NUMERICAL COMPUTATIONS FOR UNIVARIATE LINEAR MODELS

BY

GENE H. GOLUB

GEORGE P. H. STYAN

STAN-CS-236-71
SEPTEMBER 1971

COMPUTER SCIENCE DEPARTMENT
School of Humanities and Sciences
STANFORD UNIVERSITY

SU326 P30-13

Numerical Computations for Univariate Linear Models

by

Gene H. Golub*
Stanford University

and

George P. H. Styan
McGill University

Issued simultaneously as a report of the

Department of Statistics
Princeton University

* Research supported by the Atomic Energy Commission under grant No. AT(04-3)326, PA30, the National Science Foundation under grant No. GP 25912, and the Office of Naval Research under grant No. N00014-67 -A-D151-0017.

Numerical Computations for Univariate Linear Models

by

Gene H. Golub
Stanford University

and

George P. H. Styan
McGill University

Abstract

We consider the usual univariate linear model $E(y) = Xy$, $V(y) = \sigma^2 I$. In Part One of this paper X has full column rank, Numerically stable and efficient computational procedures are developed for the least squares estimation of y and the error sum of squares. We employ an orthogonal triangular decomposition of X using Householder transformations. A lower bound for the condition number of X is immediately obtained from this decomposition. Similar computational procedures are presented for the usual P-test of the general linear hypothesis $L'\gamma = 0$; $L'\gamma = m$ is also considered for $m \neq 0$. Updating techniques are given for adding to or removing from (X, y) a row, a set of rows or a column.

In Part Two, X has less than full rank. Least squares estimates are obtained using generalized inverses. The function $L'\gamma$ is estimable whenever it admits an unbiased estimator linear in y . We show how to computationally verify estimability of $L'\gamma$ and the equivalent testability of $L'y = 0$.

Key Words

Linear Models

Regression Analysis

Numerical Computations

Householder Transformations

Analysis of Variance

Matrix Analysis

Computer Algorithms

PART ONE: UNIVARIATE LINEAR MODEL WITH FULL RANK

1. Least squares estimation and error sum of squares

We consider the univariate general linear model

$$(1.1) \qquad E(\mathbf{y}) = \mathbf{X}\gamma \, ; \; V(\mathbf{y}) = \sigma^2 \mathbf{I} \, ,$$

where $E(\cdot)$ denotes mathematical expectation and $V(\cdot)$ the variance-covariance matrix. We take the design matrix $\mathbf{X}$ to be $n \times q$ of rank $q<n$ and known; in part two we relax this assumption of full column rank. The unknown vector γ of q regression coefficients is estimated by least squares from an observation $\mathbf{y}$ by minimizing the sum of squares

$$(1.2) \qquad (\mathbf{y} - \mathbf{X}\gamma)'(\mathbf{y} - \mathbf{X}\gamma) \, .$$

Prime denotes transposition; bold-face capital letters denote matrices and bold lower-case letters vectors, with rows always appearing primed.

In the case where $V(\mathbf{y}) = \sigma^2 \mathbf{A}$ in (1.1), with $\mathbf{A}$ known and positive definite, we may replace $\mathbf{y}$ by $\mathbf{Fy}$ and $\mathbf{X}$ by $\mathbf{FX}$ where $\mathbf{F}$ satisfies $\mathbf{FAF}' = \mathbf{I}$. The matrix $\mathbf{F}$ is not unique but it is possible to find an $\mathbf{F}$ which is lower triangular from the Cholesky decomposition of $\mathbf{A}$ (cf. e.g., Healy, 1968).

It is well known that the least squares estimate $\hat{\gamma}$ satisfies the normal equations

$$(1.3) \qquad \mathbf{X}'\mathbf{X}\hat{\gamma} = \mathbf{X}'\mathbf{y}$$

and is unique when $\mathbf{X}$ has full rank. The matrix $\mathbf{X}'\mathbf{X}$ is greatly influenced by roundoff errors and is often ill-conditioned: by this we mean that a relatively "small" change in X will induce a correspondingly "large" change in $(X'X)^{-1}$ and in the solution $\hat{\gamma} = (X'X)^{-1}X'y$ to (1.3). For these reasons we prefer to work with X directly rather than $X'X$ [cf. e.g., Longley (1967), Wampler (1969, 1970)].

It is possible to find an $n \times n$ orthogonal matrix P such that

$$(1.4) \qquad X = P\begin{pmatrix} R \\ 0 \end{pmatrix} \quad ; \quad P'X = \begin{pmatrix} R \\ 0 \end{pmatrix} ,$$

where R is upper triangular of order $q \times q$. This orthogonal triangular decomposition (OTD) may be made in various ways; a very stable numerical procedure (Golub, 1965) is to obtain P as the product of q Householder transformations.

A square matrix of the form $H = I - 2uu'$, where $u'u = 1$, is defined to be a Householder transformation. Clearly $H = H'$ and $HH' = H'H = H^2 = I$, so that H is a symmetric and orthogonal matrix. All but one of the characteristic roots of H are unity, the simple root being -1.

A vector x may be transformed by a Householder transformation to a vector with each element zero except for the first, i.e.,

$$(1.5) \qquad Hx = re_1 \; ; \quad r \neq 0 ,$$

say, where e_j is an $n \times 1$ vector with each component 0 except for the j-th which is 1 $(j = 1,2,\ldots,n)$. Premultiplying (1.5) by its transpose yields

$$(1.6) \qquad x'x = x'H'Hx = r^2 e_1' e_1 = r^2 .$$

Substituting $H = I - 2uu'$ in (1.5) gives

$$(1.7) \qquad x - 2(u'x)u = re_1 \quad ;$$

premultiplication by u' yields $-u'x = ru_1$, where $u_1 = e_1'u$, the first element in u . Substitution in (1.7) gives $x + 2ru_1u = re_1$, so that with $x = \{x_i\}$,

$$(1.8) \qquad 2u_1^2 = 1 - (x_1/r) \quad ; \quad 2u_i = -x_i/(ru_1) \quad , \quad i = 2,\ldots,n \ .$$

The first expression will always be computed positive if the square root of (1.6) is taken as

$$(1.9) \qquad r = -\text{sgn}(x_1)\cdot(x'x)^{1/2} \quad ,$$

where $\text{sgn}(x_1) = +1$ if $x_1 \geq 0$ and -1 otherwise. Then

$$(1.10) \qquad 2u_1^2 = 1 + (|x_1|/s) \quad ; \quad 2u_i = \text{sgn}(x_1)\cdot x_i/(su_1) \quad , \quad i = 2,\ldots,n \ ,$$

where

$$(1.11) \qquad s = +(x'x)^{1/2} \quad .$$

This gives $u'u = 1$, for $2u_i^2 = x_i^2/(2s^2u_1^2) = x_i^2/(s^2 + s|x_1|)$, $i = 2,\ldots,n$. Hence $2\sum_{i=2}^{n} u_i^2 = (s^2 - x_1^2)/(s^2 + s|x_1|) = 1 - (|x_1|/s) = 2(1-u_1^2)$. We note that H need not be computed explicitly as $Hx = x - 2(u'x)u$, for which we need only u and $u'x$ In the above form, it is necessary to compute two square roots per Householder transformation; if, however, we write $H = I - u(u'u)^{-1}u'$ then only one square root need be calculated (Businger and Golub, 1965;.

Applying this procedure with x replaced by $\underset{\sim}{X}\underset{\sim}{e}_1$, we obtain

$$(1.12) \qquad \underset{\sim}{H}\underset{\sim}{X} = (r_{11}\underset{\sim}{e}_1, \underset{\sim}{X}_1) \quad ,$$

where r_{11} replaces r , and $\underset{\sim}{X}_1$ is $n \times (q-1)$ such that $\underset{\sim}{X}_1\underset{\sim}{e}_j = \underset{\sim}{x}_{j+1} - 2(\underset{\sim}{u}'\underset{\sim}{x}_{j+1})\underset{\sim}{u}$, $j = 1,\ldots,q-1$ and $\underset{\sim}{x}_{j+1} = \underset{\sim}{X}\underset{\sim}{e}_{j+1}$. This procedure is now repeated with $\underset{\sim}{X}_1\underset{\sim}{e}_1$ as $\underset{\sim}{x}$ and a Householder transformation $\underset{\sim}{H}_1 = \underset{\sim}{I} - 2\underset{\sim}{u}_1\underset{\sim}{u}_1'$, say, with $\underset{\sim}{u}_1'\underset{\sim}{e}_1 = 0$. The last n-2 elements of $\underset{\sim}{X}_1\underset{\sim}{e}_1$ are now annihilated. So $\underset{\sim}{H}_1\underset{\sim}{H}\underset{\sim}{X}\underset{\sim}{e}_1 = r_{11}\underset{\sim}{H}_1\underset{\sim}{e}_1 = r_{11}\underset{\sim}{e}_1$, while $\underset{\sim}{H}_1\underset{\sim}{H}\underset{\sim}{X}\underset{\sim}{e}_2 = \underset{\sim}{H}_1\underset{\sim}{X}_1\underset{\sim}{e}_1$ has its last n-2 components zero. The product $\underset{\sim}{H}_1\underset{\sim}{H}$ is orthogonal. Further repetitions, annihilating at the j-th stage the last n-j elements in the j-th column of the matrix $\underset{\sim}{X}$ transformed previously by j-1 Householder transformations $(j = 1,\ldots,q)$, realizes P as the product of q Householder transformations. The matrix P is not computed explicitly. Details of this algorithm are given by Golub (1965), and Businger and Golub (1965) who also give a program in Algol 60.

