Filomat 36:12 (2022), 4195–4209
https://doi.org/10.2298/FIL2212195Y

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

An Analysis of the Mixed Least Squares-Total Least Squares Problems

Zhanshan Yang[a]

[a] *School of Mathematics and Statistics, Qinghai University for Nationalities, Xining 810007, P.R. China*

Abstract. In this paper, we first get further consideration of the first order perturbation with normwise condition number of the MTLS problem. For easy estimation, we show a lower bound for the normwise condition number which is proved to be optimal. In order to overcome the problems encountered in calculating the normwise condition number, we give an upper bound for computing more effectively and nonstandard and unusual perturbation bounds for the MTLS problem. Both of the two types of the perturbation bounds can enjoy storage and computational advantages. For getting more insight into the sensitivity of the MTLS technique with respect to perturbations in all data, we analyze the corrections applied by MTLS to the data in $Ax \approx b$ to make the set compatible and indicate how closely the data A, b fit the so-called general errors-in-variables model. On how to estimate the conditioning of the MTLS problem more effectively, we propose statistical algorithms by taking advantage of the superiority of small sample statistical condition estimation (SCE) techniques.

1. Introduction

The problem of linear parameter estimation arises in a broad class of scientific disciplines such as signal processing, automatic control, system theory, general engineering, statistics, physics, economics, biology, and medicine. It can be described by a linear equation:

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b,$$

where $a_1, \cdots, a_n$ and b denote the variables and $x = [x_1, x_2, \cdots, x_n]^T \in \mathbb{R}^n$ plays the role of a parameter vector that characterizes the special system. A basic problem of applied mathematics is to determine an estimate of the true but unknown parameters from certain measurements of the variables. This gives rise to an overdetermined set of m linear equations ($m \geq n$):

$$Ax \approx b \text{ for } A \in \mathbb{R}^{m\times n} \text{ and } b \in \mathbb{R}^m, \ m \geq n. \tag{1}$$

In the classical least squares (LS) approach the measurements A of the variables a_i (the left-hand side of (1) are assumed to be free of error and, hence, all errors are confined to the observation vector b. But this assumption is frequently unrealistic: sampling errors, human errors, modelling errors, and instrument

2020 *Mathematics Subject Classification*. Primary 65F05 ; Secondary 15A12, 65F35
Keywords. mixed least squares-total least squares, perturbation analysis, perturbation bound, small sample statistical condition estimation.
Received: 09 February 2021; Revised: 17 August 2021; Accepted: 07 March 2022
Communicated by Predrag Stanimirović
Research supported by the National Natural Science Foundation (11701456), Natural Science Foundation of Qinghai Province (2018-ZJ-717), Foundation Sciences Qinghai Nationalities University(2020XJG11, 2019XJZ10)
Email address: yangzhsh15@lzu.edu.cn (Zhanshan Yang)

errors may imply inaccuracies of the data matrix A. For those cases, TLS has been devised and amounts to fitting a "best" subspace to the measurement points (A_i^T, b_i), $i = 1, \cdots, m$, where A_i^T is the i-th row of A. Much of the literature concerns the classical TLS problem in which all variables are observed with errors, see e.g., [8], [12–14], [19–21] and so on. However in many linear parameter estimation problems, some of the variables a_i in (1) may be observed without error. For instance, in regression analysis, e.g., in curve fitting and intercept models, we often encounter such problems, as well as in system identification and signal processing applications whenever some signals can be observed without error while the other ones are disturbed by zero-mean white noise. This implies that some of the columns of A in (1) are assumed to be known exactly. To maximize the accuracy of the estimated parameters x it is natural to require that the corresponding columns of A be unperturbed since they are known exactly. In order to maintain consistency of the result when solving these problems, the classical TLS formulation can be generalized, the mixed least squares-total least squares (MTLS) problems just as posed in [1, 25]. About the approximate solution of the MTLS problem (2), one can have them by Cho-INV iteration, Rayleigh quotient iteration method, see [1, 28], the generalized TLS Algorithm GTLS [24].

Perturbation analysis is an important research area in numerical analysis. In this paper, we get further consideration of the first order perturbation with normwise condition number of the MTLS problem posed in [17]. It should be noted that computing the matrix cross product A^TA for normwise condition number is a source of rounding errors and is potentially numerical unstable. For easy estimation, we show a lower bound for the normwise condition number which is proved to be optimal. Then, we give nonstandard and unusual perturbation bounds for computing more effectively based Wei's results [14]. The upper bound are easy to compute and do not need to compute matrix cross product. The efficiency of our bounds will be demonstrated by numerical examples.

For getting more insight into the sensitivity of the MTLS technique with respect to perturbations in all data, we analyze the corrections applied by MTLS to the data in $Ax \approx b$ to make the set compatible. We also deduce the assumptions about the underlying pertubation model. Thus indicates how closely the data A, b fit the so-called general errors-in-variables model.

In practice, the efficient estimation for the condition number is difficult. Thus, practical algorithms for approximating the condition numbers are worth studying. We propose statistical algorithms by taking advantage of the superiority of the small sample statistical condition estimation (SCE), that a small number of function evaluations at perturbed arguments suffices to give a highly reliable condition estimate.

