SIAM J. MATRIX ANAL. APPL.
Vol. 10, No. 3, pp. 294–315, July 1989

003

ANALYSIS AND PROPERTIES OF THE GENERALIZED TOTAL LEAST SQUARES PROBLEM $AX \approx B$ WHEN SOME OR ALL COLUMNS IN A ARE SUBJECT TO ERROR*

SABINE VAN HUFFEL† AND JOOS VANDEWALLE†

Abstract. The Total Least Squares (TLS) method has been devised as a more global fitting technique than the ordinary least squares technique for solving overdetermined sets of linear equations $AX \approx B$ when errors occur in all data. This method, introduced into numerical analysis by Golub and Van Loan, is strongly based on the Singular Value Decomposition (SVD). If the errors in the measurements A and B are uncorrelated with zero mean and equal variance, TLS is able to compute a strongly consistent estimate of the true solution of the corresponding unperturbed set $A_0X = B_0$. In the statistical literature, these coefficients are called the parameters of a classical errors-in-variables model.

In this paper, the TLS problem, as well as the TLS computations, are generalized in order to maintain consistency of the parameter estimates in a general errors-in-variables model; i.e., some of the columns of A may be known exactly and the covariance matrix of the errors in the rows of the remaining data matrix may be arbitrary but positive semidefinite and known up to a factor of proportionality. Here, a computationally efficient and numerically reliable Generalized TLS algorithm GTLS, based on the Generalized SVD (GSVD), is developed. Additionally, the equivalence between the GTLS solution and alternative expressions of consistent estimators, described in the literature, is proven. These relations allow the main statistical properties of the GTLS solution to be deduced. In particular, the connections between the GTLS method and commonly used methods in linear regression analysis and system identification are pointed out. It is concluded that under mild conditions the GTLS solution is a consistent estimate of the true parameters of any general multivariate errors-in-variables model in which all or some subsets of variables are observed with errors.

Key words. total least squares, generalized singular value decomposition, errors in variables, numerical linear algebra

AMS(MOS) subject classifications. 15A18, 65F20

C.R. classification. G1.3

1. Introduction.

Every linear parameter estimation problem gives rise to an overdetermined set of linear equations $AX \approx B$. Whenever *both* the data matrix A and observation matrix B are *inaccurate*, the Total Least Squares (TLS) technique is appropriate for solving this set. The problem of *linear parameter estimation* arises in a broad class of scientific disciplines such as signal processing, automatic control, system theory, general engineering, statistics, physics, economics, biology, and medicine. It can be described by a linear equation:

$$a_1x_1 + \cdots + a_nx_n = b \tag{1}$$

where $a_1, \cdots, a_n$ and b denote the variables and $x = [x_1, \cdots, x_n]^T \in \mathscr{R}^n$ plays the role of a parameter vector that characterizes the special system ($\mathscr{R}$ denotes the set of real numbers). A basic problem of applied mathematics is to determine an estimate of the true but unknown parameters from certain measurements of the variables. This gives rise to an overdetermined set of m linear equations ($m \geqq n$):

$$Ax \approx b \tag{2}$$

* Received by the editors March 29, 1988; accepted for publication (in revised form) November 4, 1988. The first author is a Senior Research Assistant of the Belgian National Fund of Scientific Research.

† Electronica-Systemen-Automatisatie-Technologie Laboratory, Department of Electrical Engineering, Katholieke Universiteit Leuven, Kardinaal Mercierlaan 94, B-3030 Heverlee, Belgium (vanhuffel@kulesat.uucp and vanhuffel%kulesat.uucp@BLEKUL60).

where the ith row of the data matrix $A \in \mathcal{R}^{m\times n}$ and the vector of observations $b \in \mathcal{R}^m$ contain the measurements of the variables $a_1, \cdots, a_n$ and b, respectively.

In the classical least squares (LS) approach the measurements A of the variables a_i (the left-hand side of (2)) are assumed to be free of error and, hence, all errors are confined to the observation vector b (the right-hand side of (2)). However, this assumption is frequently unrealistic: sampling errors, human errors, modelling errors, and instrument errors may imply inaccuracies of the data matrix A. For those cases, TLS has been devised and amounts to fitting a "best" subspace to the measurement points (A_i^T, b_i), $i = 1, \cdots, m$, where A_i is the ith row of A.

Much of the literature concerns the classical TLS problem in which all variables are observed with errors (see, e.g., [31], [14], [11], [34]) with particular emphasis on the univariate linefitting problem, i.e., $n = 1$ in (1) (see, e.g., [1], [29]). If the errors on the measurements A and b are uncorrelated with zero mean and equal variance, then under mild conditions the TLS solution $\hat{x}$ of (2) is a strongly consistent estimate of the true but unknown parameters x in (1), i.e., $\hat{x}$ converges to x with probability one as the number of equations m tends to infinity. However in many linear parameter estimation problems, some of the variables a_i in (1) may be observed without error. This implies that *some of the columns* of A in (2) are assumed to be *known exactly*. For instance, every intercept model

$$\alpha + a_1x_1 + \cdots + a_nx_n = b \tag{3}$$

gives rise to an overdetermined set of equations

$$[1_m; A]\begin{bmatrix}\alpha \\ x\end{bmatrix} = b \tag{4}$$

with $1_m = [1, \cdots, 1]^T$ in which the first column of the left-hand side matrix is known exactly [11], [13]. To maximize the accuracy of the estimated parameters x it is natural to require that the corresponding columns of A be unperturbed since they are known exactly. Moreover, the errors in the remaining data may be *correlated* and *not* equally sized. In order to maintain consistency of the result when solving these problems, the classical TLS formulation can be *generalized* as follows (E denotes the expected value operator).

GENERALIZED TLS FORMULATION. Given a set of m linear equations in $n \times d$ unknowns X:

$$AX \approx B, \qquad A \in \mathcal{R}^{m\times n}, B \in \mathcal{R}^{m\times d}, \text{ and } X \in \mathcal{R}^{n\times d}, \tag{5}$$

$$\text{Partition } A = [A_1; A_2], \qquad A_1 \in \mathcal{R}^{m\times n_1}, \quad A_2 \in \mathcal{R}^{m\times n_2}, \quad n = n_1 + n_2,$$

$$X = [X_1^T; X_2^T]^T, \qquad X_1 \in \mathcal{R}^{n_1\times d}, \quad X_2 \in \mathcal{R}^{n_2\times d}. \tag{6}$$

Assume that the columns of A_1 are known exactly and that the covariance matrix $E(\Delta^T\Delta)$ of the errors $\Delta \in \mathcal{R}^{m\times(n_2+d)}$ in the perturbed data matrix $[A_2; B]$ is given by $C \in \mathcal{R}^{(n_2+d)\times(n_2+d)}$, up to a factor of proportionality. Let $C = R_C^TR_C$ be nonsingular. Then a generalized TLS (GTLS) solution of (5)–(6) is any solution of the set

$$\hat{A}X = A_1X_1 + \hat{A}_2X_2 = \hat{B} \tag{7}$$

where $\hat{A} = [A_1; \hat{A}_2]$ and $\hat{B}$ are determined such that

$$\text{Range}(\hat{B}) \subseteq \text{Range}(\hat{A}), \tag{8}$$

$$\|[\Delta\hat{A}_2; \Delta\hat{B}]R_C^{-1}\|_F = \|[A_2 - \hat{A}_2; B - \hat{B}]R_C^{-1}\|_F \text{ is minimal.} \tag{9}$$

The *problem of finding* $[\Delta\hat{A}_2; \Delta\hat{B}]$ such that (8)–(9) are satisfied, is referred to as the GTLS *problem*. Whenever the solution is *not unique*, GTLS singles out the *minimum norm* solution, denoted by $\hat{X} = [\hat{X}_1^T; \hat{X}_2^T]^T$.

An even more general GTLS formulation, that allows for correlations between the errors in each column of $[A_2; B]$, is given in [35]. It is worth noting that when all columns of A are known exactly and when $C \sim I$, the GTLS solution reduces to the ordinary Least Squares (LS) estimate. By varying n_1 from zero to n, this formulation can handle the ordinary LS problem, as well as every TLS ($C \sim I$) and GTLS problem.

Although the name "total least squares" has appeared only recently in the literature [14], this method of fitting is certainly not new and has a long history in the statistical literature where the method is known as *orthogonal regression* or *errors-in-variables* regression. Indeed, the univariate linefitting problem ($n = 1$) was already scrutinized in the previous century [1]. About 20 years ago, the technique was extended to multivariate problems ($n > 1$) and later on to multidimensional problems that deal with more than one observation vector ($d > 1$ in (5)), e.g., [31], [11]. More recently, the TLS approach to fitting has also attracted interest outside of statistics. In the field of *numerical analysis*, this problem was first studied by Golub and Van Loan [14]. Their analysis, as well as their algorithm, is strongly based on the *Singular Value Decomposition* (SVD). Geometrical insight into the properties of the SVD has brought us independently to the same concept. We have *generalized* the algorithm of Golub and Van Loan [14] to all cases in which their algorithm fails to produce a solution, described the properties of these so-called nongeneric TLS problems and proved that the proposed generalization still satisfies the TLS criteria (8)–(9) if additional constraints are imposed on the solution space [39]–[40].

Although this linear algebraic approach is quite different, it is easy to see that the multivariate errors-in-variables regression estimate, given by Gleser [11], coincides with the TLS solution given by Golub and Van Loan [14] whenever the TLS problem has a unique minimizer. The only difference between both methods is the algorithm used: Gleser's method is based on an eigenvalue-eigenvector analysis, while the TLS Algorithm uses the SVD, which is numerically more robust. Furthermore, the TLS algorithm computes the minimum norm solution whenever the TLS problem lacks a unique solution. These extensions are not considered by Gleser. Also in the field of *experimental modal analysis*, the TLS technique (more commonly known as the H_v technique), has recently been studied [25]. And finally in the field of *system identification*, Levin [26] first studied the same problem. His method, called the *eigenvector method* or the *Koopmans-Levin method* [6], computes the same estimate as the TLS Algorithm whenever the TLS problem has a unique solution.

