A note on the condition number of the scaled total least squares problem

Shaoxin Wang\textsuperscript{a}\textsuperscript{*}, Hanyu Li\textsuperscript{b}, Hu Yang\textsuperscript{b}

\textsuperscript{a}School of Statistics, Qufu Normal University, Qufu, 273165, P. R. China
\textsuperscript{b}College of Mathematics and Statistics, Chongqing University, Chongqing, 401331, P. R. China

Abstract

In this paper, we consider the explicit expressions of the normwise condition number for the scaled total least squares problem. Some techniques are introduced to simplify the expression of the condition number, and some new results are derived. Based on these new results, new expressions of the condition number for the total least squares problem can be deduced as a special case. New forms of the condition number enjoy some storage and computational advantages. We also proposed three different methods to estimate the condition number. Some numerical experiments are carried out to illustrate the effectiveness of our results.

Keywords: condition number, Fréchet derivative, the scaled total least squares problem, power method, probabilistic condition estimation method

2010 MSC: 65F35, 15A12, 15A60

1. Introduction

The scaled total least squares (STLS) problem (or technique) was proposed by Rao \cite{1} to give a unified treatment of ordinary least squares (OLS) problem, total least squares (TLS) problem and the data least squares (DLS) problem. Paige and Strakš \cite{2} reformulated the STLS problem and presented a detailed analysis of conditions that guarantee the STLS problem has a unique solution. Following their line, for $A \in \mathbb{R}^{m \times n}$ with $m > n$ and $b \in \mathbb{R}^m$, the STLS problem is given by

$$
\min \| [E \quad r] \|_F, \quad \text{subject to } \lambda b - r \in \mathcal{R}(A + E),
$$

(1.1)

where $\lambda$ is a positive real number, $\| \cdot \|_F$ denotes the Frobenius norm and $\mathcal{R}(\cdot)$ is the range space. Let $[E_S \quad r_S]$ be the solution to (1.1), then the solution to the linear system $(A + E_S)\lambda x = \lambda b - r_S$ is called the STLS solution and denoted by $x_S$. As shown in \cite{2}, when $\lambda = 1$, $\lambda \to 0$ and $\lambda \to \infty$, $x_S$ becomes the TLS solution $x_T$, OLS solution $x_O$ and DLS solution $x_D$, respectively.

The condition number gives a quantitative measurement of the maximum amplification of the resulting change in solution with respect to a perturbation in the data and has been extensively studied for too many topics to list here. For the STLS problem, Zhou et al. \cite{3} considered its perturbation analysis and presented the normwise, mixed and componentwise condition numbers. Based on the

*This work is supported by a project of Shandong Province Higher Educational Science and Technology Program (Grant No. J17KA160) and the National Natural Science Foundation of China (Grant Nos. 11671059, 11671060)

*Corresponding author

Email addresses: shuangmy@163.com, shxwang@qfnu.edu.cn (Shaoxin Wang), lihy.hy@gmail.com (Hanyu Li), hy@cqu.edu.cn (Hu Yang)

Preprint submitted to Manuscript

January 15, 2018
perturbation theory of singular value decomposition (SVD) given in [4], Li and Jia [5] gave a different approach to derive the normwise and componentwise condition numbers of the STLS problem, and the structured condition numbers were also investigated there. It should be noted that the normwise condition number in [3] contains a Kronecker product which makes it impractical to compute, especially for large-scale problems. Based on the fact that \( \|A\|_2 = \|A^T A\|_2^{1/2} \), \( \| \cdot \|_2 \) denotes the spectral norm of matrix or Euclidean norm of vector, some closed formulas and upper (or lower) bounds of normwise condition number for the TLS problem were given in [6] and [7], and these results are easy to compute and do not contain Kronecker product any more. Xie et al. [8] showed that the expressions of condition number given in [6] and [7] are mathematically equivalent. However, computing the matrix cross product \( A^T A \) is a source of rounding errors and is potentially numerical unstable [9, pp. 386]. Some progress to avoid computing \( A^T A \) was made in [6, 7, 8]. In this paper, we present a new expression of the normwise condition number for the STLS problem. The new expression is easy to compute and does not need to compute Kronecker product or matrix cross product. On the other hand, we also propose a procedure to compute the condition number, which does not need to form the explicit expression of condition number and avoid computing the Kronecker product.

The rest of the paper is organized as follows. Section 2 contains the main results of the paper. Some algorithms and numerical experiments are presented in Section 3. Concluding remarks are given in Section 4. Before proceeding to the following sections, we introduce some notation first: For any matrix \( B \), \( A \otimes B = [a_{ij}B] \) denotes the Kronecker product of \( A \) and \( B \), \( \text{vec}(\cdot) \) is a linear map defined by \( \text{vec}(A) = [a_{1,1}, \ldots, a_{m,1}, \ldots, a_{1,n}, \ldots, a_{m,n}]^T \).

