COMPUTATION OF A TEST STATISTIC IN DATA QUALITY CONTROL[*]

XIAO-WEN CHANG[†], CHRISTOPHER C. PAIGE[†], AND CHRISTIAN C. J. M. TIBERIUS[‡]

Abstract. When processing observational data, statistical testing is an essential instrument for rendering harmless incidental anomalies and disturbances in the measurements. A commonly used test statistic based on the general linear model is the generalized likelihood ratio test statistic. The standard formula given in the literature for this test statistic is not defined if the noise covariance matrix is singular, and is not suitable for computation if any of the matrices involved are ill-conditioned. Based on Paige's generalized linear least squares method [*Comm. Statist. B—Simulation Comput.*, 7 (1978), pp. 437–453], a numerically stable approach is proposed for the computation of the test statistic, as well as for the estimates of the parameter vectors, and reliable representations of the error covariance matrices for these estimates are presented. This approach allows the noise covariance matrix to be singular and can be applied directly to the linear model with linear equality constraints.

Key words. data quality control, statistical testing, generalized likelihood ratio, generalized least squares, generalized QR factorization, numerical stability

AMS subject classifications. 62J12, 62J20, 65F20, 65F25

DOI. 10.1137/030601557

1. Introduction. For a wide range of applications, data processing involves least squares (LS) estimation. The LS estimate in itself, however, does not offer any indication of the validity of the result. Once some basic conditions are met, LS estimates for the unknown parameters, for example, the position coordinates in a Global Positioning System (GPS) application, are computed. Usually these give no indication as to whether, or to what extent, the input measurements are corrupted; i.e., they do not satisfy the assumed mathematical model. In critical applications such as real-time navigation, the user has to be protected against misleading (position) information caused by one or more anomalies in the measurements underlying the resulting position estimate. The incoming data need to be carefully validated by means of statistical testing, using the redundancy in the measurement set-up and the consequent strength of the mathematical model. Only then can the integrity of the eventual position solution be assured. A discussion of statistical testing for data quality control can be found, for example, in [11, Chap. 7] (mainly for GPS data processing) and [12].

A commonly used general test statistic for model validation based on linear models is called the generalized likelihood ratio test statistic, which follows the Chi-square distribution. A direct evaluation of the standard formula for the test statistic (see (2.8)) may lead to large rounding errors in the result when any of the matrices involved are ill-conditioned (i.e., have a large ratio of maximum to minimum singular value). In particular, those methods which use the inverse of the covariance matrix

---

[†]School of Computer Science, McGill University, Montreal, QC, Canada H3A 2A7 (chang@cs.mcgill.ca, paige@cs.mcgill.ca). The research of the first author was supported by NSERC of Canada grant RGPIN217191-03 and FCAR of Quebec, Canada grant 2001-NC-66487. The research of the second author was supported by NSERC of Canada grant RGPIN9236-01.

[‡]Section of Mathematical Geodesy and Positioning, Delft University of Technology, Thijsseweg 11, 2629 JA Delft, The Netherlands (c.c.j.m.tiberius@geo.tudelft.nl).

are numerically unstable and so can lead to large errors with finite precision computation. This was illustrated in [5, section 4]. Furthermore, the formula is not defined when the noise covariance matrix is singular. To our knowledge, there has not yet been a numerically sound approach for computing this test statistic. The goal of this paper is to present a numerically stable algorithm to compute the test statistic, and the estimates of the parameter vectors as well, based on Paige's generalized linear LS method (see [5, 6, 7, 8]). Our approach can handle the case when the noise covariance matrix is singular. This will also allow us to deal with linear models with linear equality constraints without altering the algorithm.

It appears unlikely that a very ill conditioned noise covariance matrix will arise directly in standard GPS applications. But constraints do appear, and these have been handled by introducing an additional noise vector whose covariance matrix has very small diagonal elements, leading to an inexact model whose noise covariance matrix is very ill conditioned. Ill-conditioned or singular noise covariance matrices can, of course, appear in other applications.

This paper is organized as follows. In section 2 we introduce the general linear model and derive the test statistic; then in section 3 we show the dangers of some obvious ways of computing this test statistic. In section 4 we first reformulate the optimization problems for estimating the parameter vectors and give an equivalent formula for the test statistic, then present an algorithm for computing the test statistic and the estimates of the parameter vectors as well, and give reliable representations of the error covariance matrices for the estimates. The backward numerical stability of the algorithm is proven in section 5. In section 6 we show how to compute the test statistic and the estimates of parameter vectors when the noise covariance matrix may be singular. We also illustrate how linear equality constraints result in a model which effectively has a singular noise covariance matrix so that the method here can be applied directly. Finally, we give a brief summary in section 7.

Throughout this paper we work with reals only. $\Re^m$ denotes the class of real m-dimensional vectors, while $\Re^{m \times n}$ denotes the class of real $m \times n$ matrices. $\mathcal{R}(A)$ denotes the range of the matrix A. The identity matrix is denoted by I and its ith column by e_i. For a matrix A, $A^\dagger$ denotes its Moore–Penrose generalized inverse. We write $\|x\|_2 = \sqrt{x^T x}$ for vectors and $\|A\|_F = \sqrt{\sum_{ij} a_{ij}^2}$ for matrices. We use $\mathcal{E}\{\cdot\}$ to denote the expected value and cov$\{\cdot\}$ to denote the covariance, that is, cov$\{x\} = \mathcal{E}\{(x - \mathcal{E}\{x\})(x - \mathcal{E}\{x\})^T\}$ for a random vector x. If v is a normally distributed random vector with mean $\bar{v}$ and covariance $\sigma^2 V$, we write $v \sim \mathcal{N}(\bar{v}, \sigma^2 V)$. V is sometimes called the cofactor matrix in geodesy; see, for example, [3, p. 250]. If a random variable δ follows a Chi-square distribution with noncentrality parameter λ and q degrees of freedom, we write $\delta \sim \chi^2(q, \lambda)$.

2. Linear model and test statistic. Suppose we have a linear model

$$(2.1) \qquad\qquad y = Ax + v, \qquad v \sim \mathcal{N}(0, \sigma^2 V),$$

where $y \in \Re^m$ is a random measurement vector, $A \in \Re^{m \times n}$ with $m \geq n$ is a design matrix with full column rank, $x \in \Re^n$ is an unknown parameter vector, and $v \in \Re^m$ is a random noise vector following a normal distribution with zero mean and a known symmetric nonnegative definite covariance matrix $\sigma^2 V$.

The above model, however, may not properly reflect reality in all cases. The model is misspecified when, for instance, one or more measurements are corrupted by outliers. They will invalidate the results of estimation. To safeguard against

anomalous measurements, it is important to validate the data and model used. Here we restrict the misspecification to the mean of the measurement vector y, i.e., an error of additive nature. Suppose (2.1) above represents the working model, the null hypothesis H_0. In a general form the alternative hypothesis H_a then reads

$$(2.2) \qquad y = Ax + C\nabla + v, \qquad v \sim \mathcal{N}(0, \sigma^2 V),$$

where the known matrix $C \in \Re^{m \times q}$ with $1 \le q \le m - n$ specifies the type of model error that occurred, $[A, C]$ has full column rank, and $\nabla \in \Re^q$ is an unknown constant vector (see, for example, [12, section 4.1]). One special case leading to the so-called w-test statistic is that in which the cofactor matrix V is the identity matrix and there is only a possible outlier in the ith measurement, i.e., $C = e_i$. If we take $i = 1, \dots, m$, we will have m alternative hypotheses. For other special cases that may arise from GPS applications, see [11, p. 280].

Initially we assume that V is nonsingular. Following [12, section 4.1], we briefly derive the generalized likelihood ratio test statistic for testing the null hypothesis H_0 against the alternative hypothesis H_a.