Much less considered is the case in which *some* of the columns of A in (5) are *known exactly*. It is quite easy to generalize the classical TLS Algorithms, given in [14], [34], and [39], in order to compute the more general TLS estimate $\hat{X} = [\hat{X}_1^T; \hat{X}_2^T]^T$ satisfying the TLS criteria (7)–(8)–(9) with $R_C \sim I$. The technique involves computing a QR *factorization of the "known" columns* A_1 and then solving a TLS *problem of reduced dimension* [12], [34, § 1.7]. Using a generalization of the Eckart–Young-Mirsky matrix approximation theorem [13], Golub, Hoffman, and Stewart have proved that this procedure indeed finds the best rank $r\,(\leqq n)$ approximation $[A_1; \hat{A}_2; \hat{B}]$ to $[A; B]$ that leaves A_1 fixed such that

$$(10)\qquad \|[A_1;A_2;B]-[A_1;\hat{A}_2;\hat{B}]\|_F = \min_{\operatorname{rank}([A_1;\tilde{A}_2;\tilde{B}]) \leqq r} \|[A_1;A_2;B]-[A_1;\tilde{A}_2;\tilde{B}]\|_F.$$

In particular, this algorithm is able to compute the *Compensated Least Squares* (CLS) estimate as derived by Guidorzi [17] and Stoica and Söderström [33]. When the only disturbance of the input-output sequences is given by zero mean white noise of equal variance, the CLS, GTLS and eigenvector methods all give the same estimate. Observe that our TLS Algorithm, that allows for exactly known columns in A and coincides with our GTLS Algorithm in § 2 for the case that $C \sim I$, is computationally more efficient than the computation procedure presented in [33].

The *generalization* of the TLS problem, presented in this paper, that allows for *correlations* between the measurement errors in the data A and B, is inspired by a generalization of the classical errors-in-variables model discussed in [9]. As said before, the TLS solution is not very meaningful unless the errors in the measurements A and B in (5) are independently derived and equilibrated. In statistical terms, this means that the errors must be uncorrelated with zero mean and all have the same variance, i.e., the associated error covariance matrix C in (6) is $\sim I$. The best statistical approach, directly related to the classical TLS concept, is the "*errors-in-variables*" model [11] that considers all observations as coming from some unknown true values plus measurement errors. The true values are assumed to follow a linear relation (1). The estimation of the parameters in this model is a problem with a long history in the statistical literature [1], yet one with a considerable recent emphasis. Much less considered is the following *general "errors-in -variables" model*, directly related to our generalized TLS formulation given above, in which *some subset* of variables is observed *with errors*.

GENERAL ERRORS-IN-VARIABLES MODEL FORMULATION.

$$
\begin{aligned}
\underset{m\times d}{B_0} &= A_0X = \underset{m\times n_1}{A_1}\;\underset{n_1\times d}{X_1} + \underset{m\times n_2}{(A_2)_0}\;\underset{n_2\times d}{X_2}\\
A_2 &= (A_2)_0 + \Delta A_2\\
B &= B_0 + \Delta B.
\end{aligned}
\tag{11}
$$

X_1 and X_2 are the true but unknown parameters to be estimated; A_1 and $(A_2)_0$ are of full column rank. They consist of constants as well as B_0. A_1 is known but $(A_2)_0$ and B_0 not. The observations A_2 and B of the unknown values $(A_2)_0$ and B_0 contain measurement errors ΔA_2 and ΔB such that the rows of $[\Delta A_2; \Delta B]$ are independently and identically distributed (i.i.d.) with zero mean and known positive definite covariance matrix C_Δ, up to a factor of proportionality c^2, i.e.,

$$
C_\Delta = c^2 C = c^2\begin{bmatrix} C_a & C_{ab}\\ C_{ab}^T & C_b\end{bmatrix} \text{ with } C_{(n_2+d)\times(n_2+d)} \text{ known.}
\tag{12}
$$

Observe that this model requires that the rows of $[\Delta A_2; \Delta B]$ are independently derived. If this assumption is not satisfied, we can premultiply the data $[A; B]$ in advance with an appropriate $m \times m$ matrix D such that the preprocessed data $D[A; B] = [DA_1; DA_2; DB]$ satisfy the assumptions of model (11). If D is ill-conditioned, this premultiplication must be performed implicitly, as outlined in [35]. This preprocessing operation does not affect the true solution X of model (11).

Now to compute strongly consistent estimates of the true but unknown parameters X of model (11)–(12), the classical TLS Algorithm, as given in [14], [34, § 1.8], and [39] can be used whenever $C_\Delta \sim I$. However, in case that the covariance matrix C_Δ is more *general*, the classical TLS algorithm may *not* be used *straightforwardly*. To maintain consistency, the data can be *pretreated* appropriately such that the covariance matrix of the transformed data is diagonal with equal error variances (i.e., $C_\Delta = c^2 I$). The classical TLS Algorithm can then be used to solve this transformed set of equations and finally the solution of the transformed system must be converted to a solution of the original set of equations. Such transformation procedures are described in [10], [11], and [34,

§ 4.5] for the case that $n_1 = 0$ and C in (12) is positive definite. This approach, however, is not recommended in general since computing $[A;\ B]R_C^{-1}$ (with $C = R_C^T R_C$) usually leads to unnecessarily large numerical errors if R_C is ill-conditioned with respect to the solution of equations.

The *objective* of this paper is to *solve the generalized* TLS *problem*, defined above, by making use of the *Generalized* SVD (GSVD). Hereto, a computationally efficient and numerically reliable *Generalized Total Least Squares* (GTLS) *Algorithm* is developed. As shown in § 3, this algorithm is able to compute consistent estimates of the parameters in any errors-in-variables model (11) directly *without transforming* the data *explicitly*. The great advantage of the GSVD is that it replaces these transformation procedures by *one*, which is numerically reliable and can more easily handle (nearly) *singular* covariance matrices C in (12). Moreover, the GSVD reveals the *structure* of the general errors-in-variables model (11) more clearly than the usual transformation procedures. Additionally, statistical properties of the GTLS solution are deduced by proving the equivalence between the GTLS solution and alternative expressions of consistent estimators described in the statistical literature [9], [33]. The GSVD of a matrix pair (A, B) is defined as follows [15], [41].

THEOREM 1. GSVD of (A, B). *If* $A \in \mathcal{R}^{m\times n}(m \geqq n)$ *and* $B \in \mathcal{R}^{p\times n}$, *then there exist orthogonal* $T \in \mathcal{R}^{m\times m}$ *and* $W \in \mathcal{R}^{p\times p}$ *and a nonsingular* $Z \in \mathcal{R}^{n\times n}$ *such that*

$$T^T AZ = D_A \quad \textit{and} \quad W^T BZ = D_B \tag{13}$$

with

$$D_A = \operatorname{diag}(\alpha_1, \cdots, \alpha_n) \in \mathcal{R}^{m\times n}, \qquad \alpha_i \geqq 0$$

and

$$D_B = \operatorname{diag}(\beta_1, \cdots, \beta_q) \in \mathcal{R}^{p\times n}, \qquad \beta_i \geqq 0 \quad q = \min\{p, n\}$$

$$\beta_1 \geqq \cdots \geqq \beta_r > \beta_{r+1} = \cdots = \beta_q = 0, \qquad r = \operatorname{rank}(B).$$

The assumption $m \geqq n$ is not restrictive from the applications point of view. The elements of the set $\sigma(A, B) = \{\alpha_i/\beta_i,\ i = 1, \cdots, r\}$ are referred to as the ordinary *generalized singular values* of A and B. The remaining generalized singular values α_i/β_i in which α_i is nonzero (respectively, zero) and β_i is zero, are called *infinite* (respectively, *indefinite*) [3]. It is worth emphasizing that infinite generalized singular values are not necessarily badly behaved. In fact, the infinite generalized singular values of (A, B) are the zero generalized singular values of (B, A) since the roles of A and B are interchangeable. Theorem 1 is a *generalization of the ordinary* SVD in that if $B = I_n$ then $\sigma(A, B)$ equals the singular value spectrum $\sigma(A)$ of matrix A. Note that there exists an intimate theoretical link between the GSVD of the matrix pair (A, B) and the *generalized symmetric eigenvalue problem* [15]. Indeed

$$\sigma_i \in \sigma(A, B) \Leftrightarrow \sigma_i^2 \in \lambda(A^T A, B^T B) \tag{14}$$

where $\lambda(A^T A, B^T B)$ is the set of generalized eigenvalues of the matrix pair $(A^T A, B^T B)$ and correspondingly

$$A^T A z_i = \sigma_i^2 B^T B z_i \tag{15}$$

where the generalized eigenvector z_i is given by the ith column of the matrix Z in Theorem 1. This matrix diagonalizes $A^T A$ and $B^T B$ simultaneously. The value of the GSVD is that these diagonalizations can be achieved without forming $A^T A$ and $B^T B$. These connections to the generalized symmetric eigenvalue problem allow us to prove the interesting statistical properties of the GTLS solution (see § 3).

In this paper, the main emphasis is put on the *linear algebraic approach* of the GTLS problem. The statistical analysis of the GTLS solution is limited and relies on showing the correspondence of particular cases of our GTLS problem with that of others in the literature. For a more detailed statistical appraisal and consistency analysis of the GTLS solution in different applications, the reader is referred to the list of references, e.g. [2], [9], [11], [22], [33]. This paper is organized into four sections. In § 2, the GTLS Algorithm is presented. Its main difference with respect to the classical TLS Algorithm is the fact that the GSVD is used instead of the ordinary SVD. Section 3 describes the *properties* of the GTLS solution. Alternative expressions of the GTLS solution are deduced that allow us to derive the main *statistical* properties of the GTLS solution. Finally, § 4 gives the conclusions.

2. The generalized TLS Algorithm GTLS. The original motivation is to compute *consistent estimates* of the parameters in a *general errors-in-variables model* of the form (11), thereby improving the computational efficiency and numerical robustness of the methods currently used in statistics, linear regression [9], and identification [6], [30], [33]. Therefore, the classical TLS problem has been generalized in the previous section. In this section, the GTLS Algorithm which solves the GTLS problem is described in detail. As shown below, the GTLS Algorithm mainly performs *orthogonal transformations* and does not need to square or invert matrices (as done in [11]), which guarantees its *numerical reliability*. Moreover, by first performing a QR factorization [15] we need only to compute the GSVD *of a smaller submatrix*, which makes the GTLS Algorithm *computationally more efficient* than methods described in [6], [30], and [33]. The GTLS Algorithm is outlined below.