Partitioning $P = (\underset{\sim}{P}_1, \underset{\sim}{P}_2)$, with $\underset{\sim}{P}_1$ $n \times q$ and $\underset{\sim}{P}_2$ $n \times (n-q)$ gives from (1.4)

$$(1.13) \qquad \underset{\sim}{P}_1'\underset{\sim}{X} = \underset{\sim}{R} \quad , \quad \underset{\sim}{P}_2'\underset{\sim}{X} = \underset{\sim}{0} \quad ,$$

with $\underset{\sim}{P}_1'\underset{\sim}{P}_1 = \underset{\sim}{I}_q$, $\underset{\sim}{P}_1'\underset{\sim}{P}_2 = \underset{\sim}{0}$ and $\underset{\sim}{P}_2'\underset{\sim}{P}_2 = \underset{\sim}{I}_{n-q}$, since $\underset{\sim}{P}'\underset{\sim}{P} = \underset{\sim}{I}_n$. If, in the above algorithm, we simultaneously apply the q Householder transformations to the observation vector $\underset{\sim}{y}$, then we have

$$(1.14) \quad \underset{\sim}{P}'\underset{\sim}{y} = \begin{pmatrix} \underset{\sim}{P}_1'\underset{\sim}{y} \\ \\ \underset{\sim}{P}_2'\underset{\sim}{y} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{z}_1 \\ \\ \underset{\sim}{z}_2 \end{pmatrix} = \underset{\sim}{z} \quad ,$$

say. Thus $\underset{\sim}{z}_2 = \underset{\sim}{P}_2'\underset{\sim}{y}$ has expectation $E(\underset{\sim}{P}_2'\underset{\sim}{y}) = \underset{\sim}{P}_2'\underset{\sim}{X}\underset{\sim}{\gamma} = \underset{\sim}{0}$ and covariance matrix $V(P_2'y) = \sigma^2 P_2'P_2 = \sigma^2 I_{n-q}$. Hence z_2 is an easily computed vector of uncorrelated regrossion residuals and may be used to test for serial correlation (cf. e.g., Grossman and Styan, 1970). It follows that

$$(1.15) \qquad P_2P_2' + X(X'X)^{-1}X' = I_n \quad ,$$

as each term on the left-hand side is idempotent and their cross-product is 0 ; their sum is idempotent with rank the sum of the ranks n-q and q. So $P_2P_2' = I - X(X'X)^{-1}X'$ and $z_2'z_2 = y'P_2P_2'y = (y - X\hat{\gamma})'(y - X\hat{\gamma})$ is the error sum of squares S_e , say -- the minimum of (1.2). It is simply computed here as the sum of squares of the n-q elements in $z_2 = P_2'y$.

The vector of (correlated) residuals $r = y - X\hat{\gamma}$ is often essential for analysis of the linear model (cf. e.g., Draper and Smith, 1966). Though the matrix P may not be computed explicitly it can be retrieved as the product of the Householder transformations when the corresponding q u vectors have been stored (which we recommend). Hence we compute $r = P_2 z_2$, since $P_2 z_2 = P_2 P_2'y = [I - X(X'X)^{-1}X']y = y - X\hat{\gamma} = r$. However, it has been observed by Gentleman (1970) that computing r in this fashion may be numerically unstable.

We also find from (1.4) that

$$(1.16) \qquad X'X = (R',0)P'P\begin{pmatrix} R \\ 0 \end{pmatrix} = R'R \quad .$$

Substitution in (1.3) yields $R'R\hat{\gamma} = (R',0)P'y = R'z_1$, so that solving

$$(1.17) \qquad R\hat{\gamma} = z_1$$

gives $\hat{\gamma}$. This is expedited by R being upper triangular.

We note that $\underset{\sim}{R}'\underset{\sim}{R}$ is a Cholesky factorization of $\underset{\sim}{X}'\underset{\sim}{X}$, for which Healy (1968) has given a Fortran program.

The estimator $\hat{\gamma}$ has covariance matrix $V(\hat{\gamma}) = \sigma^2(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$; an unbiased estimate is $S_e(\underset{\sim}{X}'\underset{\sim}{X})^{-1}/(n-q)$ which is easily computed using (1.16) as $(\underset{\sim}{z}_2'\underset{\sim}{z}_2)\underset{\sim}{R}^{-1}(\underset{\sim}{R}^{-1})'/(n-q)$. The generalized variance (cf. e.g., Anderson, 1958) is $|V(\underset{\sim}{\hat{\gamma}})|_I = \sigma^{2q}/|X'X|$, where $|\cdot|$ denotes determinant. In optimal design theory a problem is to choose $\underset{\sim}{X}$ so that $|\underset{\sim}{X}'\underset{\sim}{X}|$ is maximized thus reducing $|V(\underset{\sim}{\hat{\gamma}})|$ as much as possible. Again using (1.16) we see that $|\underset{\sim}{X}'\underset{\sim}{X}| = |\underset{\sim}{R}'\underset{\sim}{R}| = \prod_{i=1}^{q} r_{ii}^2$, as $\underset{\sim}{R}$ is upper triangular. Hence $|V(\hat{\gamma})|$ is estimated by $[\underset{\sim}{z}_2'\underset{\sim}{z}_2/(n-q)]^q / \prod_{i=1}^{q} r_{ii}^2$.

A measure of the ill-conditioning of a matrix is its condition number which we define as the ratio of the largest and smallest nonzero singular values of the matrix. The singular values of a (possibly rectangular) matrix A are the positive square roots of the characteristic roots of $\underset{\sim}{A}'\underset{\sim}{A}$ or $\underset{\sim}{A}\underset{\sim}{A}'$. When the condition number far exceeds the rank we find (cf. Wilkinson, 1967) that the matrix is extremely ill-conditioned.

A lower bound for the condition number $\kappa(X)$ of the design matrix $\underset{\sim}{X}$ is the ratio of the largest and smallest (in absolute value) diagonal elements of $\underset{\sim}{R}$. To see this we note first that X and $\underset{\sim}{P}'\underset{\sim}{X}$ have the same singular values, due to the orthogonality of $\underset{\sim}{P}$. As $\underset{\sim}{P}'\underset{\sim}{X}$ is merely $\underset{\sim}{R}$ bordered by zeroes, $sg(X) = sg(\underset{\sim}{R})$, where $sg(\cdot)$ denotes singular value. For any square matrix A of order $n \times n$,

$$(1.18) \qquad sg_n(\underset{\sim}{A}) \le |ch_j(\underset{\sim}{A})| \le sg_1(\underset{\sim}{A}) \; ; \; j = 1,\ldots,n$$

with $ch(\cdot)$ denoting characteristic root. The subscript j indicates j-th largest. To prove (1.18) when A has real roots, let $\lambda = ch_j(A)$ with $Av = \lambda v$. Then

(1.19) $$sg_1^2(A) = ch_1(A'A) = \max_x [x'A'Ax/x'x] \geq v'A'Av/v'v = $$

$$= \lambda v'A'v/v'v = \lambda^2 .$$

Similarly $sg_n^2(A) \leq \lambda^2$. Thus

(1.20) $$\kappa(X) = \frac{sg_1(X)}{sg_q(X)} = \frac{sg_1(R)}{sg_q(R)} \geq \frac{\max|ch(R)|}{\min|ch(R)|} = \frac{\max|r_{ii}|}{\min|r_{ii}|} .$$

Other properties of $\kappa(A)$ are given by Wilkinson (1967).