Let $p(y|x)$ and $p(y|x, \nabla)$ denote the density functions of y under H_0 and H_a, respectively. Then

$$(2.3) \qquad p(y|x) = \frac{1}{(2\pi)^{\frac{m}{2}} \sigma \det(V)^{\frac{1}{2}}} \exp\left[-\frac{1}{2\sigma^2}(y - Ax)^T V^{-1}(y - Ax)\right],$$

$$(2.4) \quad p(y|x, \nabla) = \frac{1}{(2\pi)^{\frac{m}{2}} \sigma \det(V)^{\frac{1}{2}}} \exp\left[-\frac{1}{2\sigma^2}(y - Ax - C\nabla)^T V^{-1}(y - Ax - C\nabla)\right].$$

Let x_0 be the maximum likelihood estimator of x under H_0, and let $\{x_a, \nabla_a\}$ be the maximum likelihood estimator of $\{x, \nabla\}$ under H_a, respectively, i.e.,

$$p(y|x_0) = \max_x p(y|x) \quad \le \quad p(y|x_a, \nabla_a) = \max_{x, \nabla} p(y|x, \nabla).$$

We expect these values to be comparable under H_0. Obviously x_0 and $\{x_a, \nabla_a\}$ are, respectively, the solutions of the following two generalized linear least squares (GLLS) problems (also referred to as weighted linear least squares problems in some literature):

$$(2.5) \qquad \text{GLLS}_0: \quad \min_x \ (y - Ax)^T V^{-1}(y - Ax);$$

$$(2.6) \qquad \text{GLLS}_a: \quad \min_{x, \nabla} \ (y - Ax - C\nabla)^T V^{-1}(y - Ax - C\nabla).$$

The two estimators are also the best linear unbiased estimators (BLUEs) of x under H_0 and $\{x, \nabla\}$ under H_a, respectively (see, for example, [4, section 3.2]). For later use, we define the corresponding residuals

$$(2.7) \qquad r_0 \equiv y - Ax_0, \qquad r_a \equiv y - Ax_a - C\nabla_a.$$

The generalized likelihood ratio is defined to be (see, for example, [12, section 4.1])

$$\mathcal{L}(H_0, H_a) \equiv \frac{\max_x p(y|x)}{\max_{\{x, \nabla\}} p(y|x, \nabla)} = \frac{p(y|x_0)}{p(y|x_a, \nabla_a)} = \exp\left[-\frac{1}{2\sigma^2}\left(r_0^T V^{-1} r_0 - r_a^T V^{-1} r_a\right)\right],$$

and the test statistic can be defined by

$$(2.8) \qquad \delta_{\text{TS}} \equiv -2 \log_e \mathcal{L}(H_0, H_a) = \sigma^{-2}(r_0^T V^{-1} r_0 - r_a^T V^{-1} r_a) \ge 0.$$

For some equivalent formulas for δ_{TS}, see [12, section 4.1]. It can be shown that (we illustrate this in section 4; see (4.23))

$$\delta_{\mathrm{TS}} \sim \chi^2(q,0) \quad \text{under } H_0,$$

(2.9) $$\delta_{\mathrm{TS}} \sim \chi^2(q,\lambda), \quad \lambda = \sigma^{-4}\nabla^T C^T V^{-1}\mathrm{cov}\{r_0\}V^{-1}C\nabla \quad \text{under } H_a,$$

where $\mathrm{cov}\{r_0\} = \sigma^2[V - A(A^T V^{-1}A)^{-1}A^T]$. When δ_{TS} is larger than a given threshold, one decides to reject the null hypothesis H_0 in favor of the alternative hypothesis H_a. Otherwise the null hypothesis H_0 will be accepted. The threshold is usually determined by the requirements of the specific application. For a discussion on this issue for general hypothesis testing, see, for example, [4, section 4.2.1] and [12, section 1.4].

3. Computational dangers. It is easy to assume that any reasonable method can be used to compute a simple result like δ_{TS} in (2.8). However, this is not true, and we give a small artificial example to illustrate some computational dangers. On the surface the following example looks fairly harmless, since the 2-norm condition numbers of A and V are about 4.44 and 33,000, respectively:

$$y = \begin{bmatrix} 5.48223618514353 \\ 0.90878847962427 \\ 25.94493985828999 \\ 5.91432884267696 \end{bmatrix}, \; A = \begin{bmatrix} 0.73591311187945 & 1.98690117078759 \\ 0.01599725305719 & 1.85723508466859 \\ 0.15632753551635 & 2.35754116764473 \\ 0.65858764131884 & 0.21189908130823 \end{bmatrix}, \; C = \begin{bmatrix} 2.72281454895206 \\ 1.87323437713309 \\ 2.51387627488834 \\ 0.87049065185808 \end{bmatrix},$$

$$V = \begin{bmatrix} 9.140496886810 & -5.179920639550 & 22.018803142087 & -2.448166448348 \\ -5.179920639550 & 31.269615846900 & -38.726345506531 & 1.768700005165 \\ 22.018803142087 & -38.726345506531 & 244.102164709880 & 43.463631186108 \\ -2.448166448348 & 1.768700005165 & 43.463631186108 & 15.497722556410 \end{bmatrix}, \quad \sigma^2 = 1.$$

This example was chosen to have $x_0 = [1,\ 2]^T$ and (see below) a large $\delta_{\mathrm{TS}} = \delta_0 - \delta_a = 2 - 1 = 1$, which we would expect to be able to compute quite accurately. We carried out computations using MATLAB 6 on an Apple Macintosh 450MHz PowerPC G4 running OSX 10.2.6, so the floating point relative precision was $\epsilon \approx 2.22*10^{-16}$. Now from (2.5), $x_0 = (A^T V^{-1}A)^{-1}A^T V^{-1}y$ and $r_0 = y - Ax_0$, so

$$\delta_0 \equiv r_0^T V^{-1}r_0 = y^T V^{-1}y - y^T V^{-1}A(A^T V^{-1}A)^{-1}A^T V^{-1}y.$$

We computed this last expression using the following MATLAB code:

```
Vinv=inv(V); Atilde=A'*Vinv; W=Atilde*A;
delta_0=y'*(Vinv*y)-y'*(Atilde'*(inv(W)*(Atilde*y)));
```

and computed $\delta_a \equiv r_a^T V^{-1}r_a$ by replacing A with $[A,C]$ in the above. From (2.8) we see $\delta_{\mathrm{TS}} = \delta_0 - \delta_a \geq 0$. But our MATLAB result was $\delta_0 - \delta_a = -14.02906911842439$, an obviously nonsensical result. This was not caused by a fault in our programming or in MATLAB (our computations gave reasonable answers for all *reasonable* problems). It is a simple reminder that combining a sequence of seemingly reliable computations does not necessarily lead to an overall numerically stable computation.

The approach of computing x_0 above, then $r_0 = y - Ax_0$, then $\delta_0 \equiv r_0^T V^{-1}r_0$, and similarly for δ_a, gave $\delta_0 - \delta_a = 0.43885318681141$, which is meaningful but still in error by more than 50%. The numerically stable method we recommend in section 4 gives $\delta_{\mathrm{TS}} = 1.00000000078345$ and $x_0^T = [1.00000000000001 \quad 2.00000000000001]$. Since a numerically stable computation gives the exact result to a nearby problem (see section 5) we suspect that the above problem is reasonably well conditioned and see that it is important to use a proven numerically stable algorithm for computing δ_{TS}. This is all the more important in many real-time applications where we might not have the luxury of IEEE standard double precision floating point arithmetic.

The numerical instabilities of the above two "obvious" methods for δ_{TS} were revealed here by $[A,C]$ having a 2-norm condition number of about $1.6*10^6$.

