UNIFYING LEAST SQUARES, TOTAL LEAST SQUARES AND DATA LEAST SQUARES

Christopher C. Paige
School of Computer Science, McGill University,
Montreal, Quebec, Canada, H3A 2A7
paige@cs.mcgill.ca

Zdeněk Strakoš
Institute of Computer Science, Academy of Sciences of the Czech Republic,
Pod Vodárenskou věží 2, 182 07 Praha 8, Czech Republic
strakos@cs.cas.cz

Abstract The standard approaches to solving overdetermined linear systems $Ax \approx b$ construct minimal corrections to the vector b and/or the matrix A such that the corrected system is compatible. In ordinary least squares (LS) the correction is restricted to b, while in data least squares (DLS) it is restricted to A. In scaled total least squares (Scaled TLS) [15], corrections to both b and A are allowed, and their relative sizes depend on a parameter γ. Scaled TLS becomes total least squares (TLS) when $\gamma = 1$, and in the limit corresponds to LS when $\gamma \to 0$, and DLS when $\gamma \to \infty$.

In [13] we presented a particularly useful formulation of the Scaled TLS problem, as well as a new assumption that guarantees the existence and uniqueness of meaningful Scaled TLS solutions for all parameters $\gamma > 0$, making the whole Scaled TLS theory consistent. This paper refers to results in [13] and is mainly historical, but it also gives some simpler derivations and some new theory. Here it is shown how any linear system $Ax \approx b$ can be reduced to a minimally dimensioned core system satisfying our assumption. The basics of practical algorithms for both the Scaled TLS and DLS problems are indicated for either dense or large sparse systems.

Keywords: scaled total least squares, ordinary least squares, data least squares, core problem, orthogonal reduction, singular value decomposition.

Introduction

Two useful approaches to solving the overdetermined linear system

$$Ax \approx b, \quad A \text{ an } n \text{ by } k \text{ matrix, } b \text{ an } n\text{-vector}, \quad b \notin \mathcal{R}(A), \tag{1}$$

are ordinary least squares (LS, or OLS, see for example [1], [8, §5.3]) and total least squares (TLS, see [6; 7], also [1, §4.6], [8, §12.3], [11]). In LS we seek (we use $\|\cdot\|$ to denote the vector 2-norm)

$$\text{LS distance} \quad \equiv \quad \min_{r,x} \|r\| \quad \text{subject to} \quad Ax = b - r. \tag{2}$$

In TLS, G and r are sought to minimize the Frobenius (F) norm in

$$\text{TLS distance} \quad \equiv \quad \min_{r,G,x} \|[r, G]\|_F \quad \text{s. t.} \quad (A + G)x = b - r. \tag{3}$$

The opposite case to LS is the data least squares problem (DLS), see [9]. In DLS the correction is allowed only in A

$$\text{DLS distance} \quad \equiv \quad \min_{G,x} \|G\|_F \quad \text{subject to} \quad (A + G)x = b. \tag{4}$$

All these approaches can be unified by considering the following very general scaled TLS problem (Scaled TLS), see the paper [15] by B. D. Rao, who called it "weighted TLS": for a given $\gamma > 0$,

$$\text{Scaled TLS distance} \quad \equiv \quad \min_{\tilde{r},\tilde{G},\tilde{x}} \|[\tilde{r}\gamma, \tilde{G}]\|_F \quad \text{s. t.} \quad (A+\tilde{G})\tilde{x} = b-\tilde{r}. \tag{5}$$

Here the relative sizes of the corrections in A and b are determined by the real parameter $\gamma > 0$. As $\gamma \to 0$ the Scaled TLS solution approaches the LS solution, when $\gamma = 1$ (5) coincides with the TLS formulation, and as $\gamma \to \infty$ it approaches DLS. The case $\gamma \to 0$ is not completely obvious, since setting $\gamma = 0$ in (5) leads to $\tilde{G} = 0$ but allows *arbitrary* $\tilde{r}$. However consideration of very small γ should at least partially convince the reader that the LS solution is obtained. The case $\gamma = 1$ is obvious, and we see that $\gamma \to \infty$ requires $\tilde{r} \to 0$, leading to DLS. For more on Scaled TLS and DLS see also [2]. Scaling by a diagonal matrix was considered in [7], and this motivated later researchers, leading eventually to the Scaled TLS formulation in [15]. The paper [4] considered the case where only some of the columns of the data matrix are contaminated, and this also suggested a way of treating LS as well as TLS in the one formulation.

