The General Linear Model and the Generalized Singular Value Decomposition

C. C. Paige
School of Computer Science
McGill University
805 Sherbrooke Street West
Montreal, Quebec, Canada H3A 2K6

Submitted by George P. H. Styan

ABSTRACT

A new derivation is given for the generalized singular value decomposition of two matrices X and F having the same number of rows. It is shown how this decomposition reveals the structure of the general Gauss-Markov linear model $(y, X\beta, \sigma^2 FF')$, and exhibits the structure and solution of the generalized linear least squares problem used to provide the best linear unbiased estimator for the model. The decomposition is used to prove optimality of the estimator and to reveal the structure of the covariance matrix of the error of the estimator.

1. INTRODUCTION

We are concerned here with two main objects: estimation in the general Gauss-Markov linear model (GLM), and the generalized singular value decomposition (GSVD). We will use the GLM to motivate the GSVD, and the GSVD to exhibit the structure and sensitivity of the GLM. The exposition deals only with theoretical properties of these objects, but some of the formulations and ideas are motivated by good computational practice, and reference will be made to this where appropriate.

The general linear model and the resulting generalized linear least squares problems are quite familiar, but because the generalized singular value decomposition is relatively recent and somewhat unfamiliar, care will be taken to present it clearly and to indicate its importance for general linear models. The development also leads to an easy derivation of what is known as the CS decomposition of an orthogonal matrix, and since this is a useful theoretical tool, it will be included for completeness.

52 Vanderbilt Ave., New York, NY 10017 0024-3795/85/$3.30

Here $X(n \times q)$ represents a known $n \times q$ matrix and X' its transpose, β is a fixed vector of q parameters to be estimated, and b and $\hat{b}$ are estimators for β. The symbols $\mathscr{E}(v)$, $\mathscr{V}(v)$ represent the mean and variance-covariance matrix (briefly, covariance) of a random vector v, while $\mathscr{R}(X)$ and $\mathscr{N}(X)$ represent the column space and null space respectively of the matrix X, and $\|X\|^2 = \text{trace}(X'X)$. We will restrict the exposition to real quantities, but everything translates trivially to complex quantities.

The general Gauss-Markov linear model can be written

$$y = X\beta + w, \qquad \mathscr{E}(w) = 0, \qquad \mathscr{V}(w) = \sigma^2 W = \sigma^2 FF'. \tag{1.1}$$

Here the vector y, the $n \times q$ matrix X, and the $n \times n$ symmetric nonnegative definite variance-covariance matrix W are known, while w is a random error vector, σ^2 is an unknown fixed scalar, and we wish to estimate the fixed vector of parameters β. If W has rank k, then the matrix F in (1.1) will be chosen to have linearly independent columns, so F will be $n \times k$. For computational reasons it is preferable to use F rather than W (see for example [1]), and so we replace (1.1) by

$$y = X\beta + Fu, \qquad \mathscr{E}(u) = 0, \qquad \mathscr{V}(u) = \sigma^2 I_k, \tag{1.2}$$

since the two are equivalent (see for example [2]). These properties of u will be expressed as $u \sim (0, \sigma^2 I)$.

When X has linearly independent columns, the estimator $\hat{b}$ of β in (1.2) that we will consider is the solution to the algebraic generalized linear least squares problem (GLLS) [3]

$$\underset{b,v}{\text{minimize}}\ v'v \qquad \text{subject to} \quad y = Xb + Fv, \tag{1.3}$$

for y, X, and F in (1.2). This estimator has error covariance

$$\mathscr{V}(\tilde{b}), \qquad \tilde{b} \triangleq \hat{b} - \beta. \tag{1.4}$$

For general F and X it is shown in [3] that the solution $\hat{b}$ to GLLS (1.3) leads to the best linear unbiased estimator (BLUE) for any estimable linear function of β in the GLM (1.2), and that GLLS has several important computational advantages.

The name "generalized linear least squares" is used to emphasize the relation of (1.3) to the linear least squares problem (LLS) obtained when $F = I$, and in LLS the singular value decomposition (SVD) of X provides a clear picture of the structure and sensitivity of (1.2), (1.3), and (1.4) when

$F = I$; see [4] to [6]. It can also be used to give a simple proof that $\hat{b}$ in GLLS is the BLUE for β in GLM (when $F = I$ and X has linearly independent columns).

The main purpose of this exposition is to introduce the generalized singular value decomposition (GSVD) of any two matrices F and X having the same number of rows, to motivate it via GLLS and the GLM, and to show how it provides the same sort of information for these that the SVD provides for LLS and the ordinary linear model. In Section 2 we will give a new derivation of the GSVD, and a quick proof of the closely related CS decomposition of an orthogonal matrix. In Section 3 we show how the GSVD relates to the GLM. The GSVD is used in Section 3.1 to exhibit the structure of the GLM, in Section 3.2 to solve GLLS, in Section 3.3 to obtain the error covariance, and in Section 3.5 to prove optimality of the GLLS estimator $\hat{b}$. Section 3.4 introduces some necessary material on linear unbiased estimators for use in Section 3.5.