Why is the condition number important and how can we use the relationship (1.20)? Let $\tilde{\gamma}$ be the computed approximation to $\hat{\gamma}$ which satisfies (1.3). Suppose that we wish to determine an upper bound for the norm of the relative error of $\tilde{\gamma}$:

(1.21) $$\|\hat{\gamma} - \tilde{\gamma}\| / \|\hat{\gamma}\| ,$$

where $\|a\|$ indicates the Euclidean norm $(a'a)^{1/2}$. Define

(1.22) $$\tilde{r} = y - X\tilde{\gamma} ,$$

which we can compute quite accurately. Then

(1.23) $$r - \tilde{r} = -X(\hat{\gamma} - \gamma) ,$$

and hence

(1.24) $$-X'X(\hat{\gamma} - \tilde{\gamma}) = -X'\tilde{r} ,$$

since $\underset{\sim}{X}'\underset{\sim}{r} = 0$. Thus

$$(1.25) \qquad \|\hat{\underset{\sim}{\gamma}} - \tilde{\underset{\sim}{\gamma}}\| = \|(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\tilde{\underset{\sim}{r}}\| \leq ch_1[(\underset{\sim}{X}'\underset{\sim}{X})^{-1}]\|\underset{\sim}{X}'\tilde{\underset{\sim}{r}}\| = \|\underset{\sim}{X}'\tilde{\underset{\sim}{r}}\|/sg_q^2(\underset{\sim}{X}) .$$

From (1.3), $\|\underset{\sim}{X}'\underset{\sim}{X}\hat{\underset{\sim}{\gamma}}\| = \|\underset{\sim}{X}'\underset{\sim}{y}\|$, so that

$$(1.26) \qquad \|\underset{\sim}{X}'\underset{\sim}{y}\| \leq \|\hat{\gamma}\| sg_1^2(\underset{\sim}{X}) .$$

Combining (1.25) and (1.26), we have

$$(1.27) \qquad \|\hat{\underset{\sim}{\gamma}} - \tilde{\underset{\sim}{\gamma}}\| / \|\hat{\underset{\sim}{\gamma}}\| \leq [sg_1(\underset{\sim}{X}) / sg_q(\underset{\sim}{X})]^2 \|\underset{\sim}{X}'\tilde{\underset{\sim}{r}}\| / \|\underset{\sim}{X}'\underset{\sim}{y}\|$$

$$= \kappa^2(\underset{\sim}{X})\|\underset{\sim}{X}'\tilde{\underset{\sim}{r}}\| / \|\underset{\sim}{X}'\underset{\sim}{y}\| .$$

Thus we see that the condition number may be used for determining an upper bound for the relative error of $\|\tilde{\gamma}\|$. This upper bound is the product of two factors; the first of which, $\kappa^2(\underset{\sim}{X})$, is independent of $\underset{\sim}{y}$. However, the lower bound provided by (1.20) would in some circumstances give insight into the relative error. Hence, if

$$(1.28) \qquad [\max|r_{ii}| / \min|r_{ii}|]^2 \|\underset{\sim}{X}'\tilde{\underset{\sim}{r}}\| / \|\underset{\sim}{X}'\underset{\sim}{y}\|$$

is large, then it is likely that the relative error in $\|\tilde{\gamma}\|$ is large.

The numerical efficiency of the above orthogonal triangular decomposition is enhanced (cf. Golub, 1965) if the column selected for each of the q Householder transformations maximizes the corresponding sum of squares. That is, at the j-th stage (j = 1,...,q) we transform that column of the q-j+1 possibilities which maximizes the sum of squares of its last n-j+1 components. The interchanges may be summarized in a permutation matrix $\underset{\sim}{\Pi}$ postmultiplying X . Thus (1.4) becomes

$$(1.29)\quad \underset{\sim}{X} = \begin{pmatrix} \underset{\sim}{R} \\ \underset{\sim}{0} \end{pmatrix} \pi' \; ; \; P'X\pi = \begin{pmatrix} \underset{\sim}{R} \\ \underset{\sim}{0} \end{pmatrix} .$$

The vector z does not change and hence neither does S_e . The solution (1.17) changes however; substituting (1.29) into (1.3) now gives $\pi R' R \pi' \hat{\gamma} = \pi R' z_1$, so that

$$(1.30)\quad \underset{\sim}{R}(\pi'\hat{\gamma}) = \underset{\sim}{z}_1 = R\underset{\sim}{\theta} ,$$

is solved for θ , and $\hat{\gamma} = \pi\theta$. As these interchanges only rearrange the r_{ii} we still find $|X'X| = \prod_{i=1}^{q} r_{ii}^2$. The lower bound for the condition number simplifies, however, as with these interchanges $\max|r_{ii}| = |r_{11}|$, and $\min|r_{ii}| = |r_{qq}|$ so that $\kappa(X) \geq |r_{11}/r_{qq}|$.

Given the $n \times n$ matrix

$$(1.31)\quad A = \begin{bmatrix} 1, & -1, & -1, & \ldots, & -1 \\ 0, & 1, & -1, & \ldots, & -1 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix} ,$$

we see that $\max|r_{ii}| = \min|r_{ii}| = 1$, and so $\kappa(\underset{\sim}{A}) \geq 1$, since $\underset{\sim}{A} = \underset{\sim}{R}$ when no column interchanges are made. However, if column interchanges are performed then for n = 10 say, $|r_{11}| \doteq 3.162$, $|r_{nn}| \doteq .003383$ and $\kappa(\underset{\sim}{A}) \geq 934.8$. The actual value of $\kappa(\underset{\sim}{A}) = 1918.5$.

The For-ban IV programs LLSQ and DLLSQ (double-precision) in the Scientific Subroutine Package (SSP) of IBM (1968) solve the least squares problem as described above. The SSP library is available at many IBM 360 computing centers. The SSP manual gives a write-up of the procedure and indicates how $\hat{\gamma}$ and S_e are output. In addition we note that the q diagonal elements of R are output as 'AUX(q+1,...,2q)', with $\max|r_{ii}|$ = AUX(q+1) and $\min|r_{ii}|$ = AUX(2q) in absolute value. The remaining nonzero elements of R are overwritten in corresponding positions of X (input as 'A'). The vector z is overwritten on Y (input as 'B') and S_e appears in 'AUX(1)'. The solution $\hat{\gamma}$ is output as 'X'.

The number of multiplications to obtain R is about $nq^2 - q^3/3$, whereas approximately $nq^2/2$ multiplications are required to form the normal equations (1.3) with about $q^3/6$ multiplications needed to solve them. Thus when n-q is small, the number of operations is roughly the same for both algorithms, but when n-q is large, it requires about twice as many operations to use the orthogonalization procedure.

The orthogonal triangular decomposition (1.4) or (1.29) is very similar to the Gram-Schmidt decomposition. Indeed if n = q and there is no roundoff error and all r_{ii} are taken positive, then the Householder and Gram-Schmidt algorithms yield precisely the same transformation. Although the modified Gram-Schmidt process (cf. e.g., Golub, 1969) may be used for solving linear least squares problems, the computed vectors may not be truly orthogonal! The Householder transformations, however, yield vectors which are more nearly orthogonal (Wilkinson, 1965). Furthermore, not only do the first q columns of P span the same space as the columns of X, but the last n-q columns of P span the complement of the space spanned by the columns of X. As we have seen above, this is quite useful.

2. Hypothesis testing and estimation under constraints

Let us consider the general linear hypothesis

$$(2.1)\qquad L'_N \gamma_H = 0$$

for the linear model of Section 1. The contrast matrix L' is taken as $s \times q$ of full row rank $s \leq q$. If we assume that y is normally distributed then $L'\hat{\gamma}$ is $N(L'\gamma, \sigma^2 L'(X'X)^{-1}L)$, with $\hat{\gamma} = (X'X)^{-1}X'y$. The numerator of the usual F-test for (2.1) is then well known to be

$$(2.2)\qquad \hat{\gamma}'L[L'(X'X)^{-1}L]^{-1}L'\hat{\gamma} = S_h ,$$

say, the "hypothesis sum of squares". Substituting (1.16) and (1.17) into (2.2) gives

$$(2.3)\qquad S_h = z_1'(R^{-1})'L[L'R^{-1}(R^{-1})'L]^{-1}L'R^{-1}z_1 .$$

We compute $(R^{-1})'L = G$, say, by solving $R'G = L$, with R' lower triangular. We then obtain an orthogonal triangular decomposition of G, $q \times s$ $(q \geq s)$,

$$(2.4)\qquad G = (R^{-1})'L = Q\begin{pmatrix} B \\ 0 \end{pmatrix} ,$$

say, where B is upper triangular $s \times s$ and the orthogonal matrix Q is the product of s Householder transformations. Then $G'G = B'B$; partitioning $Q = (Q_1, Q_2)$, where Q_1 is $q \times s$ and Q_2 $q \times (q-s)$ gives $G = Q_1B$ from (2.4). Substitution in (2.3) yields

$$(2.5)\qquad S_h = z_1'Q_1Q_1'z_1$$

which we compute by applying the s Householder transformations of (2.4) to $\underset{\sim}{z}_1$ simultaneously with G and then summing the squares of the first s components of the transformed $\underset{\sim}{z}_1$.