ALGORITHM. GTLS

Given

- An $m \times d$ matrix B and an $m \times n$ matrix $A = [A_1; A_2]$ whose first n_1 columns A_1 have full column rank and are known exactly, $n = n_1 + n_2$ and $m \geqq n_1$.
- An $(n_2 + d) \times (n_2 + d)$ matrix C, proportional with the covariance matrix $E(\Delta^T\Delta)$ of the errors $\Delta_{m\times(n_2+d)}$ in the matrix $[A_2; B]$, or any square root R_C of C (such that $C = R_C^T R_C$).

Step 1: QR factorization and Cholesky decomposition.

1.a. If $n_1 > 0$ then begin

compute the QR factorization of $[A_1; A_2; B]$:

$$Q^T[A_1; A_2; B] = R = \begin{array}{cc} \left[\begin{array}{cc} R_{11} & R_{12} \\ 0 & R_{22} \end{array}\right] & \begin{array}{c} n_1 \\ m-n_1 \end{array} \\ \begin{array}{cc} n_1 \quad & \quad n_2+d \end{array} & \end{array} \tag{16}$$

with Q being orthogonal and R_{11} upper triangular.

if $n_1 = n$ then begin

solve $R_{11}\hat{X} = R_{12}$ by back substitution

stop

end

end

else $R_{22} \leftarrow [A; B]$

1.b. If C is given and $\not\propto I_{n_2+d}$, compute the Cholesky decomposition of C:

$$C = R_C^T R_C \tag{17}$$

Step 2: GSVD.

2.a. Compute the GSVD (or SVD if $C \sim I$) of the matrix pair (R_{22}, R_C) as in (13):

$$\begin{aligned} &T^T R_{22} Z = \operatorname{diag}(\alpha_1, \cdots, \alpha_s), \alpha_{s+1} = \cdots = \alpha_{n_2+d} = 0, \qquad s = \min\{m-n_1, n_2+d\} \\ &W^T R_C Z = \operatorname{diag}(\beta_1, \cdots, \beta_{n_2+d}) \end{aligned} \tag{18}$$

where the generalized singular values $\sigma_i = \alpha_i/\beta_i$, $i = 1, \cdots, n_2 + d$ are organized in decreasing order of magnitude (i.e., $\sigma_i \geqq \sigma_{i+1}$) and the corresponding columns, z_i, of Z are normalized to unit norm.

2.b. If not user determined, compute the rank $r(\leqq n_2)$ of the matrix pair (R_{22}, R_C) by

$$\sigma_1 \geqq \cdots \geqq \sigma_r > R_0 \geqq \sigma_{r+1} \geqq \cdots \geqq \sigma_{n_2+d} \tag{19}$$

with R_0 a user-defined rank determinator.

2.c. Solve by back substitution:

$$Z_2 \leftarrow [z_{r+1}, \cdots, z_{n_2+d}]; \; R_{11} Z_1 = -R_{12} Z_2 \tag{20}$$

Step 3: GTLS solution $\hat{X} = [\hat{X}_1^T; \hat{X}_2^T]^T$.

3.a. If $C \not\sim I_{n+d}$, $d > 1$ and $r < n_2$, orthonormalize $[\begin{smallmatrix} Z_1 \\ Z_2 \end{smallmatrix}]$ using a QR factorization:

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = Q_z R_z; \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \leftarrow Q_z \qquad Q_z^T Q_z = I_{n_2 - r + d} \text{ and } R_z \text{ upper triangular}$$

3.b. Perform Householder transformations such that

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} Q = \begin{array}{cc} \begin{bmatrix} W & Y \\ 0 & \Gamma \end{bmatrix} & \begin{array}{c} n \\ d \end{array} \\ \begin{array}{cc} n_2 - r & d \end{array} & \end{array} \tag{21}$$

with Q being orthogonal and Γ d by d upper triangular.

(22)
If Γ is nonsingular then {GTLS problem is generic}
solve $\hat{X}\Gamma = -Y$
else {GTLS problem is nongeneric}
begin
$r \leftarrow r - \rho$ where ρ is the multiplicity of σ_r
go back to Step 2.c.
end

END

The following comments are in order:

- The GSVD can readily be applied to (5) to yield the solution of the generalized TLS problem. Just as the SVD is a valuable tool for the solution and analysis of the classical TLS problem, so the GSVD plays the same role for the generalized problem. Stable numerical methods have emerged for computing the GSVD [3], [27], [32], [42]. The methods proposed in [3] and [27], being based on an implicit *Kogbetliantz approach*, have potential for systolic implementation [4]. For background information on the GSVD the reader is strongly recommended to consult [28] and [41].

- To compute the decomposition (17) of any positive semidefinite covariance matrix C, a *Cholesky decomposition with complete pivoting* can be used. This method is proven to be one of the most numerically stable methods [5, p. 3.16]. Software for computing this decomposition is readily available, notably in the LINPACK library [5, Chap. 8]. An error analysis of this Cholesky decomposition is given in [19]. Whenever C is singular, R_C is not of full row rank. Hence, indefinite generalized singular values (i.e., $\alpha_i = \beta_i = 0$ in (18)) may occur. These values and corresponding columns in Z are to be considered in the GTLS computation.
- The QR *factorization* of $[A_1; A_2; B]$ can be computed with the LINPACK routine SQRDC [5, Chap. 9]. *Pivoting* may be done within three groups of columns: the first n_1 columns A_1, the next n_2 columns A_2 and the last d columns B. Columns may not be pivoted with columns from another group. Even for full rank problems, column pivoting seems to produce more accurate solutions [21]. If pivoting is done, the columns in R_C must be permuted correspondingly if $C \not\sim I_{n_2+d}$, and the inverse permutations must be performed in the last step of the GTLS Algorithm.
- To orthonormalize the columns of $[\begin{smallmatrix} Z_1 \\ Z_2 \end{smallmatrix}]$ in Step 3.a., a QR factorization is performed. This can again be computed with the LINPACK routine SQRDC [5, Chap. 9].
- If *no columns* of A *are known exactly* ($n_2 = n$) and $C \sim I_{n+d}$, the GTLS Algorithm reduces to the classical TLS Algorithm, given in [14], [34, § 1.8.1], and [39].
- If a subset A_1 of A is known exactly and $C \sim I_{n_2+d}$, the GTLS Algorithm reduces to the TLS *Algorithm with exactly known columns*, as described in [34, § 1.8.2]. Observe also that the GTLS algorithm solves the *ordinary* LS *problem*, using a QR factorization, if all columns of A are known exactly ($n_1 = n$).
- Mostly, the matrix pair (R_{22}, R_C) has maximal rank $r = n_2$. If $r < n_2$ (e.g., when the set of equations $AX \approx B$ is *underdetermined*), the GTLS solution is *no longer unique*. In this case, GTLS singles out the *minimum norm* solution. Indeed if the solution $[\begin{smallmatrix} \hat{X} \\ -I \end{smallmatrix}]$ is deduced from an orthonormal basis $[\begin{smallmatrix} Y \\ \Gamma \end{smallmatrix}]\begin{smallmatrix} n \\ d \end{smallmatrix}$, then

$$\|\hat{X}\|_F^2 = \|\Gamma^{-1}\|_F^2 - d \quad \text{and} \quad \|\hat{X}\|_2^2 = (1 - \sigma_{\min}^2(\Gamma))/\sigma_{\min}^2(\Gamma)$$

as proved in [15, p. 422] and [34, p. 69], based on the CS decomposition [15, Thm. 2.4-1]. Hence we have to select d base vectors $[\begin{smallmatrix} Y \\ \Gamma \end{smallmatrix}]\begin{smallmatrix} n \\ d \end{smallmatrix}$ within the Range $([\begin{smallmatrix} Z_1 \\ Z_2 \end{smallmatrix}])$ such that $\|\Gamma^{-1}\|_F$ is minimized and the minimal singular value $\sigma_{\min}(\Gamma)$ of Γ is maximized. This is done by computing an orthonormal matrix Q (e.g., by using Householder transformations) such that

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} Q = \begin{bmatrix} Z_1 Q \\ Z_2 Q \end{bmatrix} \begin{matrix} n \\ d \end{matrix} = \begin{array}{c} \begin{bmatrix} W & Y \\ 0 & \Gamma \end{bmatrix} \begin{matrix} n \\ d \end{matrix} \\ \begin{matrix} n_2 - r & \quad d \end{matrix} \end{array}$$

Z_2 and $Z_2 Q$ have the same singular values. Denote by $\tilde{\gamma}_i$, $i = 1, \cdots, d$, the singular values of a submatrix $\tilde{\Gamma}$ of Z_2 or $Z_2 Q$ (obtained by deleting $n_2 - r$ columns) and by σ_i, $i = 1, \cdots, d$, the singular values of Z_2 or $Z_2 Q$, in decreasing order of magnitude. Then, the interlacing property for singular values [15, p. 286] yields

$$\tilde{\gamma}_i \leqq \sigma_i \text{ or equivalently } \tilde{\gamma}_i^{-1} \geqq \sigma_i^{-1} \; i = 1, \cdots, d.$$

Hence, $\|\tilde{\Gamma}^{-1}\|_F^2 = \sum_{i=1}^{d} \tilde{\gamma}_i^{-2}$ and $\tilde{\gamma}_d^{-2}$ are minimized if $\tilde{\gamma}_i = \sigma_i$, for all i. Since the d by d submatrix Γ, defined above, has the same singular values as Z_2, it follows directly that the TLS solution $\hat{X} = -Y\Gamma^{-1}$, computed from (21)–(22), has minimal norm $\|\hat{X}\|_2$ and $\|\hat{X}\|_F$.

Observe that the expressions of $\|\hat{X}\|_2$ and $\|\hat{X}\|_F$ are deduced from the orthonormality of $[\begin{smallmatrix} Y \\ \Gamma \end{smallmatrix}]$. Therefore the base vectors in $[\begin{smallmatrix} Z_1 \\ Z_2 \end{smallmatrix}]$ must first be orthonormalized in Step 3.a.

If $C \sim I_{n+d}$, the columns in $[\begin{smallmatrix} Z_1 \\ Z_2 \end{smallmatrix}]$ are already orthonormal since they are the right singular vectors of $R_{22} = [A; B]$, obtained from its SVD in Step 2.a.