The formulation of the Scaled TLS problem that we use is slightly different from that in (5). For any positive bounded γ, substitute in (5) $r \equiv \tilde{r}\gamma$, $x \equiv \tilde{x}$ and $G \equiv \tilde{G}$ to obtain the new formulation of the Scaled TLS problem:

$$\text{Scaled TLS distance} \quad \equiv \quad \min_{r,G,x} \|[r, G]\|_F \quad \text{s. t.} \quad (A+G)x\gamma = b\gamma-r. \tag{6}$$

We call the $x = x(\gamma)$ that minimizes this distance the *Scaled TLS solution* of (6). In analogy with (3), we call $x(\gamma)\gamma$ the *TLS solution* of (6). In (6) we could have written x instead of $x\gamma$. We chose the present form so that for positive bounded γ, the Scaled TLS solution $x = x(\gamma)$ of (6) is identical to the solution $\tilde{x}$ of (5). Thus (5) and (6) have identical distances and solutions for positive bounded γ. Therefore our results and discussions based on (6) apply fully to the Scaled TLS problem (5).

In [13, §6] we showed for (6) in the general case of complex data that as $\gamma \to 0$, $x(\gamma)$ becomes the LS solution x of (2), (Scaled TLS distance)/γ becomes the LS distance. As $\gamma \to \infty$, $x(\gamma)$ becomes the DLS solution x of (4), and the Scaled TLS distance becomes the DLS distance. The convergence of the Scaled TLS problem to the LS problem has been described in [15], and essentially in [7], for the real case.

We found that the development of our results was more simple and intuitive using the formulation (6) rather than (5). In particular, all the known TLS theory and algorithms can be applied directly to (6). The equivalence of (6) and (5) is extremely useful. This equivalence was pointed out to us by Sabine Van Huffel [10] after she read an earlier version of our work based on (6). We have not seen it stated in the literature, but it is implicit in the paper by Rao [15].

In (6), γ simply scales the right-hand side vector b (and the Scaled TLS solution $x = x(\gamma)$). Thus it is appropriate to call the formulation (6) the *Scaled* TLS problem, rather than the "weighted" TLS problem as was done in [15]. This also avoids the possibility of confusing the meaning of "weighted" here with its different meaning in "weighted least squares".

Using γ can have a statistical significance. Suppose that the elements of A are known to have independent zero-mean random errors of equal standard deviation δ_A. Suppose also that the elements of b have been observed with independent zero-mean random errors of equal standard deviation δ_b, and that the errors in b and A are independent. Then taking $\gamma = \delta_A/\delta_b$ in (6) will ensure that all the errors in that model have equal standard deviation (and so variance), and (6) is the ideal formulation for providing estimates. This agrees with the limiting behaviour described above, for clearly if $\delta_A = 0$ and $\delta_b \neq 0$, then LS is the correct choice, while if $\delta_A \neq 0$ and $\delta_b = 0$, then DLS is the correct choice. However (6) can also be useful outside any statistical context, and then γ does not have the above interpretation, see for example [14] which is summarized in our other contribution in this book.