If we test the hypothesis

$$(2.6) \qquad \underset{\sim}{L}'\underset{\sim}{\gamma} = m ,$$

where $\underset{\sim}{m}$ is a given $s \times 1$ vector, not necessarily 0 , then we proceed by computing $\underset{\sim}{L}'\hat{\underset{\sim}{\gamma}} - \underset{\sim}{m} = \underset{\sim}{h}$, say, and sum the squares of the components of $(\underset{\sim}{B}^{-1})'\underset{\sim}{h}$; we find the latter by solving $\underset{\sim}{L}'\hat{\underset{\sim}{\gamma}} - \underset{\sim}{m} = \underset{\sim}{h} = \underset{\sim}{B}'\underset{\sim}{t}$, say, for $\underset{\sim}{t}$, with B' lower triangular.

The described procedure can be improved upon when $s > q-s$. We first obtain an orthogonal triangular decomposition of $\underset{\sim}{L}$,

$$(2.7) \qquad \underset{\sim}{L} = \underset{\sim}{T}\begin{pmatrix} \underset{\sim}{U} \\ \underset{\sim}{0} \end{pmatrix}$$

say, where T is orthogonal and U upper triangular. Partitioning $T = (\underset{\sim}{T}_1, \underset{\sim}{T}_2)$, where $\underset{\sim}{T}_1$ is $q \times s$ and $\underset{\sim}{T}_2$ is $q \times (q-s)$ leads to

$$(2.8) \qquad \underset{\sim}{L}'\underset{\sim}{T}_1 = \underset{\sim}{U}' \; ; \; \underset{\sim}{L}'\underset{\sim}{T}_2 = 0 .$$

Thus $\underset{\sim}{L}'\underset{\sim}{y} = \underset{\sim}{0}$ if and only if $\underset{\sim}{y} = \underset{\sim}{T}_2\underset{\sim}{\theta}$ for some $\underset{\sim}{\theta}$, now unconstrained. Hence

$$(2.9) \qquad \min_{\underset{\sim}{L}'\underset{\sim}{\gamma} = \underset{\sim}{0}} (\underset{\sim}{y} - X\gamma)'(\underset{\sim}{y} - X\gamma) = \min_{\underset{\sim}{\theta}} (\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{T}_2\underset{\sim}{\theta})'(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{T}_2\underset{\sim}{\theta}) .$$

Using (1.4) and (1.14), we see that (2.9) reduces to

$$(2.10) \qquad \min_{\underset{\sim}{\theta}} (\underset{\sim}{z}_1 - \underset{\sim}{R}\underset{\sim}{T}_2\underset{\sim}{\theta})'(\underset{\sim}{z}_1 - \underset{\sim}{R}\underset{\sim}{T}_2\underset{\sim}{\theta}) + \underset{\sim}{z}_2'\underset{\sim}{z}_2 ,$$

so that S_h equals the first term in (2.10) which is easily computed as in Section 1 with z_1 replacing y and RT_2 replacing X . Since (cf. e.g., Good (1965), p. 89),

$$(2.11) \qquad sg_1(XT) \geq sg_1(XT_2) \ ,$$

$$(2.12) \qquad sg_q(XT) \leq sg_{q-s+1}(XT) \leq sg_{q-s}(XT_2) \ ,$$

we have

$$(2.13) \qquad \kappa(XT_2) \leq \kappa(XT) = u(x) \ .$$

Thus, by eliminating the constraints, the linear least squares problem may become better conditioned.

The least squares estimate y^* , say, of y subject to $L'\gamma = 0$ is obtained from the solution $\hat{\theta}$ to (2.10) by

$$(2.14) \qquad \gamma^* = T_2\hat{\theta} \ .$$

If the constraints have nonnull righthand side m as in (2.9) then the procedure is changed as follows. Evidently $L'\gamma = m$ holds if and only if $y = T_2\theta + T_1(U^{-1})'m = T_2\theta + T_1w$, say. We obtain w by solving $m = U'w$, with U' lower triangular. Thus y is replaced by $y - XT_1w$ and hence z_1 by $z_1 - RT_1w$ the resulting value of S_h is therefore

$$(2.15) \qquad \min_{\theta}(z_1 - RT_1w - RT_2\theta)'(z_1 - RT_1w - RT_2\theta)$$

which we compute as in Section 1 with $z_1 - RT_1w$ replacing y and RT_2 replacing X .

The relevant F-test for the hypotheses (2.1) or (2.6) is then computed as

$$(2.16) \qquad F = \frac{S_h/s}{S_e/(n-q)} ,$$

with the critical region formed by values of (2.16) exceeding the corresponding tabulated value of F with s and n−q degrees of freedom.

In some special, though common, situations the above computations simplify considerably.

If we test a single contrast in $\underset{\sim}{y}$ equal to 0 we obtain (2.1) with s = 1 . Let us write this as

$$(2.17) \qquad \underset{\sim}{\ell}'\underset{\sim}{\gamma} = 0 .$$

A particular case might be testing a single regression coefficient equal to 0 . Then $(\underset{\sim}{R}^{-1})'L = \underset{\sim}{K}$ becomes $(\underset{\sim}{R}^{-1})'\underset{\sim}{\ell} = \underset{\sim}{k}$, say, found by solving $\underset{\sim}{\ell} = \underset{\sim}{R}'\underset{\sim}{k}$ as before. Then (2.3) becomes

$$(2.18) \qquad (\underset{\sim}{\ell}'\hat{\underset{\sim}{\gamma}})^2 / \underset{\sim}{k}'\underset{\sim}{k} = S_h ,$$

and we compute the denominator in (2.18) by summing squares of components in $\underset{\sim}{k}$. The one-sided t-test for

$$(2.19) \qquad \underset{\sim}{\ell}'\underset{\sim}{\gamma} > 0$$

has critical region large positive values of $\underset{\sim}{\ell}'\hat{\underset{\sim}{\gamma}} / [\underset{\sim}{k}'\underset{\sim}{k}\, S_e/(n-q)]^{1/2}$.

Another special case occurs with s = q-1 when $\underset{\sim}{L}'\underset{\sim}{\gamma} = \underset{\sim}{0}$ if and only if

$$(2.20) \qquad \underset{\sim}{\gamma} = \underset{\sim}{\theta t} ,$$

where θ is now a scalar. The vector $\underset{\sim}{t}$ is often found upon inspection (without transforming $\underset{\sim}{L}$). For example in testing for homogeneity of coefficients of $\underset{\sim}{y}$, we have $t = \underset{\sim}{e}$, the vector with each component unity. Substituting t for $\underset{\sim}{T}_2$ in (2.10) yields

(2.21) $$\hat{\theta} = \underset{\sim}{z}_1'\underset{\sim}{R}\underset{\sim}{t} / \underset{\sim}{t}'\underset{\sim}{R}'\underset{\sim}{R}\underset{\sim}{t} ,$$

and

(2.22) $$S_h = \underset{\sim}{z}_1'\underset{\sim}{z}_1 - (\underset{\sim}{z}_1\underset{\sim}{R}\underset{\sim}{t})^2 / \underset{\sim}{t}'\underset{\sim}{R}'\underset{\sim}{R}\underset{\sim}{t} ,$$

with the denominators computed by summing squares of elements of $\underset{\sim}{R}\underset{\sim}{t}$.

3. Updating procedures

After a particular set of data has been analyzed it is often pertinent to add to or remove from $\underset{\sim}{X}$ and $\underset{\sim}{y}$ a row (or set of rows or to add to or remove from $\underset{\sim}{X}$ a column. This happens when new information becomes available or when existing experimental units have been classified as extreme, or independent variables insignificant.