Throughout this paper, the following notations are used:

- $R^{m\times n}$ denotes the set of $m \times n$ matrices with real entries.
- I_n stands for the identity matrix with order n , e_i is the i-th canonical vector.
- Single vertical bars around a matrix or vector indicate the componentwise absolute value of a matrix or vector.
- For a matrix $A \in R^{m\times n}$, A^T denotes the transpose of A, $\|A\|_2$ and $\|A\|_F$ denote the spectral norm and the Frobenius norm of A, respectively.
- We define $\mathrm{vec}(A) = [a_1^T, a_2^T, \cdots, a_n^T]^T \in R^{mn}$ and the 'unvec' operator undoes the operation.
- For a vector a, $\mathrm{diag}(a)$ is a diagonal matrix whose diagonals are given as components of a.
- The uniform continuous distribution between a and b is abbreviated $\mathcal{U}(a, b)$.
- $X^\dagger$ the Moore-Penrose inverse of X.
- For any $a, b \in R^n$, we define $\frac{a}{b} = [c_1, c_2, \cdots, c_n]^T$ by

$$c_i = \begin{cases} \frac{a_i}{b_i} & if \ \ b_i \neq 0, \\ 0 & if \ \ a_i = b_i = 0, \\ \infty & otherwise. \end{cases}$$

2. Preliminaries

First of all, we give a brief description of the MTLS problem [1, 22–25] in Frobenius norm (F-norm) stated as:

$$\begin{cases} \min\limits_{E_2,f} \|(E_2, f)\|_F, \\ \text{s.t.} \quad R(b+f) \subseteq R(A_1, A_2+E_2), \end{cases} \tag{2}$$

where $A_1 \in \mathbb{R}^{m\times n_1}$, $A_2 \in \mathbb{R}^{m\times n_2}$, and $n_1 + n_2 = n$. Let (E_2, f) be the minimizer of (2), the MTLS solution set belonging to (E_2, f) is defined by

$$\mathcal{X} = \{x = (x_1^T, x_2^T)^T | A_1x_1 + (A_2 + E_2)x_2 = b + f\}.$$

Obviously, if $n_1 = 0$, the MTLS problem will become an TLS problem, while $n_2 = 0$, it will reduce to an LS problem. We can factorize (A, b) into the QR form to solve MTLS problem:

$$Q^T(A_1, A_2, b) = R = \begin{array}{c} \begin{matrix} n_1 & n_2 & 1 \end{matrix} \\ \begin{pmatrix} R_{11} & R_{1a} & R_{1b} \\ 0 & R_{2a} & R_{2b} \end{pmatrix} \end{array} \begin{array}{c} n_1 \\ m-n_1 \end{array}, \tag{3}$$

then reduce to TLS problem $R_{2a}x_2 \approx R_{2b}$. The vector x_1 can be obtained from $R_{11}x_1 = R_{1b} - R_{1a}x_2$ by back substitution. Let $\sigma_j(A)$ denotes the j-th largest singular value of A. We know that, under the condition

$$\sigma = \sigma_{n_2}(R_{22}) > \sigma_{n_2+1}(R_{22}, R_{2b}) = \tilde{\sigma}, \tag{4}$$

the reduced TLS problem and therefore the MTLS have a unique solution [23]

$$x_{\text{MTLS}} = (A^TA - \tilde{\sigma}^2C)^{-1}A^Tb, \tag{5}$$

where $C = \begin{pmatrix} 0 & 0 \\ 0 & I_{n_2} \end{pmatrix}$.

In [1], it's proved that the MTLS problem is mathematically equivalent to the WTLS problems (6) to (7).

$$\begin{cases} \min\limits_{E,f} \|(E, f)\|_F, \\ \text{s.t.} \quad (A_\varepsilon + E)C_\varepsilon x_\varepsilon = b + f \end{cases}, \tag{6}$$

where

$$A_\varepsilon = AC_\varepsilon^{-1}, \ E = \tilde{E}C_\varepsilon^{-1} \quad \text{with} \quad C_\varepsilon = \begin{pmatrix} \varepsilon I_{n_1} & 0 \\ 0 & I_{n_2} \end{pmatrix}, \tag{7}$$

ε is a small positive number. In practical computations, with a possible choice for ε such that

$$\varepsilon_* = \mu^{\frac{1}{2}} \frac{\|A_1\|_2}{\|\tilde{A}_2\|_2} \kappa(A_1)^{-1},$$

where $\kappa(A_1) = \|A_1\|_2\|A_1^\dagger\|_2$ is the condition number of A_1, $\tilde{A}_2 = (A_2, b)$, μ is the machine precision.

Throughout the paper, let $\hat{U}^TA\hat{V} = diag(\hat{\sigma}_1, \cdots, \hat{\sigma}_n) = \hat{\Sigma}$ be the thin SVD of $A \in R^{m\times n}$, where $\hat{\sigma}_1 \geq \cdots \geq \hat{\sigma}_n$, $\hat{U} \in R^{m\times n}$, $\hat{V} \in R^{n\times n}$, and let $U^T[A, b]V = diag(\sigma_1, \cdots, \sigma_{n+1}) = \Sigma$ be the thin SVD of $[A, b] \in R^{m\times(n+1)}$, where $\sigma_1 \geq \cdots \geq \sigma_{n+1}$. Let $\check{U}^TA_\varepsilon\check{V} = diag(\check{\sigma}_1, \cdots, \check{\sigma}_n) = \check{\Sigma}$ be the thin SVD of $A_\varepsilon \in R^{m\times n}$, where $\check{\sigma}_1 \geq \cdots \geq \check{\sigma}_n$, $\check{U} \in R^{m\times n}$, $\check{V} \in R^{n\times n}$, and let $\bar{U}^T[A_\varepsilon, b]\bar{V} = diag(\bar{\sigma}_1, \cdots, \bar{\sigma}_{n+1}) = \bar{\Sigma}$ be the thin SVD of $[A_\varepsilon, b] \in R^{m\times(n+1)}$, where $\bar{\sigma}_1 \geq \cdots \geq \bar{\sigma}_{n+1}$, $\bar{U}$, $\bar{V}$, and $\bar{\Sigma}$ are partitioned as follows:

$$\bar{U} = \begin{pmatrix} \bar{U}_1, & \bar{u}_{n+1} \end{pmatrix}, \quad \bar{V} = \begin{array}{c} \begin{matrix} n & 1 \end{matrix} \\ \begin{pmatrix} \bar{V}_{11} & \bar{v}_{12} \\ \bar{v}_{21}^T & \bar{v}_{22} \end{pmatrix} \end{array} \begin{array}{c} n \\ 1 \end{array}, \quad \bar{\Sigma} = \begin{pmatrix} \bar{\Sigma}_1 & 0 \\ 0 & \bar{\sigma}_{n+1} \end{pmatrix} \tag{8}$$

We assume the genericity condition:

$$\check{\sigma}_n > \bar{\sigma}_{n+1} \tag{9}$$

to ensure the existence and uniqueness of the WTLS solution x_ε throughout this paper.