In all these formulations, if $b \in \mathcal{R}(A)$, then zero distance can be obtained via a direct solution. Otherwise TLS, and so Scaled TLS solutions can be found via the singular value decomposition (SVD). Let $\sigma_{min}(\cdot)$ denote the smallest singular value of a given matrix. To be precise,

$\sigma_{min}(M)$ will denote the j-th largest singular value of an n by j matrix M, and will be zero if $n < j$. The interlacing property for the eigenvalues of $[A, b]^T[A, b]$ and of A^TA [16, Ch2, §47, pp. 103–4] tells us that $\sigma_{min}([A, b]) \leq \sigma_{min}(A)$. When

$$\sigma_{min}([A, b]) < \sigma_{min}(A) \tag{7}$$

the n by k matrix A must have rank k, the unique solution of the TLS problem (3) is obtained from scaling the right singular vector of $[A, b]$ corresponding to $\sigma_{min}([A, b])$, and the norm of the TLS correction satisfies $\min_{r,G,x} \|[r, G]\|_F = \sigma_{min}([A, b])$, (see for example [8, §12.3]).

However when $\sigma_{min}([A, b]) = \sigma_{min}(A)$, the theory and solution methods became complicated, see for example the discussions on nongeneric problems in [11]. For this and other reasons we argued in [13] that (7) should not be used as a basis for the TLS theory.

A similar argument to that following (7) shows that when

$$\sigma_{min}([A, b\gamma]) < \sigma_{min}(A) \quad \text{for a given } \gamma > 0, \tag{8}$$

the Scaled TLS distance in (6) is $\sigma_{min}([A, b\gamma])$, but we also showed in [13] that this should not be used as a basis for the Scaled TLS theory.

In the general case, let $\mathcal{U}_{min}$ be the left singular vector subspace of A corresponding to $\sigma_{min}(A)$. We showed in [13] that a satisfactory condition for building the theory of the TLS, DLS and Scaled TLS formulations for solving (1) is the γ-independent criterion:

$$\text{the } n \times k \text{ matrix } A \text{ has rank } k, \text{ and } \; b \not\perp \mathcal{U}_{min}. \tag{9}$$

We showed in [13, Thm.3.1] that this implies

$$\sigma(\gamma) \equiv \sigma_{min}([A, b\gamma]) < \sigma_{min}(A) \quad \text{for all } \gamma \geq 0, \tag{10}$$

which of course implies (7) and (8). The stronger condition (9) is the simplest one. It can be checked using direct computations, see Sectinon 2, while the others each apparently require two SVDs.

A crucial property of the criterion (9) is that *any* linear system $Ax \approx b$ can in theory be reduced to a "core" problem satisfying (9). We show how this can be done by direct computations that can be usefully applied to all Scaled TLS and DLS problems.

This paper is necessarily short, and can be considered as an introduction to [13] which contains the full theory that has been omitted here. That paper presented a new and thorough analysis of the theoretical foundations of the Scaled TLS problem, and of its relationships to the LS and DLS problems. Here we mention some of those results, but concentrate mainly on the concept of the core problem.

The rest of the paper is organized as follows. Section 1 indicates why the formulations (3)–(6) are incomplete without the criterion (9). Section 2 shows how to handle the completely general Scaled TLS problem by reducing it to a core problem that satisfies an even stronger criterion than (9) — one which ensures the core problem is irreducible in the sense of containing no information that is redundant or irrelevant to the solution. Section 3 discusses how Scaled TLS problems can be solved computationally, and describes a simple solution to the DLS problem. Section 4 summarizes the advances and outlines the philosophy.

1. Conditions for meaningful solutions

The problem formulations (3)–(6) are not good for solving $Ax \approx b$ in certain cases. It was shown in [13, §7] that (3)–(6) are not good when n by k A does not have rank k. The formulations should at least demand the solution vectors be orthogonal to the null space. It is preferable to eliminate the null space, so now let us assume A has rank k.