We begin by considering the addition of data from m , say, further experimental units. Let $\underset{\sim}{X}_m$ and $\underset{\sim}{y}_m$ be the corresponding data of order mxq and mxl respectively. Following (1.4) and (1.14) we may write

$$(3.1) \qquad \begin{pmatrix} \underset{\sim}{I}_m & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{P}' \end{pmatrix} \begin{pmatrix} \underset{\sim}{X}_m & \underset{\sim}{y}_m \\ \underset{\sim}{X} & \underset{\sim}{y} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{X}_m & \underset{\sim}{y}_m \\ \underset{\sim}{R} & \underset{\sim}{z}_1 \\ \underset{\sim}{0} & \underset{\sim}{z}_2 \end{pmatrix} .$$

Applying q Householder transformations of order m+q to the first m+q rows of (3.1) yields

$$(3.2) \qquad \begin{pmatrix} \underset{\sim}{X}_m & \underset{\sim}{y}_m \\ \underset{\sim}{R} & \underset{\sim}{z}_1 \end{pmatrix} = \underset{\sim}{P}_1 \begin{pmatrix} \underset{\sim}{R}_1 & \underset{\sim}{z}_0^* \\ \underset{\sim}{0} & \underset{\sim}{z}_1^* \end{pmatrix} ,$$

say, where $\underset{\sim}{R}_1$ is qxq upper triangular, $\underset{\sim}{z}_0^*$ is qxl and $\underset{\sim}{z}_1^*$ is m x 1 . Hence

$$(3.3) \qquad \underset{\sim}{P}_2' \begin{pmatrix} \underset{\sim}{X}_m & \underset{\sim}{y}_m \\ \underset{\sim}{X} & \underset{\sim}{y} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{R}_1 & \underset{\sim}{z}_0^* \\ \underset{\sim}{0} & \underset{\sim}{z}_1^* \\ \underset{\sim}{0} & \underset{\sim}{z}_2 \end{pmatrix}$$

where

$$(3.4) \qquad P_2' = \begin{pmatrix} P_1' & 0 \\ 0 & I_{n-q} \end{pmatrix} \begin{pmatrix} I_m & 0 \\ 0 & P' \end{pmatrix}$$

is an orthogonal matrix formed from 2q Householder transformations, and has order m+n . The new residual sum of squares is $z_1^{*\prime} z_1^* + z_2' z_2$, i.e., the previous sum of squares, $z_2' z_2$, augmented by the sum of squares of the m components of z_1^* ,these components themselves give m additional uncorrelated residuals.

Next, suppose we wish to add a (q+1)-th variable whose n values constitute a vector x . We first compute $P'x$ by applying in turn the q Householder transformations determined by the stored vectors u (cf. residual calculations in Section 1). We need then only one further Householder transformation, H , say, of order n−q to annihilate the last n−q−1 elements in $P'x$, i.e.,

$$(3.5) \qquad \begin{pmatrix} I_q & 0 \\ 0 & H \end{pmatrix} P'(X,x) = \begin{pmatrix} R & P_1'x \\ 0 & HP_2'x \end{pmatrix} = \begin{pmatrix} R & P_1'x \\ 0 & he_1 \end{pmatrix} ,$$

where $P = (P_1, P_2)$, as in §1, and $h = x'P_2P_2'x$ -- the sum of squares of the last n−q components of $P'x$.

The procedure for removing an experimental unit is more complicated. The method given previously by Golub and Saunders (1970), may under certain circumstances prove unstable. We now give a new method which should provide a more accurate solution.

Suppose we want to remove x_i' , the i-th row of X . We seek an upper triangular matrix S , say, so that

(3.6) $$X'X - x_i x_i' = R'R - x_i x_i' = S'S = R'(I - tt')R ,$$

say, where $R't = x_i$; the vector t is easily computed since R' is lower triangular. We now construct an orthogonal matrix Q so that $Qt = ce_1$; thus $c^2 = t't = x_i'(R'R)^{-1}x_i = e_i'X(X'X)^{-1}X'e_i < 1$. We define the quasi-diagonal matrices of order qxq :

(3.7) $$Z_k = \begin{bmatrix} I_{k-1} & & \\ & \Theta_k & \\ & & I_{q-k-1} \end{bmatrix} \quad ; \quad k = 1,\ldots,q-1 ,$$

where

(3.8) $$\Theta_k = \begin{pmatrix} \cos\theta_k , & \sin\theta_k \\ -\sin\theta_k , & \cos\theta_k \end{pmatrix} \quad ; \quad k = 1,\ldots,q-1 .$$

Clearly Z_k and Θ_k are orthogonal. Let

(3.9) $$(R_\ell , t_\ell) = Z_{q-\ell}(R_{\ell-1}, t_{\ell-1}) \; ; \; \ell = 1,\ldots,q-1 ,$$

with $t_0 = t$ and $R_0 = R$. We choose θ_k so that $Z_{q-\ell}$ annihilates $e_{q-\ell+1}'t_{\ell-1}$ and hence $e_{q-\ell+1}'t_\ell = 0$; $\ell = 1,\ldots,q-1$. Then the matrix

(3.10) $$Q = Z_1 Z_2 \cdots Z_{q-1}$$

satisfies $Qt = ce_1$ and is orthogonal. From (3.6) we may write

(3.11) $S'S = R'Q'(I - c^2 e_1 e_1') Q R$,

which is positive definite if and only if $c^2 > 1$. It follows that

(3.12) $$QR = W = \begin{vmatrix} w_{11}, & w_{12}, & \cdots & w_{1,q-1}, & w_{1q} \\ w_{21}, & w_{22}, & \cdots & w_{2,q-1}, & w_{2q} \\ 0, & w_{32}, & \cdots & w_{3,q-1}, & w_{3q} \\ \vdots & \vdots & & \vdots & \vdots \\ \cdot & \cdot & \cdots & w_{q,q-1}, & w_{qq} \end{vmatrix}$$

is an upper Hessenberg matrix. Thus (3.11) becomes $S'S = W'D^2W$, with

(3.13) $$D = \begin{pmatrix} (1-c^2)^{1/2} & 0' \\ 0 & I_{q-1} \end{pmatrix} ,$$

which is real when $c^2 < 1$. We compute S by applying orthogonal transformations to the upper Hessenberg matrix DW . Let

(3.14) $S_k = Z_k^* S_{k-1}$; $k = 1, \ldots, q-1$,

with $S_0 = DW$ and Z_k^* formed as Z_k in (3.7) but with θ_k^* replacing θ_k and so chosen that Z_k^* annihilates $e_{k+1}' S_k e_k = e_{k+1}' D W e_k$ and thus $e_{k+1}' S_k e_k = 0$. Then

(3.15) $S = S_{q-1} = Z_{q-1}^* Z_{q-2}^* \cdots Z_2^* Z_1^* D W$

This procedure requires about $9q^2/2$ multiplications and $2q-1$ square roots.

The above algorithm can also be used for adding an observation but about twice as many numerical operations are required as in the procedure given by (3.3) and (3.4). We also note that the problem of deleting an observation is numerically delicate. Since

$$S'S = R'(I - tt')R \ , \tag{3.16}$$

it follows that

$$\kappa(S) \leq \kappa(R) / (1 - t't)^{1/2} \ . \tag{3.17}$$

Thus if $t't$ is close to 1 , then $\kappa(S)$ could be quite large as the right-hand side of (3.17) is attainable.

Finally suppose we wish to remove an independent variable or column of x . If it is the last then no further calculations are required; but suppose it is the first. Let

$$R = \begin{pmatrix} r_{11} & r_{12} \dots & r_{1q} \\ 0 & r_{22} \dots & r_{2q} \\ \vdots & \vdots & \vdots \\ 0 & \dots & r_{qq} \end{pmatrix} = (r_{11}e_1 \ , \bar{R}) \tag{3.18}$$

where $\bar{R}$ is $q \times (q-1)$ and has one more row than an upper Hessenberg matrix. We annihilate the elements just below the main diagonal of $\bar{R}$, i.e., r_{22}, [illegible], by applying orthogonal transformations of the type (3.7) with

$$R_k = Z_k \ R_{k-1} \ ; \qquad k = 1,\dots,q-1 \ , \tag{3.19}$$

and $\underset{\sim}{R}_0 = \bar{\underset{\sim}{R}}$ we choose Θ_k in $\underset{\sim}{Z}_k$ so that $\underset{\sim}{e}'_{k+1} \underset{\sim}{R}_{k-1} \underset{\sim}{e}_k = r_{k+1,k+1}$ is annihilated; thus $\underset{\sim}{e}'_{k+1} \underset{\sim}{R}_k \underset{\sim}{e}_k = 0$ and $\underset{\sim}{R}_{q-1}$ is the new triangular matrix sought.