From [1], we know that the unique WTLS solution is determined by

$$x_\varepsilon = C_\varepsilon^{-1}(A_\varepsilon^T A_\varepsilon - \bar{\sigma}_{n+1}^2 I)^{-1} A_\varepsilon^T b, \tag{10}$$

and

$$\begin{aligned} \lim_{\varepsilon\to 0^+} x_\varepsilon &= \lim_{\varepsilon\to 0^+} C_\varepsilon^{-1}(A_\varepsilon^T A_\varepsilon - \bar{\sigma}_{n+1}^2 I)^{-1} A_\varepsilon^T b, \\ &= (A^T A - \tilde{\sigma}^2 C)^{-1} A^T b = x_{\text{MTLS}} \end{aligned} \tag{11}$$

under the condition (4) and the assumption

$$\varepsilon^2 \|A_1^\dagger\|_2^2 \|\tilde{A}_2\|_2^2 < \frac{\sigma^2 - \tilde{\sigma}^2}{2\sigma^2}. \tag{12}$$

Throughout this paper, we assume that the conditions in Equations (4) and (12) hold.

Let $\tilde{A}_\varepsilon = A_\varepsilon + \Delta A_\varepsilon$ and $\tilde{b} = b + \delta b$, where ΔA_ε and δb are the perturbations of the input data A_ε and b respectively. Consider the perturbed WTLS problem

$$\min_{\Delta A_\varepsilon, \delta b} \|(\Delta A_\varepsilon, \delta b)\|_F \qquad \text{s.t.} \quad (A_\varepsilon + \Delta A_\varepsilon) C_\varepsilon \tilde{x}_\varepsilon = b + \delta b, \tag{13}$$

and denote the singular values of the matrix $[\tilde{A}_\varepsilon, \tilde{b}]$ by $\tilde{\sigma}_1 \geq \tilde{\sigma}_2 \geq \cdots \geq \tilde{\sigma}_{n+1} \geq 0$. If the norm $\|(\Delta A_\varepsilon, \delta b)\|_F$ of the perturbations is sufficiently small, then the well-known perturbation analysis of singular values ensures that the perturbed WTLS problem above has a unique solution $\tilde{x}_\varepsilon$ and it can be expressed as

$$\tilde{x}_\varepsilon = C_\varepsilon^{-1}(\tilde{A}_\varepsilon^T \tilde{A}_\varepsilon - \tilde{\sigma}_{n+1}^2 I)^{-1} \tilde{A}_\varepsilon^T \tilde{b}.$$

For convenience, let the change in the solution be

$$\Delta x_\varepsilon = \tilde{x}_\varepsilon - x_\varepsilon. \tag{14}$$

Kronecker product turns out to be convenient for deriving the results in this paper. For clarity we briefly state a few of its useful properties here.

Lemma 2.1. *[5] Let* $A, \ B, \ X \in R^{N\times N}$, $D \in R^{M\times N}$, $Y \in R^{N\times K}$, $E \in R^{K\times L}$, $F \in R^{M\times L}$, $P = \sum_{i=1}^{N}\sum_{j=1}^{N} E_{ij} \otimes E_{ij}^T$, *where* $E_{ij} = (e_{kl})$, $k, \ l = 1, \cdots, N$, *is a* $N \times N$ *matrix, and*

$$e_{kl} = \begin{cases} 1 & k = i \ \text{and} \ l = j, \\ 0 & k \neq i \ \text{or} \ l \neq j. \end{cases}$$

Then

(1) $vec(X)^T = Pvec(X)$;

(2) $(B \otimes A) = P^T(A \otimes B)P$;

(3) $DYE = F \Longleftrightarrow (E^T \otimes D)vec(Y) = vec(F)$.

3. First-order perturbation for the MTLS problem

We know that, by (14) using the definition of the normwise condition number of the WTLS problem and the expression of $f'(A_\varepsilon, b).(\Delta A_\varepsilon, \delta b)$, the upper bound for $\frac{\|\Delta x\|_2}{\|x\|_2}$ and the explicit formula for the MTLS condition number can be obtained with some Δx corresponding to Δx_ε, which profits from the connection between the MTLS problem and the WTLS problem.

Theorem 3.1. *[17] We consider the MTLS problem and assume that the genericity assumption holds. Setting $B_\lambda = A^TA - \tilde{\sigma}^2C$, then the condition number of x_{MTLS} of the MTLS solution is given by*

$$\kappa_{MTLS}(A,b) = \|M\|_2^{\frac{1}{2}}, \tag{15}$$

and the upper bound for $\frac{\|\Delta x\|_2}{\|x\|_2}$ is expressed by

$$\frac{\|\Delta x\|_2}{\|x\|_2} \leq \|M\|_2^{\frac{1}{2}} \frac{\|(A,b)\|_F}{\|x\|_2} \frac{\|\Delta A, \delta b)\|_F}{\|(A,b)\|_F} \equiv B_{BY}, \tag{16}$$

where M is the $n \times n$ matrix

$$M = \gamma B_\lambda^{-1}A^TAB_\lambda^{-1} + \tilde{\sigma}^2\bar{\gamma}B_\lambda^{-1}(I - 2\frac{Cxx^TC}{\bar{\gamma}})B_\lambda^{-1}, \tag{17}$$

$\bar{\gamma} = 1 + \|Cx\|_2^2$, $\gamma = 1 + \|x\|_2^2$, x is the exact solution of the MTLS problem, r is the MTLS residual.

