An algorithm for solving rank-deficient scaled total least square problems

Wei Xu[1], Sanzheng Qiao[2] and Yimin Wei[3] *
[1,2] Department of Computing and Software, McMaster University
Hamilton, Ont, L8S 4K1, Canada.
[1] xuw5@mcmaster.ca
[2] qiao@mcmaster.ca
[3] Department of Mathematics, Fudan University,
Shanghai, 200433, P.R. China. [3] ymwei@fudan.edu.cn

Abstract

The scaled total least square (STLS) problem, introduced by B.D. Rao in 1997, unifies both the total least square (TLS) and the least square (LS) problems. The STLS problems can be solved by the singular value decomposition (SVD). In this paper, we give a rank-revealing two-sided orthogonal decomposition method for solving the STLS problem. An error analysis is presented. Our numerical experiments show that this algorithm computes the STLS solution as good as the SVD method with less computation.

1 Introduction

Rao [8] unified the least squares (LS) and the total least squares (TLS) problems and introduced the scaled total least square (STLS) problem: Given $A \in \mathrm{R}^{m\times n}$ $(m > n)$, $b \in \mathrm{R}^m$, and a scalar $\lambda > 0$, find $E \in \mathrm{R}^{m\times n}$ and $r \in \mathrm{R}^m$ such that

$$\min_{(b-r)\in \mathrm{range}(A+E)} \|[E, \lambda r]\|_F.$$

Paige and Strakoš [7] suggested a slightly different but equivalent formulation:

$$\min_{(\lambda b-r)\in \mathrm{range}(A+E)} \|[E, r]\|_F. \tag{1}$$

If $[E_{\mathrm{STLS}}, r_{\mathrm{STLS}}]$ solves the above problem (1), then the solution x_{STLS} for x in $(A + E_{\mathrm{STLS}})\lambda x = \lambda b - r_{\mathrm{STLS}}$ is called the scaled total least square solution.

*The first and second authors are partially supported by the Natural Sciences and Engineering Research Council of Canada. The third author is supported by the National Natural Science Foundation of China and Shanghai Education Committee.

Obviously, when $\lambda = 1$, the STLS (1) reduces to the total least square (TLS) problem. It is also shown in [11] that $x_{\rm STLS}$ approaches to $x_{\rm LS}$, the solution of the least square (LS) problem $\min_x \|Ax - b\|_2$, as $\lambda \to 0$. In the STLS literatures [6, 7, 8], A is assumed to be of full rank. In this paper, we consider the case when A is rank-deficient.

The conditions for the existence of the STLS solution and explicit expressions of the STLS solution are given in [11]. To state the results in [11], we denote the SVD of

$$C := [A \ \lambda b] = U\Sigma V^T, \tag{2}$$

where $U \in \mathrm{R}^{m\times(n+1)}$ has orthonormal columns, $V \in \mathrm{R}^{(n+1)\times(n+1)}$ is orthogonal and $\Sigma = \mathrm{diag}(\sigma_1(C), \cdots, \sigma_{n+1}(C))$, $\sigma_1(C) \geq \sigma_2(C) \geq \cdots \geq \sigma_{k+1}(C) > \sigma_{k+2}(C) = \cdots = \sigma_{n+1}(C) = 0$, $\sigma_k(C)$ the k-th singular value of C, and $k = \mathrm{rank}(A)$. Then we partition U, Σ and V in (2):

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}, \quad U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}, \quad V = \begin{bmatrix} V_{11} & V_{12} \\ v_{21}^T & v_{22}^T \end{bmatrix}, \tag{3}$$

such that $\Sigma_1 = \mathrm{diag}(\sigma_1(C), \cdots, \sigma_k(C))$, $\Sigma_2 = \mathrm{diag}(\sigma_{k+1}(C), 0, \cdots, 0)$, U_1 and U_2 are respectively the first k columns and last $n+1-k$ columns of U, $V_{11} \in \mathrm{R}^{n\times k}$, $V_{12} \in \mathrm{R}^{n\times(n+1-k)}$, $v_{21} \in \mathrm{R}^k$, and $v_{22} \in \mathrm{R}^{n+1-k}$.

Accordingly, we denote the SVD of A as

$$A = U_A \begin{bmatrix} \Sigma_A & 0 \\ 0 & 0 \end{bmatrix} V_A^T, \quad U_A = \begin{bmatrix} U_{A1} & U_{A2} \end{bmatrix}$$

where $U_A \in \mathrm{R}^{m\times m}$ and $V_A \in \mathrm{R}^{n\times n}$ are orthogonal, $\Sigma_A = \mathrm{diag}(\sigma_1(A), \cdots, \sigma_k(A))$, $\sigma_1(A) \geq \cdots \geq \sigma_k(A) > 0$, and U_{A1} and U_{A2} are respectively the first k columns and the last $m-k$ columns of U_A.

It is shown in [11] that (1) has a unique minimal norm solution if $b \notin \mathrm{range}(A)$ and $U_{A1}^T b \neq 0$, which imply $\sigma_k(A) > \sigma_{k+1}(C)$. Then, from [11], the STLS solution can be explicitly expressed as

$$\lambda x_{\rm STLS} = -V_{12}(v_{22}^T)^+ = (V_{11}^T)^+ v_{21} = (A^T A - V_{12}\Sigma_2^2 V_{12}^T)^+(\lambda A^T b - V_{12}\Sigma_2^2 v_{22}), \tag{4}$$

where $(v_{22}^T)^+$ denotes the pseudoinverse of v_{22}^T.

For most STLS problems, the condition $U_{A1}^T b \neq 0$ is satisfied. So, in this paper, we assume $U_{A1}^T b \neq 0$. It then remains to check the condition $\sigma_k(A) > \sigma_{k+1}(C)$. As shown above, the STLS problem can be solved by using the SVD. As we know, computing the SVD is expensive. In this paper, we present an algorithm for solving the STLS problem using a rank revealing decomposition. This algorithm is more efficient than the SVD method and it is particularly efficient for the STLS problems with same coefficient matrix but multiple right hand side vectors. In section 2, we first describe a complete orthogonal decomposition (COD) [3] to illustrate the ideas behind our algorithm. Then we present a practical algorithm for solving the STLS problem using the rank revealing ULV decomposition (RRULVD) [9]. The details of computing the RRULVD are described in Section 3. A perturbation analysis of our STLS algorithm is given in Section 4 and numerical experiments are presented in Section 5.

2 Algorithm

The STLS solution expression (4) shows that to compute the solution, we need only V_{12} and v_{22}, which, from the partition of V in (3), form the null space and the right singular vector corresponding to the smallest nonzero singular value of the augmented matrix C defined in (2). It is unnecessary to compute all the individual singular values and singular vectors.