4. The algorithm. An obvious approach to solving problems GLLS_0 (2.5) and GLLS_a (2.6) is to transform the problems into their corresponding ordinary LS problems. Suppose the symmetric positive definite V has the factorization

$$V = BB^T, \tag{4.1}$$

where $B \in \Re^{m \times m}$ is nonsingular. In some applications B itself is given rather than V; then computing V can result in loss of important information and could cause the computed V to be singular even when B is nonsingular; see, for example, [1, p. 163]. Even if B is not given, we can compute the Cholesky factorization of V to find B. Then by forming $B^{-1}y$, $B^{-1}A$, and $B^{-1}C$, (2.5) and (2.6) can be transformed into

$$\text{GLLS}_0 : \quad \min(B^{-1}y - B^{-1}Ax)^T(B^{-1}y - B^{-1}Ax); \tag{4.2}$$

$$\text{GLLS}_a : \quad \min(B^{-1}y - B^{-1}Ax - B^{-1}C\nabla)^T(B^{-1}y - B^{-1}Ax - B^{-1}C\nabla). \tag{4.3}$$

For these ordinary LS problems, the QR factorization method or the normal equation method can then be used to compute the optimal solutions and the test statistic δ_{TS}. This approach is still often used in some practical areas. But the solutions may lose accuracy unnecessarily when B is ill-conditioned. Another problem is that for some applications the cofactor matrix V may be singular so that the GLLS formulations (2.5) and (2.6) (or (4.2) and (4.3)) and the test statistic δ_{TS} in (2.8) are not defined. Fortunately we can use Paige's approach to solving GLLS problems (see [5, 6, 7, 8]) to design a numerically stable method to compute the test statistic and the estimates of x under H_0 and under H_a as well. Our algorithm can handle a singular cofactor matrix V, but in order to expose the ideas quickly, we still assume that V is nonsingular in this section and the next. We will show how to handle singular V in section 6.

With (4.1), for the random noise vector v in (2.1) and (2.2) we can write

$$v \equiv Bu, \qquad u \sim \mathcal{N}(0, \sigma^2 I). \tag{4.4}$$

Then the linear models (2.1) and (2.2) can be replaced with

$$y = Ax + Bu, \qquad u \sim \mathcal{N}(0, \sigma^2 I) \qquad \text{under } \text{H}_0; \tag{4.5}$$

$$y = Ax + C\nabla + Bu, \quad u \sim \mathcal{N}(0, \sigma^2 I) \qquad \text{under } \text{H}_a. \tag{4.6}$$

Therefore problems GLLS_0 in (4.2) and GLLS_a in (4.3) can be reformulated as follows:

$$\text{GLLS}_0 : \quad \min_{u,x} \|u\|_2^2 \quad \text{subject to} \ \ y = Ax + Bu; \tag{4.7}$$

$$\text{GLLS}_a : \quad \min_{u,x,\nabla} \|u\|_2^2 \quad \text{subject to} \ \ y = Ax + C\nabla + Bu. \tag{4.8}$$

Even if B is not square or is singular, we can still find the optimal solutions for GLLS_0 and GLLS_a as long as the constraints in (4.7) and (4.8) are consistent (see section 6). Let u_0 and u_a denote the optimal u for problems GLLS_0 and GLLS_a, respectively. Then with the assumption that B is nonsingular (see (2.7))

$$u_0 = B^{-1}(y - Ax_0) = B^{-1}r_0, \qquad u_a = B^{-1}(y - Ax_a - C\nabla_a) = B^{-1}r_a. \tag{4.9}$$

When $B = I$, u_0 and u_a are just the ordinary LS residuals r_0 and r_a, respectively. So they can be called the generalized least squares (GLS) residuals. From (2.8), we have

$$\delta_{\text{TS}} = \sigma^{-2}(\|u_0\|_2^2 - \|u_a\|_2^2). \tag{4.10}$$

Introducing $C\nabla$ decreases the square of the GLS residual norm from $\|u_0\|_2^2$ to $\|u_a\|_2^2$, so δ_{TS} is a scaled measure of this decrease. Note that in the formulations (4.7), (4.8), and (4.10), no inverse of B or V appears. This is the key to the design of the numerically stable algorithm that we now develop.

First consider solving problem GLLS_a in (4.8). Let $[A, C]$ and B have the following generalized QR (GQR) factorization (see [9]):

$$(4.11) \quad P^T[A, C] = \begin{bmatrix} U_A & U_{AC} \\ 0 & U_C \\ 0 & 0 \end{bmatrix} \begin{matrix} n \\ q \\ m-n-q \end{matrix} \quad, \quad P^T B Q = \begin{bmatrix} R_A & R_{AC} & R_{A3} \\ 0 & R_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{matrix} n \\ q \\ m-n-q \end{matrix} \quad,$$
$$\quad\quad\quad\quad\quad\quad\quad n \quad\, q \quad\quad\quad\quad\quad\quad\quad n \quad\;\; q \quad\; m-n-q$$

where

$$(4.12) \quad P = [\underset{n}{P_A}, \underset{q}{P_C}, \underset{m-n-q}{P_3}] \in \Re^{m\times m} \quad \text{and} \quad Q = [\underset{n}{Q_A}, \underset{q}{Q_C}, \underset{m-n-q}{Q_3}] \in \Re^{m\times m}$$

are orthogonal ($P^{-1} = P^T$, $Q^{-1} = Q^T$), and U_A, U_C, R_A, R_C, and R_3 are nonsingular upper triangular. Multiplying the constraint equality in (4.8) by P^T from the left and using $QQ^T = I$, we have

$$P^T y = P^T A x + P^T C\nabla + P^T B Q Q^T u \quad\quad \text{under } \mathrm{H}_a.$$

Then by substituting (4.11) into the above equation and defining

$$(4.13) \quad\quad z \equiv P^T y \equiv \begin{bmatrix} z_A \\ z_C \\ z_3 \end{bmatrix} \begin{matrix} n \\ q \\ m-n-q \end{matrix} \quad , \quad w \equiv Q^T u \equiv \begin{bmatrix} w_A \\ w_C \\ w_3 \end{bmatrix} \begin{matrix} n \\ q \\ m-n-q \end{matrix} \quad ,$$

we can transform problem GLLS_a in (4.8) into

$$\min(\|w_A\|_2^2 + \|w_C\|_2^2 + \|w_3\|_2^2) \quad \text{subject to}$$

$$(4.14) \quad \begin{bmatrix} z_A \\ z_C \\ z_3 \end{bmatrix} = \begin{bmatrix} U_A \\ 0 \\ 0 \end{bmatrix} x + \begin{bmatrix} U_{AC} \\ U_C \\ 0 \end{bmatrix} \nabla + \begin{bmatrix} R_A & R_{AC} & R_{A3} \\ 0 & R_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} w_A \\ w_C \\ w_3 \end{bmatrix} \quad \text{(i.e., under } \mathrm{H}_a\text{)}.$$

Obviously the optimal solution x_a, ∇_a, $w_a \equiv [(w_A^a)^T, (w_C^a)^T, (w_3^a)^T]^T$ satisfies

$$(4.15) \quad\quad w_A^a = 0, \quad\quad w_C^a = 0, \quad\quad \begin{bmatrix} U_A & U_{AC} & R_{A3} \\ 0 & U_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} x_a \\ \nabla_a \\ w_3^a \end{bmatrix} = \begin{bmatrix} z_A \\ z_C \\ z_3 \end{bmatrix}.$$

The upper triangular system in (4.15) can easily be solved by back substitution. From (4.14) and (4.15) we have $w_3^a = w_3$ under H_a. Since $w = Q^T u \sim \mathcal{N}(0, \sigma^2 I_m)$,

$$(4.16) \quad\quad\quad\quad\quad\quad w_3^a \sim \mathcal{N}(0, \sigma^2 I_{m-n-q}) \quad \text{under } \mathrm{H}_a.$$

Thus w_3^a is a vector of uncorrelated residuals with respect to the model (4.14), which can be regarded as the extension of the linear unbiased with scalar covariance matrix using Householder transformations (LUSH) residuals to a linear model with a general noise covariance matrix. For the LUSH residuals, see [10]. Thus the GLS residual u_a (the optimal u under H_a) satisfies (see (4.13))

$$(4.17) \quad u_a = Q w_a = Q_3 w_3^a, \quad\quad \|u_a\|_2^2 = \|w_a\|_2^2 = \|w_A^a\|_2^2 + \|w_C^a\|_2^2 + \|w_3^a\|_2^2 = \|w_3^a\|_2^2.$$

Note that, unlike the elements of w_3^a, the elements of u_a are correlated, since from (4.16) and (4.17) $\operatorname{cov}\{u_a\} = \sigma^2 Q_3 Q_3^T$.