In addition, we have the relative condition number of the MTLS problem

$$\kappa_{\text{MTLS}}^{\text{rel}}(A,b) = \|M\|_2^{1/2} \frac{\|(A,b)\|_F}{\|x\|_2}. \tag{18}$$

In many applications, an upper or a lower bound would be sufficient to estimate the normwise condition number. We next present a lower bound for $\kappa_{\text{MTLS}}(A,b)$, which only involves the singular values of matrix K. Let first see the theorem [28] about perturbation of the eigenvalues of a matrix.

Theorem 3.2. *Let $A, B = A + E \in \mathbb{C}^{n\times n}$, A, B, and E are all Hermitian matrices, their eigenvalues are $\lambda_1 \geq \cdots \geq \lambda_n$, $\mu_1 \geq \cdots \geq \mu_n$ and $\epsilon_1 \geq \cdots \geq \epsilon_n$ respectively. Then*

$$|\mu_i - \lambda_i| \leq \|B - A\|_2, \quad i = 1, 2, \cdots, n.$$

Theorem 3.3. *We have*

$$\|M\|_2 \geq |\sigma_1^2 - \mu_1^2|,$$

where σ_1 be largest singular value of K, $K = B_\lambda^{-1}\begin{pmatrix} \sqrt{\gamma}A^T & \|r\|_2 I \end{pmatrix}$, $\mu_1 = \sqrt{y^Ty}$ and $y = \sqrt{2}\tilde{\sigma}B_\lambda^{-1}Cx$.

Proof. From the proof of Theorem 3.1, we have

$$\begin{aligned} M &= \gamma B_\lambda^{-1}A^TAB_\lambda^{-1} + \|r\|_2^2B_\lambda^{-2} - 2\tilde{\sigma}^2B_\lambda^{-1}Cxx^TCB_\lambda^{-1} \\ &= B_\lambda^{-1}\begin{pmatrix} \sqrt{\gamma}A^T & \|r\|_2 I \end{pmatrix}\begin{pmatrix} \sqrt{\gamma}A \\ \|r\|_2 I \end{pmatrix}B_\lambda^{-1} - \left(\sqrt{2}\tilde{\sigma}B_\lambda^{-1}Cx\right)\left(\sqrt{2}\tilde{\sigma}B_\lambda^{-1}Cx\right)^T \\ &= KK^T - yy^T, \end{aligned}$$

where $K = B_\lambda^{-1}\begin{pmatrix} \sqrt{\gamma}A^T & \|r\|_2 I \end{pmatrix}$, $y = \sqrt{2}\tilde{\sigma}B_\lambda^{-1}Cx$. And we know

$$\begin{aligned} y^Ty &= \left(\sqrt{2}\tilde{\sigma}B_\lambda^{-1}Cx\right)^T\left(\sqrt{2}\tilde{\sigma}B_\lambda^{-1}Cx\right) \\ &= 2\tilde{\sigma}^2x^T(R_{2a}^TR_{2a} - \tilde{\sigma}^2I)^{-1}\begin{pmatrix} 0 & 0 \\ -(R_{11}^{-1}R_{1a})^T & I \end{pmatrix}\begin{pmatrix} 0 & -R_{11}^{-1}R_{1a} \\ 0 & I \end{pmatrix}(R_{2a}^TR_{2a} - \tilde{\sigma}^2I)^{-1}x \\ &= 2\tilde{\sigma}^2x_2^T(R_{2a}^TR_{2a} - \tilde{\sigma}^2I)^{-1}(I + (R_{11}^{-1}R_{1a})^TR_{11}^{-1}R_{1a})(R_{2a}^TR_{2a} - \tilde{\sigma}^2I)^{-1}x_2 = \mu_1^2. \end{aligned}$$

Let σ_1 be the largest singular value of K, then by Theorem 3.2, we get the bound

$$\|M\|_2 \geq |\sigma_1^2 - \mu_1^2|.$$

□

We find that we only need to calculate the product of matrices and vectors, since B_λ^{-1} and $(R_{2a}^T R_{2a} - \tilde{\sigma}^2 I)^{-1}$ can be the intermediate result when the MTLS problem is solved. So the bound enjoys storage and computational advantages. Moreover, this bound can be attained when y is the left singular vector of K, see Theorem 2.1 in [27].

In addition, we find that computing the matrix cross product $A^T A$ for normwise condition number is a source of rounding errors and is potentially numerical unstable. In order to overcome these difficulties, we present a perturbation bound for computing more effectively without using the already well-known condition number and nonstandard and unusual perturbation bounds which will be shown in Section 4.

4. Nonstandard and unusual perturbation bounds

In this section we derive another perturbation estimates, based on Wei's results [14], for the MTLS problem.