Suppose that

$$C = \bar{P} \begin{bmatrix} \bar{L} & 0 \\ 0 & 0 \end{bmatrix} \bar{Q}^T$$

is the COD of C, where $\bar{P} \in \mathrm{R}^{m\times(n+1)}$ has orthonormal columns, $\bar{Q} \in R^{(n+1)\times(n+1)}$ is orthogonal, and $\bar{L}$ is a $(k+1)$-by-$(k+1)$ nonsingular lower triangular matrix. Let w be the right singular vector corresponding to the smallest nonzero singular value $\sigma_{k+1}(\bar{L})$ of $\bar{L}$ and

$$\bar{L} = U_{\bar{L}} \Sigma_{\bar{L}} [V_{\bar{L}1} \ w]^T$$

be the SVD of $\bar{L}$, then

$$C = \bar{P} \begin{bmatrix} U_{\bar{L}} & 0 \\ 0 & I_{m-k-1} \end{bmatrix} \begin{bmatrix} \Sigma_{\bar{L}} & 0 \\ 0 & 0 \end{bmatrix} \left(\bar{Q} \begin{bmatrix} V_{\bar{L}1} & w & 0 \\ 0 & 0 & I_{n-k} \end{bmatrix} \right)^T$$

is the SVD of C. Comparing the above SVD and the SVD in (2) and using the partition of V in (3), we have

$$\begin{bmatrix} V_{12} \\ v_{22}^T \end{bmatrix} = \bar{Q} \begin{bmatrix} w & 0 \\ 0 & I_{n-k} \end{bmatrix}.$$

Partitioning $\bar{Q} = [\bar{Q}_1 \ \bar{Q}_2]$ such that $\bar{Q}_1$ and $\bar{Q}_2$ are respectively the first $k+1$ and the last $n-k$ columns of $\bar{Q}$, we get

$$\begin{bmatrix} V_{12} \\ v_{22}^T \end{bmatrix} = [\bar{Q}_1 w \ \ \bar{Q}_2] \, .$$

It is shown in [10] that when $U_{A1}^T b \neq 0$, V_{11} is of full rank and v_{22} is a nonzero vector. Then we can find an $(n-k+1)$-by-$(n-k+1)$ Householder matrix H, such that $\widetilde{Q} = [\bar{Q}_1 w \ \ \bar{Q}_2]H$ and $\widetilde{Q}(n+1, \ 2:n-k+1) = 0$. That is $v_{22}^T H = \|v_{22}\|_2 [1, 0, ..., 0]$. Thus, from (4),

$$\begin{aligned} \lambda x_{\mathrm{STLS}} &= -V_{12}(v_{22}^T)^+ = -V_{12}(v_{22}^T v_{22})^{-1} v_{22} = -\|v_{22}\|_2^{-2} (V_{12}H)(v_{22}^T H)^T \\ &= -\|v_{22}\|_2^{-1} V_{12} H \,[1, 0, ..., 0]^T \\ &= -\widetilde{Q}(1:n, 1)/\widetilde{Q}(n+1, 1). \end{aligned} \tag{5}$$

Note that $\widetilde{Q}(n+1, 1) = \|v_{22}\|_2 \neq 0$, since v_{22} is a nonzero vector.

Now, we have described a COD method for computing the STLS solution. This method has the following issues to be dealt with. First, the COD is sensitive to perturbations and rounding errors when the matrix is rank deficient. Second, we still need to compute the right singular vector corresponding to the smallest nonzero singular value of C. Third, we may want to check the solution existence condition $\sigma_k(A) > \sigma_{k+1}(C)$, recalling that $\sigma_k(A)$ and $\sigma_{k+1}(C)$ are the smallest nonzero singular values of A and C respectively. To alleviate these problems, we propose a rank revealing ULV decomposition [9] (RRULVD) algorithm. The RRULVD of $A \in \mathrm{R}^{m \times n}$ is defined as

$$A = P_A \begin{bmatrix} L_A & \\ H_A & F_A \end{bmatrix} Q_A^T, \tag{6}$$

where L_A and F_A are lower triangular, L_A is of order k, the numerical rank of A, $\|F_A\|_2 \approx \sigma_{k+1}(A)$ and $\|H_A\|_2$ is sufficiently small so that $\|F_A\|_2 + \|H_A\|_2 \approx \sigma_{k+1}(A)$. Thus RRULVD reveals the numerical rank of A. When both $\|H_A\|_2$ and $\|F_A\|_2$ are small, the RRULVD can be viewed as an approximation of the COD of a rank-deficient matrix. In addition, in the next Section, we will show that in the computation of the decomposition of A, we get an estimate for $\sigma_k(A)$. Moreover, the RRULVD can be efficiently updated when a column λb is appended to A. Also, in updating the decomposition, we can get estimates for $\sigma_{k+1}(C)$ and the corresponding right singular vector. All the information needed for computing the STLS solution and checking the condition $\sigma_k(A) > \sigma_{k+1}(C)$ can be obtained. Letting

$$C := [A \ \lambda b] = P_C \begin{bmatrix} L_C & \\ H_C & F_C \end{bmatrix} Q_C^T \tag{7}$$

be the updated RRULVD after λb is appended to A, we present the following algorithm. The details of computing the RRULVD, the crucial part of the algorithm, is given in the next section.

Algorithm 1 (STLS Algorithm based on RRULVD)
Given A, b, λ, this algorithm computes STLS solution x_{STLS} using the RRULVD.

1. *Compute the RRULVD (6) and an estimate of* $\sigma_k(A)$*;*
2. *Append* λb *to* A*, update the RRULVD, and compute the estimates for* $\sigma_{k+1}(C)$ *and the corresponding right singular vector* w*;*
3. **if** $(\sigma_k(A) = \sigma_{k+1}(C))$ *quit* **end***;*
4. *Partition* $Q_C = [Q_{C1} \ Q_{C2}]$ *such that* Q_{C1} *and* Q_{C2} *contain the first* $k+1$ *and the last* $n-k$ *columns of* Q_C *respectively;*
5. *Find a Householder matrix* H *such that* $\widetilde{Q} = [Q_{C1}w \ Q_{C2}]H$ *and* $\widetilde{Q}(n+1, \ 2 : n-k+1) = 0$*;*
6. $\lambda x_{\mathrm{STLS}} = -\widetilde{Q}(1 : n, \ 1)/\widetilde{Q}(n+1, \ 1)$.

3 Computing RRULVD

In [2] and [4], two RRULVD algorithms for a rank-deficient matrix A are presented. Both of them first apply the QL decomposition to A, then some techniques are used to reveal the rank. Although the QL decomposition can be applied in our case, it is inaccurate and unstable due to the rank deficiency and the rounding errors. The RRULVD algorithm presented in this section is based on Stewart's method [9]. It is a column updating scheme in that the RRULVD of A is efficiently updated when a column is added to A.