PART Two: UNIVARIATE LINEAR MODEL WITH LESS THAN FULL RANK

4. Least squares estimation and error sum of squares

We consider now the univariate general linear model (1.1),

$$(4.1) \qquad E(y) = X\gamma \; , \quad V(y) = \sigma^2 I \; ,$$

with the design matrix X of rank $r < q < n$. We obtain the same normal equations as (1.3),

$$(4.2) \qquad X'X\hat{\gamma} = X'\gamma$$

which are consistent; their solution, however, may not be unique. Consider a solution to (4.2) which we may write

$$(4.3) \qquad \hat{\gamma} = (X'X)^{-}X'y \; ,$$

where $(\cdot)^{-}$ denotes generalized inverse. We follow Pringle and Rayner (1971) and define a generalized inverse of a matrix A_m , $m \times n$, as any matrix A^- satisfying

$$(4.4) \qquad AA^-A = A \; .$$

Evidently A^- has order $n \times m$. Such a generalized inverse exists but is not unique in general; if, however, A^- satisfies (4.4) and

$$(4.5) \qquad A^-AA^- = A^- \; ,$$

$$(4.6) \qquad (AA^-)' = AA^- \; ,$$

$$(4.7) \qquad (A^-A)' = A^-A \; ,$$

then we write $\underset{\sim}{A}^- = \underset{\sim}{A}'$, <u>the pseudo-inverse</u> of $\underset{\sim}{A}$. When we only require that (4.4) is satisfied we will write $\underset{\sim}{A}^- = g_1(\underset{\sim}{A})$ -- a <u>g_1-inverse</u> of $\underset{\sim}{A}$. Similarly when (4.4) and (4.5) are satisfied, $\underset{\sim}{A}^- = g_{12}(\underset{\sim}{A})$; (4.4), (4.5), and (4.6): $\underset{\sim}{A}^- = g_{123}(\underset{\sim}{A})$. The pseudo-inverse $\underset{\sim}{A}^+ = g_{1234}(\underset{\sim}{A})$. The solution $\hat{\underset{\sim}{\gamma}}_0$, say, to (4.3) which minimizes $\hat{\underset{\sim}{\gamma}}'\hat{\underset{\sim}{\gamma}}$ equals $\underset{\sim}{X}^+\underset{\sim}{y}$ as is shown, for example, by Peters and Wilkinson (1970). Our concern, however, focuses more on estimable functions of $\underset{\sim}{y}$, rather than $\underset{\sim}{y}$ <u>per se</u> so we will not discuss here computation of $\hat{\underset{\sim}{\gamma}}_0$. We define an <u>estimable function</u> of $\underset{\sim}{Y}$ as a vector $\underset{\sim}{L}'\underset{\sim}{\gamma}$ which admits an unbiased estimator of the form $\underset{\sim}{K}'\underset{\sim}{y}$, where $\underset{\sim}{L}'$ is $s \times q$, say, and $\underset{\sim}{K}'$, $s \times n$. The least squares estimate is then $\underset{\sim}{L}'\hat{\underset{\sim}{\gamma}} = \underset{\sim}{L}'(\underset{\sim}{X}'\underset{\sim}{X})^-\underset{\sim}{X}'\underset{\sim}{y}$ so that $\underset{\sim}{K}' = \underset{\sim}{L}'(\underset{\sim}{X}'\underset{\sim}{X})^-\underset{\sim}{X}'$. We shall see (Section 5) that when $\underset{\sim}{L}'\underset{\sim}{y}$ is estimable, $\underset{\sim}{L}'(\underset{\sim}{X}'\underset{\sim}{X})^-\underset{\sim}{X}'$ is unique for all $(\underset{\sim}{X}*\underset{\sim}{X})' = g_1(\underset{\sim}{X}'\underset{\sim}{X})$. Rather than form $\underset{\sim}{X}*\underset{\sim}{X}$, find a $g_1(\underset{\sim}{X}'\underset{\sim}{X})$ and then postmultiply it by X' , we compute a $g_{123}(\underset{\sim}{X})$ directly, noting that $\underset{\sim}{G}$ is a $g_1(\underset{\sim}{A})$ if and only if it can be written as $(\underset{\sim}{A}'\underset{\sim}{A})^-\underset{\sim}{A}'$ for some $g_1(\underset{\sim}{A}'\underset{\sim}{A}) = (\underset{\sim}{A}'\underset{\sim}{A})^-$ [Pringle and Rayner (1971), p. 26].

We proceed as in Section 1 to orthogonally transform $\underset{\sim}{X}$ by Householder transformations with column interchanges. If $\underset{\sim}{X}$ has rank r then after r Householder transformations we obtain, cf. (1.29),

$$(4.8) \qquad \underset{\sim}{X} = \underset{\sim}{P}\begin{pmatrix} \underset{\sim}{R} & \underset{\sim}{S} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix}\underset{\sim}{\Pi}' \quad ; \quad \underset{\sim}{P}'\underset{\sim}{X}\underset{\sim}{\Pi} = \begin{pmatrix} \underset{\sim}{R} & \underset{\sim}{S} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix} ,$$

where $\underset{\sim}{R}$ is upper triangular, $r \times r$, $\underset{\sim}{S}$ is $r \times (q-r)$, and $\underset{\sim}{\Pi}$ is a permutation matrix of order $q \times q$. We now claim that

$$(4.9) \qquad \underset{\sim}{X}^* = \underset{\sim}{\Pi}\begin{pmatrix} \underset{\sim}{R}^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix}\underset{\sim}{P}' = g_{123}(\underset{\sim}{X}) .$$

We have $\underset{\sim}{X}\underset{\sim}{X}^* = \underset{\sim}{P}\begin{pmatrix} \underset{\sim}{I} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix}\underset{\sim}{P}'$, clearly symmetric. Hence $\underset{\sim}{X}\underset{\sim}{X}^*\underset{\sim}{X} = \underset{\sim}{P}\begin{pmatrix} \underset{\sim}{R} & \underset{\sim}{S} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix}\underset{\sim}{\Pi}' = \underset{\sim}{X}$,

while $\underset{\sim}{X}^*\underset{\sim}{X}\underset{\sim}{X}^* = \underset{\sim}{\Pi}\begin{pmatrix} \underset{\sim}{R}^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix}\underset{\sim}{P}' = \underset{\sim}{X}^*$ so that (4.9) is proved. The solution

$\hat{\gamma} = \underset{\sim}{X}^*\underset{\sim}{y}$ to (4.2) afforded by (4.9) is often called a basic solution as it contains at most q-r nonzero elements.

Thus (4.9) accomodates our purposes; moreover we do not have a stronger g-inverse than is needed. As in Section 1 we partition $P = (\underset{\sim}{P}_1, \underset{\sim}{P}_2)$, but now let $\underset{\sim}{P}_1$ be nx r and $\underset{\sim}{P}_2$ nx(n-r). From (4.8) it follows, cf. (1.13), that

(4.10) $\underset{\sim}{P}_1'\underset{\sim}{X}\underset{\sim}{\Pi} = (\underset{\sim}{R}, \underset{\sim}{S})$

(4.11) $\underset{\sim}{P}_2'\underset{\sim}{X} = \underset{\sim}{0}$.

Following (1.14) we now write

(4.12) $\underset{\sim}{P}'\underset{\sim}{y} = \begin{pmatrix} \underset{\sim}{P}_1'\underset{\sim}{y} \\ \underset{\sim}{P}_2'\underset{\sim}{y} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{z}_1 \\ \underset{\sim}{z}_2 \end{pmatrix} = \underset{\sim}{z}$,

where $\underset{\sim}{z}_1$ is now $r \times 1$ and $\underset{\sim}{z}_2$ (n-r) x1 . Thus $\underset{\sim}{z}_2$ is again a vector of uncorrelated residuals; moreover

(4.13) $\underset{\sim}{P}_2\underset{\sim}{P}_2' + \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^-\underset{\sim}{X}' = \underset{\sim}{I}_n$,

as in (1.15), with $\underset{\sim}{P}_2\underset{\sim}{P}_2'$ idempotent rank n-r and $\underset{\sim}{X}(\underset{\sim}{X}^*\underset{\sim}{X})^-\underset{\sim}{X}^*$ symmetric idempotent rank r . By (4.11) their cross-product is $\underset{\sim}{0}$ and so their sum is idempotent rank (n-r)+r = n and hence $\underset{\sim}{I}_n$ as claimed. Thus

(4.14) $$\underset{\sim}{z}_2' \underset{\sim}{z}_2 = \underset{\sim}{y}'(\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}')\underset{\sim}{y}$$

is the residual sum of squares, computed as the sum of squares of the $n-r$ components in $\underset{\sim}{z}_2$.

The vector of (correlated) residuals $\underset{\sim}{r} = \underset{\sim}{y} - \underset{\sim}{X}\hat{\underset{\sim}{\gamma}} = (\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}')\underset{\sim}{y} = \underset{\sim}{P}_2 \underset{\sim}{P}_2' \underset{\sim}{y}$ as in Section 1, and using (4.13) it follows that (4.14) equals $\underset{\sim}{r}'\underset{\sim}{r}$.