Wei [14] derive the perturbation bounds for the TLS solutions, with or without minimal length. We can use the result in WTLS problems and get the following theorem:

Theorem 4.1. *Consider the WTLS problem (6) to (7). Assume that the genericity condition (9) holds. Partition $\bar{V}$ as in (8) and let $A'_\varepsilon \in R^{m\times n}$, $b' \in R^m$, and $(A'_\varepsilon, b') = (A_\varepsilon, b) + (\Delta A_\varepsilon, \delta b)$ with $\|(\Delta A_\varepsilon, \delta b)\| = \eta_\varepsilon \leq \frac{1}{6}(\check{\sigma}_n - \bar{\sigma}_{n+1})$, and let A'_ε, (A'_ε, b') be $\check{U}'^T A'_\varepsilon \check{V}' = diag(\check{\sigma}'_1, \cdots, \check{\sigma}'_n) = \check{\Sigma}'$ be the thin SVD of $A'_\varepsilon \in R^{m\times n}$, where $\check{\sigma}'_1 \geq \cdots \geq \check{\sigma}'_n$, $\check{U}' \in R^{m\times n}$, $\check{V}' \in R^{m\times n}$, and let $\bar{U}'^T[A'_\varepsilon, b']\bar{V}' = diag(\bar{\sigma}'_1, \cdots, \bar{\sigma}'_{n+1}) = \bar{\Sigma}'$ be the thin SVD of $[A'_\varepsilon, b'] \in R^{m\times(n+1)}$, Partition $\bar{V}'$ conformally with $\bar{V}$ as in (8), and replace $\bar{V}_{ij}$ by $\bar{V}'_{ij}$. for $i, j = 1, 2$. Define $C_\varepsilon x'_\varepsilon = ((\bar{V}'_{11})^T))^\dagger(\bar{v}'_{21})^T$. Then one has the following estimates:*

$$\begin{aligned} \|C_\varepsilon x_\varepsilon - C_\varepsilon x'_\varepsilon\| &\leq \frac{\eta_\varepsilon + \bar{\sigma}_{n+1}}{\check{\sigma}_n - \bar{\sigma}_{n+1}}(3 + 5\|C_\varepsilon x_\varepsilon\|) \\ &\leq 6\frac{\eta_\varepsilon + \bar{\sigma}_{n+1}}{\check{\sigma}_n - \bar{\sigma}_{n+1}}(\sqrt{1 + \|C_\varepsilon x_\varepsilon\|^2}). \end{aligned} \tag{19}$$

Then one can attain the following theorem for the MTLS problem by (11) with the same way

Theorem 4.2. *Consider the MTLS problem (2). Assume that the genericity condition (4) holds. Let $A' \in R^{m\times n}$, $b' \in R^m$, and $(A', b') = (A, b) + (\Delta A, \delta b)$ with $\|(\Delta A, \delta b)\| = \eta \leq \frac{1}{6}(\sigma - \tilde{\sigma})$, and let R'_{22}, (R'_{22}, R'_{2b}) be $\hat{U}'^T R'_{22} \hat{V}' = diag(\hat{\sigma}'_1, \cdots, \hat{\sigma}'_{n_2}) = \hat{\Sigma}'$ be the thin SVD of $R'_{22} \in R^{(m-n_1)\times n_2}$, where $\hat{\sigma}'_1 \geq \cdots \geq \hat{\sigma}'_{n_2}$, $\hat{U}' \in R^{(m-n_1)\times n_2}$, $\hat{V}' \in R^{n_2\times n_2}$, and let $U'^T[R'_{22}, R'_{2b}]V' = diag(\sigma'_1, \cdots, \sigma'_{n_2+1}) = \Sigma'$ be the thin SVD of $[R'_{22}, R'_{2b}] \in R^{(m-n_1)\times(n_2+1)}$, Partition V' as in (8). Define $x'_2 = ((V'_{11})^T))^\dagger(v'_{21})^T$. Let $x'_2 = x_2 + \Delta x_2$, then one has the following estimates:*

$$\begin{aligned} \|x_2 - x'_2\| &\leq \frac{\eta + \tilde{\sigma}}{\sigma - \tilde{\sigma}}(3 + 5\|x_2\|) \\ &\leq 6\frac{\eta + \tilde{\sigma}}{\sigma - \tilde{\sigma}}(\sqrt{1 + \|x_2\|^2}), \end{aligned} \tag{20}$$

where $\sigma = \sigma_{n_2}(R_{22})$, $\tilde{\sigma} = \sigma_{n_2+1}(R_{22}, R_{2b})$.

The proof is easy and we omit it.

We turn now to the problem of obtaining perturbation bounds of Q and R, which is needed for obtaining the perturbation bound of $\|x_1 - x'_1\|$, for the QR factorization of the matrix A. Stewart [26] show us that

Theorem 4.3. *Let $A = QR$, where A has rank n and $Q^TQ = I$. Let $\Delta A \in R^{m\times n}$ satisfy $\|A\|\|\Delta A\| < \frac{1}{2}$. Then there are matrices $\Delta Q \in R^{m\times n}$ and $\Delta R \in R^{n\times n}$ such that $A + \Delta A = (Q + \Delta Q)(R + \Delta R)$, $(Q + \Delta Q)^T(Q + \Delta Q) = I$. Let $\tau = n\|R^{-1}\|\Big[1 + \frac{1}{\sqrt{2}}\kappa(R)\Big]$, $\varsigma = n(2 + \sqrt{2})\kappa(R)$. Define the operator $\mathbf{T}$ that maps the space of upper triangular matrices into the space of symmetric, let $\mathbf{F}_1 = \mathbf{T}^{-1}[R^T(Q^T\Delta A) + (\Delta A^TQ)R]$. If $\tau\|\mathbf{F}_1\| < \frac{1}{4}$, then there is a unique solution of*

$$\mathbf{T}\Delta R = R^T(Q^T\Delta A) + ((\Delta A)^TQ)R + (\Delta A)^T(\Delta A) - (\Delta R)^T(\Delta R) \tag{21}$$

that satisfies

$$\|\Delta R\| < 2\|F_1\| \le 2\varsigma\|\Delta A\|. \tag{22}$$

Moreover $A + \Delta A = (Q + \Delta Q)(R + \Delta R)$, where $Q + \Delta Q$ has orthonormal columns

$$\|\Delta Q\| < \frac{3\kappa(A)\frac{\|\Delta A\|}{\|A\|}}{1 - 2\kappa(A)\frac{\|\Delta A\|}{\|A\|}}, \tag{23}$$

and

$$\frac{\|\Delta R\|}{\|A\|} \le \frac{\|\Delta A\|}{\|A\|} + \|\Delta Q\|\Big(1 + \frac{\|\Delta A\|}{\|A\|}\Big). \tag{24}$$

At this point, it is natural that the perturbation domain of $\|x_1 - x_1'\|$ is coming.