We assume that the RRULVD (6) of A is available and a column a is appended to A. We will show how the RRULVD of A can be efficiently updated to the RRULVD of the augmented matrix $[A\ a]$ where $a := \lambda b$.

Let

$$y = P_A^T a = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix},$$

then, from (6), we have

$$[A\ a] = PLQ^T := P_A \begin{bmatrix} L_A & 0 & y_1 \\ H_A & F_A & y_2 \end{bmatrix} \begin{bmatrix} Q_A^T & 0 \\ 0 & 1 \end{bmatrix} \tag{8}$$

What we need to do next is to triangularize L and update P and Q correspondingly. While there are several ways of eliminating y_1 and y_2, we want to choose one that keeps the rank revealing structure as much as possible. If $\|y_2\|_2$ is small, then $\mathrm{rank}(A) = \mathrm{rank}([A\ a])$ and the RRULVD can be updated by postmultiplying a sequence of rotations to eliminate y_1 and y_2. If $\|y_2\|_2$ is not too small, we premultiply a sequence of plane rotations $G_{i,j}$ to transform y_2 into ηe_1, $\eta = \|y_2\|_2$. Since each rotation $G_{i,j}$ creates a bulge in the lower triangular F_A, we simultaneously postmultiply a sequence of rotations $K_{i,j}$ to restore the lower triangular structure. The following figure illustrates this process. For simplicity, we only show the part $[H_A\ \ F_A\ \ y_2]$ that is modified.

$$\begin{bmatrix} h & h & h & f & & & & y_2 \\ h & h & h & f & f & & & y_2 \\ h & h & h & f & f & f & & y_2 \\ h & h & h & f & f & f & f & y_2 \end{bmatrix} \xrightarrow{G_{3,4}} \begin{bmatrix} h & h & h & f & & & & y_2 \\ h & h & h & f & f & & & y_2 \\ h & h & h & f & f & f & f & y_2 \\ h & h & h & f & f & f & f & 0 \end{bmatrix} \xrightarrow{K_{6,7}}$$

$$\begin{bmatrix} h & h & h & f & & & & y_2 \\ h & h & h & f & f & & & y_2 \\ h & h & h & f & f & f & & y_2 \\ h & h & h & f & f & f & f & 0 \end{bmatrix} \xrightarrow{G_{2,3}} \begin{bmatrix} h & h & h & f & & & & y_2 \\ h & h & h & f & f & f & & y_2 \\ h & h & h & f & f & f & & 0 \\ h & h & h & f & f & f & f & 0 \end{bmatrix} \xrightarrow{K_{5,6}}$$

$$\begin{bmatrix} h & h & h & f & & & & y_2 \\ h & h & h & f & f & & & y_2 \\ h & h & h & f & f & f & & 0 \\ h & h & h & f & f & f & f & 0 \end{bmatrix} \xrightarrow{G_{1,2}} \begin{bmatrix} h & h & h & f & f & & & \eta \\ h & h & h & f & f & & & 0 \\ h & h & h & f & f & f & & 0 \\ h & h & h & f & f & f & f & 0 \end{bmatrix} \xrightarrow{K_{4,5}}$$

$$\begin{bmatrix} h & h & h & f & & & & \eta \\ h & h & h & f & f & & & 0 \\ h & h & h & f & f & f & & 0 \\ h & h & h & f & f & f & f & 0 \end{bmatrix}.$$

The above figure also shows that the entries of the updated H_A and F_A stay small. The procedure is given as follows.

Algorithm 2 (Triangularization 1.)

Given the decomposition (8), this algorithm transforms y_2 into ηe_1 and updates P and Q while keeping H_A and F_A small and F_A lower triangular.

1. **for** $i = m : -1 : k+2$
2. *Generate the rotation* $G_{i-1,i}$ *to zero out the ith entry using the* $(i-1)$*th entry in* y*;*
3. *Apply* $G_{i-1,i}$ *to the rows* $i-1$ *and* i *of* L*;*
4. *Update* P *by applying* $G^T_{i-1,i}$ *to the columns* i *and* $i-1$ *of* P*;*
5. *Generate the rotation* $K_{i+n-m-1,i+n-m}$ *to restore the lower triangular structure of* F_A*;*
6. *Apply* $K_{i+n-m-1,i+n-m}$ *to the columns of* $i+n-m-1$ *and* $i+n-m$ *of* L*;*
7. *Update* Q *by applying* $K_{i+n-m-1,i+n-m}$ *to the columns* $i+n-m-1$ *and* $i+n-m$ *of* Q*;*
8. **end**

After the above procedure, the L matrix in the decomposition (8) of the augmented matrix has the following structure:

$$\begin{bmatrix} l & & & & & & y \\ l & l & & & & & y \\ l & l & l & & & & y \\ h & h & h & f & & & \eta \\ h & h & h & f & f & & 0 \\ h & h & h & f & f & f & 0 \end{bmatrix}. \tag{9}$$

Now, we triangularize the matrix in (9) by postmultiplying plane rotations. At the same time, we want to keep the rank revealing structure as much as possible. The following example shows how it works:

$$\begin{bmatrix} l & & & & & & y \\ l & l & & & & & y \\ l & l & l & & & & y \\ h & h & h & f & & & \eta \\ h & h & h & f & f & & 0 \\ h & h & h & f & f & f & 0 \end{bmatrix} \overset{K_{1,7}}{\longrightarrow} \begin{bmatrix} l & & & & & & 0 \\ l & l & & & & & y \\ l & l & l & & & & y \\ \eta & h & h & f & & & \eta \\ h & h & h & f & f & & h \\ h & h & h & f & f & f & h \end{bmatrix} \overset{K_{2,7}}{\longrightarrow}$$

$$
\begin{bmatrix} l & & & & & & 0 \\ l & l & & & & & 0 \\ l & l & l & & & & y \\ \eta & \eta & h & f & & & \eta \\ h & h & h & f & f & & h \\ h & h & h & f & f & f & h \end{bmatrix} \overset{K_{3,7}}{\longrightarrow} \begin{bmatrix} l & & & & & & 0 \\ l & l & & & & & 0 \\ l & l & l & & & & 0 \\ \eta & \eta & \eta & f & & & \eta \\ h & h & h & f & f & & h \\ h & h & h & f & f & f & h \end{bmatrix} \overset{K_{4,7}}{\longrightarrow}
$$

$$
\begin{bmatrix} l & & & & & & 0 \\ l & l & & & & & 0 \\ l & l & l & & & & 0 \\ \eta & \eta & \eta & \eta & & & 0 \\ h & h & h & f & f & & f \\ h & h & h & f & f & f & f \end{bmatrix} \overset{K_{5,7}}{\longrightarrow} \begin{bmatrix} l & & & & & & 0 \\ l & l & & & & & 0 \\ l & l & l & & & & 0 \\ \eta & \eta & \eta & \eta & & & 0 \\ h & h & h & f & f & & 0 \\ h & h & h & f & f & f & f \end{bmatrix} \overset{K_{6,7}}{\longrightarrow} \begin{bmatrix} l & & & & & & 0 \\ l & l & & & & & 0 \\ l & l & l & & & & 0 \\ \eta & \eta & \eta & \eta & & & 0 \\ h & h & h & f & f & & 0 \\ h & h & h & f & f & f & 0 \end{bmatrix}
$$

As shown in the above figure, if $\eta = \|y_2\|_2$ is not small, the row $k+1$ of L is not small and the numerical rank of the augmented matrix may be $k+1$. The following is the algorithm.

