{R}(W) \oplus N(X') = R^n$ and $\underset{\sim}{R}(W) \subset V^{-1}[\mathscr{E}]$ if and only if W is such that

$$\underset{\sim}{R}(VW) \subset \underset{\sim}{R}(X) \quad \text{and} \quad r(W'X) = r(W) = r(X). \tag{6.5}$$

If W satisfies (6.5), then Corollary 4.1 implies that a b.l.u. estimator for an $\mathscr{L}$-estimable $\lambda'\beta$ ($\lambda'\beta$ is $\mathscr{L}$-estimable if and only if $\lambda \in \underset{\sim}{R}(X')$) is given by $\lambda'\hat{\beta}$ for any $\hat{\beta}$ satisfying $W'X\hat{\beta} = W'Y$. Of the matrices W satisfying (6.5) the subset composed of matrices of the form $V^- X$ has been investigated by Zyskind and Martin [17]. It may also be noted that Y is in $\underset{\sim}{R}(X) + \underset{\sim}{R}(V)$ with probability one and thus (6.4) is satisfied with probability one if $\underset{\sim}{r}(X'W) = \underset{\sim}{r}(X)$ and $\underset{\sim}{R}(VW) \subset \underset{\sim}{R}(X)$. Thus, the conditions in (6.5) could be relaxed to $\underset{\sim}{R}(VW) \subset \underset{\sim}{R}(X)$ and $\underset{\sim}{r}(W'X) = \underset{\sim}{r}(X)$.

EXAMPLE 6. Consider the situation in the previous example and suppose $X\beta$ is partitioned in the form $X\beta = X_1\beta_1 + X_2\beta_2$ and that interest is in b.l.u. estimators for parametric functions of the form $\lambda'\beta_2$. Let T be any $n \times k$ matrix such that $\underset{\sim}{R}(T) = \underset{\sim}{N}(X_1')$, let $\mathscr{C} = R^k$, let $\overline{\mathscr{C}} = \{\rho' T' Y : \rho \in \mathscr{C}\}$, and note that

$$T'Y \sim F_k(T'X_2\beta_2, \sigma^2 T'VT, \Omega).$$

The choice of the matrix T leads to two interesting implications concerning the linear space $\overline{\mathscr{C}}$. First, it is easily seen that $\lambda'\beta_2$ is $\overline{\mathscr{C}}$-estimable if and only if $\lambda \in \underset{\sim}{R}(X_2'T)$ and thus if and only if $\lambda'\beta_2$ is $\mathscr{L}$-estimable. Second, from Theorem 2 it follows that $\rho \in \mathscr{C}$ is a $T'VT$-min element if and only if $T'VT\rho \in T'[\mathscr{E}]$; and if this last condition is true, then

$$VT\rho \in \underset{\sim}{R}(X_1) + \mathscr{E} \subset \underset{\sim}{R}(X).$$

Thus, Example 1 implies that $\overline{\mathscr{C}}$-best estimators are in fact b.l.u. estimators. Therefore, the linear space $\overline{\mathscr{C}}$ may be used for obtaining b.l.u. estimators for $\mathscr{L}$-estimable functions of the form $\lambda'\beta_2$. For example, using Corollary 4.1 as in Example 5 let W satisfy

$$\underset{\sim}{R}(T'VTW) \subset \underset{\sim}{R}(T'X_2) \quad \text{and} \quad \underset{\sim}{r}(W) = \underset{\sim}{r}(T'X_2) = r(W'T'X_2).$$

It follows that $\lambda'\hat{\beta}_2$ is a b.l.u. estimator for any $\mathscr{L}$-estimable $\lambda'\beta_2$ provided that $W'T'X_2\hat{\beta}_2 = W'T'Y$. In the event that $\underset{\sim}{r}(T'V) = \underset{\sim}{r}(T)$ the matrix W may be taken as $(T'VT)^- T'X_2$ and the equations reduce to

$$X_2'T(T'VT)^- T'X_2\hat{\beta}_2 = X_2'T(T'VT)^- T'Y.$$

It may be noted when $V = I$ that $T(T'T)^{-}T'$ is the projection on $N(X_1')$ along $\underline{R}(X_1)$ and that these last equations agree with those given for the same situation by Zyskind *et al.* [15].