5. Estimating estimable functions and testing testable hypotheses

As mentioned in Section 4 we are not directly concerned with the estimation per se of y. We define $L'\gamma$ to be an estimable function of γ whenever it admits an unbiased estimator which is linear in y, $K'y$, say. Thus

$$(5.1) \qquad L'\gamma = E(K'y) = K'X\gamma$$

holds for all γ. Hence

$$(5.2) \qquad L' = K'X \ .$$

As in Section 3 we take L' to be $s \times q$, but now relax the assumption of full row rank taking $r(L) = t \leq r$. We obtain

$$(5.3) \qquad r\begin{pmatrix} L' \\ X \end{pmatrix} = r(X) \ ,$$

directly from (5.2). Substituting (4.8) into (5.3) gives

$$(5.4) \qquad r\begin{bmatrix} L'\Pi \\ P\begin{pmatrix} R, S \\ 0, 0 \end{pmatrix} \end{bmatrix} \quad r\begin{pmatrix} L_1', L_2' \\ R \ , \ S \end{pmatrix} = r(R) = r(X) = r,$$

where we partition

$$(5.5) \qquad L'\Pi = (L_1', L_2') \ ,$$

with L_1' $s \times r$, and L_2' $s \times (q-r)$. The matrix $L'\Pi$ is the contrast matrix L' with its columns permuted according to the interchanges which rearrange the columns of X to make the first r columns linearly independent. Then L_1' are the corresponding r columns of L' or $L'\Pi$.

We apply $v \geq r$ Householder transformations of order s+r, whose product is V', say, so that

$$(5.6) \qquad V'\begin{pmatrix} L_1', & L_2' \\ R, & S \end{pmatrix}\Pi_1 = \begin{pmatrix} T & U \\ 0 & 0 \end{pmatrix} ,$$

where Π_1 is a permutation matrix, and T is upper triangular $v \times v$. If (5.6) is achieved at the r-th stage, i.e., v = r, then $L_{UN}'\gamma$ is estimable. If not, then $L'\gamma$ is not estimable.

An alternative procedure which is often easy to verify theoretically follows and is included for completeness.

THEOREM 5.1. The function $L'\gamma$ is estimable if and only if

$$(5.7) \qquad L'(X'X)^{-}X'X = L'$$

for any $(X'X)^{-} = g_1(X'X)$.

Proof. We show that (5.2) and (5.7) are equivalent. Clearly (5.7) implies 5.2); conversely

$$(5.8) \qquad L'(X'X)^{-}X'X = K'X(X'X)^{-}X'X = K'X = L' ,$$

since $X(X^*X)^{-}X^*X = X$ [cf. Pringle and Rayner (1971), p. 26].

Q.E.D.

We may use (5.7) to computationally verify estimability as follows. Substituting (4.8) and (4.9) into (5.7), with $X^* = (X'X)^{-}X'$ gives

$$(5.9) \qquad L'\Pi \begin{pmatrix} I & R^{-1}S \\ 0 & 0 \end{pmatrix} = L' .$$

Substituting (5.5) into (5.9) yields

$$(5.10) \qquad L_1' R^{-1} S = L_2'$$

To verify (5.10), therefore, we solve $RW = S$ for W, say, which equals $R^{-1}S$, with R upper triangular. We then examine $L_1'W - L_2'$ and if close enough to 0 conclude $L'\gamma$ estimable.

For the remainder of this section we will assume $L'\gamma$ estimable. From (4.3),

$$(5.11) \qquad L'\hat{\gamma} = L'(X'X)^{-}X'y = L'X^{*}y ,$$

where $X^{*} = (X^{*}X)^{-}X' = g_{123}(X)$, cf. (4.9). Thus

$$(5.12) \qquad L'\hat{\gamma} = L'\Pi \begin{pmatrix} R^{-1} & 0 \\ 0 & 0 \end{pmatrix} P'y = L_1' R^{-1} z_1 ,$$

using (4.12) and (5.5). We compute $L'\hat{\gamma}$, therefore, by solving $Rw = z_1$ for w, say, which equals $R^{-1}z_1$, with R upper triangular. We then premultiply by L_1' which contains the r columns of L' corresponding to the r linearly independent columns of X which yielded R. We note that $L'\hat{\gamma}$ is uniquely determined by (5.11) for any $(X^{*}X)^{-} = g_1(X'X)$. To see this, set $L' = K'X$ from (5.2), so that $L'(X'X)^{-}X' = K'X(X'X)^{-}X' = K'X(X'X)^{+}X' = L'(X'X)^{+}X'$, since $X(X^{*}X)^{-}X^{*}$ is unique [cf. Pringle and Rayner (1971), p. 25].

We define the general linear hypothesis

(5.13) $\quad \underset{\sim}{L}'\underset{\sim}{y} = \underset{\sim}{0}$

as testable whenever $\underset{\sim}{L}'\underset{\sim}{\gamma}$ is estimable. The numerator of the usual F-test for testable (5.13) is then, cf. (2.2),

(5.14) $\quad \hat{\underset{\sim}{\gamma}}'\underset{\sim}{L}[\underset{\sim}{L}'(\underset{UN}{X'X})^{-}\underset{\sim}{L}]^{-}\underset{\sim}{L}'\hat{\underset{\sim}{\gamma}} = S_h$.

To see that (5.14) is invariant over choices of $(\underset{\sim}{X}^*\underset{\sim}{X})^-$, notice that $\underset{\sim}{L}'(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{L} = K'X(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{K} = K'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{+}\underset{\sim}{X}'\underset{\sim}{K} = \underset{\sim}{L}'(\underset{\sim}{X}'\underset{\sim}{X})^{+}\underset{\sim}{L}$ from (5.2). Moreover, (5.14) is also-invariant over choices of $[\underset{\sim}{L}'(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{L}]^{-}$; writing $\underset{\sim}{X}^* = (\underline{x^*x})^{-}\underset{\sim}{X}'$ we find that (5.14) may be written

(5.15) $\quad \underset{\sim}{y}'(\underset{\sim}{X}^*)'\underset{\sim}{L}[\underset{\sim}{L}'\underset{\sim}{X}^*(\underset{\sim}{X}^*)'\underset{\sim}{L}]^{-}\underset{\sim}{L}'\underset{\sim}{X}^*\underset{\sim}{y} = S_h$,

using (5.7) and (5.11). Sh is uniquely defined since for any $\underset{\sim}{A}$, $\underset{\sim}{A}(\underset{\sim}{A}'\underset{\sim}{A})^{-}\underset{\sim}{A}'$ is unique [cf. Pringle and Rayner (1971), p. 25].

To compute S_h we see from (4.9) and (5.11) that (5.15) *may* be written

(5.16) $\quad S_h = \underset{\sim}{z}_1'(\underset{\sim}{R}^{-1})'\underset{\sim}{L}_1[\underset{\sim}{L}_1'\underset{\sim}{R}^{-1}(\underset{\sim}{R}^{-1})'\underset{\sim}{L}_1]^{-}\underset{\sim}{L}_1'\underset{\sim}{R}^{-1}\underset{\sim}{z}_1$.

We obtain an orthogonal triangular decomposition of

(5.17) $$G = (\underset{\sim}{R}^{-1})'\underset{\sim}{L}_1 = \underset{\sim}{Q}\begin{pmatrix} \underset{\sim}{B} & \underset{\sim}{C} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix}\underset{\sim}{\pi}_2' ,$$

say, where $\underset{\sim}{B}$ is upper triangular $t \times t$, with $t = r(\underset{\sim}{L}) = r(\underset{\sim}{L}_1)$ by (5.10). . The orthogonal matrix $\underset{\sim}{Q}$ is the product of t Householder transformations, while the permutation matrix $\underset{\sim}{\pi}_2$ rearranges the columns of $\underset{\sim}{L}_1$, $r \times s$, to make the first t linearly independent. Substituting (5.17) into (5.16) yields

$$S_h = \underset{\sim}{z}_1' \underset{\sim}{G} \underset{\sim}{G}^* \underset{\sim}{z}_1 , \tag{5.18}$$

where $\underset{\sim}{G}^* = g_{123}(\underset{\sim}{G})$ is given by

$$\underset{\sim}{G}^* = \underset{\sim}{\Pi}_2 \begin{pmatrix} \underset{\sim}{B}^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix} \underset{\sim}{Q}' \tag{5.19}$$

We partition $\underset{\sim}{Q} = (\underset{\sim}{Q}_1, \underset{\sim}{Q}_2)$, where $\underset{\sim}{Q}_1$ is $r \times t$ and $\underset{\sim}{Q}_2$ $r \times (r-t)$. [If $t = r$, $\underset{\sim}{Q}_1 \equiv \underset{\sim}{Q}$.] Then (5.18) reduces to

$$S_h = \underset{\sim}{z}_1' \underset{\sim}{Q}_1 \underset{\sim}{Q}_1' \underset{\sim}{z}_1 , \tag{5.20}$$

as at (2.5). We compute (5.20) by applying the t Householder transformations of $\underset{\sim}{Q}$ in (5.17) to $\underset{\sim}{z}_1$ simultaneously with $\underset{\sim}{G}$ and then summing the squares of the first t components of the transformed $\underset{\sim}{z}_1$.