Algorithm 3 (Triangularization 2.)

Given the matrix in (9), this algorithm restores the lower triangular structure.

1. **for** $i = 1 : n$

2. *Generate the rotation $K_{i,n+1}$ to eliminate the ith entry in y using the ith diagonal element;*

3. *Apply $K_{i,n+1}$ to the columns i and $n+1$ to L;*

4. *Update Q by applying $K_{i,n+1}$ to the columns i and $n+1$ of Q;*

5. **end**

Applying these two processes, we can add a column to A and restore the triangular structure of L while keeping its rank revealing structure as much as possible. The structure of the new lower triangular matrix shows that the numerical rank of the augmented matrix is either $k+1$ or k. So, the remaining problem is to determine the numerical rank.

We use the deflation proposed in [9] to find the numerical rank. Assume that w is the right singular vector corresponding to the smallest singular value, denoted by $\sigma_{k+1}(L)$, of the order $k+1$ leading principal submatrix of L, or the large block. Then a product K of plane rotations can be found such that

$$
Kw = \|w\|_2 \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} = \|w\|_2 e_n.
$$

Consequently, we obtain

$$\sigma_{k+1}(L) = \|Lw\|_2 = \|GLK^T Kw\|_2 = \|(GLK^T)e_n\|_2,$$

where G is an orthogonal matrix such that GLK^T is lower triangular. This equation shows that the $(k+1, k+1)$-entry of L, equals $\sigma_{k+1}(L)$.

Suppose that w is an approximation of the right singular vector corresponding to $\sigma_{k+1}(L)$, then the $(k+1, k+1)$-entry is an approximation of $\sigma_{k+1}(L)$. Therefore, given a tolerance for the numerical rank, if $L_{n,n}$ is larger than the tolerance then the numerical rank of L is $k+1$, otherwise, the numerical rank is at most k.

The following figures depict the deflation procedure. We first find a sequence of planes rotations $K_{i+1,i}$ such that $K_{k+1,k} \cdots K_{2,1} w = \|w\|_2 e_n$:

$$\begin{matrix} w \\ w \\ w \\ w \end{matrix} \overset{K_{2,1}}{\longrightarrow} \begin{matrix} 0 \\ w \\ w \\ w \end{matrix} \overset{K_{3,2}}{\longrightarrow} \begin{matrix} 0 \\ 0 \\ w \\ w \end{matrix} \overset{K_{4,3}}{\longrightarrow} \begin{matrix} 0 \\ 0 \\ 0 \\ \|w\|_2 \end{matrix}.$$

Next, we postmultiply L with $K_{i+1,i}^T$ and simultaneously find $G_{i+1,i}$ to restore its lower triangularity:

$$\begin{matrix} l & 0 & 0 & 0 \\ l & l & 0 & 0 \\ l & l & l & 0 \\ l & l & l & l \end{matrix} \overset{K_{2,1}^T}{\longrightarrow} \begin{matrix} l & \tilde{l} & 0 & 0 \\ l & l & 0 & 0 \\ l & l & l & 0 \\ l & l & l & l \end{matrix} \overset{G_{2,1}}{\longrightarrow} \begin{matrix} l & 0 & 0 & 0 \\ l & l & 0 & 0 \\ l & l & l & 0 \\ l & l & l & l \end{matrix} \overset{K_{3,2}^T}{\longrightarrow} \begin{matrix} l & 0 & 0 & 0 \\ l & l & \tilde{l} & 0 \\ l & l & l & 0 \\ l & l & l & l \end{matrix}$$

$$\overset{G_{3,2}}{\longrightarrow} \begin{matrix} l & 0 & 0 & 0 \\ l & l & 0 & 0 \\ l & l & l & 0 \\ l & l & l & l \end{matrix} \overset{K_{4,3}^T}{\longrightarrow} \begin{matrix} l & 0 & 0 & 0 \\ l & l & 0 & 0 \\ l & l & l & \tilde{l} \\ l & l & l & l \end{matrix} \overset{G_{4,3}}{\longrightarrow} \begin{matrix} l & 0 & 0 & 0 \\ l & l & 0 & 0 \\ l & l & l & 0 \\ l & l & l & l \end{matrix}$$

After this procedure, we determine the numerical rank of L by comparing its $(k+1, k+1)$-entry with the tolerance.

In the following algorithm for the deflation, we use Van Loan's 2-norm condition estimator [5] to compute the approximations of the smallest singular value and its corresponding right singular vector.

Algorithm 4 (Deflation)

1. *Using Van Loan's condition estimator to compute an approximation of the smallest singular value, denoted by $\sigma_{k+1}(L)$, of the order $k+1$ leading principal submatrix of L and its corresponding right singular vector w;*

2. **if** *$\sigma_{k+1}(L) > tol$, return $k+1$ as the numerical rank of L, $\sigma_{k+1}(L)$ as an approximation of $\sigma_{k+1}(C)$* **end***;*

3. **for** $i = 1 : k$

4. *Generate* $K_{i+1,i}$ *to eliminate* w_i *using* w_{i+1}*;*

5. *Apply* $K_{i+1,i}^T$ *to the columns* i *and* $i+1$ *of* L*;*

6. *Update* Q *by apply* $K_{i+1,i}$ *to the columns* i *and* $i+1$ *of* Q*;*

7. *Generate* $G_{i+1,i}$ *to eliminate the* $(i, i+1)$*-entry of* L *using the* $(i+1, i+1)$*-entry;*

8. *Apply* $G_{i+1,i}$ *to the rows* i *and* $i+1$ *of* L *to restore the lower triangular structure of* L*;*

9. *Update* P *by applying* $G_{i+1,i}^T$ *to the columns* i *and* $i+1$ *of* P*;*

10. **end**

11. Do refinement (Algorithm 5, optional, described soon).

Since from (4) the STLS solution can be expressed by the null space and the right singular vector corresponding to the smallest nonzero singular value of C, the accuracy of the STLS solution computed by the above algorithm depends on the quality of the approximations of $\sigma_{k+1}(L)$ and w. It is shown in [1] that the quality of the subspaces obtained by the RRULVD algorithm depends on the quality of the condition estimator on the lower triangular matrix L. Thus we propose the following improvement on the approximations of $\sigma_{k+1}(L)$ and w.