Now consider solving problem GLLS_0 in (4.7). Note that the difference between GLLS_0 and GLLS_a is that the constraint equality in GLLS_0 does not have the ∇ term. Thus from (4.14) it follows that GLLS_0 in (4.7) can be transformed into

$$
\min(\|w_A\|_2^2 + \|w_C\|_2^2 + \|w_3\|_2^2)
$$

(4.18)   subject to
$$
\begin{bmatrix} z_A \\ z_C \\ z_3 \end{bmatrix} = \begin{bmatrix} U_A \\ 0 \\ 0 \end{bmatrix} x + \begin{bmatrix} R_A & R_{AC} & R_{A3} \\ 0 & R_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} w_A \\ w_C \\ w_3 \end{bmatrix}.
$$

It follows that the optimal solution x_0, $w_0 \equiv [(w_A^0)^T, (w_C^0)^T, (w_3^0)^T]^T$ satisfies

(4.19)
$$
w_A^0 = 0, \qquad \begin{bmatrix} U_A & R_{AC} & R_{A3} \\ 0 & R_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} x_0 \\ w_C^0 \\ w_3^0 \end{bmatrix} = \begin{bmatrix} z_A \\ z_C \\ z_3 \end{bmatrix}.
$$

This upper triangular system can also be solved by back substitution. Note that the matrix coefficients and the right-hand sides of both (4.15) and (4.19) were found from the one GQR in (4.11). Similar to w_3^a, $\begin{bmatrix} w_C^0 \\ w_3^0 \end{bmatrix}$ is a vector of uncorrelated residuals, since from (4.18), (4.19), (4.13), and (4.4),

(4.20)
$$
\begin{bmatrix} w_C^0 \\ w_3^0 \end{bmatrix} = \begin{bmatrix} w_C \\ w_3 \end{bmatrix} \sim \mathcal{N}(0, \sigma^2 I_{m-n}) \quad \text{under } \mathrm{H}_0.
$$

Thus the GLS residual u_0 (the optimal u under H_0) satisfies

(4.21) $u_0 = Q w_0 = Q_C w_C^0 + Q_3 w_3^0$, $\|u_0\|_2^2 = \|w_0\|_2^2 = \|w_C^0\|_2^2 + \|w_3^0\|_2^2$.

From (4.15) and (4.19) we observe that $w_3^a = w_3^0$. By using the equivalent formula of δ_{TS} (see (4.10)), from (4.17) and (4.21) we obtain a result not involving differences:

(4.22) $\delta_{\mathrm{TS}} = \sigma^{-2}(\|u_0\|_2^2 - \|u_a\|_2^2) = \sigma^{-2}\|w_C^0\|_2^2$.

From (4.20) we see $w_C^0 \sim \mathcal{N}(0, \sigma^2 I_q)$ under H_0, and we know $w_C \sim \mathcal{N}(0, \sigma^2 I_q)$; see (4.13), (4.4). Combining (4.19) with (4.14), we have under H_a,

$$
\begin{bmatrix} R_C & R_{C3} \\ 0 & R_3 \end{bmatrix} \begin{bmatrix} w_C^0 \\ w_3^0 \end{bmatrix} = \begin{bmatrix} z_C \\ z_3 \end{bmatrix} = \begin{bmatrix} U_C \nabla \\ 0 \end{bmatrix} + \begin{bmatrix} R_C & R_{C3} \\ 0 & R_3 \end{bmatrix} \begin{bmatrix} w_C \\ w_3 \end{bmatrix},
$$

so $w_C^0 = R_C^{-1} U_C \nabla + w_C \sim \mathcal{N}(R_C^{-1} U_C \nabla, \sigma^2 I_q)$ under H_a. Therefore from (4.22) and the definition of the Chi-square distribution, we have

(4.23) $\delta_{\mathrm{TS}} \sim \chi^2(q, 0)$ under H_0; $\delta_{\mathrm{TS}} \sim \chi^2(q, \sigma^{-2}\|R_C^{-1} U_C \nabla\|_2^2)$ under H_a.

It can be verified from (4.11) that the noncentrality parameter $\sigma^{-2}\|R_C^{-1} U_C \nabla\|_2^2$ is just the λ given in (2.9).

The above derivation leads to the following algorithm to compute the test statistic, the BLUE x_0 under H_0, and the BLUE x_a under H_a.

ALGORITHM.

Compute the GQR factorization of $[A, C]$ and B (4.11).

Solve $\begin{bmatrix} R_C & R_{C3} \\ 0 & R_3 \end{bmatrix} \begin{bmatrix} w_C^0 \\ w_3^0 \end{bmatrix} = \begin{bmatrix} z_C \\ z_3 \end{bmatrix}$ (see (4.19)).

Compute $\delta_{\mathrm{TS}} = \sigma^{-2}\|w_C^0\|_2^2$ (see (4.22)).

If δ_{TS} is smaller than a given threshold, accept the null hypothesis,
 and solve $U_A x_0 = z_A - R_{AC} w_C^0 - R_{A3} w_3^0$ (see (4.19))
else accept the alternative hypothesis,

 and solve $\begin{bmatrix} U_A & U_{AC} \\ 0 & U_C \end{bmatrix} \begin{bmatrix} x_a \\ \nabla_a \end{bmatrix} = \begin{bmatrix} z_A - R_{A3} w_3^0 \\ z_C - R_{C3} w_3^0 \end{bmatrix}$ (see (4.15)).

In the following we give a few comments about the above algorithm.

The GQR factorization (4.11) can be computed using Householder transformations. P and Q do not need to be formed in the computation. Note that R_A is not involved in the computation. Thus it is not required to be upper triangular, or even square (R_A may not be square when B is not square; see section 6). After we compute the QR factorization of A, we apply P^T to B and obtain

$$(4.24) \qquad P^T B = \begin{bmatrix} P_A^T B \\ P_C^T B \\ P_3^T B \end{bmatrix}.$$

Then we determine Q by the RQ factorization:

$$\begin{bmatrix} P_C^T B \\ P_3^T B \end{bmatrix} Q = \begin{bmatrix} 0 & R_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix}.$$

Although both R_{AC} and R_{A3} are involved in computing x_0, and R_{A3} is involved in computing x_a, we do not need to explicitly compute them. In fact, what we need to compute are $R_{AC} w_C^0 + R_{A3} w_3^0$ and $R_{A3} w_3^0$. Note that by (4.11),

$$R_{AC} w_C^0 + R_{A3} w_3^0 = P_A^T B Q \begin{bmatrix} 0 \\ w_C^0 \\ w_3^0 \end{bmatrix}, \qquad R_{A3} w_3^0 = P_A^T B Q \begin{bmatrix} 0 \\ 0 \\ w_3^0 \end{bmatrix},$$

where $P_A^T B$ has been computed in (4.24). The above computation can be performed by using only matrix-vector multiplication.