If $d = 1$, Γ is a scalar. To minimize $\|\hat{X}\|_F$, this scalar must be maximized. This can be accomplished by (21) such that the last column $[\begin{smallmatrix} Y \\ \Gamma \end{smallmatrix}]\begin{smallmatrix} n \\ 1 \end{smallmatrix}$ has the largest $(n+1)$th component of all unit vectors within the Range $([\begin{smallmatrix} Z_1 \\ Z_2 \end{smallmatrix}])$. Therefore, the columns of $[\begin{smallmatrix} Z_1 \\ Z_2 \end{smallmatrix}]$ need not be orthogonal.

If $r = n$, $\hat{X}$ is unique. The columns of $[\begin{smallmatrix} Z_1 \\ Z_2 \end{smallmatrix}]$ need not be orthonormalized since the GTLS solution $\hat{X}$ is invariant with respect to any base transformation P in its solution space. Indeed, with $[\begin{smallmatrix} Y \\ \Gamma \end{smallmatrix}]$ in (21) a basis (not necessarily orthogonal) of the TLS solution space, Range $([\begin{smallmatrix} \hat{X} \\ -I \end{smallmatrix}])$, we have that $[\begin{smallmatrix} Y \\ \Gamma \end{smallmatrix}]P = [\begin{smallmatrix} YP \\ \Gamma P \end{smallmatrix}]$. Hence the GTLS solution $\hat{X} = -(YP)(\Gamma P)^{-1} = -YPP^{-1}\Gamma^{-1} = -Y\Gamma^{-1}$ remains invariant.

- *If* Γ *in* (21) is nonsingular (respectively, *singular*), the GTLS solution is called generic (respectively, *nongeneric*). As shown in § 3, Γ can only be singular when A is (nearly) rank-deficient or when the set of equations $AX \approx B$ is highly incompatible. In this case, the generic GTLS solution does not exist but the GTLS computations are generalized in order to solve these *nongeneric* GTLS *problems* in the same way as the nongeneric TLS problem [39], [40].
- If A_1 does *not* have *full column rank*, i.e., rank $(A_1) = r_A < n_1$, we can always replace A_1 with a matrix having r_A independent columns selected from A_1, apply the GTLS Algorithm and set the coefficients of $\hat{X}_1$ corresponding to this missing column to zero without changing either the rank of the result or the norm of the difference (9). Note, however, that the GTLS solution $\hat{X}$ no longer has minimal norm in this case.
- Finally, observe that we only need to compute *a few vectors* z_i associated with the *smallest* generalized singular values of (R_{22}, R_C) in order to obtain the GTLS solution $\hat{X}$. Moreover, we only need to compute a *basis* of the solution space Range $([\begin{smallmatrix} Z_1 \\ Z_2 \end{smallmatrix}])$. Indeed, as proven before the GTLS solution $\hat{X}$ is invariant with respect to any base transformation P in its solution space.

Based on these properties, we were able to improve the efficiency of the TLS computations by computing the SVD only "*partially*" [36]. This results in the development of an improved algorithm *Partial Total Least Squares* (PTLS) [38]. PTLS is about two times faster than the classical TLS algorithm [14], [39], while the same accuracy can be maintained. The same modifications could be applied to the generalized SVD and GTLS Algorithms insofar as they are based on the QR Algorithm [15, § 8.2].

3. Properties of the generalized TLS solution. In this section a number of theorems are proven that *link the* GTLS *solution* with alternative *expressions of consistent estimators* given in literature. These links allow us to deduce the *main statistical properties* of the GTLS solution as shown below.

Throughout this section we will make use of the following notation and assumptions:

(23a) Consider the *set of equations* given in (5)–(6), and assume that A_1 has full column rank n_1 and is known exactly.

(23b) The n by n *identity matrix* is denoted by I_n.

(23c) Let the *covariance matrix* $E(\Delta^{*T}\Delta^*)$ of the errors Δ^* in $[A; B] = [A_1; A_2; B]$ be given by C^*, up to a factor of proportionality:

$$C^* = \begin{array}{cc} \begin{bmatrix} C_a^* & C_{ab}^* \\ C_{ab}^{*T} & C_b^* \end{bmatrix} & \begin{array}{c} n \\ d \end{array} \\ \begin{array}{cc} n & \quad d \end{array} & \end{array} = \begin{array}{cc} \begin{bmatrix} 0 & 0 \\ 0 & C \end{bmatrix} & \begin{array}{c} n_1 \\ n_2+d \end{array} \\ \begin{array}{cc} n_1 & n_2+d \end{array} & \end{array}$$

where

$$C = \begin{array}{cc} \begin{bmatrix} C_a & C_{ab} \\ C_{ab}^T & C_b \end{bmatrix} & \begin{array}{c} n_2 \\ d \end{array} \\ \begin{array}{cc} n_2 & \quad d \end{array} & \end{array}$$

is the positive definite covariance matrix $E(\Delta^T\Delta)$ of the errors Δ in $[A_2; B]$, up to a factor of proportionality.

(23d) Let R_C^* be any *square root* of C^*, defined by $C^* = R_C^{*T}R_C^*$ and partitioned as follows:

$$R_C^* = \begin{array}{cc} \begin{bmatrix} R_{Ca}^* & R_{Cab}^* \\ R_{Cba}^* & R_{Cb}^* \end{bmatrix} & \begin{array}{c} n \\ d \end{array} \\ \begin{array}{cc} n & \quad d \end{array} & \end{array} = \begin{array}{cc} \begin{bmatrix} 0 & 0 \\ 0 & R_C \end{bmatrix} & \begin{array}{c} n_1 \\ n_2+d \end{array} \\ \begin{array}{cc} n_1 & n_2+d \end{array} & \end{array}$$

where R_C is any square root of C, i.e., $C = R_C^T R_C$.

(23e) Denote by

$$\hat{X}_{n\times d} = \begin{bmatrix} \hat{X}_1 \\ \hat{X}_2 \end{bmatrix} \begin{array}{c} n_1 \\ n_2 \end{array}$$

the GTLS solution, as computed by the GTLS algorithm (given in § 2).

(23f) Let $\sigma(A, B)$ (respectively, $\lambda(A, B)$) be the set of *generalized singular values* σ_i (respectively, *generalized eigenvalues* λ_i) of the matrix pair (A, B), organized in decreasing order of magnitude, i.e., $\sigma_i \geqq \sigma_{i+1}$ (respectively, $\lambda_i \geqq \lambda_{i+1}$). Analogously, if $B \sim I$, then $\sigma(A)$ (respectively, $\lambda(A)$) denotes the ordinary singular values (respectively, ordinary eigenvalues) of A.

Based on the properties of the (generalized) eigenvalue decomposition and the (generalized) SVD [15, § 8.6], the following links between the different problems can be established (assume $C = C^*$ and $m \geqq n$ for simplicity):

$$\text{The generalized eigenvalue problem: } [A;B]^T[A;B]z_i = \lambda_i C z_i$$

$$\Updownarrow$$

$$\text{The ordinary eigenvalue problem: } C^{-1}[A;B]^T[A;B]z_i = \lambda_i z_i$$

$$\Updownarrow$$

The generalized SVD of $([A;B], R_C)$:

$$T^T[A;B]Z = \operatorname{diag}(\alpha_1, \cdots, \alpha_{n+d}), \qquad \sigma_i^2 = \lambda_i = \alpha_i^2/\beta_i^2$$

$$W^T R_C Z = \operatorname{diag}(\beta_1, \cdots, \beta_{n+d}), \; z_i = i\text{th column of } Z$$

$$\Updownarrow$$

The symmetric eigenvalue problem: $R_C^{-T}[A;B]^T[A;B]R_C^{-1}v_i=\lambda_i v_i$ and $z_i=R_C^{-1}v_i$

$$\Updownarrow$$

The ordinary SVD of $[A;B]R_C^{-1}=\sum_{i=1}^{n+d}\sigma_i u_i v_i^T \quad \sigma_i^2=\lambda_i \quad z_i=R_C^{-1}v_i.$

These links are used to prove the main theorems in this section.

THEOREM 2. *Consider the notation and assumptions* (23). *Let* $r^*=n_1+r\leqq n$ *be the rank of* $([A_1;\,A_2;\,B],\,R_C^*)$ *as computed by the* GTLS *Algorithm from* (19), *and assume that* Γ *in* (21) *is nonsingular; then*

$$\operatorname{Range}\left(\begin{bmatrix}\hat{X}\\ -I_d\end{bmatrix}\right)\subseteq \operatorname{Range}(Z^*) \textit{ such that } \|\hat{X}\|_F \textit{ and } \|\hat{X}\|_2 \textit{ are minimal}$$

where $Z^*_{(n+d)\times(n+d-r^*)}$ *contains the vectors* z_i *associated with the* $(n+d-r^*)$ *smallest generalized singular values, obtained from the* GSVD (13) *of the matrix pair* $([A;\,B],\,R_C^*)$.

Proof. Set $s=n+d-r^*$. Let Σ_2 be a diagonal matrix containing on its diagonal the s smallest generalized singular values of the GSVD (13) of the matrix pair $([A;\,B],\,R_C^*)$ and let

$$Z^*=\begin{bmatrix}Z_1\\ Z_2\end{bmatrix}\begin{matrix}n_1\\ n_2+d\end{matrix}$$

be the s corresponding columns of the nonsingular matrix $Z_{(n+d)\times(n+d)}$. Now use the link (15) between the GSVD of $([A;\,B],\,R_C^*)$ and the generalized symmetric eigenvalue problem $([A;\,B]^T[A;\,B],\,C^*)$:

$$[A;B]^T[A;B]\begin{bmatrix}Z_1\\ Z_2\end{bmatrix}=C^*\begin{bmatrix}Z_1\\ Z_2\end{bmatrix}\Sigma_2^2. \tag{24}$$

Compute the QR factorization (16) of $[A;\,B]$.