7. $\mathscr{A}$-best estimators. Kruskal [4] and Zyskind [14] both use the notion of invariant subspaces for investigating when simple least squares estimators (i.e., b.l.u. estimators when $V = I$) are also b.l.u. estimators. Thomas [13] uses an equivalent notion to describe necessary and sufficient conditions under which b.l.u. estimators are the same for two different nonsingular covariance matrices. This question of when b.l.u. estimators are the same under different covariance structures is basically the same question as asking when $\mathscr{A}$-best estimators exist, and the results in [4] and [14] provide the basis for the present section. Additionally, Eaton [3] recently discussed when GM estimators (see [3]) exist for a multivariate linear model situation in essentially the same manner as pursued in this section, and Theorem 1.2 in [3] with slight modifications is the same as Corollary 5.2 below. We begin by stating the following lemma, the proof of which is obvious.

LEMMA 2. *Two linear operators* $\mathbf{T}$ *and* $\mathbf{\Sigma}$ *on* $\mathscr{A}$ *and a subspace* $\mathscr{M}$ *of* $\mathscr{A}$ *are such that* $\mathbf{T}[\mathscr{M}] \subset \mathbf{\Sigma}^{-1}[\mathscr{M}]$ *if and only if* $\mathscr{M}$ *is an invariant subspace of the linear operator* $\mathbf{\Sigma T}$.

Lemma 2 is quite interesting. For example, suppose $\mathscr{M}$ is a subspace of $\mathscr{A}$ such that $\underline{R}(\mathbf{\Sigma}_\theta) \cap \mathscr{M} \subset \mathscr{E}$ for all $\theta \in \Omega$ and that $\mathbf{T}$ is a linear operator on $\mathscr{A}$. If $\mathscr{M}$ is an invariant subspace of $\mathbf{\Sigma}_\theta \mathbf{T}$, then

$$\mathbf{T}[\mathscr{M}] \subset \mathbf{\Sigma}_\theta^{-1}[\mathscr{M}] = \mathbf{\Sigma}_\theta^{-1}[\underline{R}(\mathbf{\Sigma}_\theta) \cap \mathscr{M}] \subset \mathbf{\Sigma}_\theta^{-1}[\mathscr{E}];$$

and so, $\mathbf{T}[\mathscr{M}] \subset \mathscr{K}(\mathbf{\Sigma}_\theta)$. Thus, if $\mathscr{M}$ is an invariant subspace of $\mathbf{\Sigma}_\theta \mathbf{T}$ for all $\theta \in \Omega$ it follows that $\mathbf{T}[\mathscr{M}] \subset \mathscr{K}$. This last observation and what may be considered as a partial converse are given in the next theorem.

THEOREM 5. *To each subspace* $\mathscr{C} \subset \mathscr{K}$ *there exists a linear operator* $\mathbf{T}$ *such that* $\mathbf{T}[\mathscr{E}] \oplus \mathscr{C} \cap \mathscr{A}_0 = \mathscr{C}$ *and such that* $\mathscr{E}$ *is an invariant subspace of* $\mathbf{\Gamma T}$ *for all* $\mathbf{\Gamma} \in \mathscr{V}$. *Conversely, if* $\mathscr{M}$ *is a subspace of* $\mathscr{A}$ *and* $\mathbf{T}$ *is a linear operator on* $\mathscr{A}$ *such that* $\underline{R}(\mathbf{\Gamma}) \cap \mathscr{M} \subset \mathscr{E}$ *for all* $\mathbf{\Gamma} \in \mathscr{V}_1$ *(*$\mathscr{V}_1$ *forms a spanning set for* $\mathscr{V}$*) and such that* $\mathscr{M}$ *is an invariant subspace of* $\mathbf{\Gamma T}$ *for all* $\mathbf{\Gamma} \in \mathscr{V}_1$, *then* $\mathbf{T}[\mathscr{M}] \subset \mathscr{K}$.