Sometimes one is interested in the covariance matrices $\mathrm{cov}\{x_0\}$ under H_0 and $\mathrm{cov}\{x_a\}$ under H_a. First, we consider computing $\mathrm{cov}\{x_0\}$ under H_0. From (4.14), without the ∇ term on the right-hand side, and (4.19), we obtain

$$\begin{bmatrix} U_A & R_{AC} & R_{A3} \\ 0 & R_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} x_0 - x \\ w_C^0 - w_C \\ w_3^0 - w_3 \end{bmatrix} = \begin{bmatrix} R_A w_A \\ 0 \\ 0 \end{bmatrix}.$$

This shows that $w_3^0 - w_3 = 0$ and $w_C^0 - w_C = 0$, so

$$U_A(x_0 - x) = R_A w_A.$$

Since $w = Q^T u \sim \mathcal{N}(0, \sigma^2 I)$ (see (4.13) and (4.4)) we have

$$(4.25) \qquad U_A \cdot \mathrm{cov}\{x_0\} \cdot U_A^T = \sigma^2 R_A R_A^T.$$

We see that here R_A is needed, although it is not needed in computing the test statistic and the estimates x_0 and x_a. If $\text{cov}\{x_0\}$ will be involved in other computations, it might be useful to have an upper triangular R_A. Notice that, if R_A is upper triangular, $\sigma U_A^{-1} R_A$ is the Cholesky factor of $\text{cov}\{x_0\}$. Computing this, however, could lead to loss of accuracy if U_A is ill-conditioned. In many applications we may not need to compute $\text{cov}\{x_0\}$ explicitly. In general (4.25) gives the most reliable representation of $\text{cov}\{x_0\}$.

Now we consider computing $\text{cov}\{x_a\}$ under H_a. From (4.14) and (4.15) we obtain

$$\begin{bmatrix} U_A & U_{AC} & R_{A3} \\ 0 & U_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} x_a - x \\ \nabla_a - \nabla \\ w_3^a - w_3 \end{bmatrix} = \begin{bmatrix} R_A & R_{AC} \\ 0 & R_C \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w_A \\ w_C \end{bmatrix}.$$

This shows that $w_3^a - w_3 = 0$, so

$$\begin{bmatrix} U_A & U_{AC} \\ 0 & U_C \end{bmatrix} \begin{bmatrix} x_a - x \\ \nabla_a - \nabla \end{bmatrix} = \begin{bmatrix} R_A & R_{AC} \\ 0 & R_C \end{bmatrix} \begin{bmatrix} w_A \\ w_C \end{bmatrix}.$$

Again since $w \sim \mathcal{N}(0, \sigma^2 I)$, we have the representation of $\text{cov}\left\{ \begin{bmatrix} x_a \\ \nabla_a \end{bmatrix} \right\}$:

$$(4.26) \quad \begin{bmatrix} U_A & U_{AC} \\ 0 & U_C \end{bmatrix} \cdot \text{cov}\left\{ \begin{bmatrix} x_a \\ \nabla_a \end{bmatrix} \right\} \cdot \begin{bmatrix} U_A & U_{AC} \\ 0 & U_C \end{bmatrix}^T = \sigma^2 \begin{bmatrix} R_A & R_{AC} \\ 0 & R_C \end{bmatrix} \begin{bmatrix} R_A & R_{AC} \\ 0 & R_C \end{bmatrix}^T,$$

and $\text{cov}\{x_a\}$ can be found by solving this. If we want to get the reliable representation of $\text{cov}\{x_a\}$, which is similar to (4.25), we find two orthogonal matrices G_1 and G_2 such that

$$(4.27) \quad G_1^T \begin{bmatrix} U_A & U_{AC} \\ 0 & U_C \end{bmatrix} = \begin{bmatrix} \tilde{U}_A & 0 \\ \tilde{U}_{CA} & \tilde{U}_C \end{bmatrix}, \qquad G_1^T \begin{bmatrix} R_A & R_{AC} \\ 0 & R_C \end{bmatrix} G_2 = \begin{bmatrix} \tilde{R}_A & 0 \\ \tilde{R}_{CA} & \tilde{R}_C \end{bmatrix},$$

where G_1 is a product of Givens rotations, G_2 is a product of Householder transformations, and $\tilde{U}_A$ and $\tilde{R}_A$ are still upper triangular. This computation can be implemented to make full use of the structures of the relevant matrices for efficiency. Combining (4.26) and (4.27) shows

$$\begin{bmatrix} \tilde{U}_A & 0 \\ \tilde{U}_{CA} & \tilde{U}_C \end{bmatrix} \cdot \text{cov}\left\{ \begin{bmatrix} x_a \\ \nabla_a \end{bmatrix} \right\} \cdot \begin{bmatrix} \tilde{U}_A & 0 \\ \tilde{U}_{CA} & \tilde{U}_C \end{bmatrix}^T = \sigma^2 \begin{bmatrix} \tilde{R}_A & 0 \\ \tilde{R}_{CA} & \tilde{R}_C \end{bmatrix} \begin{bmatrix} \tilde{R}_A & 0 \\ \tilde{R}_{CA} & \tilde{R}_C \end{bmatrix}^T.$$

The (1,1) block of the two sides of this equality gives the representation of $\text{cov}\{x_a\}$:

$$\tilde{U}_A \cdot \text{cov}\{x_a\} \cdot \tilde{U}_A^T = \sigma^2 \tilde{R}_A \tilde{R}_A^T.$$

5. Numerical stability of the algorithm. In this section we will prove the numerical stability of the computation of the test statistic δ_{TS}, the BLUE x_0 of x under H_0, and the BLUE x_a under H_a. The proof extends the approach in [7].

Let the computed triangular matrices in the GQR factorization (4.11) be denoted by $\hat{U}$ and $\hat{R}$. We can show (see, for example, [2, Lem. 19.3, Thm. 19.4]) that there exist orthogonal $\tilde{P} \in \Re^{m \times m}$ and $\tilde{Q} \in \Re^{m \times m}$ such that

$$(5.1) \quad \tilde{P}^T [A + \Delta A_1, C + \Delta C_1] = \begin{bmatrix} \hat{U}_A & \hat{U}_{AC} \\ 0 & \hat{U}_C \\ 0 & 0 \end{bmatrix},$$

$$\|\Delta A_1\|_F = O(\epsilon)\|A\|_F, \qquad \|\Delta C_1\|_F = O(\epsilon)\|A\|_F,$$

$$(5.2) \qquad \tilde{P}^T(B + \Delta B_1)\tilde{Q} = \begin{bmatrix} \hat{R}_A & \hat{R}_{AC} & \hat{R}_{A3} \\ 0 & \hat{R}_C & \hat{R}_{C3} \\ 0 & 0 & \hat{R}_3 \end{bmatrix}, \quad \|\Delta B_1\|_F = O(\epsilon)\|A\|_F,$$

where ϵ is the unit roundoff, and for simplicity we use $O(\epsilon)$ instead of the product of ϵ and a constant dependent only on the dimensions of the problem. Here and throughout we ignore any $O(\epsilon^2)$ terms.

Note that $z = P^T y$ (see (4.13)). We can show that the computed z satisfies (see, for example, [2, Lem. 19.3])

$$\hat{z} = \tilde{P}^T(y + \Delta y), \qquad \|\Delta y\|_2 = O(\epsilon)\|y\|_2,$$

and the computed solution $\hat{x}_0$, $\hat{w}_C^0$, $\hat{w}_3^0$ of (4.19) satisfies (see [2, Thm. 8.5])

$$(5.3)$$

$$\begin{bmatrix} \hat{U}_A + \Delta U_A & \hat{R}_{AC} + \Delta R_{AC} & \hat{R}_{A3} + \Delta R_{A3} \\ 0 & \hat{R}_C + \Delta R_C & \hat{R}_{C3} + \Delta R_{C3} \\ 0 & 0 & \hat{R}_3 + \Delta R_3 \end{bmatrix} \begin{bmatrix} \hat{x}_0 \\ \hat{w}_C^0 \\ \hat{w}_3^0 \end{bmatrix} = \begin{bmatrix} \hat{z}_A \\ \hat{z}_C \\ \hat{z}_3 \end{bmatrix},$$

$$\|\Delta U_A\|_F = O(\epsilon)\|\hat{U}_A\|_F, \qquad \left\| \begin{bmatrix} \Delta R_{AC} & \Delta R_{A3} \\ \Delta R_C & \Delta R_{C3} \\ 0 & \Delta R_3 \end{bmatrix} \right\|_F = O(\epsilon) \left\| \begin{bmatrix} \hat{R}_{AC} & \hat{R}_{A3} \\ \hat{R}_C & \hat{R}_{C3} \\ 0 & \hat{R}_3 \end{bmatrix} \right\|_F.$$