We argue that (3)–(6) are best restricted to problems of the form (1) satisfying (9). Suppose the data can be transformed so that

$$\left[\begin{array}{c|c} \tilde{b} & \tilde{A} \end{array}\right] = P^T \left[\begin{array}{c|c} b & AQ \end{array}\right] = \left[\begin{array}{c|c|c} b_1 & A_{11} & 0 \\ \hline 0 & 0 & A_{22} \end{array}\right], \tag{11}$$

where P and Q are orthogonal. The approximation problem $Ax \approx b$ then represents two *independent* approximation problems:

$$A_{11}x_1 \approx b_1, \qquad A_{22}x_2 \approx 0, \qquad x \equiv Q \left[\begin{array}{c} x_1 \\ x_2 \end{array}\right], \tag{12}$$

in that the solution to each of these has no effect upon, and can be found independently of, the other. In this case the non-core problem $A_{22}x_2 \approx 0$ has the solution $x_2 = 0$, and only $A_{11}x_1 \approx b_1$ need be solved.

If $b \perp \mathcal{U}_{min}$, see (9), then orthogonal P and Q clearly exist giving (11) where A_{22} contains all the singular values of A equal to $\sigma_{min}(A)$. Then it was shown in [13, §7] that (3)–(6) applied directly to the combined problem $Ax \approx b$ can give meaningless solutions. But even when they give meaningful solutions these minimum singular values are irrelevant, and should be removed from the problem, since rounding errors effectively introduce a nonzero vector below b_1 in (11), and so cause these irrelevant singular values to contaminate the solution. Although (2) in theory gives $x_2 = 0$, this last comment suggests we might gain by insisting on (9) for LS too.

The criterion (9) leads to a clear and consistent theory, and ensures that the minimum singular value of A is relevant to the solution. Fortu-

nately there is an elegant transformation which produces the minimally dimensioned core problem obeying (9) from any given $[b, A]$.

2. The core problem within $Ax \approx b$

Here we answer the following question. Given any n by k matrix A and n-vector b, how can the problem $Ax \approx b$ be split into two independent problems as in (11) and (12), giving a trivial problem $A_{22}x_2 \approx 0$ of maximal dimensions, and the minimally dimensioned core problem $A_{11}x_1 \approx b_1$ satisfying (9)? We can set $x_2 = 0$, and then each of the formulations (2)–(6) gives a unique meaningful solution to $A_{11}x_1 \approx b_1$.

We answer this by choosing orthogonal matrices P and Q to produce the following real bidiagonal matrix, see for example [8, §5.4.3–5, pp. 251–254]. In the usual case of $n \times k$ A with $n > k$ we obtain, (where the bottom 0 either represents a zero vector or is nonexistent):

$$[\tilde{b}, \tilde{A}] \equiv P^T \left[\begin{array}{c|c} b & A \end{array}\right] \begin{bmatrix} 1 & \\ & Q \end{bmatrix} = \left[\begin{array}{c|cccc} \beta_1 & \alpha_1 & & & \\ & \beta_2 & \alpha_2 & & \\ & & \cdot & \cdot & \\ & & & \beta_k & \alpha_k \\ & & & & \beta_{k+1} \\ & & & & 0 \end{array}\right]. \quad (13)$$

Remember the Scaled TLS solution requires some knowledge of the SVD of $[b\gamma, A]$. This SVD can quickly be computed from this bidiagonal form for any choice of γ, see for example [8, §8.6.2, pp. 452–456].

There are two ways this algorithm can terminate prematurely: with a $\beta_j = 0$, giving (11) and (12) with $A_{11}x_1 = b_1$ being a compatible system; or with $\alpha_j = 0$, giving an incompatible system $A_{11}x_1 \approx b_1$ in (12).

The computations described in [8, §5.4.3–5, pp. 251–254] are designed for dense matrices. If we have large sparse $[b, A]$, then we could consider the iterative bidiagonalization of Golub and Kahan in [5]. This iterative bidiagonalization is the basis for the LSQR algorithm in [12] which solves large sparse LS (as well as consistent) problems. In theory after j and a half steps this produces the first $j + 1$ columns of P, the first j columns of Q, and the leading $j + 1$ by $j + 1$ block of the right-hand side in (13). Björck [1, §7.6.5, pp.310-311] suggested applying this iterative bidiagonalization to the TLS problem, see also [3, §4.1]. Now we see this approach is also applicable to solving the Scaled TLS problem, as well as (at least in theory) delivering the core problem, for any large sparse linear system $Ax \approx b$. The adaptation of LSQR for solving large sparse Scaled TLS or DLS problems using finite precision computations will be further investigated. See Section 3 for the DLS solution using (13).