PROOF. To establish the first statement let $\mathscr{B}$ be a subspace such that $\mathscr{C} = \mathscr{A}_0 \cap \mathscr{C} \oplus \mathscr{B}$. Assume $\mathscr{B} \neq \{0\}$($\mathscr{B} = \{0\}$ is trivial) and let $\{t_i\}$ be a basis for $\mathscr{B}$. For each i let $t_i = x_i + z_i$ where $x_i \in \mathscr{E}$ and $z_i \in \mathscr{A}_0$ and let $\{x_j'\}$ be such that $\{x_i, x_j'\}$ forms a basis for $\mathscr{E}$. Let $\mathbf{T}$ be any linear operator on $\mathscr{A}$ such that $\mathbf{T}x_i = t_i$ and $\mathbf{T}x_j' = 0$ and note that $\mathbf{T}[\mathscr{E}] = \mathscr{B}$. For $\theta \in \Omega$ and $x = \sum_i \alpha_i x_i + \sum_j \beta_j x_j'$ it follows that

$$\mathbf{\Sigma}_\theta \mathbf{T}x = \textstyle\sum_i \alpha_i \mathbf{\Sigma}_\theta \mathbf{T}x_i + \sum_j \beta_j \mathbf{\Sigma}_\theta \mathbf{T}x_j' = \sum_i \alpha_i \mathbf{\Sigma}_\theta t_i.$$

Since (t_i, Y) is $\mathscr{A}$-best it follows that $\mathbf{\Sigma}_\theta t_i \in \mathscr{E}$. Thus, $\mathscr{E}$ is an invariant subspace of $\mathbf{\Sigma}_\theta \mathbf{T}$ for all $\theta \in \Omega$ and hence $\mathscr{E}$ is an invariant subspace of $\mathbf{\Gamma T}$ for all $\mathbf{\Gamma} \in \mathscr{V}$. The converse follows immediately.

COROLLARY 5.1. *To every $\mathscr{A}$-estimable function there exists an $\mathscr{A}$-best estimator if and only if there exists a linear operator* $\mathbf{T}$ *such that* $\mathbf{T}[\mathscr{E}]+\mathscr{A}_0 = \mathscr{A}$ *and such that $\mathscr{E}$ is an invariant subspace of* $\mathbf{\Gamma T}$ *for all* $\mathbf{\Gamma}$ *in some spanning set for* $\mathscr{V}$.

COROLLARY 5.2. *Suppose there exists* $\mathbf{\Gamma}_0 \in \mathscr{V}$ *such that* $\mathbf{\Gamma}_0 \geqq 0$ *and such that* $\mathbf{\Gamma}_0$ *is invertible. To every $\mathscr{A}$-estimable function there exists an $\mathscr{A}$-best estimator if and only if $\mathscr{E}$ is an invariant subspace of* $\mathbf{\Gamma}\mathbf{\Gamma}_0^{-1}$ *for all* $\mathbf{\Gamma}$ *in some spanning set for* $\mathscr{V}$.

To illustrate the preceding, two examples follow. The first example considers when b.l.u. estimators exist in a mixed linear model situation and the second considers a relatively simple situation involving quadratic estimators. For the second example, i.e., Example 8, a more complete discussion may be found in [11] and for results under more general circumstances see [9]

EXAMPLE 7. Consider a mixed linear model with β restricted only through the covariance matrix and suppose there is a positive definite matrix $V_0 \in \mathscr{V}$. Clearly, $\mathscr{E} = \underline{R}(X)$ and Corollary 5.2 implies that to each $\mathscr{A}$-estimable $\lambda'\beta$ there exists a b.l.u. estimator if and only if $\underline{R}(X)$ is an invariant subspace of $V_i V_0^{-1}$ for $i = 1, 2, \cdots, m$.

EXAMPLE 8. Suppose U is an $n \times 1$ random vector distributed according to a multivariate normal with zero mean and covariance matrix $vV+\sigma^2 I$. Assume that $\Omega = \{(v, \sigma^2)\}$ contains a non-void open set and that $vV+\sigma^2 I$ is nonsingular for each $\theta \in \Omega$. Let $(\mathscr{A}, (-, -))$ denote the inner product space of $n \times n$ real symmetric matrices with the trace inner product and let $Y = UU'$ so that $\mathscr{A}$ is the linear space of quadratic estimators. For $\theta = (v, \sigma^2)$ note that $\mu_\theta = vV+\sigma^2 I$ and that

$$A \in \mathscr{A} \Rightarrow \Sigma_\theta A = 2\sigma^4 A + 2v\sigma^2(VA+AV)+2v^2 VAV.$$