2. Main results

As stated in the Introduction, the TLS can be treated as a special case of the STLS problem. An interesting result is that we can solve the STLS problem by finding the solution to a special TLS problem. When \( \lambda = 1 \), we get the following TLS problem

\[
\begin{aligned}
\min \left\| \begin{bmatrix} E \\ r \end{bmatrix} \right\|_F, & \text{ subject to } b - r \in \mathcal{R}(A + E).
\end{aligned}
\]

(2.1)

It is easy to check that when \( x_S \) is the solution of (1.1) then \( \lambda x_S \) is the TLS solution to the following TLS problem

\[
\begin{aligned}
\min \left\| \begin{bmatrix} E \\ r \end{bmatrix} \right\|_F, & \text{ subject to } \lambda b - r \in \mathcal{R}(A + E).
\end{aligned}
\]

(2.2)

Let the SVDs of the matrices \( [A \lambda b] \) and \( A \) be

\[
U^T \begin{bmatrix} A \\ \lambda b \end{bmatrix} V = \Sigma, \quad \hat{U}^T A\hat{V} = \hat{\Sigma},
\]

where \( U = [u_1, \ldots, u_{n+1}] \in \mathbb{R}^{m \times (n+1)} \), \( V = [v_1, \ldots, v_n] \in \mathbb{R}^{(n+1) \times (n+1)} \), \( \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_{n+1}) \) with \( \sigma_1 \geq \cdots \geq \sigma_{n+1} \geq 0 \), and \( \hat{\Sigma} = \text{diag}(\hat{\sigma}_1, \ldots, \hat{\sigma}_n) \) with \( \hat{\sigma}_1 \geq \cdots \geq \hat{\sigma}_n \geq 0 \). Analogous to the Golub-Van Loan condition [10] for TLS problem to guarantee the existence and uniqueness of a solution, Zhou et al. [3] presented the following sufficient condition to ensure the STLS problem has a unique solution

\[
\hat{\sigma}_n > \sigma_{n+1}.
\]

Therefore, by (2.3) the TLS solution to (2.2) is

\[
\lambda x_S = (A^T A - \sigma_{n+1} I_n)^{-1} A^T (\lambda b),
\]
which gives

\[ x_S = (A^T A - \sigma_{n+1}^2 I_n)^{-1} A^T b. \] (2.4)

Let \( \Delta A \) and \( \Delta b \) be the corresponding perturbations to \( A \) and \( b \), then we have the following perturbed STLS problem

\[
\min \left\| \begin{bmatrix} E & r \end{bmatrix} \right\|_F, \quad \text{subject to } \lambda (b + \Delta b) - r \in \mathcal{R} ((A + \Delta A) + E).
\] (2.5)

For the perturbed STLS problem, Zhou et al. [3] and Li and Jia [5] presented two different approaches to show that when the perturbation \([\Delta A, \Delta b] \) is sufficiently small, the perturbed STLS problem admits a unique solution \( x_{PS} \). We take the result given in [5] as the following theorem with some modifications of symbols.

**Theorem 2.1.** Under the assumption (2.3), if \( \| [\Delta A, \Delta b] \|_F \) is small enough, then the perturbed STLS problem (2.5) has the unique solution \( x_{PS} \). Moreover,

\[
\Delta x = x_{PS} - x_S = K \left[ \text{vec}(\Delta A) \Delta b \right] + \mathcal{O} \left( \| [\Delta A, \Delta b] \|_F^2 \right)
\] (2.6)

where

\[ K = M^{-1} \left( \left[ \begin{array}{c} 2 \| r \|_2^2 A^T r r^T - A^T \end{array} \right] \left( \left[ \begin{array}{c} x_S^T \setminus 1 \end{array} \right] \otimes I_m \right) - \left[ I_n \otimes r^T \ 0_{n \times m} \right] \right) \]

with \( M = A^T A - \sigma_{n+1}^2 I_n \) and \( r = Ax_S - b \).

Li and Jia [5] also presented a very detailed comparison of the above results and those given in [3] and showed that their perturbation estimate is the same as that given in [3]. In addition, according to Theorem 2.1 we can deduce that if we set

\[ F : \mathbb{R}^{m \times n} \times \mathbb{R}^m \rightarrow \mathbb{R}^n \]

\[
\left[ A \ \lambda b \right] \rightarrow x_S = M^{-1} A^T b
\]

then the map \( F \) is Fréchet differentiable at \([A \lambda b] \) under the assumption (2.3) and the Fréchet derivative of \( F \) at \([A \lambda b] \) is given by

\[ DF(A, \lambda b) := K. \]

According to the definition of condition number given in [11] and [12], the relative normwise condition number of STLS problem is given by

\[
\kappa_{r,F}(A, \lambda b) = \lim_{\delta \to 0} \sup_{\| \Delta A, \lambda b \|_F < \delta} \frac{\| F(A+\Delta A, \lambda(b+\Delta b)) - F(A, \lambda b) \|_2}{\| F(A, \lambda b) \|_2 \| \Delta A, \lambda b \|_F },
\] (2.7)

When \( F \) is Fréchet differentiable, \( \kappa_{r,F}(A, \lambda b) \) reduces to

\[
\kappa_{r,F}(A, \lambda b) = \frac{\| DF(A, \lambda b) \|_2 \| [A \lambda b] \|_F }{\| F(A, \lambda b) \|_2},
\]

and \( \kappa_F(A, \lambda b) = \| DF(A, \lambda b) \|_2 \) is the absolute condition number. We should remark that the definition of condition number given by (2.7) can be extended to a more general sense. Wang and Yang [13] presented a unified definition of condition number to cope with the conditioning of equality constrained indefinite least squares problem, which include the normwise, mixed and componentwise condition numbers as its special cases, for further discussions see [13, 14].