Theorem 4.4. *Consider the MTLS problem (2). Assume that the conditions in Theorems 4.2 and 4.3 hold. Let $x_1' = x_1 + \Delta x_1$, then one has the following estimates:*

$$\begin{aligned}\|x_1 - x_1'\| &\le \frac{\acute{o}(\eta + \tilde{o})}{\sigma - \tilde{\sigma}}(3 + 5\|x_2\|) + \|R_{11}^{-1}\|(1 + \|x\|^2)2\varsigma\eta + O(\eta^2)\\ &\le \frac{6\acute{o}(\eta + \tilde{o})}{\sigma - \tilde{\sigma}}(\sqrt{1 + \|x_2\|^2}) + \|R_{11}^{-1}\|(1 + \|x\|^2)2\varsigma\eta + O(\eta^2),\end{aligned} \tag{25}$$

or

$$\begin{aligned}\|x_1 - x_1'\| &\le \frac{\acute{o}(\eta + \tilde{o})}{\sigma - \tilde{\sigma}}(3 + 5\|x_2\|) + \|R_{11}^{-1}\|\Big(\eta + \frac{3\kappa(A)\frac{\eta}{\|A\|}}{1 - 2\kappa(A)\frac{\eta}{\|A\|}}(\|A\| + \eta)\Big) + O(\eta^2)\\ &\le \frac{6\acute{o}(\eta + \tilde{o})}{\sigma - \tilde{\sigma}}(\sqrt{1 + \|x_2\|^2}) + \|R_{11}^{-1}\|\Big(\eta + \frac{3\kappa(A)\frac{\eta}{\|A\|}}{1 - 2\kappa(A)\frac{\eta}{\|A\|}}(\|A\| + \eta)\Big) + O(\eta^2),\end{aligned} \tag{26}$$

where $\acute{o}$ is the maximum singular value of the matrix $R_{11}^{-1}R_{12}$

Proof. From equation $R_{11}x_1 = R_{1b} - R_{1a}x_2$ and genericity condition (4), we have

$$x_1 = R_{11}^{-1}(R_{1b} - R_{1a}x_2)$$

and

$$x_1' = x_1 + \Delta x_1 = (R_{11} + \Delta R_{11})^{-1}((R_{1b} + \Delta R_{1b}) - (R_{1a} + \Delta R_{1a})(x_2 + \Delta x_2)).$$

Only retaining the first-order terms gives

$$x_1' \quad = \quad x_1 + R_{11}^{-1}(\Delta R_{1b} - R_{1a}\Delta x_2 - \Delta R_{1a}x_2) - R_{11}^{-1}\Delta R_{11}R_{11}^{-1}(R_{1b} - R_{1a}x_2) + O(\|(\Delta A, \delta b)\|^2),$$

in which

$$\begin{aligned}&R_{11}^{-1}(\Delta R_{1b} - R_{1a}\Delta x_2 - \Delta R_{1a}x_2) - R_{11}^{-1}\Delta R_{11}R_{11}^{-1}(R_{1b} - R_{1a}x_2)\\ &= -R_{11}^{-1}R_{1a}\Delta x_2 + R_{11}^{-1}\Big[\Delta R_{1b} - \Delta R_{1a}x_2 - \Delta R_{11}R_{11}^{-1}(R_{1b} - R_{1a}x_2)\Big]\\ &= -R_{11}^{-1}R_{1a}\Delta x_2 + R_{11}^{-1}(\Delta R_{1b} - \Delta R_{1a}x_2 - \Delta R_{11}x_1)\\ &= -R_{11}^{-1}R_{1a}\Delta x_2 + R_{11}^{-1}\begin{pmatrix} -x_1^T\otimes I & -x_2^T\otimes I & I\end{pmatrix}\begin{pmatrix} vec(\Delta R_{11})\\ vec(\Delta R_{1a})\\ vec(\Delta R_{1b})\end{pmatrix}.\end{aligned}$$

Hence,

$$\begin{aligned}\|x_1 - x_1'\| &\leq \|R_{11}^{-1}R_{1a}\|\|\Delta x_2\| + \|R_{11}^{-1}\|(1+\|x\|^2)\|\Delta R\|_F + O(\|(\Delta A, \delta b)\|^2)\\ &\leq \|R_{11}^{-1}R_{1a}\|\frac{\eta+\tilde{\sigma}}{\sigma-\tilde{\sigma}}(3+5\|x_2\|) + \|R_{11}^{-1}\|(1+\|x\|^2)2\varsigma\eta + O(\eta^2)\\ &\leq \|R_{11}^{-1}R_{1a}\|\frac{6(\eta+\tilde{\sigma})}{\sigma-\tilde{\sigma}}(\sqrt{1+\|x_2\|^2}) + \|R_{11}^{-1}\|(1+\|x\|^2)2\varsigma\eta + O(\eta^2),\end{aligned}$$

where we use $\|vec(A)\| = \|A\|_F$ and $\|A\|^2 = \|AA^T\|$.

Also there is another bound can be found by (23) and (24). Because we have

$$\|\Delta R\| \leq \eta + \frac{3\kappa(A)\frac{\eta}{\|A\|}}{1-2\kappa(A)\frac{\eta}{\|A\|}}(\|A\|+\eta),$$

so there comes the result (26). □

It is not difficult to find that (25) and (26) can enjoy storage and computational advantages, and a good estimation for the perturbation of the MTLS solution is provided by it with this point being well illustrated by examples.