2. THE GENERALIZED SINGULAR VALUE DECOMPOSITION

The derivation of the generalized singular value decomposition in [10] is not straightforward, and so we give a more direct presentation here. To motivate it as an obvious generalization of the singular value decomposition, we consider the following special case involving the general linear model.

When F is nonsingular we can write $F^{-1}y = F^{-1}X\beta + u$ in (1.2) and (1.3) (see for example [4] and [7]) and use the singular value decomposition of $F^{-1}X$ to exhibit some of the structure and sensitivity of the problem. Thus there exist orthogonal matrices U $(n \times n)$ and V $(q \times q)$ giving

$$U'F^{-1}XV = \underset{n \times q}{D} = \begin{pmatrix} D_1 & 0 \\ 0 & 0 \end{pmatrix}, \tag{2.1}$$

where D_1 $(r \times r)$ is a positive definite diagonal matrix with the same rank as X. When F is not square this is not possible, and using $F^{+}X$, where F^{+} is the Moore-Penrose pseudo-inverse, is theoretically invalid; see [8] and [9]. Even when F is nonsingular it is computationally risky to work with $F^{-1}X$ (see [3]), and so in general it makes sense to reformulate (2.1) with no inverse:

$$\begin{gathered} XV = FUD = [XV_1, 0] = [FU_1D_1, 0], \\ V = [V_1, V_2], \qquad U = [U_1, U_2], \end{gathered} \tag{2.2}$$

where U_1 and V_1 each have r columns.

This is essentially the generalized singular value decomposition of X and nonsingular F having the same number of rows [10], and in this case it is equivalent to the singular value decomposition of $F^{-1}X$. We see the orthogonal transformations have separated the column spaces of X and F into:

(i) a common column space $\mathscr{R}(FU_1) = \mathscr{R}(XV_1)$ of maximum dimension,
(ii) a maximum number of zero columns XV_2, and
(iii) the remaining $\mathscr{R}(FU_2)$ having no nonzero element in the common column space.

The key object is the common column space, and in it we have arranged for the corresponding columns of XV_1 and FU_1 to be parallel, with the ratio of column norms being the singular values of $F^{-1}X$. This can be done for any common column space; in fact we have the following.

LEMMA 2.1. *Let A and B be $n \times q$ matrices such that $\mathscr{R}(A) = \mathscr{R}(B)$ has dimension q. Then there exist orthogonal matrices U and V and positive diagonal $q \times q$ matrices S and C such that*

$$AUS = BVC, \qquad S^2 + C^2 = I.$$

Proof. We have $A = BG$ for unique $q \times q$ G with singular value decomposition $G = VDU'$, so $AU = BVD$. Taking $S^2 = (D^2 + I)^{-1}$ and $C = DS$ gives the result. ■

Clearly corresponding columns of AU and BV are parallel. We will now develop the generalized singular value decomposition for general X and F having the same number of rows. One form of this will be a decomposition like (2.2) with properties (i) (ii) and (iii) above, only now we will also have the equivalent of (ii) for F and the equivalent of (iii) for X.

Thus consider orthogonal U and V giving

$$FU = [FU_1, FU_2, FU_3], \qquad XV = [XV_1, XV_2, XV_3]. \tag{2.3}$$

These may be chosen so $FU_3 = 0$, $XV_1 = 0$, while $[FU_1, FU_2]$ has linearly independent columns, as does $[XV_2, XV_3]$. The U and V may also be chosen so that

$$\mathscr{R}(FU_2) = \mathscr{R}(XV_2) = \mathscr{R}(F) \cap \mathscr{R}(X) \tag{2.4}$$

is the common space. It was pointed out by an Editor that the decomposition of F in (2.3) corresponds to the decomposition of L in the unpublished Ph.D.

dissertation by Alalouf [21, Lemma 1.3.13, pp. 19–20], though the equivalent decomposition of X is not stated there.

Now from Lemma 2.1 we can also ensure $FU_2S = XV_2C$, giving

$$FU\begin{pmatrix} 0 & & \\ & S & \\ & & I \end{pmatrix} = XV\begin{pmatrix} I & & \\ & C & \\ & & 0 \end{pmatrix}, \tag{2.5}$$

$$S^2 + C^2 = I,$$

which is the generalization of (2.2). This could be taken as the generalized singular value decomposition of F and X; however, we will transform this into an equivalent form which will be more useful here. First we note the following.