Clearly, $\mathscr{E} = \text{sp}\{V, I\}$ and $\mathscr{V} = \text{sp}\{\mathbf{\Gamma}_1, \mathbf{\Gamma}_2, \mathbf{\Gamma}_3\}$ where $\mathbf{\Gamma}_1 A = A, \mathbf{\Gamma}_2 A = VA+AV$, and $\mathbf{\Gamma}_3 A = VAV$ for all $A \in \mathscr{A}$. Since $\mathbf{\Gamma}_1$ is the identity operator, Corollary 5.2 implies that to each $\mathscr{A}$-estimable function there exists an $\mathscr{A}$-best estimator if and only if $\mathscr{E}$ is an invariant subspace of $\mathbf{\Gamma}_2$ and $\mathbf{\Gamma}_3$. This condition is easily seen to be equivalent to $V^2 \in \mathscr{E}$. Moreover, it may be verified that $V^2 \in \mathscr{E}$ if and only if V has no more than two distinct eigenvalues. Therefore, a necessary and sufficient condition for an $\mathscr{A}$-best estimator to exist for each $\mathscr{A}$-estimable function is that $V^2 \in \mathscr{E}$ or equivalently that the matrix V have no more than two distinct eigenvalues.

REFERENCES

[1] AITKEN, A. C. (1934). On least squares and linear combinations of observations. *Proc. Roy. Soc. Edinburgh* **55** 42–48.

[2] BARANKIN, E. W. (1949). Locally best unbiased estimates. *Ann. Math. Statist.* **20** 477–501.

[3] EATON, M. L. (1970). Gauss–Markov estimation for multivariate linear models: A coordinate free approach. *Ann. Math. Statist.* **41** 528–538.

[4] KRUSKAL, W. (1968). When are Gauss–Markov and least squares estimators identical? A coordinate-free approach. *Ann. Math. Statist.* **39** 70–75.

[5] LEHMANN, E. L. and SCHEFFÉ, H. (1950). Completeness, similar regions and unbiased estimation—Part 1. *Sankhyā* **10** 305–340.

[6] RAO, C. R. (1965). *Linear Statistical Inference and Its Applications*. Wiley, New York.
[7] RAO, C. R. (1967). Least squares theory using an estimated dispersion matrix and its applications to measurement of signals. *Proc. Fifth Berkeley Symp. Math. Statist.* **1** 355–372. Univ. of California Press.
[8] SEELY, J. (1969). Estimation in finite-dimensional vector spaces with application to the mixed linear model. Ph.D. dissertation, Iowa State Univ.
[9] SEELY, J. (1969). Quadratic subspaces and completeness for a family of normal distributions. Technical Report No. 17, Department of Statistics, Oregon State Univ.
[10] SEELY, J. (1970). Linear spaces and unbiased estimation. *Ann. Math. Statist.* **41** 1725–1734.
[11] SEELY, J. and ZYSKIND, G. (1969). Linear spaces and minimum variance unbiased estimation. Technical Report No. 12, Department of Statistics, Oregon State Univ.
[12] STEIN, C. (1950). Unbiased estimators with minimum variance. *Ann. Math. Statist.* **21** 406–415.
[13] THOMAS, D. H. (1968). When do minimum variance estimators coincide? (abstract). *Ann. Math. Statist.* **39** 1365.
[14] 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.
[15] ZYSKIND, G., KEMPTHORNE, O., WHITE, R. F., DAYHOFF, E. E. and DOERFLER, T. R. (1964). Research on Analysis of Variance and Related Topics. Aerospace Research Laboratories, Technical Report 64-193, Wright–Patterson Air Force Base, Ohio.
[16] ZYSKIND, G. and MARTIN, F. (1966). A general Gauss–Markoff theorem in the case of any non-negative covariance matrix of observations (abstract). *Ann. Math. Statist.* **37** 1862.
[17] ZYSKIND, G. and MARTIN, F. (1969). On best linear estimation and a general Gauss–Markov theorem in linear models with arbitrary non-negative covariance structure. *SIAM J. Appl. Math.* **17** 1190–1202.