5. General errors-in-variables model

Because we have already established the formulas for the MTLS, the work is rather straightforward. Comparing (5) with the LS solution shows that $\tilde{\sigma}$ and C determines the differences between both solutions, in a sense, we could say that $\tilde{\sigma}$ measures the difference between both solutions, as well as the degree of incompatibility of the set $AX \approx b$ and thus indicates how closely the data A, b fit the so-called general errors-in-variables model, which is exactly expressed below

Proposition 5.1. *Consider the linear system (2), let $x = (x_1^T, x_2^T)^T$ be any n-dimensional column vector with $x_2 \in R^{n_2}$. Then*

$$M(x) = (\Delta A_2, \delta b) = (b - Ax)(x_2^T x_2 + 1)^{-1}(x_2^T, I)$$

has the minimal Frobenius norm which makes

$$(A_1, A_2 + \Delta A_2)x = b - \delta b$$

consistent.

Proof. For any fixed x, let the matrix $(A_1, A_2 + \Delta A_2)$ be such that

$$(A_1, A_2 + \Delta A_2)x = b - \delta b$$

is consistent. Then

$$\begin{aligned}A_1x_1 + A_2x_2 + \Delta A_2x_2 &= b - \delta b\\ \Delta A_2x_2 - \delta b &= b - (A_1x_1 + A_2x_2)\\ \begin{pmatrix}\Delta A_2 & \delta b\end{pmatrix}\begin{pmatrix}x_2\\ -1\end{pmatrix} &= b - Ax.\end{aligned}$$

Among all those $(\Delta A_2, \delta b)$s',

$$\begin{pmatrix}\Delta A_2 & \delta b\end{pmatrix} = (b - Ax)(x_2^T x_2 + 1)^{-1}(x_2^T, I) \tag{27}$$

has the minimal Frobenius norm. □

This implies that the MTLS residual is orthogonal to the MTLS approximate subspace $R([\hat{A}, \hat{b}])$, and the length of the MTLS residual follows immediately

Proposition 5.2. *Let x be a solution of the MTLS problem (2) with corresponding correction matrix $(\Delta A_2, \delta b)$, and r be the MTLS residual, then*

$$\|r\|_2 = \tilde{\sigma}\sqrt{1 + \|x_2\|_2^2},$$

$$\Delta A_2 = \frac{r x_2^T}{1 + \|x_2\|_2^2},$$

$$\delta b = -\frac{r}{1 + \|x_2\|_2^2},$$

and

$$\|\Delta A_2\|_F^2 = \frac{\|(\Delta A_2, \delta b)\|_F^2 \|x_2\|_2^2}{1 + \|x_2\|_2^2} = \tilde{\sigma}^2 \frac{\|x_2\|_2^2}{1 + \|x_2\|_2^2},$$

$$\|\delta b\|_2^2 = \frac{\|(\Delta A_2, \delta b)\|_F^2}{1 + \|x_2\|_2^2} = \tilde{\sigma}^2 \frac{1}{1 + \|x_2\|_2^2}.$$

Proof. By Liu [1], the MTLS problem (2) is equivalent to the following optimization problem

$$\min_x \frac{\|b - Ax\|_2^2}{1 + x^T C x} = \min_x \frac{\|b - Ax\|_2^2}{1 + x_2^T x_2} = \tilde{\sigma}^2,$$

thereby

$$\|r\|_2 = \tilde{\sigma}\sqrt{1 + \|x_2\|_2^2}.$$

(27) tells us that

$$\Delta A_2 = \frac{r x_2^T}{1 + \|x_2\|_2^2},$$

$$\delta b = -\frac{r}{1 + \|x_2\|_2^2},$$

and

$$\Delta A_2^T \Delta A_2 = \tilde{\sigma}^2 \frac{x_2 x_2^T}{x_2^T x_2 + 1},$$

$$\delta b^T \delta b = \tilde{\sigma}^2 \frac{1}{x_2^T x_2 + 1},$$

then we have

$$\|(\Delta A_2, \delta b)\|_F^2 = \tilde{\sigma}^2,$$

and using the above three relationships reduce to

$$\|\Delta A_2\|_F^2 = \frac{\|(\Delta A_2, \delta b)\|_F^2 \|x_2\|_2^2}{1 + \|x_2\|_2^2},$$

$$\|\delta b\|_2^2 = \frac{\|(\Delta A_2, \delta b)\|_F^2}{1 + \|x_2\|_2^2}.$$

□

In this section, we find that the parameters ($\tilde{\sigma}$ and C) are useful tools for getting more insight into the sensitivity of both techniques with respect to perturbations.

6. Small sample statistical condition estimation

Although the expressions of the condition numbers presented are explicit, they involve the solution and their computation is intensive when the problem size is large. Thus, practical algorithms for approximating the condition numbers are worth studying. We propose statistical algorithms by taking advantage of the superiority of the small sample statistical condition estimation(SCE) techniques.

We apply SCE that a small number of function evaluations at perturbed arguments suffices to give a highly reliable condition estimate. Based on SCE method, we present a practical method for estimating the condition numbers for the MTLS problem.