Lemma 2.2. *If* $[A, B]$ *has linearly independent columns and* $[B, K]$ *has linearly independent columns, and*

$$\mathcal{R}(B) = \mathcal{R}([A, B]) \cap \mathcal{R}([B, K]), \tag{2.6}$$

then $[A, B, K]$ *has linearly independent columns.*

Proof. If $[A, B, K]$ has linearly dependent columns, then since A and $[B, K]$ each have linearly independent columns, there is a nonzero vector $z \in \mathcal{R}(A) \subset \mathcal{R}([A, B])$ which also satisfies $z \in \mathcal{R}[(B, K])$. But from (2.6) this means $z \in \mathcal{R}(B)$, the common column space, which is a contradiction, since $[A, B]$ has linearly independent columns. ∎

From this lemma and the choice of U and V in (2.3) [see (2.4)], we have that $[FU_1, FU_2, XV_3]$ has linearly independent columns, and so there exists an orthogonal matrix Q giving, with (2.5),

$$Q'\left[FU_1, FU_2C^{-1} = XV_2S^{-1}, XV_3\right] = \begin{pmatrix} R_1 & R_{12} & R_{13} \\ & R_2 & R_{23} \\ & & R_3 \\ & & 0 \end{pmatrix} = \begin{pmatrix} R \\ 0 \end{pmatrix} \tag{2.7}$$

where R is upper triangular and nonsingular with positive diagonal elements. By multiplying each of U_1 and V_3 by further orthogonal matrices and applying further orthogonal matrices from the left, we could also make R_1 and R_3 diagonal with positive elements. This leads to the following key result.

THEOREM 2.3. *For any real matrices $F(n\times k)$ and $X(n\times q)$ there exist orthogonal matrices $Q(n\times n)$, $U(k\times k)$, and $V(q\times q)$ so that*

$$Q'[F,X]\left(\begin{array}{c|c} U & 0 \\ 0 & V\end{array}\right)=\begin{pmatrix}R\\0\end{pmatrix}\left(\underbrace{\begin{matrix} I & & \\ & C & \\ & & 0\end{matrix}}_{k}\,\middle|\,\underbrace{\begin{matrix} 0 & & \\ & S & \\ & & I\end{matrix}}_{q}\right), \tag{2.8}$$

$$C^2+S^2=I,$$

where R is nonsingular and upper triangular with positive diagonal elements, and C and S are positive definite diagonal matrices. We can further choose R to have the form in (2.7) with diagonal R_1 and R_3.

This is the form of the generalized singular value decomposition we will use here. It has the same form as that in [10], and is more general and in a computationally more desirable form than the original decomposition in [11]. It can be computed using orthogonal transformations alone; see [12], [13], and [14]. The first two methods arise from the constructive derivation in [10], while the third approaches the problem more directly in a manner somewhat related to the development here.

The normalization in Lemma 2.1 has been chosen so that the rows of the matrix multiplying R in (2.8) are orthogonal, and so the singular values of R are the nonzero singular values of $[F, X]$. Because of this, Theorem 2.3 specializes to another very useful theoretical and computational tool, which we now exhibit. This is not used elsewhere in the paper, but we include it because it is so closely related to the generalized singular value decomposition and follows simply from the previous theorem, and because it is very useful in statistical as well as other areas; see for example [20].

THEOREM 2.4. *For any n rows $[P_{11}, P_{12}]$ of a $(k+q)$-square orthogonal matrix P, with P_{11} $n\times k$ and P_{12} $n\times q$, there exist orthogonal matrices Q $(n\times n)$, U $(k\times k)$, and V $(q\times q)$ such that*

$$Q'[P_{11},P_{12}]\begin{pmatrix} U & 0 \\ 0 & V\end{pmatrix}=\left(\underbrace{\begin{matrix} I & & \\ & C & \\ & & 0\end{matrix}}_{k}\,\middle|\,\underbrace{\begin{matrix} 0 & & \\ & S & \\ & & I\end{matrix}}_{q}\right), \tag{2.9}$$

$$C^2+S^2=I,$$

where C and S are positive definite diagonal matrices.