For the convenience of presentation, we summarize the above discussion in the following theorem.
\textbf{Theorem 2.2.} Under the assumption (2.3), the relative normwise condition number of STLS problem defined by (2.7) is

\[ \kappa_{rF}(A, \lambda b) = \frac{\|K\|_2 \|A \lambda b\|_F}{\|F(A, \lambda b)\|_2}, \]

and its absolute condition number is given by

\[ \kappa_F(A, \lambda b) = \|K\|_2. \quad (2.8) \]

It should be noted that the Kronecker product enlarges the order of matrix and may make it impractical to explicitly forming \( K \) when \( m \) and \( n \) are large. For the TLS problem, adjoint techniques are employed to eliminate the Kronecker product in [8, 6]. Following their step and going to further, we give the following theorem to simplify the condition number of the STLS problem, which is also the main result of our paper.

Considering the relationship between relative and absolute condition numbers, in the following we only focus on the simplification of \( \kappa_F(A, \lambda b) \).

\textbf{Theorem 2.3.} The absolute condition number \( \kappa_F(A, \lambda b) \) for STLS problem has the following two equivalent forms

\[ \kappa_{F1}(A, \lambda b) = \|M^{-1}(1 + \|r\|_2^2)A^T A - A^T r r^T - x S r^T A + \|r\|_2^2 I_n) M^{-1}\|_2^{1/2}, \quad (2.9) \]

\[ \kappa_{F2}(A, \lambda b) = \left\| M^{-1} \begin{pmatrix} A^T, \ |A|_2^2 A^T \left( I_m - \frac{1}{\|r\|_2^2} r r^T \right), \ |r|_2^2 \left( I_n - \frac{1}{\|r\|_2^2} A^T r r^T \right) \right\|_2. \quad (2.10) \]

\textbf{Proof.} For a real matrix \( X, \|X\|_2 = \|X^T X\|_2^{1/2} = \|X X^T\|_2^{1/2} \) holds. Analogous to [6], we have

\[ \|K\|_2 = \|KK^T\|_2^{1/2}. \]

Since \( M \) is symmetric, by the equality vec\((AXB) = (B^T \otimes A)\)vec\((X) \) [15, Chapt.4] we can get

\[ KK^T = M^{-1} \left( \begin{pmatrix} 2 & 0 \\ \|r\|_2^2 & 1 \end{pmatrix} A^T r r^T - A^T \right) \left( \begin{pmatrix} x S^T \\ -I_m \end{pmatrix} \right) \left( I_m + \|r\|_2^2 A^T - x S r^T A + \|r\|_2^2 I_n) M^{-1} \right). \]

\[ = M^{-1} \left( 1 + \|x S\|_2^2 \right) A^T - A^T r r^T - x S r^T A + \|r\|_2^2 I_n) M^{-1} \quad (2.11) \]

\[ = M^{-1} \left( A^T \right) I_n \left( 1 + \|x S\|_2^2 \right) \left( I_m - \frac{r r^T}{\|r\|_2^2 I_n} \right) \quad (2.12) \]

Since

\[ \left( \begin{pmatrix} 1 & \frac{1}{\|r\|_2^2} r r^T \\ -x S^T \end{pmatrix} \left( \begin{pmatrix} x S^T \\ -I_m \end{pmatrix} \right) \right) = \left( \begin{pmatrix} I_m & 0 \\ 0 & I_n \end{pmatrix} \right) \left( \begin{pmatrix} I_m & \frac{1}{\|r\|_2^2} r r^T \\ 0 & I_n \end{pmatrix} \right) \quad (2.13) \]

and

\[ I_m + \|x S\|_2^2 \left( I_m - \frac{1}{\|r\|_2^2} r r^T \right) = \left( \begin{pmatrix} I_m & \|x S\|_2^2 \left( I_m - \frac{1}{\|r\|_2^2} r r^T \right) \\ \|x S\|_2^2 \left( I_m - \frac{1}{\|r\|_2^2} r r^T \right) \end{pmatrix} \right) \quad (2.14) \]

we substitute (2.14) and (2.13) into (2.12) and get

\[ KK^T = M^{-1} W W^T M^{-1}, \quad (2.15) \]
where
\[
W = \left[ A^T \|x_S\|_2 A^T \left( I_m - \frac{1}{\|r\|_2} r r^T \right) \|r\|_2 \left( I_n - \frac{1}{\|r\|_2} A^T r x_S^T \right) \right].
\]

Using the formula \(\|K\|_2 = \|K K^T\|_2^{1/2}\), again we complete the proof with (2.11) and (2.15). \(\square\)

**Remark 2.1.** From Theorem 2.3, we can see that the orders of the matrices in (2.8), (2.9) and (2.10) are \(n \times m(n+1)\), \(n \times n\) and \(n \times (2m+n)\), respectively. When \(m\) and \(n\) are comparable and large, the two equivalent forms given by Theorem 2.3 no longer contain a Kronecker product, and thus have some superiorities in storage and practical computation. But as pointed out in [6] and [9], the matrix cross product \(K K^T\) may lead to large rounding errors, so \(\kappa_{F2}(A, \lambda b)\) is preferable for numerical stability.