To simplify the discussion, let L_C be the order $k+1$ leading principal submatrix of L. We first use Van Loan's method [5] to get an approximation of y, the right singular vector of L_C^T. Then we solve the linear system $L_C x = y$. Now, $\|x\|_2$ is an approximation of $\sigma_{k+1}(L_C)$, the smallest singular value of L_C, and $w = x/\|x\|$ is its corresponding right singular vector. Since L_C is lower triangular, the linear systems can be solved cheaply. Furthermore, the accuracy of the approximation of the smallest singular value and its right singular vector is improved significantly, especially when $\sigma_{k+1}(L_C)$ is large. Table 1 compares the results before and after the improvement. As shown in the first three columns in Table 1 we generate three random matrices of different orders with specified smallest singular values and the relative gaps between the two smallest singular values, that is gap $= (\sigma_k(L_C) - \sigma_{k+1}(L_C))/\sigma_{k+1}(L_C)$. The fourth column gives the approximations of the singular values, σ_{vl}, using Van Loan's method without improvement. The fifth column shows the approximations, σ_m, with improvement. Also, the sixth and seventh columns of Table 1 are cosines of the angles between the exact singular vectors and the vectors obtained from Van Loan's method without and with improvement, that is $\cos\theta_{vl} = \|v_{k+1}^T v_{vl}\|_2$ and $\cos\theta_m = \|v_{k+1}^T v_m\|_2$ where v_{k+1}, v_{vl}, and v_m are right singular vectors corresponding to $\sigma_{k+1}(L_C)$, σ_{vl}, and σ_m, respectively. From Table 1, we see that it improves the singular value and singular vector estimates significantly. In particular, when the singular value is large, for example, $\sigma_{k+1}(L_C) = 10$, the relative error in the approximated singular value is almost 30% for the original version while it is only 5% after the improvement.

n	gap	$\sigma_{k+1}(L_C)$	σ_{vl}	σ_m	$\cos\theta_{vl}$	$\cos\theta_m$
16	10	0.0280	0.0320	0.0280	1.0000	1.0000
16	2	0.0280	0.0435	0.0281	0.9655	0.9990
64	5	1.0000	1.8189	1.0135	0.8874	0.9962
128	1.5	10.0000	13.4835	10.5153	0.9051	0.9697

Table 1: Comparison of the smallest singular value and singular vector estimates computed by Van Loan's method with and without improvement.

The accuracy of the STLS solution depends not only on the quality of the computed singular values and singular vectors, but also on the quality of the null space. Thus the remaining problem is to improve the null space approximation by making the off diagonal block H_C in (7) small. To motivate the refinement technique [9], we assume that L has numerical rank $k+1$ and consider the order $k+2$ leading principal submatrix:

$$T_{k+2} = \begin{bmatrix} L_C & 0 \\ h^T & \xi \end{bmatrix},$$

of L, where L_C is a $(k+1)\times(k+1)$ lower triangular matrix, h^T is a row vector of order $k+1$ and ξ is a scalar.

Now suppose we find an orthogonal matrix G partitioned according to the partition of T_{k+2} such that T_{k+2} is transformed into a block upper triangular matrix:

$$\begin{bmatrix} G_{11} & g_{12} \\ g_{21}^T & g_{22} \end{bmatrix} \begin{bmatrix} L_C & 0 \\ h^T & \xi \end{bmatrix} = \begin{bmatrix} \tilde{L}_C & \tilde{h} \\ 0 & \tilde{\xi} \end{bmatrix}. \tag{10}$$

Specifically, G can be a product of a sequence of rotations which eliminate h using the rows of L_C from bottom to top. It then follows that

$$\xi\, g_{12} = \tilde{h}$$

and

$$g_{21}^T L_C \ + \ g_{22} h^T \ = \ 0.$$

Let $\sigma_{k+1}(L_C)$ be the smallest singular value of L_C, then

$$\sigma_{k+1}(L_C)\|g_{21}^T\|_2 \leq \|g_{21}^T L_C\|_2 = \|g_{12} h^T\|_2 \leq \|h\|_2,$$

that is

$$\|g_{21}^T\|_2 \leq \frac{\|h\|_2}{\sigma_{k+1}(L_C)}.$$

Applying this inequality to $\xi\, g_{12} = \tilde{h}$, we obtain

$$\|\tilde{h}\|_2 = \|\xi\, g_{12}\|_2 \leq \frac{|\xi|}{\sigma_{k+1}(L_C)}\|h\|_2.$$

In other words, $\|\tilde{h}\|_2$ is decreased by a factor of $|\xi|/\sigma_{k+1}(L_C)$, which is the ratio between the smallest singular values of T_{k+2} and L_C. Next, we postmultiply an orthogonal matrix K to restore the lower triangular structure:

$$\begin{bmatrix} \tilde{L}_C & \tilde{h} \\ 0 & \tilde{\xi} \end{bmatrix} K = \begin{bmatrix} \hat{L}_C & 0 \\ \hat{h}^T & \hat{\xi} \end{bmatrix}.$$

Specifically, K can be a product of a sequence of rotations that eliminate $\tilde{h}$ using the columns of $\tilde{L}_C$ from left to right. Now, we can conclude from the above analysis that $\|\hat{h}\|_2 \leq (|\xi|/\sigma_{k+1}(L_C))^2 \|h\|_2$. When the ratio $|\xi|/\sigma_{k+1}(L_C)$ is small, $\|\hat{h}\|_2$ is much smaller than $\|h\|_2$. Hence if we apply this refinement to the last $m-k-1$ rows of L, the off diagonal elements in those rows become significantly small. Consequently, the quality of the null space is improved. The refinement algorithm is as follows.

Algorithm 5 (Refinement)

Given the lower triangular matrix L produced by the first 10 steps in Algorithm 3, this algorithm applies refinement on its last $(m-k-1)$ rows, where $k+1$ is the numerical rank of L.