Proof. Putting $[F, X] = [P_{11}, P_{12}]$ in (2.8) and using $PP' = I$ gives $RR' = I_n$, so $R = I_n$ in Theorem 2.3, and the result follows. ∎

COROLLARY 2.5. *There also exists orthogonal $W(m \times m)$, $m = q + k - n$ such that*

$$\left(\begin{array}{c|c} Q' & 0 \\ \hline 0 & W' \end{array}\right) P \left(\begin{array}{c|c} U & 0 \\ 0 & V \end{array}\right) = \left(\begin{array}{ccc|ccc} I & & & 0 & & \\ & C & & & S & \\ & & 0 & & & I \\ \hline 0 & & & I & & \\ & S & & & -C & \\ & & I & & & 0 \end{array}\right). \tag{2.10}$$

Proof. From the transpose of (2.9), there exist orthogonal $\tilde{U}$, $\tilde{Q}$, and W such that

$$\left(\begin{array}{c|c} \tilde{Q}' & 0 \\ \hline 0 & W' \end{array}\right) \begin{pmatrix} P_{11} \\ P_{21} \end{pmatrix} \tilde{U} = \left(\begin{array}{ccc} I & & \\ & \tilde{C} & \\ & & 0 \\ \hline 0 & & \\ & \tilde{S} & \\ & & I \end{array}\right).$$

But in (2.9), $Q'P_{11}U$ gives the SVD of P_{11}, so $\tilde{Q}'P_{11}\tilde{U} = Q'P_{11}U$, and therefore $\tilde{C} = C$ and $\tilde{S} = S$, and there are orthogonal $\bar{U}$ and $\bar{Q}$ such that $\tilde{U}\bar{U} = U$, $\tilde{Q}\bar{Q} = Q$, leading to the first n rows and k columns in (2.10). The last block follows from orthogonality. ∎

This is called the CS decomposition [12], with C and S corresponding loosely to cosine and sine. This decomposition is implicit in [15] and was stated explicitly in [16] for the case $k = n$, and in [10] for the general case.

3. THE GSVD AND THE GENERAL LINEAR MODEL

We will show how effective the generalized singular value decomposition is in dealing with the general Gauss-Markov linear model. First we will see how it reveals the structure of the model, then how it can be used to solve the generalized linear least squares problem, which we know provides best linear unbiased estimators for the model. We will then use it to reveal the structure of the error covariance, and to give another proof that the generalized linear least squares problem does provide estimators with the required optimal properties. All these results are easy to derive and understand because the generalized singular value decomposition lays bare the essential elements quite clearly.

3.1. *Structure of the General Linear Model*

In order to reveal the structure in the model we apply the orthogonal transformations in (2.8) to the model (1.2) to give

$$Q'y = Q'XVV'\beta + Q'FUU'u, \qquad U'u \sim (0, \sigma^2 I), \tag{3.1}$$

which on writing $\beta_i = V_i'\beta$, $u_i = U_i'u$, becomes

$$\begin{pmatrix} Q_1'y \\ Q_2'y \end{pmatrix} = \begin{pmatrix} R \\ 0 \end{pmatrix} \left\{ \begin{pmatrix} 0 & & \\ & S & \\ & & I \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} I & & \\ & C & \\ & & 0 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \right\}. \tag{3.2}$$

It follows that $Q_2'y = 0$, corresponding to $y \in \mathcal{R}([X, F])$; otherwise the model is incorrect.

We then have from (2.7)

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \triangleq Q_1'y = \begin{pmatrix} R_1 & R_{12} & R_{13} \\ & R_2 & R_{23} \\ & & R_3 \end{pmatrix} \begin{pmatrix} u_1 \\ S\beta_2 + Cu_2 \\ \beta_3 \end{pmatrix},$$
$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \sim (0, \sigma^2 I), \tag{3.3}$$

which decomposes the model into the following structural elements:

(i) β_1 has no effect on y and cannot be estimated. Clearly β_2 and β_3 do affect y, and we write, partitioning V as in (3.1) and (3.2),

$$\beta = \beta^{(n)} + \beta^{(e)}, \qquad \beta^{(n)} = V_1\beta_1, \qquad \beta^{(e)} = V_2\beta_2 + V_3\beta_3 \tag{3.4}$$

and refer to $\beta^{(e)}$ as the *estimable part of* β, and $\beta^{(n)}$ as the nonestimable part. Of course from (2.8), $\mathcal{R}(V_1) = \mathcal{N}(X)$ and $\beta^{(n)} \in \mathcal{N}(X)$.

(ii) We have the *deterministic equation*

$$y_3 = R_3\beta_3, \tag{3.5}$$

where y_3 is not a random vector and so β_3 can be determined exactly. This is an important possibility in the general linear model which is not exhibited by the ordinary linear model. It is only possible if F does not have full row rank, that is, if W in (1.1) is singular.

(iii) β_2 can be estimated directly from

$$d_2 \triangleq R_2^{-1}(y_2 - R_{23}\beta_3) = S\beta_2 + Cu_2, \qquad u_2 \sim (0, \sigma^2 I), \tag{3.6}$$

where of course d_2 is known once y is observed.

(iv) The random vector u_1 can be found exactly by solving the system of equations in (3.3) once y is observed. Since $u_1 \sim (0, \sigma^2 I)$ this is a vector of uncorrelated regression residuals (see [7]) and can be used in estimating σ^2 or in testing for serial correlation; see for example [17].