It should be noted that when \(\lambda = 1\), we get the TLS problem from (1.1). Based on Theorems 2.2 and 2.3, different expressions of the condition number for TLS problem follow

\[
\kappa_{TLSF}(A, b) = \left\| M^{-1} \left( \frac{2}{\|r\|_2} A^T \mathbb{I}^T - A^T \right) \left( \begin{bmatrix} x_T^T & -1 \end{bmatrix} \otimes I_m - \left[ I_n \otimes r^T \right] \begin{bmatrix} 0_{n \times m} \end{bmatrix} \right) \right\|_2,
\]

\[
\kappa_{TLSF_1}(A, b) = \left\| M^{-1} \left( 1 + \|x_T\|_2^2 \right) A^T A - A^T \mathbb{I}^T - x_T r^T A + \|r\|_2^2 I_n \right\}_2^{1/2},
\]

and

\[
\kappa_{TLSF_2}(A, b) = \left\| M^{-1} \left[ A^T \|x_T\|_2 A^T \left( I_m - \frac{1}{\|r\|_2} r r^T \right) \|r\|_2 \left( I_n - \frac{1}{\|r\|_2} A^T r x_T^T \right) \right] \right\}_2.
\]

We note that \(\kappa_{TLSF_1}(A, b)\) was an intermediate result of Theorem 1 in [6, Equation 3.8], and \(\kappa_{TLSF}(A, b)\) is given by Jia and Li [7, Theorem 2]. Based on \(M x_T = A^T b\) and its variants, Baboulin and Gratton [6] also showed that

\[
\kappa_{TLSF_1}(A, b) = \left\| \left( 1 + \|x_T\|_2^2 \right) M^{-1} \left( A^T A + \sigma_{n+1} \left( I_n - \frac{2}{\|x_T\|_2^2} x_T x_T^T \right) \right) M^{-1} \right\}_2^{1/2},
\]

and suggested that when the TLS problem is solved by the SVD method, the computation of (2.18) can be further simplified. But their simplified expression needs the SVDs of both \(A\) and \([A, b]\), this may be expensive. Jia and Li [7] further showed that only the SVD of \([A, b]\) will be enough. Based on the SVDs of \(A\) and/or \([A, b]\), some computable upper and lower bounds of the condition number were also presented in [6] and [7]. Furthermore, it can be easily checked that

\[
A^T A + \sigma_{n+1} \left( I_n - \frac{2}{1 + \|x_T\|_2^2} x_T x_T^T \right)
\]

is positive definite. Xie et al. [8, Remark 2] suggested that to use Cholesky decomposition to further simplify the expression of condition number, but no explicit expression was given there. According to Remark 2.1, our new expression \(\kappa_{TLSF_2}(A, b)\) needs less storage space, and does not need to calculate Cholesky decomposition. We only need to calculate the product of matrices and vectors, since \(M^{-1}\) can be the intermediate result when the TLS problem is solved with its normal equation. So we may say that the \(\kappa_{TLSF_2}(A, b)\) is a new result on the condition number of TLS problem, and enjoys storage and computational advantages.

As in [3] and [5], when \(\lambda \to 0\), we get \(\sigma_{n+1} \to 0\). Therefore, \((A^T A - \sigma_{n+1})^{-1} \to (A^T A)^{-1}\) and \(x_S\) converges to \(x_O\). When the matrix \(A\) has full column rank, from Theorems 2.2 and 2.3 and by the equality \(A^T r = 0\), we get the following three equivalent expressions of the condition number for OLS problem

\[
\kappa_{OLS}(A, b) = \left\| \left( A^T A \right)^{-1} \left( -A^T \left( \begin{bmatrix} x_T^T & -1 \end{bmatrix} \otimes I_m \right) - \left[ I_n \otimes r^T \right] \begin{bmatrix} 0_{n \times m} \end{bmatrix} \right) \right\|_2.
\]
\[ \kappa_{\text{OLS}F_1}(A, \lambda b) = \left\| (A^T A)^{-1} \left( (1 + \|x_0\|^2_2) A^T A + \|r\|^2_2 I_n \right) (A^T A)^{-1} \right\|_2, \tag{2.19} \]

and

\[ \kappa_{\text{OLS}F_2}(A, \lambda b) = \left\| (A^T A)^{-1} \left[ A^T \|x_0\|_2 A^T \|r\|_2 I_n \right] \right\|_2. \tag{2.20} \]

With a little algebra, we can check that \( \kappa_{\text{OLS}F}(A, b) \) can be rewritten as follows

\[ \kappa_{\text{OLS}F}(A, b) = \left\| \left[ -(x_0^T \otimes A^T) - (A^T A)^{-1} \otimes r^T \ A^T \right] \right\|_2, \tag{2.21} \]

where \( A^T = (A^T A)^{-1} A^T \) is the Moore-Penrose inverse of matrix \( A \) (see [16, 17]). It should be noted that (2.21), (2.19) and (2.20) have been given by Li and Wang [14] in investigating the condition number for indefinite least squares problem.