1. **while** $m > k+1$
2.   **for** $i = m-1 : -1 : 1$
3.   *Generate rotation* $G_{i,m}$ *to eliminate the* (m,i)*-entry of* L *using its* (i,i)*-entry;*
4.   *Apply* $G_{i,m}$ *to the rows* i *and* m *of* L*;*
5.   *Update* P *by applying* $G_{i,m}^T$ *to the columns* i *and* m *of* P*;*
6.   **end**
7.   **for** $i = 1 : m-1$
8.   *Generate rotation* $K_{i,m}$ *to eliminate the* (i,m)*-entry of* L *using its* (i,i)*-entry;*
9.   *Apply* $K_{i,m}$ *to the columns* i *and* m *of* L*;*
10.   *Update* Q *by applying* $K_{i,m}$ *to the columns* i *and* m *of* Q*;*
11.   **end**
12. $m = m-1;$
13. **end**

In summary, to compute the RRULVD of A, starting with the RRULVD of the first column of A, we append the columns of A, one column at a time, and update the RRULVD using Algorithms 1, 2, and 3. Then, we append λb to A and update the RRULVD. Refinement Algorithm 5 may be applied in updating to improve the quality of the null space. Since only one right singular vector and the null space of C are required for solving the STLS problems, updating P is unnecessary when we compute the RRULVD of C.

4 Perturbation Analysis

Algorithm 1 first computes an RRULVD:

$$C := [A\ \lambda b] = P_C \begin{bmatrix} L_C & 0 \\ H_C & F_C \end{bmatrix} Q_C^T, \tag{11}$$

where the blocks H_C and F_C are introduced by rounding errors and approximations. Then the algorithm computes the STLS solution using the truncated RRULVD as the COD:

$$P_C \begin{bmatrix} L_C & 0 \\ 0 & 0 \end{bmatrix} Q_C^T =: [\widehat{A}\ \lambda\widehat{b}] = \widehat{C}. \tag{12}$$

Since H_C and F_C are introduced by rounding errors, we assume that

$$E := C - \widehat{C} = -P \begin{bmatrix} 0 & 0 \\ H_C & F_C \end{bmatrix} Q^T,$$

is small, specifically,

$$\|H_C\|_2 + \|F_C\|_2 = c\,u\,\|C\|_2 =: \eta, \tag{13}$$

where c is a moderate constant and u is the unit of roundoff. What is the difference between the solution corresponding to $C = [A\ \lambda b]$ and that of $\widehat{C} = [\widehat{A}\ \lambda\widehat{b}]$? In this section, we derive an upper bound for the error $\|x_{\text{STLS}} - \widehat{x}_{\text{STLS}}\|_2$, where x_{STLS} and $\widehat{x}_{\text{STLS}}$ denote the solutions corresponding to C and $\widehat{C}$ respectively.

Before deriving the error bound, it is necessary to verify the existence condition. From (13), it follows that

$$\begin{aligned} & \sigma_k(\widehat{A}) - \sigma_{k+1}(\widehat{C}) \\ = \ & \sigma_k(A) - \sigma_{k+1}(C) + \sigma_k(\widehat{A}) - \sigma_k(A) + \sigma_{k+1}(C) - \sigma_{k+1}(\widehat{C}) \\ \geq \ & \sigma_k(A) - \sigma_{k+1}(C) - 2\eta. \end{aligned}$$

Thus, if $\sigma_k(A) - \sigma_{k+1}(C) > 2\eta$, then the existence condition $\sigma_k(\widehat{A}) > \sigma_{k+1}(\widehat{C})$ for the perturbed STLS problem is satisfied.

Now, we derive the error bound. Using the SVD (2) of C and the partitions (3), we define

$$E_A := A - U_2\Sigma_2 V_{12}^T = U_1\Sigma_1 V_{11}^T \quad \text{and} \quad \lambda\, e_b := \lambda b - U_2\Sigma_2 v_{22} = U_1\Sigma_1 v_{21}.$$

Then, from (4), it can be verified that

$$\lambda\, x_{\text{STLS}} = (V_{11}^T)^+ v_{21} = \lambda E_A^+ e_b. \tag{14}$$

Note that when $\sigma_k(A) > \sigma_{k+1}(C)$, V_{11} is of full column rank [10], implying that $I = V_{11}^+ V_{11} = V_{11}^T (V_{11}^T)^+$. Consequently,

$$E_A x_{\text{STLS}} = U_1\Sigma_1 V_{11}^T x_{\text{STLS}} = \lambda^{-1} U_1\Sigma_1 V_{11}^T (V_{11}^T)^+ v_{21} = \lambda^{-1} U_1 \Sigma_1 v_{21} = e_b.$$

Similarly, letting $\widehat{C} = \widehat{U}\widehat{\Sigma}\widehat{V}^T$ be the SVD of $\widehat{C}$, partitioning $\widehat{U}$, $\widehat{\Sigma}$, and $\widehat{V}$ according to (3), and defining

$$E_{\widehat{A}} := \widehat{A} - \widehat{U}_2\widehat{\Sigma}_2\widehat{V}_{12}^T = \widehat{U}_1\widehat{\Sigma}_1\widehat{V}_{11}^T \quad \text{and} \quad \lambda e_{\hat{b}} := \lambda\widehat{b} - \widehat{U}_2\widehat{\Sigma}_2\widehat{v}_{22} = \widehat{U}_1\widehat{\Sigma}_1\widehat{v}_{21},$$

we have the solution

$$\widehat{x}_{\mathrm{STLS}} = E_{\widehat{A}}^{+} e_{\hat{b}}. \tag{15}$$

Comparing the two solutions (14) and (15), we get

$$\begin{aligned}
x_{\mathrm{STLS}} - \widehat{x}_{\mathrm{STLS}} &= x_{\mathrm{STLS}} - E_{\widehat{A}}^{+} e_{\hat{b}} \\
&= x_{\mathrm{STLS}} - E_{\widehat{A}}^{+} E_{\widehat{A}} x_{\mathrm{STLS}} + E_{\widehat{A}}^{+} E_{\widehat{A}} x_{\mathrm{STLS}} - E_{\widehat{A}}^{+} e_b - E_{\widehat{A}}^{+}(e_{\hat{b}} - e_b) \\
&= x_{\mathrm{STLS}} - E_{\widehat{A}}^{+} E_{\widehat{A}} x_{\mathrm{STLS}} + E_{\widehat{A}}^{+} E_{\widehat{A}} x_{\mathrm{STLS}} - E_{\widehat{A}}^{+} E_A x_{\mathrm{STLS}} - E_{\widehat{A}}^{+}(e_{\hat{b}} - e_b) \\
&= (I - E_{\widehat{A}}^{+} E_{\widehat{A}}) x_{\mathrm{STLS}} + E_{\widehat{A}}^{+}(E_{\widehat{A}} - E_A) x_{\mathrm{STLS}} - E_{\widehat{A}}^{+}(e_{\hat{b}} - e_b).
\end{aligned}$$