(v) The random vector u_3 has no effect on y. Of course, if we insist that F have linearly independent columns, then (2.8) and (3.2) show u_3 has no elements.

The transformed model (3.3) also reveals the sensitivity of the observed y_i to the unknown β_i and random u_i, especially if we have arranged for R_1 and R_3 to be diagonal. If we assume this, then briefly we can see if R_3 has some very small singular values, the corresponding elements of β_3 will not affect y_3 strongly, and it will be correspondingly difficult to compute these elements accurately in the presence of additional measurement or rounding errors. Of course β_3 will usually affect y_2 and y_1, but only y_3 is used in computing β_3. Similarly, if R_1 has some very small singular values, then the corresponding elements of u_1 will not be strongly reflected in y_1, and it will be difficult to compute these elements accurately in the presence of errors. Of course β_3 etc. affect y_1 too, and inaccuracies here will also lead to inaccuracies in computing u_1. A similar argument holds for $d_2 = S\beta_2 + Cu_2$. On top of this we see that if S has some very small diagonal elements, then the corresponding elements of C will be close to unity, and the effect of some elements of β_2 on d_2 will tend to be very small compared with some elements of u_2.

3.2. *Solution of the Generalized Linear Least Squares Problem*

Since the linear constraints in (1.3) correspond exactly to the linear model we transformed in (3.1) to (3.3), the generalized singular value decomposition transforms the generalized linear least squares problem (1.3) to the following simpler form: minimize $v_1'v_1 + v_2'v_2$ subject to v_1, v_2, b_2, b_3 satisfying

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} R_1 & R_{12} & R_{13} \\ 0 & R_2 & R_{23} \\ 0 & 0 & R_3 \end{pmatrix} \begin{pmatrix} v_1 \\ Sb_2 + Cv_2 \\ b_3 \end{pmatrix}, \tag{3.7}$$

where we have assumed in (1.3) $y \in \mathscr{R}([X, F])$, so that a solution exists, and have written $b_i = V_i'b$, $v_i = U_i'v$.

Let $\hat{v}_1, \hat{v}_2, \hat{b}_2, \hat{b}_3$ solve (3.7). For $\hat{v} = V_1\hat{v}_1 + V_2\hat{v}_2 + V_3\hat{v}_3$ in (1.3) we take $\hat{v}_3 = 0$, and we note that $b_1 = V_1'b$, which cannot affect the outcome, does not appear in the formulation. Thus if we take $\hat{b}_1 = 0$, we will obtain the minimum 2-norm solution $\hat{b}$ in (1.3) with general X, and $\hat{b} = V_2\hat{b}_2 + V_3\hat{b}_3$ will then be an estimator for $\beta^{(e)}$ in (3.4). Next we see the constraints can be solved for $\hat{b}_3$,

$$d_2 = Sb_2 + Cv_2, \tag{3.8}$$

and $\hat{v}_1$, and these three vectors are then fixed. The only freedom we have in minimizing the sum of squares in (3.7) is in choosing b_2 and v_2 in (3.8). The solution to (3.7) is therefore given by

$$\hat{v}_2 = 0, \qquad S\hat{b}_2 = d_2, \tag{3.9}$$

so the generalized singular value decomposition makes the solution obvious. Of course there are more direct ways of solving the problem; see [3].

3.3. *The Error Covariance*

The generalized linear least squares problem (3.7) gave estimates $\hat{v}_1$, $\hat{v}_2 = 0$, $\hat{b}_2$, and $\hat{b}_3$ satisfying [see (3.3)]

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = R\begin{pmatrix} \hat{v}_1 \\ S\hat{b}_2 \\ \hat{b}_3 \end{pmatrix} = R\begin{pmatrix} u_1 \\ S\beta_2 + Cu_2 \\ \beta_3 \end{pmatrix}, \qquad \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \sim (0, \sigma^2 I), \tag{3.10}$$

which on being transformed by R^{-1} gives the important results

$$\begin{aligned} \hat{b}_3 &= \beta_3, & \hat{v}_1 &= u_1, \\ S(\hat{b}_2 - \beta_2) &= Cu_2, & u_2 &\sim (0, \sigma^2 I). \end{aligned} \tag{3.11}$$

Thus β_3 is given exactly, β_1 cannot be estimated, and the estimator $\hat{b}_2$ of β_2 has error

$$\tilde{b}_2 = \hat{b}_2 - \beta_2, \qquad \tilde{b}_2 \sim (0, \sigma^2 S^{-2} C^2). \tag{3.12}$$

The generalized singular value decomposition has therefore not only decomposed the space of $\beta = V_1\beta_1 + V_2\beta_2 + V_3\beta_3$ into its deterministic part $V_3\beta_3$, its nonestimable part $V_1\beta_1$, and its estimable part $V_2\beta_2$, but it has also decomposed the error $\tilde{b}_2$ of the generalized linear least squares estimator into uncorrelated elements for which the relative sizes of the variances are known.