For the computation of $\delta_{\mathrm{TS}} = \sigma^{-2}\|w_C^0\|_2^2$, it is easy to show that the computed δ_{TS} satisfies (see [2, sect. 3.1])

$$(5.4) \qquad \hat{\delta}_{\mathrm{TS}} = (\sigma + \Delta\sigma)^{-2}\|\hat{w}_C^0\|_2^2, \qquad \Delta\sigma = O(\epsilon)\sigma.$$

Combining (5.1), (5.2), (5.3), and (5.4), we can easily verify that $\hat{\delta}_{\mathrm{TS}}$ and $\hat{x}_0$ are the exact test statistic and the BLUE of x under H_0, respectively, for the initial data

$$\tilde{y} \equiv y + \Delta y, \quad \|\tilde{y} - y\|_2 = O(\epsilon)\|y\|_2,$$

$$\tilde{A} \equiv A + \Delta A_1 + \tilde{P} \begin{bmatrix} \Delta U_A \\ 0 \\ 0 \end{bmatrix}, \quad \|\tilde{A} - A\|_F = O(\epsilon)\|A\|_F,$$

$$\tilde{B} \equiv B + \Delta B_1 + \tilde{P} \begin{bmatrix} \Delta R_{AC} & \Delta R_{A3} \\ \Delta R_C & \Delta R_{C3} \\ 0 & \Delta R_3 \end{bmatrix} \tilde{Q}^T, \quad \|\tilde{B} - B\|_F = O(\epsilon)\|B\|_F,$$

$$\tilde{\sigma} \equiv \sigma + \Delta\sigma, \quad |\tilde{\sigma} - \sigma| = O(\epsilon)|\sigma|.$$

Thus the computations of δ_{TS} and x_0 are numerically stable. It is pleasing that $\hat{\delta}_{\mathrm{TS}}$ and $\hat{x}_0$ are exact for initial data with the same small perturbations.

Similarly, we can easily show that the computation of the BLUE x_a of x under H_a is numerically stable.

6. Extension to the singular case. In this section we assume that the rank of the cofactor matrix V is $k \le m$. Suppose V still has the factorization (4.1), where $B \in \Re^{m \times k}$ has full column rank. If B is not given, we can use diagonal pivoting to compute the Cholesky factorization of V (see, for example, [1, section 2.2.2]) to give (4.1), where the strictly upper triangular part of B is zero, i.e., $b_{ij} = 0$ for $i < j$, or to compute the "reverse" Cholesky factorization of V to give (4.1), where $b_{ij} = 0$ for $i > m - k + j$. So we still have models (4.5) and (4.6).

If $y \notin \mathcal{R}([A, B])$, then y could not have come from the model (2.1), and this has to be checked. For simplicity we assume that the two models are consistent so that unique optimal solutions $\{x_0, u_0\}$ and $\{x_a, \nabla_a, u_a\}$ exist for problems GLLS$_0$ in (4.7) and GLLS$_a$ in (4.8), respectively. Now (see [5, section 3]), the solution of GLLS$_0$ results in the BLUE of any estimable function of x under H$_0$. Also since B has full column rank we have $B^\dagger B = I$, and writing $v = Bu$ in (2.1) we see that in (4.5) and (4.6),

$$(6.1) \qquad u = B^\dagger v \sim \mathcal{N}(0, \sigma^2 I_k).$$

From this we will show that we can still use the formula for the test statistic δ_{TS} in (4.10). Unlike u_0 and u_a in (4.9), $y = Ax_0 + Bu_0$, etc. lead to

$$u_0 = B^\dagger(y - Ax_0), \qquad u_a = B^\dagger(y - Ax_a - C\nabla_a).$$

In the following we will make a sequence of orthogonal transformations to compute GQR factorizations involving A, C, and B. Based on these, we then discuss the computation of δ_{TS}, x_0, and x_a.

Just as in (4.11), we find the QR factorization of $[A, C]$:

$$(6.2) \qquad \bar{P}^T[A, C] = \begin{bmatrix} U_A & U_{AC} \\ 0 & U_C \\ 0 & 0 \end{bmatrix} \begin{matrix} n \\ q \\ m-n-q \end{matrix} \quad .$$
$$\phantom{(6.2) \qquad \bar{P}^T[A, C] = } \begin{matrix} n & q \end{matrix}$$

Then we apply $\bar{P}^T$ to B to get

$$(6.3) \qquad \bar{P}^T B = \begin{bmatrix} B_A \\ B_C \\ B_3 \end{bmatrix} \begin{matrix} n \\ q \\ m-n-q \end{matrix} \quad .$$
$$\phantom{(6.3) \qquad \bar{P}^T B = } \begin{matrix} k \end{matrix}$$

We would like to transform B_3 into a nonsingular upper triangular matrix. First, find the RQ factorization of B_3 with row pivoting:

$$(6.4) \qquad \Pi_3^T B_3 Q = \begin{bmatrix} 0 & \bar{R}_3 \end{bmatrix} m-n-q,$$
$$ \begin{matrix} k-l & l \end{matrix}$$

where Π_3 is a permutation matrix, Q is orthogonal, and $\bar{R}_3$ is a full-column-rank trapezoidal matrix with zeros in its strictly lower left triangle. Then find the QR factorizations of $\bar{R}_3$:

$$(6.5) \qquad \check{P}_3^T \bar{R}_3 = \begin{bmatrix} R_3 \\ 0 \end{bmatrix} \begin{matrix} l \\ m-n-q-l \end{matrix} \quad ,$$
$$\phantom{(6.5) \qquad \check{P}_3^T \bar{R}_3 = } \begin{matrix} l \end{matrix}$$

where $\check{P}_3$ is orthogonal and R_3 is nonsingular upper triangular. Combining (6.4) and (6.5) leads to

$$(6.6) \qquad \check{P}_3^T \Pi_3^T B_3 Q = \begin{bmatrix} 0 & R_3 \\ 0 & 0 \end{bmatrix} \begin{matrix} l \\ m-n-q-l \end{matrix} \quad .$$
$$\phantom{(6.6) \qquad \check{P}_3^T \Pi_3^T B_3 Q = } \begin{matrix} k-l & l \end{matrix}$$

Defining

$$P \equiv \bar{P} \begin{bmatrix} I_n & 0 & 0 \\ 0 & I_q & 0 \\ 0 & 0 & \Pi_3 \check{P}_3 \end{bmatrix},$$

we then have from (6.2), (6.3), and (6.6) that

$$(6.7) \qquad P^T[A, C] = \begin{bmatrix} U_A & U_{AC} \\ 0 & U_C \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{matrix} n \\ q \\ l \\ m-n-q-l \end{matrix} \quad , \quad P^T B Q \equiv \begin{bmatrix} \bar{B}_A & R_{A3} \\ \bar{B}_C & R_{C3} \\ 0 & R_3 \\ 0 & 0 \end{bmatrix} \begin{matrix} n \\ q \\ l \\ m-n-q-l \end{matrix} \quad .$$
$$\qquad \qquad \qquad \quad \begin{matrix} n & q \end{matrix} \qquad \qquad \qquad \qquad \qquad \begin{matrix} k-l & l \end{matrix}$$

Here $\bar{B}_C$ does not have any special structure.