Obviously, $\|(I - E_{\widehat{A}}^{+} E_{\widehat{A}}) x_{\mathrm{STLS}}\|_2 \leq \|x_{\mathrm{STLS}}\|_2$. From $E_{\widehat{A}} = \widehat{A} - \widehat{U}_2\widehat{\Sigma}_2\widehat{V}_{12}^T$, we have

$$\sigma_k(E_{\widehat{A}}) \geq \sigma_k(\widehat{A}) - \|\widehat{U}_2\widehat{\Sigma}_2\widehat{V}_{12}^T\|_2 \geq \sigma_k(\widehat{A}) - \sigma_{k+1}(\widehat{C}) \geq \sigma_k(A) - \sigma_{k+1}(C) - 2\eta,$$

which implies that

$$\|E_{\widehat{A}}^{+}\|_2 = (\sigma_k(E_{\widehat{A}}))^{-1} \leq \frac{1}{\sigma_k(A) - \sigma_{k+1}(C) - 2\eta}, \tag{16}$$

since $\mathrm{rank}(E_{\widehat{A}}) = k$. Furthermore, we have

$$\begin{aligned}
\|E_{\widehat{A}} - E_A\|_2 &= \|\widehat{A} - A - \widehat{U}_2\widehat{\Sigma}_2\widehat{V}_{12}^T + U_2\Sigma_2 V_{12}^T\|_2 \\
&\leq \|\widehat{A} - A\|_2 + \|\widehat{U}_2\widehat{\Sigma}_2\widehat{V}_{12}^T\|_2 + \|U_2\Sigma_2 V_{12}^T\|_2 \\
&\leq \|\widehat{C} - C\|_2 + \sigma_{k+1}(\widehat{C}) + \sigma_{k+1}(C) \\
&\leq \eta + \sigma_{k+1}(C) + \sigma_{k+1}(\widehat{C}) \qquad (17) \\
&\leq 2\eta + 2\sigma_{k+1}(C) \qquad (18)
\end{aligned}$$

and

$$\begin{aligned}
\|e_{\hat{b}} - e_b\|_2 &= \|\widehat{b} - b - \lambda^{-1}\widehat{U}_2\widehat{\Sigma}_2\widehat{v}_{22} + \lambda^{-1}U_2\Sigma_2 v_{22}\|_2 \\
&\leq \|\widehat{b} - b\|_2 + \lambda^{-1}(\sigma_{k+1}(\widehat{C}) + \sigma_{k+1}(C)) \qquad (19) \\
&= \eta + \lambda^{-1}(2\sigma_{k+1}(C) + \eta). \qquad (20)
\end{aligned}$$

Putting the above three inequalities (16), (18), and (20) together, we get

$$\begin{aligned}
\|x_{\mathrm{STLS}} - \widehat{x}_{\mathrm{STLS}}\| &\leq \|x_{\mathrm{STLS}}\|_2 + \|E_{\widehat{A}}^{+}\|_2 \|E_{\widehat{A}} - E_A\|_2 \|x_{\mathrm{STLS}}\|_2 + \|E_{\widehat{A}}^{+}\|_2 \|e_{\hat{b}} - e_b\|_2 \\
&\leq \|x_{\mathrm{STLS}}\|_2 + \frac{2\sigma_{k+1}(C) + 2\eta}{\sigma_k(A) - \sigma_{k+1}(C) - 2\eta}\|x_{\mathrm{STLS}}\|_2
\end{aligned}$$

$$
\begin{aligned}
& + \frac{\lambda^{-1}(2\sigma_{k+1}(C)+\eta)+\eta}{\sigma_k(A)-\sigma_{k+1}(C)-2\eta} \\
= \; & \frac{\sigma_k(A)+\sigma_{k+1}(C)}{\sigma_k(A)-\sigma_{k+1}(C)-2\eta}\|x_{\mathrm{STLS}}\|_2 + \frac{\lambda^{-1}(2\sigma_{k+1}(C)+\eta)+\eta}{\sigma_k(A)-\sigma_{k+1}(C)-2\eta} \\
= \; & \frac{\sigma_k(A)\|x_{\mathrm{STLS}}\|_2 + \sigma_{k+1}(C)(\|x_{\mathrm{STLS}}\|_2 + 2\lambda^{-1}) + (\lambda^{-1}+1)\eta}{\sigma_k(A)-\sigma_{k+1}(C)-2\eta} \\
< \; & \frac{(\sigma_k(A)+\sigma_{k+1}(C))(\|x_{\mathrm{STLS}}\|_2 + \lambda^{-1})+\eta}{\sigma_k(A)-\sigma_{k+1}(C)-2\eta},
\end{aligned}
$$

since $\eta < \sigma_k(A) - \sigma_{k+1}(C)$. The above argument is valid for any small perturbation E. Thus we obtain the following theorem.

Theorem 4.1 *Suppose that $C = [A\ \lambda b]$ and $\widehat{C} = C + E =: [\widehat{A}\ \lambda\widehat{b}]$ and $\|E\|_2 \approx c\,u\,\|C\|_2 =: \eta$, where c is a moderate constant and u is the unit of roundoff. Let x_{STLS} and $\widehat{x}_{\mathrm{STLS}}$ be the STLS solutions corresponding to C and $\widehat{C}$ respectively, then*

$$
\|x_{\mathrm{STLS}} - \widehat{x}_{\mathrm{STLS}}\|_2 \leq \frac{(\sigma_k(A)+\sigma_{k+1}(C))(\|x_{\mathrm{STLS}}\|_2 + \lambda^{-1})+\eta}{\sigma_k(A)-\sigma_{k+1}(C)-2\eta},
$$

provided that $\sigma_k(A) - \sigma_{k+1}(C) > 2\eta$.

This theorem shows that if the perturbation $\eta = \|E\|_2$ is small, we can expect a small error $\|x_{\mathrm{STLS}} - \widehat{x}_{\mathrm{STLS}}\|_2$ as long as $\sigma_k(A)$ and $\sigma_{k+1}(C)$ are not very close to each other. If $\sigma_k(A)$ is very close to $\sigma_{k+1}(C)$, the computed solution $\widehat{x}_{\mathrm{STLS}}$ may be very different from the exact solution x_{STLS}. Moreover, as λ approaches to zero, both $\sigma_{k+1}(C)$ and $\sigma_{k+1}(\widehat{C})$ approach to zero as fast as λ does. Specifically, $\lim_{\lambda\to 0}\sigma_{k+1}(C)/\lambda = \|r\|_2$, where r is the residual of the least square problem $Ax \approx b$ [11]. Thus, from (16), (17), and (19), the inequality in the theorem reduces to

$$
\|x_{\mathrm{STLS}} - \widehat{x}_{\mathrm{STLS}}\|_2 \leq (1 + \frac{\eta}{\sigma_k(A)})\|x_{\mathrm{STLS}}\|_2 + \frac{\eta}{\sigma_k(A)}(1 + \|\hat{r}\|_2 + \|r\|_2).
$$

It shows that the difference between x_{STLS} and $\widehat{x}_{\mathrm{STLS}}$ is independent of the scalar λ, when λ approaches to zero.