**Remark 2.2.** In [5], the authors also considered the linear structured condition number for the STLS problem. Although we can also make some progress on finding the compact form, like (2.9), of the linear structured condition number through the method given in [18], the final expression may enjoy some computational advantage and is still very complicated. So in this paper we will not consider the normwise structured condition number for the STLS problem, for more research on structured condition number (see, e.g., [19, 20, 21]).

### 3. Numerical experiment

In this part, we mainly focus on the computation of the condition number for STLS problem via its different forms. We note that the main task of calculating the condition number is to find the maximum eigenvalue of a matrix. For a large matrix, iterative techniques are always preferred in finding its extreme eigenvalues, a standard reference is [22, Chapt. 10]. Since the product of matrix and vector can be used to cancel the Kronecker product which coincides with the main step of the power method [22, pp. 365], Baboulin and Gratton [6] proposed an iteration of the power method to compute the condition number for TLS problem. Similar to [6, Proposition 3], an iterative procedure can also be established.

To apply the power method, from equation (2.8) we get

\[ K^T y = \left( \left( \frac{2}{\|r\|_2^2} A^T r r^T - A^T \right) \left[ \begin{array}{c} x \\ -1 \end{array} \right] \otimes I_m \right) - \left[ I_n \otimes r^T \ 0_{n \times n} \right] M^{-1} y. \]

Since computing the inverse of a matrix is not desired, we may set \( z = M^{-1} y \) and get \( z \) by solving the linear system \( M z = y \). Thus, we can obtain

\[ K^T y = \left( \begin{array}{c} x \\ -1 \end{array} \right) \otimes I_m \left( \frac{2}{\|r\|_2^2} A^T r r^T - A \right) \left[ \begin{array}{c} I_n \otimes r \\ 0_{n \times n} \end{array} \right] z = \vec{\left( \begin{array}{c} w x^T - w \\ rz^T - w \end{array} \right)} = \vec{\left( \begin{array}{c} w x^T - rz^T - w \end{array} \right)}, \]

where \( w = \left( \frac{2}{\|r\|_2^2} A^T r r^T - A \right) z \). We present the following algorithm for computing the condition number (2.8), which circumvents the Kronecker product. The derivation of Algorithm 1 is very similar to the Algorithm 1 in [6], so we omit some details.
Algorithm 1 Power method
Given the initial $y \in \mathbb{R}^n$.

\begin{algorithm}
\For{$i = 1, 2, \ldots$}
  \begin{enumerate}
  \item $[A_p, b_p] \leftarrow [wx^T - rz^T - w]$
  \item $v \leftarrow \| [A_p, b_p] \|_F$
  \item $[A_p, b_p] \leftarrow \frac{1}{v} [A_p, b_p]$
  \item $y \leftarrow M^{-1} \left( 2 \parallel r \parallel_2^2 A^T r x - A^T r x_S^T \right) (A_p x - b_p) - A_p^T r$
  \end{enumerate}
\end{algorithm}

Remark 3.1. Algorithm 1 is used to compute $v$, the maximum eigenvalue of $KK^T$, so we use $\sqrt{v}$ as the condition number of STLS problem. Moreover, the power method converges if $v$ is dominant and the initial vector $y$ has a component in the direction of the corresponding dominant eigenvector [22, pp. 366]. The choice of $y$ is usually implied by applications or by using random vector.

We also present a probabilistic condition estimation (PCE) method. This method is based on the probabilistic spectral norm estimator proposed by Hochstenbach in [23], which gives an interval containing the spectral norm of a matrix with high probability. The PCE method has been used to estimate the condition number of various problems (see [14, 24]). For the convenience of presentation, we summarize the method given in [23] as the following lemma.

Lemma 3.1. Let $U(S_{p-1})$ be the uniform distribution over unit sphere $S_{p-1}$ in $\mathbb{R}^p$, and $A \in \mathbb{R}^{m \times n}$. If we choose a random vector $z$ from $U(S_{p-1})$, then by the probabilistic spectral norm estimator we have

$$
\begin{align*}
\alpha &\leq \|A\|_2, \\
\|A\|_2 &\leq \beta \text{ with probability at least } 1 - \epsilon, \\
\beta/\alpha &\leq 1 + \theta,
\end{align*}
$$

where $\alpha$ is the guaranteed lower bound, $\beta$ is the probabilistic upper bound, $\epsilon$ and $\theta$ are two use-chosen parameters.

Based on the above lemma, we can propose the following algorithm to give a sharp estimate of the condition number for the STLS problem.