This emphasizes a key property of the generalized singular value decomposition (2.8) of F and X, and that is that R in (2.8) contains the common part of F and X. Thus while u_2 in (3.10) leads to the random part of $\hat{b}_2$ via both X and F, R does not affect the sensitivity of $\hat{b}_2$ to u_2. Instead R cancels out, and only the positive diagonal matrices C and S are involved in the key relation (3.11) between u_2 and $\hat{b}_2$.

Of course, in the computation of $\hat{b}_2$ from

$$\begin{pmatrix} R_2 & R_{23} \\ 0 & R_3 \end{pmatrix} \begin{pmatrix} S\hat{b}_2 \\ \hat{b}_3 \end{pmatrix} = \begin{pmatrix} y_2 \\ y_3 \end{pmatrix}, \tag{3.13}$$

we see that part of R does affect the sensitivity of $\hat{b}_2$ to both rounding and additional measurement errors in y, and so the sensitivity of the numerical problem of computing $\hat{b}_2$ differs from the sensitivity of $\hat{b}_2$ to the random vector u_2.

3.4. *Unbiased Linear Estimators*

Some basic theory is needed in order to consider optimality of estimators. The concept of estimating estimable functions of the unknown vector β in the model (1.2) is standard and may help in a theory based for example on generalized inverses, but in the present analysis it is much simpler and quite sufficient to estimate the estimable part $\beta^{(e)}$ of β in (3.4) alone. Of course $\beta^{(e)}$ is an estimable linear function of β, but it is the only one we need consider here, and it contains everything we can find out about β from the linear model. Thus it will be sufficient to seek an unbiased estimator for $\beta^{(e)}$. For completeness we will restate the definition of an estimator and of an unbiased estimator in terms of $\beta^{(e)}$ alone.

DEFINITION 3.1. An *estimator* of the estimable part $\beta^{(e)}$ of the vector of q parameters β in the general linear model

$$y = X\beta + Fu, \qquad u \sim (0, \sigma^2 I)$$

$$\beta = \beta^{(e)} + \beta^{(n)}, \qquad \beta^{(e)} \in \mathscr{R}(X'), \quad \beta^{(n)} \in \mathscr{N}(X), \tag{3.14}$$

is any deterministic vector function of y having q elements. A linear estimator of $\beta^{(e)}$ is any linear function $b \triangleq a + Ay$ of y with fixed vector a and matrix A having q rows.

DEFINITION 3.2. An estimator b of the estimable part $\beta^{(e)}$ of β in (3.14) is said to be *unbiased* if

$$\mathscr{E}(b - \beta^{(e)}) = 0 \tag{3.15}$$

for all feasible $\beta^{(e)}$ in (3.14).

Here the last line emphasizes that (3.15) does not have to hold for all q-vectors $\beta^{(e)}$, only those of the form $\beta^{(e)} = V_2\beta_2 + V_3\beta_3$.

For the present we consider the solution of the generalized linear least squares problem. Let $\hat{b}_2$ and $\hat{b}_3$ be as given in (3.10); then since R is nonsingular, $V_2\hat{b}_2 + V_3\hat{b}_3$ is a linear estimator for $\beta^{(e)}$. When $\hat{b}$ gives the minimum 2-norm solution to (1.3), we have

$$\hat{b} = V_2\hat{b}_2 + V_3\hat{b}_3, \qquad \beta = V_1\beta_1 + \beta^{(e)}, \qquad \beta^{(e)} = V_2\beta_2 + V_3\beta_3, \tag{3.16}$$

and so from (3.11)

$$\hat{b} - \beta^{(e)} = V_2(\hat{b}_2 - \beta_2) = V_2 S^{-1} C u_2, \tag{3.17}$$
$$\mathscr{E}(\hat{b} - \beta^{(e)}) = 0,$$

and $\hat{b}$ is a linear unbiased estimator for $\beta^{(e)}$. If we choose any other solution to (1.3) then $\mathscr{E}(b - \beta^{(e)}) = V_1 b_1 \neq 0$ and we do not have an unbiased estimator.