5 Numerical Experiments

In the STLS formulation (1), a scalar λ is introduced to the right side vector b. The residual to be minimized is $[E, r]$, same as the TLS problem. In this section, we compare STLS with TLS. The STLS problem is solved by the RRULVD method presented in the previous section, whereas the TLS problem is solved by the SVD method.

All of our numerical experiments were performed in MATLAB on a Sun SPARC workstation Ultra 10 using double precision. The rank deficient matrices were generated as the product

$$
A = U \begin{bmatrix} \Sigma & 0 \\ 0 & Z \end{bmatrix} V^T,
$$

where $U \in \mathrm{R}^{m\times n}$ and $V \in \mathrm{R}^{n\times n}$, $(m > n)$ are random matrices with orthonormal columns, Σ diagonal of order k, whose diagonal elements are random variables uniformly distributed over $[0, 1]$, and Z a zero matrix of order $n-k$. The right-hand side vectors were generated as random vectors uniformly distributed over $[0, 1]$. The random perturbations E and r on A and b respectively were constructed by

$$E = \xi\, \mathrm{randn}(m, n), \qquad r = \xi\, \mathrm{randn}(m, 1),$$

where ξ is a parameter controlling the magnitude of the perturbations, and the entries of E and r are random variables normally distributed with zero mean and variance one. In all examples, we set $\xi = 3\times 10^{-8}$ and the numerical rank tolerance to 2×10^{-5}. Since the perturbations are smaller than the numerical rank tolerance, all matrices are numerically rank deficient.

To compare STLS and TLS, we denote θ_S and θ_T as the angles between b and Ax_{STLS} and between b and Ax_{TLS}, respectively, that is $\cos\theta_S := \|b^T Ax_{\mathrm{STLS}}\|_2/(\|Ax_{\mathrm{STLS}}\|_2\|b\|_2)$ and $\cos\theta_T := \|b^T Ax_{\mathrm{TLS}}\|_2/(\|Ax_{\mathrm{TLS}}\|_2\|b\|_2)$. Also, we denote the residuals $res_S := \|[E_{\mathrm{STLS}},\ r_{\mathrm{STLS}}]\|_F$, which is equal to $\sigma_{k+1}(C)$ [11], and $res_T := \|[E_{\mathrm{TLS}},\ r_{\mathrm{TLS}}]\|_F = \sigma_{k+1}(C)$ [10]. Note that θ_T and res_T are independent of λ.

Tables 2, 3, and 4 show the results for three rank-deficient problems of various sizes. From the results, we can see that

- For small values of λ, Ax_{STLS} is closer to b than Ax_{TLS} is, and the STLS residual is much smaller than the TLS residual;
- When λ is small, θ_S is insensitive to the change of λ;
- When $\lambda > 1$, $\cos\theta_S$ may be smaller than $\cos\theta_T$. See, for example, $\lambda = 5$ in Tables 2 and 4.

We note that

- In theory, when $\lambda = 1$, $x_{\mathrm{STLS}} = x_{\mathrm{TLS}}$. The differences in the tables when $\lambda = 1$ are due to the different algorithms used to solve the STLS problem and the TLS problem. However, we can see that the corresponding values are in the same magnitude order.
- For large values of λ, large vectors λb are appended to A to form C. Consequently, the right singular vectors corresponding to $\sigma_{k+1}(C)$ of the resulting matrices C vary little. Recall that the STLS solution depends on the right singular vector and the null space. Thus, the STLS solutions vary little for large values of λ.

Conclusion: Choose $\lambda < 1$.

References

[1] Ricardo. D. Fierro and James R. Bunch. Bounding the subspaces from rank revealing two-sided orthogonal decompositions. *SIAM J Matrix Anal. Appl.*, 16(1995), 743–759.

λ	$\cos\theta_S$	$\cos\theta_T$	res_S	res_T
0.01	0.9428	0.5724	0.01057	0.5805
0.1	0.9428	0.5724	0.1057	0.5805
1	0.8696	0.5724	0.6324	0.5805
5	0.3789	0.5724	0.9920	0.5805

Table 2: Comparison of the STLS solution with the TLS solution for a 30-by-20 matrix A of rank 18.

λ	$\cos\theta_S$	$\cos\theta_T$	res_S	res_T
0.01	0.9428	0.2610	0.0164	0.8050
0.1	0.9428	0.2610	0.1632	0.8050
1	0.6073	0.2610	0.9489	0.8050
5	0.8916	0.2610	1.0450	0.8050

Table 3: Comparison of the STLS solution with the TLS solution for a 64-by-48 matrix A of rank 43.

λ	$\cos\theta_S$	$\cos\theta_T$	res_S	res_T
0.01	0.9110	0.6080	0.0380	2.4507
0.1	0.8419	0.6080	0.9804	2.4507
1	0.7772	0.6080	5.0317	2.4507
5	0.4463	0.6080	5.0019	2.4507

Table 4: Comparison of the STLS solution with the TLS solution for a 256-by-120 matrix A of rank 105.

[2] Ricardo D. Fierro, Per Christian Hansen and Sϕren Kirk Hansen. UTV tools: Matlab templates for rank-revealing UTV decomposition. *Numerical Algorithms*, 20(1999), 165–194.

[3] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd Ed., The Johns Hopkins University Press, Baltimore, MD, 1996.

[4] Sabine Van Huffel and Hongyuan Zha. An efficient total least squares algorithm based on a Rank-Revealing two-sided orthogonal decomposition. *Numerical Algorithms*, 4(1993), 101–133.

[5] Charles Van Loan. On estimating the condition of eigenvalues and eigenvectors. *Linear Algebra and its Applications*, 88/89(1987), 715–732.

[6] Christopher C. Paige and Zdeněk Strakoš. Bounds for the least squares distance using scaled total least squares. *Numer. Math.* 91(2002), 93–115.

[7] Christopher C. Paige and Zdeněk Strakoš. Scaled total least squares fundermentals. *Numer. Math.* 91(2002), 117–146.

[8] B.D. Rao. Unified treatment of LS, TLS and Truncated SVD methods using a weighted TLS framework. *Recent Advanced in Total Least Squares Techniques and Errors-in-Variables Modeling*, edited by S. Van Huffel. SIAM, Philadelphia PA, 1997, 11–20.

[9] G.W. Stewart. Matrix Algorithms, volume I: Basic Decompositions. SIAM, Philadelpha, 1998.

[10] Musheng Wei. The analysis for the total least square problem with more than one soluiton. *SIAM J. Matrix Anal. Appl.*, 13(1992), 746–763.

[11] Wei Xu, Yimin Wei and Sanzheng Qiao. An analysis of rank-deficient scaled total least squares problem, *Technical Report No. CAS 03-10-SQ*, McMaster University, Hamilton, Ont. Canada.