Since Q is orthogonal, the generalized eigenvalues Σ_2^2 and corresponding eigenvectors $[\begin{smallmatrix}Z_1\\ Z_2\end{smallmatrix}]$ of $([A;\,B]^T[A;\,B],\,C^*)$ and $(R^TR,\,C^*)$ coincide. Hence, (24) yields

$$R^TR\begin{bmatrix}Z_1\\ Z_2\end{bmatrix}=\begin{bmatrix}R_{11}^TR_{11} & R_{11}^TR_{12}\\ R_{12}^TR_{11} & R_{12}^TR_{12}+R_{22}^TR_{22}\end{bmatrix}\begin{bmatrix}Z_1\\ Z_2\end{bmatrix}=\begin{bmatrix}0 & 0\\ 0 & C\end{bmatrix}\begin{bmatrix}Z_1\\ Z_2\end{bmatrix}\Sigma_2^2$$

$$\text{or } R_{11}^T(R_{11}Z_1+R_{12}Z_2)=0 \tag{25}$$

$$R_{12}^T(R_{11}Z_1+R_{12}Z_2)+R_{22}^TR_{22}Z_2=CZ_2\Sigma_2^2. \tag{26}$$

Since A_1 has full column rank, R_{11} is nonsingular. Hence, the columns of R_{11} span the whole n_1-dimensional space $\mathscr{R}^{n_1}$ and thus, (25) is only satisfied if

$$R_{11}Z_1+R_{12}Z_2=0 \tag{27}$$

Substituting (27) into (26), we obtain

$$R_{22}^TR_{22}Z_2=CZ_2\Sigma_2^2. \tag{28}$$

Equation (28) implies that the generalized eigenvalues Σ_2^2 and corresponding eigenvectors Z_2 are obtained from the symmetric eigenvalue decomposition of $(R_{22}^TR_{22},\,C)$ or equivalently, from the GSVD of $(R_{22},\,R_C)$. It is precisely this matrix Z_2 which is computed in Step 2 of the GTLS Algorithm. Once Z_2 is computed, Z_1 is obtained from (27), as also done in Step 2.c of our GTLS algorithm

$$Z_1 = -R_{11}^{-1} R_{12} Z_2. \tag{29}$$

Hence, $[\begin{smallmatrix} Z_1 \\ Z_2 \end{smallmatrix}]$, as computed in Step 2 of our GTLS algorithm, equals precisely the vectors z_i in Z^*, associated with the $(n+d-r^*)$ smallest generalized singular values obtained from the GSVD (13) of the matrix pair $([A; B], R_C^*)$.

Since the orthonormalization of $[\begin{smallmatrix} Z_1 \\ Z_2 \end{smallmatrix}]$ in Step 3.a of our GTLS algorithm (if needed) does not change its range and since Range $([\begin{smallmatrix} Y \\ \Gamma \end{smallmatrix}]) \subseteq$ Range $([\begin{smallmatrix} Z_1 \\ Z_2 \end{smallmatrix}])$ from (21)–(22), it follows that

$$\text{Range}\left(\begin{bmatrix} \hat{X} \\ -I_d \end{bmatrix}\right) = \text{Range}\left(\begin{bmatrix} Y \\ \Gamma \end{bmatrix}(-\Gamma^{-1})\right) \subseteq \text{Range}\,(Z^*) = \text{Range}\left(\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}\right)$$

As proven in § 2, $\hat{X} = -Y\Gamma^{-1}$ has minimal norm $\|\hat{X}\|_F$ and $\|\hat{X}\|_2$. □

This theorem allows us to deduce the following *relationships* between the GSVD of the matrix pairs $([A; B], R_C^*)$, (R, R_C^*) and (R_{22}, R_C) (with R, R_{22} defined in (16)).

COROLLARY 1. *Consider the notation and assumptions* (23) *and let* (16) *be the* QR *factorization of* $[A_1; A_2; B]$; *then*

$$\forall i = 1, \cdots, n_2 + d: \sigma_i \in \sigma(R_{22}, R_C) \Leftrightarrow \sigma_{n_1+i} \in \sigma([A; B], R_C^*) \Leftrightarrow \sigma_{n_1+i} \in \sigma(R, R_C^*)$$

and the $(n_2 + d)$ *corresponding vectors* $Z_{R_{22}}^*$ *of* (R_{22}, R_C), Z_{AB}^* *of* $([A; B], R_C^*)$ *and* Z_R^* *of* (R, R_C^*) *obtained from their respective* GSVD (13) *are, up to a normalization factor, related by*

$$Z_{AB}^* = Z_R^* = \begin{bmatrix} -R_{11}^{-1} R_{12} Z_{R_{22}}^* \\ Z_{R_{22}}^* \end{bmatrix}$$

Proof. The proof follows straightforwardly from the proof of Theorem 2. □

Theorem 2 and Corollary 1 imply that the GTLS solution can also be computed from the GSVD of $([A; B], R_C^*)$, namely, from the vectors z_i corresponding to its smallest generalized singular values, as follows:

$$\text{Range}\left(\begin{bmatrix} \hat{X} \\ -I \end{bmatrix}\right) = \text{Range}\,(Z^*) \quad \text{and} \quad \hat{X} = -Z_1^* Z_2^{*-1}$$

where

$$Z^* = \begin{array}{c} \begin{bmatrix} Z_1^* \\ Z_2^* \end{bmatrix} \end{array}\begin{array}{l} n \\ d \end{array}$$

are the vectors z_i associated with the d smallest generalized singular values obtained from the GSVD (13) of $([A; B], R_C^*)$.

For the *one-dimensional* GTLS problem (i.e., $d = 1$ in (5)) with A of *full column rank* and Γ in (21) nonsingular (this is the problem mostly considered in literature), this means that the GTLS solution can also be computed from the vector z_{n+1} corresponding to the smallest singular value σ_{n+1} of the GSVD (13) of $([A; B], R_C^*)$ or equivalently, from the generalized eigenvector z_{n+1} corresponding to the smallest generalized eigenvalue σ_{n+1}^2 of the matrix pair $([A; B]^T[A; B], C^*)$. However computing the GTLS solution in this way requires the GSVD of a *larger* matrix pair than the matrix pair (R_{22}, R_C) used in the GTLS Algorithm. It is evident that our GTLS algorithm is *computationally more efficient* than the Koopmans–Levin method described in [6], and the compensated LS estimation method described in [33]. These methods are based on computing the GSVD of the matrix pair $([A; B], R_C^*)$ even if R_C^* is given by

$$\begin{bmatrix} 0 & 0 \\ 0 & I_{n_2+d} \end{bmatrix}.$$

Since $R_C \sim I$ in the latter case, the solution can be computed by the TLS algorithm with exactly known columns, given in [34, § 1.8.2], which first performs a QR factorization of the known columns and then proceeds with an ordinary SVD of the submatrix R_{22}.

In Theorem 1 we assumed that Γ in (21) is *nonsingular*, i.e., the GTLS solution is *generic*. Analogously to the classical TLS problem (see [34], [39]), we can deduce conditions that guarantee the *existence and uniqueness* of the generic GTLS solution. Hereto, we apply the results of Theorem 2.

THEOREM 3. Existence and uniqueness of the generic GTLS solution. *Consider the notation and assumptions* (23). *Let $m \geqq n$ and denote by σ' (respectively, σ) the smallest (respectively, $(n+1)$th) generalized singular value of (A, R_{Ca}^*) (respectively, $([A; B], R_C^*)$), then*

$$\sigma' > \sigma \Rightarrow \textit{the GTLS solution } \hat{X} = \begin{bmatrix} \hat{X}_1 \\ \hat{X}_2 \end{bmatrix} \textit{ is unique and generic.}$$

Proof. Compute the QR factorization (16) of $[A_1; A_2; B]$ and let R be partitioned as follows:

$$R = \begin{array}{c} \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix} \begin{array}{l} n_1 \\ m-n_1 \end{array} \\ \begin{array}{cc} n_1 & n_2+d \end{array} \end{array} = \begin{array}{c} \begin{bmatrix} R_{11} & R_{1a} & R_{1b} \\ 0 & R_{2a} & R_{2b} \end{bmatrix}. \\ \begin{array}{ccc} n_1 & n_2 & d \end{array} \end{array} \tag{30}$$

Denote

$$R_{Ca}^* = \begin{array}{c} \begin{bmatrix} 0 & 0 \\ 0 & R_{Ca} \end{bmatrix} \begin{array}{l} n_1 \\ n_2 \end{array}. \\ \begin{array}{cc} n_1 & n_2 \end{array} \end{array}$$

Since

$$A^TA = \begin{bmatrix} R_{11} & 0 \\ R_{1a} & R_{2a} \end{bmatrix} \begin{bmatrix} R_{11} & R_{1a} \\ 0 & R_{2a} \end{bmatrix},$$

we can prove analogous relationships between the GSVD of

$$(A, R_{Ca}^*), \quad \left(\begin{bmatrix} R_{11} & R_{1a} \\ 0 & R_{2a} \end{bmatrix}, R_{Ca}^* \right),$$

and (R_{2a}, R_{Ca}) as given in Corollary 1. This and Corollary 1 imply that

$$\sigma' = \sigma'_{n_2} = \min \{ \sigma(R_{2a}, R_{Ca}) \} \quad \text{and} \quad \sigma = \sigma_{n_2+1} \in \sigma(R_{22}, R_C). \tag{31}$$

Since R_{Ca} and R_C are nonsingular, we can use the link with the corresponding ordinary SVD problems (as given in the beginning of this section). Hence, (31) yields

$$\sigma' = \sigma'_{n_2} = \min \{ \sigma(R_{2a} R_{Ca}^{-1}) \}, \quad \sigma = \sigma_{n_2+1} \in \sigma(R_{22} R_C^{-1}), \quad \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = R_C^{-1} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}, \tag{32}$$

where

$$\begin{array}{c}\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}\begin{matrix} n_2 \\ d \end{matrix}\end{array} \qquad \left(\text{respectively, } \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}\begin{matrix} n_2 \\ d \end{matrix}\right)$$

are the d vectors z_i (respectively, right singular vectors v_i) associated with the d smallest (generalized) singular values σ_i, obtained from the GSVD (18) of (R_{22}, R_C) (respectively, SVD of $R_{22}R_C^{-1}$). The assumption $\sigma' > \sigma$ guarantees that the TLS solution $\tilde{X}_2 = -V_1V_2^{-1}$ of the classical TLS problem $R_{22}R_C^{-1}[\begin{smallmatrix} X_2 \\ -I \end{smallmatrix}] \approx 0$, is unique and generic (i.e., V_2 nonsingular) according to the existence and uniqueness theorems [34, Thms. 1-1, and 1-2], [39] of the classical generic TLS solution. Since the GTLS solution $\hat{X}_2$ of the GTLS problem, $R_{22}[\begin{smallmatrix} X_2 \\ -I \end{smallmatrix}] \approx 0$ with corresponding error covariance matrix $C = R_C^TR_C$, is given by $\hat{X}_2 = -Z_1Z_2^{-1}$ (Z_1, Z_2 defined by (32)), $\hat{X}_2$ and $\tilde{X}_2$ are related by

$$R_C^{-1}\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = R_C^{-1}\begin{bmatrix} \tilde{X}_2 \\ -I \end{bmatrix}(-V_2) = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} \hat{X}_2 \\ -I \end{bmatrix}(-Z_2).$$

Since R_C is nonsingular, the existence and uniqueness of the generic TLS solution $\tilde{X}_2$ imply that the generic GTLS solution $\hat{X}_2$ is unique and generic (i.e., Z_2 nonsingular). Since A_1 has full column rank, R_{11} is nonsingular. This implies that $\hat{X}_1 = R_{11}^{-1}R_{12}[\begin{smallmatrix} Z_1 \\ Z_2 \end{smallmatrix}]Z_2^{-1}$ (using (20)–(21)–(22)) exists and is unique. Thus, the GTLS solution

$$\hat{X} = \begin{bmatrix} \hat{X}_1 \\ \hat{X}_2 \end{bmatrix}$$

of the GTLS problem is unique and generic. □

Whenever Γ in (21) is *singular*, the GTLS problem is called *nongeneric*. Using Theorem 3 this happens when $\sigma' \leqq \sigma$, i.e., when *A is (nearly) rank-deficient* ($\sigma' \approx 0$) or when *the set of equations $AX \approx B$* (or at least one subset $AX_i \approx B_i$) *is highly incompatible* ($\sigma' \approx \sigma$) (see also [34, § 1.6], [39], and [40]).