We will want to compare $\hat{b}$ with other unbiased estimators, so let b be any linear estimator of $\beta^{(e)}$, then from Definition 3.1, (3.1) and (3.2), writing $\tilde{R} = R\,\mathrm{diag}(I, S, I)$,

$$V'b = a + A\begin{pmatrix} \tilde{R}^{-1} & 0 \\ 0 & I \end{pmatrix} Q'y = a + [A_1, A_2, A_3]\begin{pmatrix} u_1 \\ \beta_2 + S^{-1}Cu_2 \\ \beta_3 \end{pmatrix} \tag{3.18}$$

for some a and $A = [A_1, A_2, A_3, A_4]$ having q rows. Since $y \in \mathscr{R}([F, X])$,

A_4 does not contribute. It follows from Definition 3.2 that for b to be an unbiased estimator of $\beta^{(e)}$, one must have

$$\begin{aligned}\mathscr{E}(V'b) &= a + A_2\beta_2 + A_3\beta_3 = V'\beta^{(e)} \\ &= \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \begin{pmatrix} A_{12} & A_{13} \\ A_{22} & A_{23} \\ A_{32} & A_{33} \end{pmatrix}\begin{pmatrix} \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix}\end{aligned} \tag{3.19}$$

for all feasible $\beta^{(e)}$ in the model, that is, for all β_2 and β_3. It follows that $A_{22} = I$, $A_{33} = I$, and the remaining a_i and A_{ij} in (3.19) are zero, so that from (3.18) and (3.17).

$$\begin{aligned} b - \beta^{(e)} &= VA_1u_1 + V_2S^{-1}Cu_2, \qquad \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \sim (0, \sigma^2 I), \\ &= \hat{b} - \beta^{(e)} + VA_1u_1, \end{aligned} \tag{3.20}$$

and the choice of A_1 characterizes the different linear unbiased estimators b for $\beta^{(e)}$.

3.5. *Optimality of the Estimator*

Now that we have shown that the minimum 2-norm solution $\hat{b}$ of the generalized linear least squares problem (1.3) is a linear unbiased estimator for $\beta^{(e)}$ in (3.14), we will show that it is optimal. Consider the mean squared error of a linear estimator $b = a + Ay$ of $\beta^{(e)}$,

$$\mathscr{E}(\|b - \beta^{(e)}\|^2) = a'a + 2a'\mathscr{E}(Ay - \beta^{(e)}) + \mathscr{E}(\|Ay - \beta^{(e)}\|^2), \tag{3.21}$$

where by equating the derivative with respect to a to zero we get the standard result:

LEMMA 3.3. *The minimum mean squared error linear estimator is unbiased.*

Thus in seeking the minimum of (3.21) we need only consider estimators of the form b in (3.20). Such a b has error covariance

$$\begin{aligned}\mathscr{V}(b - \beta^{(e)}) &= VA_1A_1'V' + V_2S^{-2}C^2V_2' \\ &= \mathscr{V}(\hat{b} - \beta^{(e)}) + VA_1A_1'V', \end{aligned} \tag{3.22}$$

and all eigenvalues of this are minimized by taking $A_1 = 0$ in (3.20), which gives $b = \hat{b}$. Thus $\hat{b}$ is the minimum error covariance linear unbiased estimator. Now from (3.21) and (3.22)

$$\mathscr{E}(\|b - \beta^{(e)}\|^2) = \operatorname{trace} \mathscr{V}(b - \beta^{(e)})$$
$$= \mathscr{E}(\|\hat{b} - \beta^{(e)}\|^2) + \|A_1\|^2, \tag{3.23}$$

which again is minimized by taking $A_1 = 0$. We summarize these results as follows.

THEOREM 3.4. *The minimum 2-norm solution $\hat{b}$ to the generalized linear least squares problem*

$$\underset{b,\,v}{\text{minimize}}\ v'v \qquad \text{subject to} \quad y = Xb + Fv \tag{3.24}$$

is a linear unbiased estimator for the estimable part $\beta^{(e)}$ of the vector of parameters β in the general Gauss-Markov linear model

$$y = X\beta + Fu, \qquad u \sim (0, \sigma^2 I),$$
$$\beta = \beta^{(e)} + \beta^{(n)}, \qquad \beta^{(e)} \in \mathscr{R}(X'), \quad \beta^{(n)} \in \mathscr{N}(X). \tag{3.25}$$

$\hat{b}$ is the minimum mean squared error linear estimator for $\beta^{(e)}$, and for any other linear unbiased estimator b for $\beta^{(e)}$,

$$\mathscr{V}(b - \beta^{(e)}) - \mathscr{V}(\hat{b} - \beta^{(e)}) \tag{3.26}$$

is nonnegative definite.