Algorithm 2 PCE method

\begin{algorithm}
\begin{enumerate}
\item Getting the start vector $v \in \mathbb{R}^{2m+n}$ from $U(S_{2m+n-1})$.
\item Compute the matrix in (2.10) and let
\[ \tilde{K} = M^{-1} \left[ A^T \|x\|_2^2 A^T \left( I_m - \frac{1}{\|r\|_2^2} r r^T \right) \|r\|_2 \left( I_n - \frac{1}{\|r\|_2^2} A^T r x_S^T \right) \right]. \]
\item Compute $\alpha$ and $\beta$ of $\tilde{K}$ by probabilistic spectral norm estimator [23].
\item $\kappa_{pceF}(A, \lambda b) = \frac{\alpha + \beta}{2}$.
\end{enumerate}
\end{algorithm}

It should be pointed out that Step 2 can be done by the Matlab function normprob.m which can be downloaded from [http://www.win.tue.nl/~hochsten/eigenvaluetools/](http://www.win.tue.nl/~hochsten/eigenvaluetools/). In the practical implementation of Algorithm 2 and as suggested in [23], explicitly forming matrix $\tilde{K}$ may not be
necessary, because what we really need is the product of a random vector with \( \hat{K} \) and \( \hat{K}^T \). Hence again, some techniques in solving linear system can be employed to reduce the computational burden, especially for large scale problems. More derivation and extension on the probabilistic condition estimation can be found in [23, 25].

In the recent paper [26], the authors proposed to use the small sample condition estimation (SCE) method [27] to estimate the condition numbers of TLS problem, and proposed two ways to estimate the normwise condition number. The SCE method used in [26] can be directly applied to estimate the normwise condition number of STLS problem, so we adapt their Algorithm 2 needing less CPU time than Algorithm 1 with some modifications as Algorithm 3.

Algorithm 3 SCE method

1. Generate \( k \) vectors \( z_1, z_2, \cdots, z_k \in \mathbb{R}^n \) with entries in the uniform continuous distribution on the interval \((0, 1)\), where \( k \) is the sample size. Orthogonalize these vectors via modified Gram-Schmidt orthogonalization process.

2. Approximate the Willis factors \( \omega_n \) and \( \omega_k \) by

\[
\omega_n \approx \sqrt{\frac{2}{\pi (p - \frac{1}{2})}} \quad \text{and} \quad \omega_k \approx \sqrt{\frac{2}{\pi (k - \frac{1}{2})}}.
\]

3. For \( i = 1, \cdots, k \), compute

\[
\kappa_i = \| z_i^T M^{-1} [(1 + \| x_S \|^2_2) A^T A - A^T x_S^T x_S - x_S r^T A + r^2 I_n] M^{-1} z_i \|^2_2.
\]

4. Estimate the absolute normwise condition number by

\[
\kappa_{sceF}(A, b) = \frac{\omega_k}{\omega_n} \sqrt{\sum_{i=1}^{k} \kappa_i^2}.
\]

Remark 3.2. In the implementation of Algorithm 3, we need to compute \( M^{-1} z_i \), which is usually done by solving the linear system

\[ M y = z_i. \]

Since \( M \) is positive definite, the preconditioned conjugate gradient (PCG) method can be employed [28]. Moreover, in practice the sample size \( k = 3 \) will give a relative high accurate estimation of condition number. Diao et al. [26] showed that the computational cost of SCE method is \( O(mn^2 + n^2) \). Here, we need to make some comments on the computational complexity of Algorithms 2 and 3. As pointed out in [18, Section 4.1], it is not easy to give an exact comparison of the computational complexity of these two algorithms. Because the probabilistic spectral norm estimator is based on Lanczos iteration method, and the dimension of Krylov space is automatically determined by \( \epsilon \) [23], whereas the SCE method mainly depends on matrix-vector product and one orthonormalization procedure. So instead of counting flops we report the CPU time to compare the efficiency of Algorithms 2 and 3.

Example 3.1. Since investigating the influence of different forms on the computation of condition number for STLS problem is our purpose, we construct the following random STLS problem, which is similar to [6]. Let \( [A \lambda b] \) be defined by

\[
\begin{bmatrix}
A & \lambda b
\end{bmatrix} = Y \begin{bmatrix}
D & 0
\end{bmatrix} Z^T \in \mathbb{R}^{m \times (n+1)}, \quad Y = I_m - 2yy^T, \quad Z = I_{n+1} - 2zz^T,
\]

8
where \( y \in \mathbb{R}^m \), \( z \in \mathbb{R}^{n+1} \) are random unit vectors, and \( D = \text{diag}(n, n-1, \cdots, 1, 1 - e_p) \) for given parameter \( e_p \). Due to the interlacing property [29, pp. 178], we get

\[
\hat{\sigma}_n - \sigma_{n+1} \leq \sigma_n - \sigma_{n+1} = e_p.
\]

Thus \( e_p \) gives a measure of the distance of the problem to nongenericity, and the solution \( x_S \) is given by (2.4). By varying \( \lambda \), \( e_p \) and the order of matrix, we report the CPU time in seconds for computing the condition number of STLS problem with different forms. All the computations are performed in Matlab R2010b on a PC with Intel i5-6600M CPU 3.30 GHz and 4.00 GB RAM.