As in section 4, defining

$$(6.8) \qquad z \equiv P^T y \equiv \begin{bmatrix} z_A \\ z_C \\ z_3 \\ z_4 \end{bmatrix} \begin{matrix} n \\ q \\ l \\ m-n-q-l \end{matrix} \quad , \quad w \equiv Q^T u \equiv \begin{bmatrix} w_{AC} \\ w_3 \end{bmatrix} \begin{matrix} k-l \\ l \end{matrix} \quad ,$$

we transform problem GLLS$_a$ in (4.8) into (cf. (4.14))

$$\min(\|w_{AC}\|_2^2 + \|w_3\|_2^2)$$

$$(6.9) \qquad \text{subject to} \quad \begin{bmatrix} z_A \\ z_C \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} U_A \\ 0 \\ 0 \\ 0 \end{bmatrix} x + \begin{bmatrix} U_{AC} \\ U_C \\ 0 \\ 0 \end{bmatrix} \nabla + \begin{bmatrix} \bar{B}_A & R_{A3} \\ \bar{B}_C & R_{C3} \\ 0 & R_3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w_{AC} \\ w_3 \end{bmatrix}.$$

Here the ∇ term will disappear when we transform problem GLLS$_0$ in (4.7) by the same transformations. Note that z_4 in (6.9) must be a zero vector, since the original model (4.6) is consistent. Obviously the optimal solution x_a, ∇_a, $w_a \equiv [(w_{AC}^a)^T, (w_3^a)^T]^T$ satisfies (cf. (4.15))

$$(6.10) \qquad w_{AC}^a = 0, \qquad \begin{bmatrix} U_A & U_{AC} & R_{A3} \\ 0 & U_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} x_a \\ \nabla_a \\ w_3^a \end{bmatrix} = \begin{bmatrix} z_A \\ z_C \\ z_3 \end{bmatrix}.$$

Then it follows that the GLS residual u_a (the optimal u under H$_a$) satisfies

$$(6.11) \qquad u_a = Q w_a, \qquad \|u_a\|_2^2 = \|w_a\|_2^2 = \|w_3^a\|_2^2.$$

Now we consider solving problem GLLS$_0$ in (4.7). From (6.9) we observe that GLLS$_0$ can be transformed into

$$\min(\|w_{AC}\|_2^2 + \|w_3\|_2^2)$$

$$(6.12) \qquad \text{subject to} \quad \begin{bmatrix} z_A \\ z_C \\ z_3 \end{bmatrix} = \begin{bmatrix} U_A \\ 0 \\ 0 \end{bmatrix} x + \begin{bmatrix} \bar{B}_A & R_{A3} \\ \bar{B}_C & R_{C3} \\ 0 & R_3 \end{bmatrix} \begin{bmatrix} w_{AC} \\ w_3 \end{bmatrix}.$$

Note that (6.12) is just (6.9) except that the ∇ term and z_4 rows have been removed. In order to estimate x, we have to transform $\bar{B}_C$ into a nonsingular upper triangular matrix. As we did to B_3 (see (6.6)), we can find orthogonal $\check{P}_C \in \Re^{q \times q}$ and $\check{Q}_C \in \Re^{(k-l) \times (k-l)}$, and a permutation $\Pi_C \in \Re^{q \times q}$, such that

$$(6.13) \qquad \check{P}_C^T \Pi_C^T \bar{B}_C \check{Q}_C = \begin{bmatrix} 0 & R_C \\ 0 & 0 \end{bmatrix} \begin{matrix} p \\ q-p \end{matrix} \, , \\ \phantom{(6.13) \qquad \check{P}_C^T \Pi_C^T \bar{B}_C \check{Q}_C} {\scriptstyle k-l-p \quad p}$$

where R_C is nonsingular upper triangular. Let

$$(6.14) \qquad \check{P}_C^T \Pi_C^T z_C \equiv \begin{bmatrix} \check{z}_C \\ \tilde{z}_C \end{bmatrix} \begin{matrix} p \\ q-p \end{matrix} \, , \qquad \check{P}_C^T \Pi_C^T R_{C3} \equiv \begin{bmatrix} \check{R}_{C3} \\ \tilde{R}_{C3} \end{bmatrix} \begin{matrix} p \\ q-p \end{matrix} \, ,$$

$$(6.15) \qquad \bar{B}_A \check{Q}_C \equiv [\, R_A \, , R_{AC} \,] \, n \, , \qquad \check{Q}_C^T w_{AC} \equiv \begin{bmatrix} w_A \\ w_C \end{bmatrix} \begin{matrix} k-l-p \\ p \end{matrix} \, . \\ \phantom{(6.15) \qquad \bar{B}_A \check{Q}_C \equiv [\,} {\scriptstyle k-l-p \quad p}$$

Therefore from (6.12) with (6.13)–(6.15) we observe that problem GLLS$_0$ can be further transformed into

$$\min(\|w_A\|_2^2 + \|w_C\|_2^2 + \|w_3\|_2^2)$$

$$(6.16) \qquad \text{subject to} \quad \begin{bmatrix} z_A \\ \check{z}_C \\ \tilde{z}_C \\ z_3 \end{bmatrix} = \begin{bmatrix} U_A \\ 0 \\ 0 \\ 0 \end{bmatrix} x + \begin{bmatrix} R_A & R_{AC} & R_{A3} \\ 0 & R_C & \check{R}_{C3} \\ 0 & 0 & \tilde{R}_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} w_A \\ w_C \\ w_3 \end{bmatrix}.$$

Notice that the bottom block equation in (6.16) can uniquely determine the estimate of w_3, since R_3 is nonsingular. Since the equations in (6.16) are consistent, the third block equation can be removed. Then obviously the optimal solution x_0, $w_0 \equiv [(w_A^0)^T, (w_C^0)^T, (w_3^0)^T]^T$ satisfies

$$(6.17) \qquad w_A^0 = 0, \qquad \begin{bmatrix} U_A & R_{AC} & R_{A3} \\ 0 & R_C & \check{R}_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} x_0 \\ w_C^0 \\ w_3^0 \end{bmatrix} = \begin{bmatrix} z_A \\ \check{z}_C \\ z_3 \end{bmatrix}.$$

Then from the second equality in (6.8) and the second equality in (6.15), we observe that the GLS residual u_0 (the optimal u under H$_0$) satisfies

$$(6.18) \qquad u_0 \equiv Q \begin{bmatrix} \check{Q}_C & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} w_A^0 \\ w_C^0 \\ w_3^0 \end{bmatrix}, \qquad \|u_0\|_2^2 = \|w_C^0\|_2^2 + \|w_3^0\|_2^2.$$

As in section 4, $w_3^a = w_3^0$ (see (6.10) and (6.17)), so using the formula of the test statistic in (4.10) we have from (6.11) and (6.18) that

$$(6.19) \qquad \delta_{\mathrm{TS}} = \sigma^{-2}(\|u_0\|_2^2 - \|u_a\|_2^2) = \sigma^{-2}\|w_C^0\|_2^2,$$

so that once again introducing $C\nabla$ decreases the square of the GLS residual from $\|u_0\|_2^2$ to $\|u_a\|_2^2$, and δ_{TS} is a scaled measure of this decrease.