The main theoretical importance of the reduction (13) is that if $\beta_j\alpha_j \neq 0$, $j = 1, \ldots, k$, then our criterion (9) holds for the reduced bidiagonal matrix. We now prove this in an extended version of [13, Thm.8.1].

Theorem 1 *Suppose n by k A has SVD $A = \sum_{i=1}^{k} u_i\sigma_i v_i^T$, and there exist orthogonal matrices P and Q giving (13) with*

$$\beta_j\alpha_j \neq 0, \quad j = 1, \ldots, k. \tag{14}$$

Then we have a stronger condition than (9) for this b and A:

$$\mathit{rank}(A) = k; \qquad b^T u_i \neq 0, \quad i = 1, \ldots, k, \tag{15}$$

and no nontrivial split of the form (11) can be obtained with orthogonal P and Q, showing $Ax \approx b$ is the minimally dimensioned core problem. Also the k singular values of A are distinct and nonzero; the $k+1$ singular values of $[b, A]$ are distinct, and all nonzero if and only if $\beta_{k+1} \neq 0$.

Proof $\tilde{A}$ and A have the same singular values, as do $[\tilde{b}, \tilde{A}]$ and $[b, A]$, and $\tilde{A} = P^T AQ$ has the SVD $\tilde{A} = \sum_{i=1}^{k} \tilde{u}_i\sigma_i\tilde{v}_i^T \equiv \sum_{i=1}^{k} P^T u_i\sigma_i v_i^T Q$, so

$$b^T u_i = b^T PP^T u_i = \tilde{b}^T\tilde{u}_i, \quad i = 1, \ldots, k.$$

Write $\tilde{A} \equiv [b_1, A_1]$, then $\tilde{A}^T\tilde{A}$ is $k \times k$ tridiagonal with nonzero next to diagonal elements, and $A_1^T A_1$ remains when the first row and column are deleted. Thus the eigenvalues of $A_1^T A_1$ strictly separate those of $\tilde{A}^T\tilde{A}$, see [16, Ch.5, §37, p.300], and the singular values of A_1 strictly separate those of $\tilde{A}$. Thus $\tilde{A}$, and so A, has distinct singular values. A similar argument holds for $[b, A]$. A clearly has rank k, and $[b, A]$ has rank $k+1$ if and only if $\beta_{k+1} \neq 0$. Suppose σ is a singular value of $\tilde{A}$ with singular vectors u and v such that

$$\tilde{b}^T u = \beta_1 e_1^T u = 0, \quad u\sigma = \tilde{A}v, \quad \sigma v^T = u^T\tilde{A}, \quad \|u\| = \|v\| = 1,$$

then $0 = e_1^T u\sigma = e_1^T\tilde{A}v = \alpha_1 e_1^T v$, and $e_1^T v = 0$. Writing $v = \begin{pmatrix} 0 \\ q \end{pmatrix}$ shows

$$\tilde{A}v = A_1 q = u\sigma, \quad u^T A_1 = \sigma q^T, \quad \|u\| = \|q\| = 1,$$

so σ is also a singular value of A_1. This is a contradiction since the singular values of A_1 strictly separate those of $\tilde{A}$, so (15) holds.

Finally if (11) could exist with nontrivial A_{22}, then b would be orthogonal to a left singular vector subspace of A, which (15) has proven is impossible. ■

Thus we need not derive results for the most general possible $[b\gamma, A]$. We can instead assume (9). Any more general $Ax \approx b$ problem can be

reduced to a core problem that satisfies (15) (and so (9)) by applying the reduction (13) and stopping at the first zero β_j or α_j. Suppose the resulting core data is $[b_1, A_{11}]$, see (11). Then the theorem also showed that A_{11} has no multiple singular values, so any singular value repeats must appear in A_{22}.