Gallo [9] used a statistical approach to prove under which condition his estimate (which equals our GTLS solution as proven by Theorem 4 below) is generic. These statistical results agree with our algebraic approach.

Now using the previously proven theorems, the correspondence between particular cases of our GTLS solution and alternative expressions of consistent estimators described in the literature can be proved, as done in the following theorem. These equivalences allow us to derive the main statistical properties of the GTLS solution in different statistical situations.

THEOREM 4. *Consider the notation and assumptions* (23). *Let* $\sigma' = \min\{\sigma(A, R_{Ca}^*)\}$ *and assume further that* $\sigma = \sigma_{n+1} = \cdots = \sigma_{n+d} \in \sigma([A; B], R_C^*)$ *has multiplicity d.*

If $\sigma' > \sigma$, the GTLS *solution $\hat{X}$ is given by*

$$\hat{X} = (A^TA - \sigma^2C_a^*)^{-1}(A^TB - \sigma^2C_{ab}^*). \tag{33}$$

Proof. Using the link (15) between the GSVD of $([A; B], R_C^*)$ and the symmetric generalized eigenvalue problem $([A; B]^T[A; B], C^*)$, we obtain

$$([A;B]^T[A;B] - \sigma^2C^*)\begin{bmatrix} Y \\ \Gamma \end{bmatrix} = 0 \tag{34}$$

where $\left[\begin{smallmatrix} Y \\ \Gamma \end{smallmatrix}\right]_d^n$ are d base vectors of the d-dimensional eigensubspace associated with the smallest generalized eigenvalue σ^2 of $([A;B]^T[A;B], C^*)$ of multiplicity d. Partitioning C^* in (34) yields

$$\begin{bmatrix} A^TA-\sigma^2C_a^* & A^TB-\sigma^2C_{ab}^* \\ B^TA-\sigma^2C_{ab}^{*T} & B^TB-\sigma^2C_b^* \end{bmatrix}\begin{bmatrix} Y \\ \Gamma \end{bmatrix}=0. \tag{35}$$

Since σ^2 has multiplicity d, the left-hand-side matrix of (35) has rank n and solutions to (35) will be determined by equations corresponding to any n linearly independent rows of that matrix. Consider the first n equations of (35):

$$(A^TA-\sigma^2C_a^*)Y+(A^TB-\sigma^2C_{ab}^*)\Gamma=0; \tag{36}$$

then, the assumption $\sigma'^2 > \sigma^2$ guarantees that $A^TA - \sigma^2C_a^*$ is invertible and also guarantees that Γ^{-1} exists. Hence, (36) yields

$$-Y\Gamma^{-1}=(A^TA-\sigma^2C_a^*)^{-1}(A^TB-\sigma^2C_{ab}^*). \tag{37}$$

Theorem 1 and the assumptions above imply that the GTLS solution space

$$\text{Range}\left(\begin{bmatrix} \hat{X} \\ -I \end{bmatrix}\right)=\text{Range}\left(\begin{bmatrix} Y \\ \Gamma \end{bmatrix}\right)$$

and thus $\hat{X} = -Y\Gamma^{-1}$ since the GTLS solution is invariant with respect to any base transformation in its solution space (see § 2). Hence, (37) yields (33). □

Theorem 4 allows us to derive the main *statistical properties* of the GTLS solution.

First, assume that *none of the columns of* A are *known exactly* ($n_1 = 0$ and $C^* = C$).

- If $C \sim I_{n+d}$, the expression (33) reduces to

$$\hat{X}=(A^TA-\sigma^2I_n)^{-1}A^TB, \tag{38}$$

which is a well-known expression of the *classical* TLS *solution* as proved in [14] for $d = 1$ and in [34, § 2.2] for $d \geqq 1$. The consistency, distributional, and asymptotic properties of the classical TLS estimate have been proved by Gleser for any $d \geqq 1$ [11]. Assuming that the rows of the error matrix $[\Delta A; \Delta B]$ in (11) are independently and identically distributed (i.i.d.) zero mean vectors with common covariance matrix $C \sim I_{n+d}$ and that $\lim_{m\to\infty} (1/m)A_0^TA_0$ exists and is positive definite, Gleser has proved that the TLS solution $\hat{X}$ is a *strongly consistent* estimate of the true but unknown parameters X of the corresponding unperturbed model $A_0X = B_0$. This result holds, regardless of the common distribution of the errors. When this common error distribution has finite fourth moments, $\hat{X}$ is shown to be *asymptotically normally distributed*. Expressions for the covariance matrix of this distribution are given in [11], as well as large-sample approximate confidence regions.

While Gleser assumes that the elements of A_0 and B_0 in model (11) are fixed (i.e., the functional equations model [23]), Kelly [22] considers the case in which these elements are *random* (called the structural equations model [23]). More specifically, Kelly assumes that the rows of $[A_0; B_0]$ are i.i.d. with common mean vector and common covariance matrix. By calculating the influence function of Gleser's errors-in-variables estimator (which equals the classical TLS solution), Kelly is able to derive an explicit expression for the asymptotic covariance matrix of this estimator in the *structural* equations model.

Finally, Aoki and Yue [2] have studied the statistical properties of the TLS solution (called the solution of the eigenvector method or the Koopmans–Levin method) of Toeplitz-like sets of equations arising in *autoregressive moving average* (ARMA) *modelling* and *system identification*. These models are given by

(39) $$y(t)+a_1 y(t-1)+\cdots+a_{n_a}y(t-n_a)=b_1u(t-1)+\cdots+b_{n_b}u(t-n_b)$$

where the $\{u(t)\}$ and $\{y(t)\}$ are the input and output sequences, respectively, and $\{a_j\}$ and $\{b_j\}$ are the unknown constant parameters of the system. The *observations* at the input and output are assumed to be perturbed by mutually independent *white noise* sequences (i.e., i.i.d. random variables) with *zero mean* and *equal variances*. If sufficient observations are taken, (39) gives rise to an overdetermined set of equations of the form (2) where the corresponding error covariance matrix on the data $\sim I_{n+d}$. Assuming that the given system is stable (i.e., the polynomial $1+\sum_{i=1}^{n_a} a_i z^i$ has all zeros outside the unit circle) and the input sequence $\{u(t)\}$ is uniformly bounded, the TLS solution of this set is *strongly consistent* if and only if the matrix

$$\lim_{m\to\infty}(1/m)\sum_{t=1}^{m} s_t s_t^T,$$

where $s_t=[-y(t+n_a-1),\cdots,-y(t),u(t+n_b-1),\cdots,u(t)]^T$, is positive definite [2] (observe that these conditions are analogous to the consistency conditions imposed by Gleser). This is the case if the input sequence is persistently exciting of order n_a+n_b (i.e., $\lim_{m\to\infty}(1/m)\sum_{t=1}^{m} q_t q_t^T$, where

$$q_t=[u(t+n_a+n_b-1),\cdots,u(t)]^T,$$

is positive definite) and if the polynomials $1+\sum_{i=1}^{n_a} a_i z^i$ and $\sum_{i=1}^{n_b} b_i z^i$ are relatively prime. The first condition is always satisfied if the input is white noise, whereas the second condition means that (39) is a minimal realization of the input-output sequences. Additionally, Aoki and Yue have given explicit expressions for the mean square error of the TLS estimates as a function of the observation noise variances and the number of observations. As demonstrated in [6], the accuracy of these TLS estimates is comparable with that of the joint output method [30] and superior to all other methods described by Söderström [30]. Moreover the TLS method based on the SVD is numerically much more robust and plays an important role in ARMA modelling. See [2], [6], [30] and [34] for more details.

- Now if *the error covariance matrix* C_Δ has the *more general* form c^2C—where C is known and positive definite—Gleser proposed [10], [11] to *transform the original data* $[A;B]$ to new data $[A;B]R_C^{-1}$ (where $C=R_C^TR_C$, R_C upper triangular) such that the error covariance matrix corresponding to the transformed system $\sim I_{n+d}$. Computing the classical TLS solution of this transformed system and converting this result back to the original set, *preserves consistency* of the result. It is easy to prove that this estimate equals our GTLS solution that can be obtained straightforwardly *without pretransforming* the original data. Indeed, in the transformed system the TLS estimate is obtained from the eigenvectors V corresponding to the smallest eigenvalues Λ of the data matrix $[A;B]R_C^{-1}$:

(40) $$R_C^{-1}[A;B]^T[A;B]R_C^{-1}V=V\Lambda.$$

Substituting R_C^{-1} by Z, we obtain the corresponding generalized eigenvector equations:

$$[A;B]^T[A;B]Z = CZ\Lambda. \tag{41}$$

The GTLS solution is computed straightforwardly from the eigenvectors Z, using a GSVD. The transformation formulas given by Gleser [10], [11] just transform the solution obtained from (40) to the GTLS solution obtained from (41). Observe that Gleser needs to compute the inverse of the square root of the covariance matrix that may cause numerical problems especially when R_C is ill-conditioned. This inversion is avoided in our GTLS algorithm.