It is useful to compare this result with the corresponding one in [3]. In that paper it is shown that if $c'\beta$ is an estimable linear function of β in (3.25), and b is *any* solution to (3.24), then $c'b$ is the best linear unbiased estimator for $c'\beta$. Since $c'\beta$ being estimable implies there exists a vector d such that $c' = d'X$, we see

$$c'b = c'\hat{b} \quad \text{and} \quad c'\beta = c'\beta^{(e)}.$$

It follows that here we take $c'\hat{b}$ as our estimator for $c'\beta$. A desirable aspect of the present approach of estimating $\beta^{(e)}$ is that it forces us to take the unique minimum 2-norm solution $\hat{b}$ of (3.24), whereas any solution will do if we are considering estimable linear functions $c'\beta$. It is not only preferable from a theoretical point of view to have a unique solution, but numerically it is wise to seek $\hat{b}$, since solutions b of large 2-norm are likely to have larger rounding errors which tend to remain after the inner product $c'b$ is computed.

4. CONCLUSION

We have given a straightforward derivation of the generalized singular value decomposition and shown its theoretical usefulness in exhibiting the structure of both the general Gauss-Markov linear model and the generalized linear least squares problem that provides an estimator for this model, as well as the error covariance of this estimator. We have also used it to prove optimality of this estimator.

Since the generalized singular value decomposition can be used to solve the generalized linear least squares problem, it can be used in practice to compute optimal estimates. However, computing the generalized singular value decomposition is necessarily an iterative process, and there are much quicker direct ways of solving GLLS; see for example [3]. These direct methods are very much faster if the matrices in the model have some sparsity structure (see for example [18]), or if we wish to update the solution by adding or discarding information (see for example [19]), and we would rarely consider the generalized singular value decomposition for such problems. Nevertheless, if the problem is not too large, then since computing power is generally so cheap, it will often make sense to consider using the generalized singular value decomposition for obtaining optimal estimates, at least when code for it becomes part of widely available and reliable subroutine packages. This is because the added information on structure and sensitivity that it provides can be very helpful in understanding the problem.

This work was supported by the National Sciences and Engineering Research Council of Canada under Grant A8652.

REFERENCES

1 G. H. Golub, Matrix decompositions and statistical calculations, in *Statistical Computations* (R. C. Milton and J. A. Nelder, Eds.), Academic, New York, 1969, pp. 365–397.

2 T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, 1958, pp. 25–26.

3 S. Kourouklis and C. C. Paige, A constrained least squares approach to the general Gauss-Markov linear model, *J. Amer. Statist. Assoc.* 76:620–625 (1981).
4 C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, N.J., 1974.
5 G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins U.P., Baltimore, 1983.
6 D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression Diagnostics*, Wiley, New York, 1980, Chapter 3.
7 G. H. Golub and G. P. H. Styan, Numerical computations for univariate linear models, *J. Statist. Comput. Simulation* 2:253–274 (1973).
8 G. Zyskind and F. B. Martin, On best linear estimation and a general Gauss-Markov theorem in linear models with arbitrary nonnegative covariance structure, *SIAM J. Appl. Math.* 17:1190–1202 (1969).
9 C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed., Wiley, New York, 1973, pp. 294–302.
10 C. C. Paige and M. A. Saunders, Towards a generalized singular value decomposition, *SIAM J. Numer. Anal.* 18:398–405 (1981).
11 C. F. Van Loan, Generalizing the singular value decomposition, *SIAM J. Numer. Anal.* 13:76–83 (1976).
12 G. W. Stewart, Computing the CS decomposition of a partitioned orthonormal matrix, *Numer. Math.* 40:297–306 (1982).
13 C. F. Van Loan, Computing the CS and generalized singular value decompositions, Report CS 604, Computer Science Dept., Cornell Univ., Ithaca, N.Y., 1984.
14 C. C. Paige, Computing the generalized singular value decomposition, *SIAM J. Sci. Stat. Comput.*, to appear.
15 C. Davis and W. M. Kahan, The rotation of eigenvectors by a perturbation. III, *SIAM J. Numer. Anal.* 7:1–46 (1970).
16 G. W. Stewart, On the perturbation of pseudo-inverses, projections and linear least squares problems, *SIAM Rev.* 19:634–662 (1977).
17 S. I. Grossman and G. P. H. Styan, Optimality properties of Theil's BLUS residuals, *J. Amer. Statist. Assoc.* 67:672–673 (1972).
18 C. C. Paige, Fast numerically stable computations for generalized linear least squares problems, *J. SIAM Numer. Anal.* 16:165–171 (1979).
19 C. C. Paige, Numerically stable computations for general univariate linear models, *Comm. Statist. B — Comput. and Simulation* 7:437–453 (1978).
20 C. Van Loan, On Stewart's singular value decomposition for partitioned orthogonal matrices, Report STAN-CS-79-767, Computer Science Dept., Stanford Univ., Stanford, Calif., 1979.
21 I. S. Alalouf, Estimability and testability in linear models, Ph.D. Thesis, Dept. of Mathematics, McGill Univ., Montréal, 1975.

Received 26 March 1984; revised 24 April 1985