Now we look at the distribution of δ_{TS}. Using the fact that orthogonal transformations of a unit covariance random vector do not change its covariance matrix, we see from (6.16) and (6.17) that $w_C^0 = w_C$ and from (6.15), (6.8), and (6.1) that

$$w_C^0 \sim \mathcal{N}(0, \sigma^2 I_p) \quad \text{under H}_0.$$

Under H_a the transformed model has the form (6.9). Let (cf. (6.14))

$$\check{P}_C^T \Pi_C^T U_C \equiv \begin{bmatrix} \check{U}_C \\ \tilde{U}_C \end{bmatrix} \begin{matrix} p \\ q-p \end{matrix} \; ;$$

then (6.9) with the z_4 rows removed can be transformed into (cf. (6.16))

$$(6.20) \qquad \begin{bmatrix} z_A \\ \check{z}_C \\ \tilde{z}_C \\ z_3 \end{bmatrix} = \begin{bmatrix} U_A \\ 0 \\ 0 \\ 0 \end{bmatrix} x + \begin{bmatrix} U_{AC} \\ \check{U}_C \\ \tilde{U}_C \\ 0 \end{bmatrix} \nabla + \begin{bmatrix} R_A & R_{AC} & R_{A3} \\ 0 & R_C & \check{R}_{C3} \\ 0 & 0 & \tilde{R}_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} w_A \\ w_C \\ w_3 \end{bmatrix}.$$

Then from (6.17) and (6.20) we see $R_C w_C^0 = \check{z}_C = \check{U}_C \nabla + R_C w_C$, so

$$w_C^0 \sim \mathcal{N}(R_C^{-1} \check{U}_C \nabla, \sigma^2 I_p) \quad \text{under } H_a.$$

Thus δ_{TS} in (6.19) is the desired test statistic, since (cf. (4.23))

$$\delta_{\mathrm{TS}} \sim \chi^2(p, 0) \quad \text{under } H_0; \qquad \delta_{\mathrm{TS}} \sim \chi^2(p, \sigma^{-2} \| R_C^{-1} \check{U}_C \nabla \|_2^2) \quad \text{under } H_a.$$

From (6.13) we see that p is the rank of $\bar{B}_C$, and this can be determined in computing (6.13). Here we give a theoretical formula for it. Suppose for P in (6.7)

$$P \equiv [P_1, P_2, P_3] \quad \text{such that} \quad \begin{bmatrix} P_2^T \\ P_3^T \end{bmatrix} BQ = \begin{bmatrix} \bar{B}_C & R_{C3} \\ 0 & R_3 \\ 0 & 0 \end{bmatrix} \begin{matrix} q \\ l \\ m-n-q-l \end{matrix} .$$

Thus it follows that

$$(6.21) \quad \mathrm{rank}([P_2, P_3]^T BQ) = \mathrm{rank}(\bar{B}_C) + \mathrm{rank}(R_3), \quad \mathrm{rank}(P_3^T BQ) = \mathrm{rank}(R_3) = l.$$

But

$$(6.22)$$
$$\mathrm{rank}([P_2, P_3]^T BQ) = \mathrm{rank}((P_2 P_2^T + P_3 P_3^T)B), \quad \mathrm{rank}(P_3^T BQ) = \mathrm{rank}(P_3 P_3^T B),$$

and from the QR factorization of $[A, C]$ in (6.7)

$$(6.23) \qquad P_2 P_2^T + P_3 P_3^T = I - AA^\dagger, \quad P_3 P_3^T = I - [A, C][A, C]^\dagger.$$

Substituting (6.22) and (6.23) into (6.21), we obtain

$$p = \mathrm{rank}(\bar{B}_C) = \mathrm{rank}((I - AA^\dagger)B) - \mathrm{rank}((I - [A, C][A, C]^\dagger)B).$$

Notice that when B is nonsingular,

$$p = \mathrm{rank}(I - AA^\dagger) - \mathrm{rank}(I - [A, C][A, C]^\dagger) = (m - n) - (m - n - q) = q.$$

This is the case we discussed in section 4.

We could discuss the computation of the covariance matrices for the estimates in the same way as we did in section 4. But for brevity we omit this discussion.

Since the theory and algorithm here can handle a singular noise covariance matrix, we can now easily handle linear equality constraints without needing to develop a new theory or algorithm. Suppose the two hypotheses are

$$H_0: \quad y = Ax + v, \quad v \sim \mathcal{N}(0, \sigma^2 V), \quad \text{subject to} \quad Ex = d;$$
$$H_a: \quad y = Ax + C\nabla + v, \quad v \sim \mathcal{N}(0, \sigma^2 V), \quad \text{subject to} \quad Ex = d,$$

where $[A, C]$ has full column rank and the symmetric nonnegative definitive V has the factorization (4.1) with B being of full column rank. Then we can apply our algorithm directly to the following GLLS problems to find the test statistic and the estimates as well:

$$\text{GLLS}_0: \quad \min \|u\|_2^2 \quad \text{subject to} \quad \begin{bmatrix} y \\ d \end{bmatrix} = \begin{bmatrix} A \\ E \end{bmatrix} x + \begin{bmatrix} B \\ 0 \end{bmatrix} u;$$

$$\text{GLLS}_a: \quad \min \|u\|_2^2 \quad \text{subject to} \quad \begin{bmatrix} y \\ d \end{bmatrix} = \begin{bmatrix} A \\ E \end{bmatrix} x + \begin{bmatrix} C \\ 0 \end{bmatrix} \nabla + \begin{bmatrix} B \\ 0 \end{bmatrix} u.$$

Note that $\left[\begin{smallmatrix} A \\ E \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} A & C \\ E & 0 \end{smallmatrix}\right]$, and $\left[\begin{smallmatrix} B \\ 0 \end{smallmatrix}\right]$ have full column rank.

7. Summary. The standard formula for the generalized likelihood ratio test statistic is not suitable for numerical computation when any of A, $[A, C]$, and the noise covariance matrix $\sigma^2 V$ in (2.1) and (2.2) are ill-conditioned. Also the formula is not defined when the noise covariance matrix is singular. To overcome these problems, we gave a new formula for the test statistic after reformulating the two GLLS problems for estimating the parameter vectors. The new formula is mathematically equivalent to the standard one, in which the noise covariance matrix is nonsingular, and is well defined when the noise covariance matrix is singular. With this formula, a numerically stable algorithm based on Paige's GLLS method was proposed to compute it. The computations of the test statistic and the BLUEs of the parameter vectors are closely connected—they are obtained from the solutions of two upper triangular linear equations, (4.15) and (4.19). We also showed how to compute reliable representations of the error covariance matrices for the BLUEs; see (4.25) and (4.26). In our algorithm, we did not assume that B has any special structure. If B comes from the Cholesky factorization of the noise covariance matrix, then we can use the structure of B to design a faster algorithm by following the approach in [8].

REFERENCES

[1] Å. BJÖRCK, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.
[2] N. J. HIGHAM, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., SIAM, Philadelphia, 2002.
[3] B. HOFMANN-WELLENHOF, H. LICHTENEGGER, AND J. COLLINS, *GPS Theory and Practice*, 4th ed., Springer, New York, 1997.
[4] K. R. KOCH, *Parameter Estimation and Hypothesis Testing in Linear Models*, 2nd ed., Springer, Berlin, 1999.
[5] S. KOUROUKLIS AND C. C. PAIGE, *A constrained least squares approach to the general Gauss-Markov linear model*, J. Amer. Statist. Assoc., 76 (1981), pp. 620–625.
[6] C. C. PAIGE, *Numerically stable computations for general univariate linear models*, Comm. Statist. B—Simulation Comput., 7 (1978), pp. 437–453.
[7] C. C. PAIGE, *Computer solution and perturbation analysis of generalized linear least squares problems*, Math. Comp., 33 (1979), pp. 171–183.
[8] C. C. PAIGE, *Fast numerically stable computations for generalized linear least squares problems*, SIAM J. Numer. Anal., 16 (1979), pp. 165–171.

[9] C. C. Paige, *Some aspects of generalized QR factorizations*, in Reliable Numerical Computation, M. G. Cox and S. Hammarling, eds., Oxford University Press, New York, 1990, pp. 73–91.

[10] G. P. H. Styan, *LUSH and other uncorrelated residuals from linear models*, Canad. J. Statist., 1 (1973), pp. 131–132.

[11] P. J. Teunissen, *Quality control and GPS*, in GPS for Geodesy, 2nd ed., P. J. Teunissen and A. Kleusberg, eds., Springer, Berlin, 1998, pp. 271–318.

[12] P. J. G. Teunissen, *Testing Theory: An Introduction*, Delft University Press, Delft, 2000.