First, we compare two "naive" methods, that is, we first compute the explicit form of the matrices in (2.8) and (2.10) and then compute its spectral norms by the built-in function \texttt{norm}(. , 2). We repeat it 200 times for one group of settings, and report the mean values of CPU time in Table 1. From Table 1, we can see that when \( m = 500, n = 300 \), computing (2.8) becomes very time consuming due to the large order of matrix, but (2.10) still works well. Moreover, we can also find that the CPU time for computing (2.8) is always smaller than that for (2.10).

Second, Algorithm 1 provides a method for avoiding the Kronecker product. So with respect to accuracy and running time, we give a comparison of the efficiency of computing condition number for four different methods: (1) use the formula (2.10) exactly(\texttt{EXA}); (2) use Algorithm 1(\texttt{PW}). (3) use Algorithm 2(\texttt{PCE}). For Algorithm 2, the user-chosen parameters are given by \( \epsilon = 0.001, \delta = 0.01; \) (4) use Algorithm 3, and the sample size \( k = 3 \). In Algorithm 1, the initial vector \( y \) is a random vector with elements from standard normal distribution, and the algorithm terminates when the difference between two successive values of \( v \) is less than \( 10^{-8} \) or the number of iterations exceeds 500. To show the accuracy of estimation, we use formula (2.10) as the benchmark, and define the following ratios

\[
\text{ratio}_1 = \frac{\text{PW}}{\text{EXA}}, \quad \text{ratio}_2 = \frac{\text{PCE}}{\text{EXA}}, \quad \text{ratio}_3 = \frac{\text{PCE}}{\text{EXA}}.
\]

The ratios are plotted in Figure 1, we only report the case \( \lambda = 5 \) and \( e_p = 0.1 \), since the results for other cases are similar. From Figure 1, we can see that both PW and PCE methods give very accurate estimates of the condition number. The SCE method also gives acceptable estimates, since the ratios are contained in the interval \((0.1, 10)\) [9]. This coincides with the results in [26]. Thus, if the accuracy of estimation is required, PW and PCE methods are preferred.

Now, we turn to the running time of these four different methods with 200 replications for each case. The numerical results are presented in Table 2. From Table 2, we can find that, in most cases, these four methods perform very efficient in estimating the condition number of STLS problem. We also note that the PCE method is the most efficient especially for the large order cases, and \texttt{EXA}, PW and SCE have comparable performance, except for the case \( \lambda = 0.05 \) and \( e_p = 0.001 \), in which the variances of PW’s CPU time are 0.3457, 16.5316 and 758.6435. An explicit derivation for the underlying reason may be unavailable now, so we give some investigation through simulation.
Figure 1: Accuracy of condition number estimators; the first row $m = 200, n = 150$, the second row $m = 500, n = 300$, the third row $m = 1000, n = 700$.

Table 2: Average CPU time in seconds for EXA, PW, PCE and SCE methods

| $\lambda$ | $e_p$ | EXA     | PW     | PCE     | SCE     |
|-----------|-------|---------|---------|---------|---------|
| 0.05      | 0.1   | 0.0073  | 0.0491  | 0.4071  |         |
|           | 0.001 | 0.0070  | 0.0495  | 0.4093  |         |

| $\lambda = 5$ | $e_p$ | EXA     | PW     | PCE     | SCE     |
|---------------|-------|---------|---------|---------|---------|
| 0.1           | 0.001 | 0.0071  | 0.0487  | 0.3810  |         |
| 0.1           | 0.001 | 0.0073  | 0.0811  | 0.5421  |         |
| 0.1           | 0.001 | 0.0033  | 0.0049  | 0.0211  |         |
| 0.1           | 0.001 | 0.0019  | 0.0110  | 0.1136  |         |

| 0.001        |       |         |         |         |         |
|---------------|-------|---------|---------|---------|---------|
| 0.001        | 0.001 | 0.0072  | 0.0487  | 0.4409  |         |
| 0.001        | 0.001 | 0.0056  | 0.0649  | 0.4861  |         |
| 0.001        | 0.001 | 0.0118  | 0.0141  | 0.0298  |         |
| 0.001        | 0.001 | 0.0021  | 0.0111  | 0.1043  |         |
According to Remark 3.1, we generate 20 groups of data, and use 100 different random vectors as the initial vector for each group of data to test its influence on the CPU time of PW method. We give the box-plot of the results as Figure 2. From Figure 2, we find that when $m$ and $n$ are large, the PW method may perform unstable for some groups of data and has a lot of outliers which lead to its large mean values. However, we cannot conclude that the PW method is inefficient in estimating the condition number of STLS problem. Because, for most groups of data and with variant initial vectors, the PW method performs quite well. This may imply that our construction tends to give an ill-posed STLS problem for small $\lambda$ and $e_p$.

From our numerical experiment, we suggest that for moderate scale STLS problems computing the condition number via (2.10) is recommended. The reason is that compared with (2.8), (2.10) not only avoids computing a Kronecker product and saves storage space but also needs less CPU time and preserves high accuracy. When the coefficient matrix of the STLS problem is large, the PCE method can give highly accurate estimates of the condition number and needs less CPU time.