If we test the hypothesis

$$\underset{\sim}{L}' \underset{\sim}{\gamma} = m \tag{5.21}$$

and L' is $s \times q$ with row rank $t < s$ then $\underset{\sim}{m}$ must satisfy the same s-t restrictions that apply to the rows of $\underset{\sim}{L}'$, i.e., (5.21) must be <u>consistent</u>. Then the numerator sum of squares is uniquely given by

$$(\hat{\underset{\sim}{\gamma}}' \underset{\sim}{L} - \underset{\sim}{m}')[\underset{\sim}{L}'(\underset{\sim}{X}'\underset{\sim}{X})^- \underset{\sim}{L}]^- (\underset{\sim}{L}' \hat{\underset{\sim}{\gamma}} - \underset{\sim}{m}) = S_h ; \tag{5.22}$$

following (5.15) and (5.16) we see that

$$\underset{\sim}{L}'(\underset{\sim}{X}'\underset{\sim}{X})^- \underset{\sim}{L} = \underset{\sim}{L}_1' \underset{\sim}{R}^{-1} (\underset{\sim}{R}^{-1})' \underset{\sim}{L}_1 = \underset{\sim}{G}'\underset{\sim}{G} \tag{5.23}$$

for which we want a gl-inverse. We use

<u>LEMMA 5.1</u>. If $A'' = g_{123}(A)$, then

$$A^*(A^*)' = g_{12}(A'A) \quad . \tag{5.24}$$

<u>Proof</u>. From (4.4), (4.5) and (4.6) we have

$$AA^*A = A \quad , \quad A^*AA^* = A^* \quad , \quad AA^* = (A^*)'A' \quad . \tag{5.25}$$

Hence $A^*(A^*)'A'A = A^*AA^*A = A^*A$. Thus $A'A[A^*(A^*)'A'A] = A' \; AA^*A = A'A$ and $[A^*(A^*)'A'A]A^*(A^*)' = A^* \; AA^*(A^*)' = A^*(A^*)'$.

Q.E.D.

From Lemma 5.1 we obtain

$$\begin{aligned} G^*(G^*)' &= [L'(X'X)^-L]^- \\ &= \Pi_{21}B^{-1}(B^{-1})'\Pi_{21}' \end{aligned} \tag{5.26}$$

from (5.19), where we partition $\Pi_2 = (\Pi_{21}, \Pi_{22})$, with Π_{21}, $s \times t$, identifying t linearly independent columns of L_1, $r \times$. Hence

$$S_h = (\hat{\gamma}'L - m')\Pi_{21}B^{-1}(B^{-1})'\Pi_{21}'(L'\hat{\gamma} - m) \quad . \tag{5.27}$$

First $L'\hat{\gamma} - m$ is computed and rearranged to form $\Pi_{21}'(L'\hat{\gamma} - m) = h$, say. Then $h = B'k$ is solved for k, where B' is lower triangular. Finally S_h is found as the sum of squares of the components in $k = (B^{-1})'h = (B^{-1})'\Pi_{21}'(L'\hat{\gamma} - m)$.

The relevant F-test for the hypotheses (5.13) or (5.21) is then computed as

(5.28) $$F = \frac{S_h/t}{S_e/(n-r)} ,$$

cf. (2.19, with the critical region formed by values of (5.28) exceeding the corresponding tabulated value of F with t and n-r degrees of freedom.

The above procedures simplify slightly when the contrast matrix $\underset{\sim}{L}'$, $s \times q$, has full rank $s < r = r(\underset{\sim}{X})$. In that case (5.23) becomes non-singular and the results of Lemma 5.1 are not needed. We use

LEMMA 5.2. *When, $\underset{\sim}{L}'\underset{\sim}{\gamma}$ is estimable,*

(5.29) $$r[\underset{\sim}{L}'(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{L}] = r(\underset{\sim}{L}) ,$$

where $r(\cdot)$ *denotes rank.*

Proof. Using (5.7), $r(\underset{\sim}{L}) = r[\underset{\sim}{L}'(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{X}] \leq r[\underset{\sim}{L}'(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'] = r[\underset{\sim}{L}'(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{X}\{(\underset{\sim}{X}'\underset{\sim}{X})^{-}\}'\underset{\sim}{L}] = r[\underset{\sim}{L}'(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{L}] \leq r(\underset{\sim}{L})$.

Q.E.D.

When $\underset{\sim}{L}'$, $s \times q$, has full row rank $s \leq r$ the decomposition (5.17) becomes

(5.30) $$\underset{\sim}{G} = \underset{\sim}{Q}\begin{pmatrix} \underset{\sim}{B} \\ \underset{\sim}{0} \end{pmatrix}\underset{\sim}{\pi}'_{21}$$

say, where $\underset{\sim}{\pi}_{21}$ is now $s \times s$ and may equal $\underset{\sim}{I}_s$ (no column interchanges). Formula (5.27) applies with essentially no change.

We defer discussion of updating techniques for the less than full rank case and extensions to multivariate models to a further paper. A computer program in Fortran IV for the IBM 360 is being developed for the procedures discussed in this paper.

Acknowledgments

The work of the first author was completed whilst a visitor at Princeton University and Imperial College. He thanks Professor G. S. Watson and Professor D. R. Cox for the generous provision of facilities. The second author thanks Vo Van Tinh and Laurel L. Ward for helpful discussions.

References

Anderson, T. W. (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York.

Businger, Peter, and Golub, Gene H. (1965). Linear least squares solutions by Householder transformations. Numer. Math., 7, 269-276.

Draper, N. R., and Smith, H. (1966). @lied Regression Analysis. Wiley, New York.

Gentleman, M. (1970). Personal communication.

Golub, Gene H. (1965). Numerical methods for solving linear least squares problems. Numer. Math., 7, 206-216.

Golub, Gene H. (1969). Matrix decompositions and statistical calculations. Statistical Computation (R. C. Milton and J. A. Nelder, eds.), Academic Press, New York, 365-397.

Golub, Gene H., and Saunders, Michael (1970). Linear least squares and quadratic programming. Integer and Nonlinear Programming (J. Abadie, ed.), North-Holland, Amsterdam/American Elsevier, New York, 229-256.

Good, Irving John (1965). The Estimation of Probabilities: An Essay on Modern Bayesian Methods. Research Monograph No. 30, The M.I.T. Press, Cambridge, Mass.

Grossman, Stanley I., and Styan, George P. H. (1970). Uncorrelated regression residuals and singular values. Technical Report No. 32, Inst. Math. Stud. Soc. Sci., Stanford, 16 pp. (Optimality properties of Theil's BLUS residuals, J. Amer. Statist. Assoc., forthcoming.)

Healy, M. J. R. (1968). Triangular decomposition of a symmetric matrix. Appl. Statist. (J. Roy. Statist. Soc. Ser. C), 17, 195-197.

I.B.M. (1968). System/360 Scientific Subroutine Package (360A-CM-03X) Version III Programmer's Manual (Fourth Edition). IBM Manual H20-0205-3, IBM, White Plains, N. Y.

Longley, James W. (1967). An appraisal of least squares programs for the electronic computer from the point of view of the user. J. Amer. Statist. Assoc., **62**, 819-841.

Peters, G., and Wilkinson, J. H. (1970). The least squares problem and pseudo-inverses. The Computer Journal, **13**, 309-316.

Pringle, R. M., and Rayner, A. A. (1971). Generalized Inverse Matrices with Applications to Statistics. Griffin, London.

Wampler, Roy H. (1969). An evaluation of linear least squares computer programs. J. Res. Nat. Bur. Standards Sect. B, **73**, 59-90.

Wampler, Roy H. (1970). A report on the accuracy of some widely used least squares computer programs. J. Amer. Statist. Assoc., **65**, 549-565.

Wilkinson, J. H. (1965). Error analysis of transformations based on the use of matrices of the form $I - 2ww^H$. Error in Digital Computation: Volume 2 (L. B. Rall, ed.), Wiley, New York, 77-101.

Wilkinson, J. H. (1967). The solution of ill-conditioned linear equations. Mathematical Methods for Digital Computers: Volume 2 (A. Ralston and H. S. Wilf, eds.), Wiley, New York, 65-93.