Even if the true covariance matrix C_Δ is not known but an *estimate* S of it, up to an unknown factor of proportionality, the GTLS solution $\hat{X}$ can still be consistent [7]. An experiment wherein several observations for each row of $[A_0; B_0]$ in (11) are available is an example of this case. Assume that the estimator S is distributed as a multiple of a Wishart matrix with $\eta^{-1}m$ degrees of freedom independently of A and B, where η is a fixed positive number. Further assume that $d = 1$ in (11) and that the rows of the error matrix $[\Delta A; \Delta b]$ are i.i.d. as a multivariate normal random variable with zero mean and covariance matrix C_Δ so that $E(S) \sim C_\Delta$. Under these assumptions, Fuller [7] has proved that $\sqrt{m}(\hat{X} - X)$ converges in distribution to a normal random variable with zero mean and computed an explicit expression for the covariance matrix of this variable.

In ARMA *modelling* and system identification, the situation of known noise covariance functions is treated in [24] and [8]. In [8] Furuta and Paquet discuss the case of *correlated noise* when all the correlation functions are known, up to a factor of proportionality. The suggested procedures, based on solving a generalized eigenvalue problem, are extensions of the eigenvector method described by Levin [26]. It is easy to see that these solutions coincide with our GTLS solution whenever the GTLS problem has a unique minimizer. Hence, the results in [24] and [8] can be applied straightforwardly to the GTLS solution. Conditions for strong consistency of the GTLS estimates in *multi-input multi-output* system models of the form (39) have been derived in [24] and are similar to those for strong consistency of the TLS estimates in the single-input single-output case [2] (see above).

Consider now the case that n_1 *columns* A_1 of A are *known exactly*.

- If $\mathbf{C} \sim \mathbf{I}_{n_2+d}$, (33) reduces to

$$\hat{X} = \left(A^TA - \sigma^2\begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}\right)^{-1} A^TB. \tag{42}$$

Equation (42) equals the expression of the *compensated least squares estimate*, derived by Stoica and Söderström [33]. Assume that the given system (39) is stable and that the polynomials $1 + \sum_{i=1}^{n_a} a_i z^i$ and $\sum_{i=1}^{n_b} b_i z^i$ are relatively prime. If the observations at the *input* are *noise-free*, persistently exciting of order $n_a + n_b$ and independent of the observation noise, and if the observations at the *output* are perturbed by *zero mean white noise of equal variance*, (42) is a *consistent* estimate of the parameters a_i and b_i in the ARMA model (39). Additionally, Stoica and Söderström [33] have proved that this estimate is *asymptotically Gaussian distributed* and an expression of the covariance matrix is explicitly given. As shown before, our GTLS Algorithm is, however, computationally more efficient than the computation procedure presented in [33].

- If the *error covariance matrix* has the *general* form c^2C—where C is known and positive definite—the statistical properties of the GTLS solution for the one-dimensional problem ($d = 1$) can be derived from Gallo [9]. Indeed, Theorem 4 proves the correspondence between our GTLS solution and the consistent estimate derived by Gallo in [9, Thm. 1]. This link allows us to investigate the properties of our GTLS solution as an estimator of the parameters in the general one-dimensional errors-in-variables model, given by (11) for $d = 1$. More specifically, when the joint distribution of the errors possesses finite fourth moment and when

$$\frac{1}{\sqrt{m}} \min\{\lambda(A_0^T A_0)\} \to \infty \quad \text{and} \quad \frac{(\min\{\lambda(A_0^T A_0)\})^2}{\max\{\lambda(A_0^T A_0)\}} \to \infty \quad \text{as } m \to \infty, \tag{43}$$

Gallo has proved that the GTLS *solution* $\hat{X}$ is a *weakly consistent* estimate of the parameters X in model (11). This property holds, regardless of the joint distribution of the errors. Observe that the conditions (43) are less restrictive than those assumed by Gleser (i.e., $\lim_{m\to\infty}(1/m)A_0^T A_0$ exists and is positive definite).

Finally, in ARMA *modelling* and system identification, systems of the form (39) whose *inputs* are observed exactly and whose observed *outputs* are disturbed by zero mean *correlated noise*, have been treated in [20] and [16] for the case that $n_a = n_b$. James, Souter, and Dixon have used the same basic principle of *bias correction* as Stoica and Söderström in [33] and have derived an expression of the form (33). Grosjean and Foulard have *extended* the eigenvector method of Levin [26] to the identification of *multi-input multi-output* systems whose *outputs* are disturbed by *correlated noise*. Assuming that the order $n_a = n_b$ of the system and the correlation functions of the output noise are known, these estimates coincide with the GTLS solution. Conditions for *strong consistency* of the GTLS solution in these cases have been described and are similar to those given by Stoica and Söderström in [33]. For more details, see [24] and [16]. Observe that multi-input multi-output systems in [16] and [24] are treated as s multi-input one output problems where s is the number of outputs, i.e., as s one-dimensional GTLS problems ($d = 1$).

Since Theorem 4 proves the link between our GTLS solution and Gallo's estimate, the following *alternative expressions for our* GTLS *solution* can be deduced straightforwardly from Theorem 1 of [9].

THEOREM 5. *Consider the notation and assumptions* (23). *Let* $\sigma' = \min\{\sigma(A, R_{Ca}^*)\}$ *and assume further that* $\sigma = \sigma_{n+1} = \cdots = \sigma_{n+d} \in \sigma([A; B], R_C^*)$ *has multiplicity* d. *If* $\sigma' > \sigma$, *the* GTLS *solution* $\hat{X}$ *is given by*

$$\hat{X}_2 = (A_2^T P A_2 - \sigma^2 C_a)^{-1}(A_2^T P B - \sigma^2 C_{ab}), \tag{44}$$

$$\hat{X}_1 = (A_1^T A_1)^{-1} A_1^T (B - A_2 \hat{X}_2) \tag{45}$$

with

$$P = I_m - A_1 (A_1^T A_1)^{-1} A_1^T$$

and

$$\hat{X}_2 = (R_{2a}^T R_{2a} - \sigma^2 C_a)^{-1}(R_{2a}^T R_{2b} - \sigma^2 C_{ab}), \tag{46}$$

$$\hat{X}_1 = R_{11}^{-1}(R_{1b} - R_{1a}\hat{X}_2) \tag{47}$$

where the matrices R_{ij} are defined by the QR *factorization of* $[A_1; A_2; B]$:

$$[A_1;A_2;B]=\mathrm{QR}=\begin{array}{cc}[Q_1 & Q_2]\\ n_1 & m-n_1\end{array}\begin{array}{c}\begin{bmatrix} R_{11} & R_{1a} & R_{1b} \\ 0 & R_{2a} & R_{2b}\end{bmatrix}\begin{array}{l} n_1 \\ m-n_1\end{array}\\ \begin{array}{ccc} n_1 & n_2 & d\end{array}\end{array} \tag{48}$$

Proof. Equations (44) and (45) follow straightforwardly from Theorem 1 of [9]. Substituting A_1 in the definition of P by Q_1R_{11} from (48) yields

$$P=I_m-Q_1Q_1^T. \tag{49}$$

Now substitute (49) in (44) and replace A_1, A_2, and B by their equivalents Q_1R_{11}, $Q_1R_{1a}+Q_2R_{2a}$ and $Q_1R_{1b}+Q_2R_{2b}$ obtained from the QR factorization. Making use of the orthonormality of Q, we obtain

$$A_2^TPA_2=A_2^TA_2-R_{1a}^TR_{1a}=R_{2a}^TR_{2a}, \tag{50}$$

$$A_2^TPB=A_2^TB-R_{1a}^TR_{1b}=R_{2a}^TR_{2b}, \tag{51}$$

$$(A_1^TA_1)^{-1}A_1^TB=R_{11}^{-1}Q_1^TB=R_{11}^{-1}R_{1b}, \tag{52}$$

$$(A_1^TA_1)^{-1}A_1^TA_2=R_{11}^{-1}Q_1^TA_2=R_{11}^{-1}R_{1a}, \tag{53}$$

which prove (46) and (47). □

In case that the GTLS *solution* is either *nonunique* or *nongeneric* or when the *subset* A_1 of A in the GTLS problem (5)–(6) is *rank-deficient*, the GTLS solution is *no longer consistent* but is in fact a *biased* estimator of the parameters in (11). Our approach in these situations can be justified as follows. All these cases mentioned above refer to the presence of *multicollinearities*, i.e., there is a (nearly) exact linear relation among the columns of A in the model (11). The consequences are well known; in particular, coefficient estimates obtained by ordinary LS or TLS (without rank reduction) tend to be inflated and can have extremely large variances. One way of handling the multicollinearity problem and *stabilizing* the coefficients is *reducing the rank* of the data matrix A and amounts to filtering out the smallest (generalized) singular values from the estimator. This approach has been adopted in our GTLS Algorithm and is *similar* to the biased estimation techniques: *principal component* regression [18] and *latent root* regression [43], used in linear regression (see also [37] and [40]).

Observe however that in these cases of nonuniqueness or nongenericity the GTLS solution *no longer equals* the solution obtained by applying the *usual transformation procedures* (see, e.g., [11] or [34, § 4.5]). This is evident from (40) and (41). Indeed, the minimum norm or nongeneric TLS solution computed in the transformed system of equations, i.e.,

$$[A;B]R_C^{-1}\begin{bmatrix}\hat{X}\\ -I\end{bmatrix}\approx 0,$$

and converted back to a solution of the original set, does not coincide with the minimum norm or nongeneric GTLS solution of the original set of equations. The properties of the GTLS solution in these cases are not yet fully analyzed.

Summarizing, *consistency results of the* GTLS *solution* have been derived in this section:

- For *any multidimensional* ($d \geqq 1$) *errors-in-variables model* given by (11), in which *none of the columns* of A are *known exactly* ($n_1 = 0$) (based on [11]); and

- For any one-dimensional ($d = 1$) errors-in-variables model given by (11) in which some columns of A are known exactly ($n_1 \geqq 0$) (based on [9]). However the authors strongly believe that the consistency results of Gallo and Gleser can be generalized in order to prove consistency of the GTLS solution for any multidimensional errors-in-variables model given by (11) in which some columns of A are known exactly ($n_1 \geqq 1$).