4. Concluding remark

In this paper, we present some new expressions of the condition number for the STLS problem. The new expressions do not contain a Kronecker product, and make it possible to store the condition number in the computer for large scale problems. The new and compact forms of the condition numbers for the STLS and TLS problem are of certain interest for the practitioners from other areas like engineering, statistics and so on. This is because the new forms need less storage space and is very easy to use. In addition, to avoid explicitly forming the matrix in the expression of condition number, some iterative methods are also introduced. We also present some numerical experiments to check the proposed algorithms, and find that our algorithms have very good performance for most of our settings. However, in our experiment, we note that when $\lambda$ and $e_p$ are small, and the matrix is large, the power method can be very inefficient. Due to the difficulty of theoretical justification, only some simulations are given to explore the underlying reason, which is not enough and further investigation
should be carried out in the future. To be on the safe side, we recommend using $\kappa_{F^2}(A, \lambda b)$ or PCE method to compute the condition number of STLS problem in practical applications.

References

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

[2] C.C. Paige, Z. Strakós, Scaled total least squares fundamentals, Numer. Math. 91(1)(2002)117-146.

[3] L. Zhou, L. Lin, Y. Wei, S. Qiao, Perturbation analysis and condition numbers of scaled total least squares problems, Numer. Algorithms 51(3)(2009)381-399.

[4] J.G. Sun, A note on simple non-zero singular values, J. Comput. Math. 6(1988)258-266.

[5] B. Li, Z. Jia, Some results on condition numbers of the scaled total least squares problem, Linear Algebra Appl. 435(3)(2011)674-686.

[6] M. Baboulin, S. Gratton, A contribution to the conditioning of the total least-squares problem, SIAM J. Matrix Anal. Appl. 32(3)(2011) 685-699.

[7] Z. Jia, B. Li, On the condition number of the total least squares problem, Numer. Math. 125(1)(2013)61-87.

[8] P. Xie, Y. Wei, H. Xiang, Perturbation Analysis and Randomized Algorithms for Large-Scale Total Least Squares Problems, arXiv preprint arXiv:1401.6832, 2014.

[9] N.J. Higham, Accuracy and Stability of Numerical Algorithms, 2 ed., SIAM, Philadelphia, 2002.

[10] G.H. Golub, C.F. Van Loan, An analysis of the total least squares problem, SIAM J. Numer. Anal. 17(6)(1980)883-893.

[11] J.R. Rice, A theory of condition, SIAM J. Numer. Anal. 3(2)(1966) 287-310.

[12] A.J. Geurts, A contribution to the theory of condition, Numer. Math. 39(1)(1982) 85-96.

[13] S. Wang, H. Yang, On the condition number of equality constrained indefinite least squares problem, arXiv preprint arXiv:1611.05949, 2016.

[14] H. Li, S. Wang, On the partial condition numbers for the indefinite least squares problem, arXiv preprint arXiv:1605.05164, 2016.

[15] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge University Press, New York, 1991

[16] A. Ben-Israel, T.N.E. Greville, Generalized Inverses: Theory and Applications, second ed., Springer, Business Media, 2003.

[17] G. Wang, Y. Wei, S. Qiao, Generalized Inverses: Theory and Computations, Science Press, Beijing, 2004.
[18] H Li, S. Wang, Partial condition number for the equality constrained linear least squares problem, Calcolo (2017). doi:10.1007/s10092-017-0221-8

[19] S.M. Rump, Structured perturbations part I: Normwise distances, SIAM J. Matrix Anal. Appl. 25(1)(2003) 1-30.

[20] S.M. Rump, Structured perturbations part II: Componentwise distances, SIAM J. Matrix Anal. Appl. 25(1)(2003) 31-56.

[21] D.J. Higham, N.J. Higham, Backward error and condition of structured linear systems, SIAM J. Matrix Anal. Appl. 13(1)(1992)162-175.

[22] G.H. Golub, C.F. Van Loan, Matrix Computation, fourth ed., Johns Hopkins University press, 2013.

[23] M.E. Hochstenbach, Probabilistic upper bounds for the matrix two-norm, J. Sci. Comput. 57(3)(2013) 464-476.

[24] S. Wang, H. Yang, H. Li, Condition numbers for the nonlinear matrix equation and their statistical estimation, Linear Algebra Appl. 482(2015) 221-240.

[25] S.W. Gaaf, M.E. Hochstenbach, Probabilistic bounds for the matrix condition number with extended Lanczos bidiagonalization, SIAM J. Sci. Comput. 37(5)(2015)581-601.

[26] H.A. Diao, Y. Wei, P. Xie, Small sample statistical condition estimation for the total least squares problem. Numer. Algor. (2016) 1-21.

[27] C. Kenney, A. Laub, Small-sample statistical condition estimates for general matrix functions. SIAM J. Sci. Comput. 15(1994) 36-61.

[28] Å. Björck, P. Heggernes, P. Matstoms, Methods for large scale total least squares problems. SIAM J. Matrix Anal. Appl. 22(2) (2000) 413-429.

[29] A. Björck, Numerical methods for least squares problems, SIAM, Philadelphia, 1996.