In theory we need not insist on (15), because a problem only satisfying (9) will in theory give the same solution and distance as it would if it were reduced to one satisfying (15). But in practice it is preferable to carry out the reduction (13) leading to (15), see Section 3.

3. Computing Scaled TLS and DLS solutions

In order to compute either Scaled TLS solutions or the DLS solution for given data $[b, A]$, we recommend first carrying out a reduction of the form (13) to the core problem in Section 2 — unless there are clear reasons for not doing so. The reasons for doing so are hard to reject. For general data we will not know if the formulations (2)–(6) have unique meaningful solutions, but the reduction will give us a subproblem for which this is so. Even if we know the original data satisfies (9), it is (from the computational point of view) highly preferable to remove all the irrelevant information from our data as early in the solution process as possible, and this is exactly what the transformation (13) does. In any case we still need some sort of SVD of the data, and this will usually first perform a reduction as costly as that in (13). But (13) allows us to find the SVD of $[b\gamma, A]$ easily for different choices of γ and so is the obvious choice. There are excellent fast and accurate algorithms for finding all or part of the SVD of (13) with β_1 replaced by $\beta_1\gamma$. We can find just the smallest singular value and its singular vectors, from which the solution vector $x(\gamma)$ can be simply attained. If we have some idea of the accuracy of our data, then when we use numerically reliable orthogonal transformations in (13), we will have a good idea of what element of (13) (if any) we can set to zero to stop the computation as soon as possible. Thus the crucial decisions can be made *before* any SVD computations are carried out. This is more efficient, but it is almost certainly more reliable to make such decisions from (independent) orthogonal transformations of the original data than from the elements of singular vectors, (see for example [11, p.23]). The remaining computations for Scaled TLS are fairly obvious. Finally (13) leads to a solution to the DLS problem (4), which we now describe. The theory here is simpler than that in [13, §9].

Suppose that the core part $[\tilde{b}, \tilde{A}]$ of the transformed $[b, A]$ has the form in (13) with (14). We will solve the DLS problem for this reduced,

or core data. Now Theorem 1 proved (9) holds. If $\beta_{k+1} = 0$ the DLS distance is zero, and the solution is obvious. Otherwise, writing

$$[\tilde{b}|\tilde{A}] \equiv \left[\begin{array}{c|c} \beta_1 & \alpha_1 e_1^T \\ \hline 0 & A_2 \end{array}\right] \equiv P^T[b|AQ], \quad \tilde{G} \equiv \left[\begin{array}{c} g^T \\ \hline G_2 \end{array}\right] \equiv P^T G Q, \quad \tilde{x} \equiv Q^T x,$$

we see for this reduced data that the DLS problem (4) becomes

$$\min_{g,G_2,\tilde{x}} \{\|g\|^2 + \|G_2\|_F^2\} \quad \text{s. t.} \quad \left[\begin{array}{c|c} \beta_1 & \alpha_1 e_1^T + g^T \\ 0 & A_2 + G_2 \end{array}\right] \left[\begin{array}{c} -1 \\ \tilde{x} \end{array}\right] = 0.$$

Since β_1 is nonzero, $\tilde{x} \neq 0$, and the minimum $\|G_2\|_F$ in $(A_2 + G_2)\tilde{x} = 0$ is $\sigma_{min}(A_2)$, with $\tilde{x}$ proportional to the right singular vector v of A_2 corresponding to $\sigma_{min}(A_2)$. But then $e_1^T v \neq 0$ (otherwise $\sigma_{min}(A_2)$ would also be a singular value of $\tilde{A}$) and we can take $g = 0$ so that

$$\tilde{x}_D = v\beta_1/(\alpha_1 e_1^T v), \qquad \sigma_D = \sigma_{min}(A_2), \tag{16}$$

are the DLS solution and distance in (4) for the reduced data $[\tilde{b}, \tilde{A}]$. The smallest singular value and its right singular vector of the nonsingular bidiagonal matrix A_2 are relatively easy to find, see for example [8, §8.6.2, pp. 452–456].