4. Conclusions. Every *linear parameter estimation* problem arising in signal processing, system identification, automatic control, or in general engineering, statistics, and medicine, gives rise to an *overdetermined* set of linear equations $AX \approx B$ that are usually solved with the ordinary least squares method. Very often, errors occur in *both* A and B. For those cases, the *Total Least Squares* (TLS) technique was devised as a better method of fitting. This method introduced into numerical analysis by Golub and Van Loan, is strongly based on the *Singular Value Decomposition* (SVD). If the errors on the measurements A and B are *uncorrelated with zero mean and equal variance*, TLS is able to compute a *strongly consistent* estimate of the true solution of the corresponding unperturbed set $A_0X = B_0$. In this paper the TLS problem is *generalized* to maintain consistency of the solution in the following cases: first of all, some columns of A may be error-free and second, the errors on the remaining data may be *correlated* and not equally sized provided the covariance matrix of the errors on the rows of the remaining data matrix is known, up to a factor of proportionality. Here, a numerically reliable *Generalized* TLS *algorithm* GTLS, based on the Generalized Singular Value Decomposition (GSVD), is developed. This GSVD avoids transforming the data A, B explicitly and is *numerically more robust* with respect to ill-conditioned covariance matrices. This explains the better numerical performance of the GTLS Algorithm with respect to the explicit transformation procedures. Moreover, by first performing a QR factorization, the GTLS Algorithm only needs to compute the GSVD of a smaller submatrix. This makes the GTLS Algorithm *computationally more efficient* than other methods described in literature.

Additionally, the correspondence between the GTLS solution and alternative expressions of consistent estimators, described in the literature, is proven. From these relations, the main statistical properties of the GTLS solution are deduced. In particular, the equivalence between the GTLS method and the errors-in-variables regression estimator, well known in statistics, is shown. It is concluded that under mild conditions the GTLS solution is a *consistent* estimate of the true parameters of any *general multivariate errors-in-variables model* in which *all or some subset* of variables are observed *with errors*. Furthermore, it is shown that the GTLS algorithm computes the same estimate as the eigenvector method, also called the Koopmans–Levin method, and the Compensated Least Squares (CLS) method. These methods, commonly used in system identification, were developed in order to provide consistent parameter estimates in ARMA *modelling* using noise-corrupted data. If the only disturbances in the observed outputs and the inputs (if they cannot be measured exactly) are given by mutually independent *zero mean white noise sequences of equal variance*, the CLS, TLS, and eigenvector methods all compute strongly consistent estimates. Using the GTLS Algorithm, consistency can be maintained in cases where the disturbances are *not* necessarily white provided the covariance matrix of the correlated noise on the input-output data is known, up to a factor of proportionality.

Acknowledgment. The author would like to thank Hongyuan Zha for his helpful comments.

REFERENCES

[1] R. J. ADCOCK, *A problem in least squares*, The Analyst, 5 (1878), pp. 53–54.

[2] M. AOKI AND P. C. YUE, *On a priori error estimates of some identification methods*. IEEE Trans. Automat. Contr. AC-15 (1970), pp. 541–548.

[3] Z. BAI, *The direct* GSVD *algorithm and its parallel implementation*, Tech. Report CS-TR-1901, Dept. Computer Science, Univ. of Maryland, College Park, MD, August 1987.

[4] R. P. BRENT, F. T. LUK, AND C. F. VAN LOAN, *Computation of the generalized singular value decomposition using mesh-connected processors*, TR-CS-83-563, Dept. Computer Science, Cornell Univ. Ithaca, New York, 1983.

[5] J. J. DONGARRA, J. R. BUNCH, C. B. MOLER, AND G. W. STEWART, LINPACK *User's Guide*, SIAM Society for Industrial and Applied Mathematics, Philadelphia, PA, 1979.

[6] K. V. FERNANDO AND H. NICHOLSON, *Identification of linear systems with input and output noise: The Koopmans-Levin method*, IEE Proc. D, 132 (1985), pp. 30–36.

[7] W. A. FULLER, *Properties of some estimators for the errors-in-variables model*, Ann. Statist., 8 (1980), pp. 407–422.

[8] K. FURUTA AND J. G. PAQUET, *On the identification of time-invariant discrete processes*, IEEE Trans. Automat. Control, AC-15 (1970), pp. 153–155.

[9] P. P. GALLO, *Consistency of regression estimates when some variables are subject to error*, Comm. Statist.–Theory Methods, 11 (1982), pp. 973–983.

[10] L. J. GLESER, *Calculation and simulation in errors-in-variables regression problems*, Mimeograph Series 78-5, Dept. of Statistics, Purdue Univ., West Lafayette, IN, 1978.

[11] ———, *Estimation in a multivariate "errors in variables" regression model: Large sample results*, Ann. Statist., 9 (1981), pp. 24–44.

[12] G. H. GOLUB, personal communication, November 1983.

[13] G. H. GOLUB, A. HOFFMAN, AND G. W. STEWART, *A generalization of the Eckart–Young–Mirsky matrix approximation theorem*, Linear Algebra Appl., 88/89 (1987), pp. 322–327.

[14] G. H. GOLUB AND C. F. VAN LOAN, *An analysis of the total least squares problem*, SIAM J. Numer. Anal., 17 (1980), pp. 883–893.

[15] ———, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 1983.

[16] A. GROSJEAN AND C. FOULARD, *Extensions of the Levin's method (or eigenvector method) for the identification of discrete, linear, multivariable, stochastic, time invariant, dynamic systems*, in Proc. 4th IFAC Symposium on Identification and System Parameter Estimation, Tbilisi, USSR, 1976, pp. 2003–2010.

[17] R. P. GUIDORZI, *Canonical structures in the identification of multivariable systems*, Automatica, 11 (1975), pp. 361–374.

[18] D. M. HAWKINS, *On the investigation of alternative regressions by principal component analysis*, Appl. Statist., 22 (1973), pp. 275–286.

[19] N. J. HIGHAM, *Analysis of the Cholesky decomposition of a semi-definite matrix*, Numerical Analysis Report no. 128, Univ. of Manchester, January 1987; in Reliable, Numerical Computation, M. Cox and S. Hammarling, eds., Oxford University Press.

[20] P. N. JAMES, P. SOUTER, AND D. C. DIXON, *Suboptimal estimation of the parameters of discrete systems in the presence of correlated noise*, Electron. Lett. 8 (1972), pp. 411–412.

[21] L. S. JENNINGS AND M. R. OSBORNE, *A direct error analysis for least squares*, Numer. Math., 22 (1975), pp. 322–332.

[22] G. KELLY, *The influence function in the errors in variables problem*, Ann. Statist., 12 (1984), pp. 87–100.

[23] M. G. KENDALL AND A. STUART, *The Advanced Theory of Statistics*, Fourth edition, Griffin, London, 1979.

[24] Ü. KOTTA, *Structure and parameter estimation of multivariable systems using the eigenvector method*, in Proc. 5th IFAC Symposium on Identification and System Parameter Estimation, Darmstadt, FRG, 1979, pp. 453–458.

[25] J. LEURIDAN, D. DE VIS, H. VAN DER AUWERAER, AND F. LEMBREGTS, *A comparison of some frequency response function measurement techniques*, in Proc. 4th Internat. Modal Analysis Conference, Los Angeles, CA, February 3-6, 1986, pp. 908–918.

[26] M. J. LEVIN, *Estimation of a system pulse transfer function in the presence of noise*, IEEE Trans. Automat. Control, AC-9 (1964), pp. 229–235.

[27] C. C. PAIGE, *Computing the generalized singular value decomposition*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 1126–1146.

[28] C. C. PAIGE AND M. A. SAUNDERS, *Towards a generalized singular value decomposition*, SIAM J. Numer. Anal., 18 (1981), pp. 398–405.

[29] K. PEARSON, *On lines and planes of closest fit to points in space*, Philos. Mag., 2 (1901), pp. 559–572.

[30] T. SÖDERSTRÖM, *Identification of stochastic linear systems in presence of input noise*, Automatica, 17 (1981), pp. 713–725.

[31] P. SPRENT, *Models in Regression and Related Topics*, Methuen, London, 1969.

[32] G. W. STEWART, *A method for computing the generalized singular value decomposition*, in Matrix Pencils, B. Kagstrom and A. Ruhe, eds., Springer-Verlag, Berlin, New York, 1983, pp. 207–220.

[33] P. STOICA AND T. SÖDERSTRÖM, *Bias correction in least squares identification*, Internat. J. Control, 35 (1982), pp. 449–457.

[34] S. VAN HUFFEL, *Analysis of the total least squares problem and its use in parameter estimation*, Ph.D. dissertation, Dept. of Electrical Engineering, Katholieke Universiteit Leuven, Belgium, June, 1987.

[35] ———, *The generalized total least squares problem: Formulation, algorithm and properties*, in Proc. NATO Advanced Study Institute on Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms, Leuven, Belgium, August 1–12, 1988.

[36] S. VAN HUFFEL, J. VANDEWALLE, AND A. HAEGEMANS, *An efficient and reliable algorithm for computing the singular subspace of a matrix, associated with its smallest singular values*, J. Comput. Appl. Math., 19 (1987), pp. 313–330.

[37] S. VAN HUFFEL AND J. VANDEWALLE, *Algebraic relationships between classical regression and total least-squares estimation*, Linear Algebra Appl., 93 (1987), pp. 149–162.

[38] ———, *The partial total least squares algorithm*, J. Comput. and Appl. Math., 21 (1988), pp. 333–341.

[39] ———, *Analysis and solution of the nongeneric total least squares problem*, SIAM J. Matrix Anal. Appl., 9 (1988), pp. 360–372.

[40] ———, *Comments on the solution of the nongeneric total least squares problem*, Internal Report ESAT-KUL-88/3, ESAT Lab., Dept. of Electrical Engineering, Katholieke Universiteit, Leuven, February 1988.

[41] C. F. VAN LOAN, *Generalizing the singular value decomposition*, SIAM J. Numer. Anal., 13 (1976), pp. 76–83.

[42] ———, *Computing the CS and generalized singular value decomposition*, Tech. Report CS-604, Department of Computer Science, Cornell University, Ithaca, NY, 1984.

[43] J. T. WEBSTER, R. F. GUNST, AND R. L. MASON, *Latent root regression analysis*, Technometrics, 16 (1974), pp. 513–522.