4. Summary and Conclusions

The philosophy behind our approach is radically different from that of previous TLS, Scaled TLS or DLS work known to us. The Scaled TLS formulation (6) makes it easy to analyze and solve the Scaled TLS problem (it shows the Scaled TLS problem is just the TLS problem with its right-hand side b scaled by γ, so all the TLS artillery is available). But more importantly than that, the approach of reducing a problem $Ax \approx b$ to its "core" problem (Section 2) and solving that core problem simplifies our understanding of the area. It also simplifies the development of algorithms, while unifying the theoretical problems in the area. Crucial to all this is the (γ-independent) criterion (9) for Scaled TLS (also TLS, DLS and even possibly LS) problems, that was introduced in [13]. The key is that *any* Scaled TLS (or LS or TLS or DLS) problem can in theory be transformed by *direct* orthogonal transformations into two independent problems: a (possibly nonexistent) trivial problem, and a core problem, where the core problem automatically satisfies (9). Solving the core problem then solves the original problem. Thus no complicated conditions such as (7) or (9) need be tested, and no special cases need be treated. All the decisions can be made by examining the sizes of elements in the orthogonally transformed data. Both theory and computations can thus be unified, simplified and clarified.

Acknowledgments

This work was supported by NSERC of Canada Grant OGP0009236 and by the GA AS CR under grant A2030801. Part of this work was performed while Zdenek Strakos was visiting Emory University, Atlanta, GA, U.S.A.

References

[1] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM Publications, Philadelphia PA, 1996.

[2] G. Cirrincione, *A Neural Approach to the Structure of Motion Problem*, PhD thesis, LIS INPG Grenoble, 1998.

[3] R. D. Fierro, G. H. Golub, P. C. Hansen and D. P. O'Leary, *Regularization by truncated total least squares*, SIAM J. Sci. Comput., 18:1223–1241, 1997.

[4] G. H. Golub, A. Hoffman and G. W. Stewart, *A generalization of the Eckart-Young-Mirsky matrix approximation theorem*, Linear Algebra Appl., 88/89:317–327, 1987.

[5] G. H. Golub and W. Kahan, *Calculating the singular values and pseudo-inverse of a matrix*, J. SIAM, Series B, Numer. Anal., 2:205–224, 1965.

[6] G. H. Golub and C. Reinsch, *Singular value decomposition and least squares solutions*, Numerische Mathematik, 14:403-420, 1970. Also in "Handbook for Automatic Computation Vol. 2: Linear Algebra", by J. H. Wilkinson and C. Reinsch, (eds.), pp. 134-151, Springer, New York, 1971.

[7] G. H. Golub and C. F. Van Loan, *An analysis of the total least squares problem*, SIAM J. Numer. Anal., 17:883–893, 1980.

[8] ———, *Matrix Computations*, The Johns Hopkins University Press, Baltimore MD, third ed. 1996.

[9] R. D. D. Groat and E. M. Dowling, *The data least squares problem and channel equalization*, IEEE Trans. Signal Processing, 42(1):407–411, 1993.

[10] S. Van Huffel. Personal communication, June 1999.

[11] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM Publications, Philadelphia PA, 1991.

[12] C. C. Paige and M. A. Saunders, *LSQR: An algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Software, 8:43–71, 1982.

[13] C. C. Paige and Z. Strakoš, *Scaled total least squares fundamentals*, to appear in Numerische Mathematik.

[14] C. C. Paige and Z. Strakoš, *Bounds for the least squares distance using scaled total least squares*, to appear in Numerische Mathematik.

[15] B. D. Rao, *Unified treatment of LS, TLS and truncated SVD methods using a weighted TLS framework*, In: S. Van Huffel (editor), *Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modelling*, pp. 11–20, SIAM Publications, Philadelphia PA, 1997.

[16] J. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford England, 1965.