Given a differentiable function $f : R^p \to R$, we are interested in the sensitivity at some input vector x. Let z be a unit vector and δ be a small positive number. For the perturbation z of x, the gradient of f at $x \in R^p$ is the row vector $\nabla f(x) = (\partial f(x)/\partial x_1, \partial f(x)/\partial x_2, \cdots,$
$\partial f(x)/\partial x_p)$, and the Taylor expansion of f has the form

$$f(x + \delta z) = f(x) + \delta \nabla f(x) z + O(\delta^2).$$

It is easy to see that up to first-order in δ, we have

$$|f(x + \delta z) - f(x)| \approx \delta \nabla f(x) z,$$

then the local sensitivity can be measured by $\|\nabla f(x)\|_2$. We can see that the norm of the gradient can measure the local sensitivity of f appropriately. If z is selected uniformly and randomly from the unit sphere S_{p-1} in R^p, which is denoted by $z \in \mathcal{U}(S_{p-1})$. Then from [9], the expected value of the condition estimator $\xi = \frac{|\nabla f(x) z|}{\omega_p}$ satisfies that

$$E(\xi) = \|\nabla f(x)\|_2,$$

where ω_p is the Wallis factor which only depends on p. For $\theta > 1$ we have

$$Prob\Big(\frac{\|\nabla f(x)\|_2}{\theta} \le \xi \le \theta \|\nabla f(x)\|_2\Big) \ge 1 - \frac{2}{\pi\theta} + O(\frac{1}{\theta^2}).$$

Thus ξ is a linear or first-order condition estimate in the sense that the chance of a catastrophically low or high estimate is inversely proportional to the size of the error. In practice, the Wallis factor can be approximated accurately [9] by

$$\omega_p \approx \sqrt{\frac{2}{\pi(p - \frac{1}{2})}}. \tag{28}$$

While this is good, there are some situations in which we need more reliability. One way to achieve this is to use more function evaluations to get different values $\xi^{(1)}, \xi^{(2)}, \cdots, \xi^{(m)}$ corresponding to independently randomly generated vectors $z^{(1)}, z^{(2)}, \cdots, z^{(m)}$ in S_{n-1} and then to take the average

$$\xi(m) \equiv \frac{\xi^{(1)} + \xi^{(2)} + \cdots + \xi^{(m)}}{m}.$$

This is the so called "averaged small-sample statistical method" [9] and it can be shown that for $\theta > 1$,

$$Prob\Big(\frac{\|\nabla f(x)\|_2}{\theta} \le \xi(m) \le \theta \|\nabla f(x)\|_2\Big) \ge 1 - \frac{1}{m!}\Big(\frac{2m}{\pi\theta}\Big)^m + O(\frac{1}{\theta^{m+1}}),$$

with asymptotic equality as $\theta \to +\infty$ or $m \to +\infty$. Thus $\xi(m)$ is an mth-order condition estimator. In condition number estimation, usually we are interested in finding an estimate that is accurate to a factor

of 10 ($\theta = 10$). We can use multiple samples of z, denoted z_j, to increase the accuracy of the condition estimator. From [9], we know that

$$E\Big(\sqrt{|\nabla f(x)z_1|^2 + |\nabla f(x)z_2|^2 + \cdots + |\nabla f(x)z_1|^k}\Big) = \frac{\omega_p}{\omega_k}\|\nabla f(x)\|_2.$$

Therefore, we can define the subspace condition estimator as

$$\nu(k) = \frac{\omega_k}{\omega_p}\sqrt{|\nabla f(x)z_1|^2 + |\nabla f(x)z_2|^2 + \cdots + |\nabla f(x)z_k|^2},$$

where $(z_1, z_2, \cdots, z_k)$ is orthonormalized after $z_1, z_2, \cdots, z_k$ are selected uniformly and randomly from $\mathcal{U}(S_{p-1})$. As shown in [9], these condition estimators give better results than the averaged statistical estimators and are analytically very tractable. Usually, at most two or three samples are sufficient for high accuracy. As an illustration, for $k = 3$, the estimator $\nu(3)$ has probability 0.9989 of being within a relative factor of 10 of the true condition number $\|\nabla f(x)\|_2$. These estimates are generally very accurate for $\theta > 10$.

These results can be conveniently extended to vector-valued or matrix-valued functions through viewing f as a map from $\mathbb{R}^s$ to $\mathbb{R}^t$ by means of the operations vec and unvec to transform between matrices and vectors, where each of the t entries of f is a scalar-valued function. The unvec operation is defined as $A = \text{unvec}(v)$ which sets the entries of A to $a_{ij} = v_{i+(j-1)n}$ for $v = (v_1, v_2, \cdots, v_{mn}) \in \mathbb{R}^{1\times mn}$.

Here, we use SCE to give an algorithm for computing the normwise condition number. Denote κ_j the condition number of the function $z_j^T x$, where $z_j's$ are random orthogonal vectors selected uniformly and randomly from the unit sphere in n dimensions. From [3], we know that κ_j can be computed by $\|M_j\|_2^{\frac{1}{2}}$

$$M_j = z_j^T\Big(\gamma B_\lambda^{-1}A^TAB_\lambda^{-1} + \tilde{\sigma}^2\bar{\gamma}B_\lambda^{-1}(I - 2\frac{Cxx^TC}{\bar{\gamma}})B_\lambda^{-1}\Big)z_j, \tag{29}$$

taking the technique and notations adopted in [10], we see that

$$\tilde{\kappa}_{\text{MTLS}} = \frac{\omega_q}{\omega_n}\sqrt{\sum_{j=1}^{q} M_j^2} \tag{30}$$

is an estimate for $\kappa_{\text{MTLS}}(A, b)$.

We use the results above to give the SCE-based method for estimating the condition of the solution to the MTLS problem under the genericity condition (4). Inputs of the method are the matrix $A \in R^{m\times n}$ and the vector $b \in R^m$, and the output is the statistical estimate for the absolute condition number. In Algorithm SCE-NCE the integer $q \geq 1$ refers to the number of SCE samples.

Algorithm 1 SCE-NCE: Subspace condition estimate for $\kappa_{\text{MTLS}}(A, b)$ of MTLS solution

1: Generate q vectors $z_1, z_2, \cdots, z_q \in R^n$ with entries in the uniform continuous distribution $\mathcal{U}(0, 1)$. Orthonormalize the vectors using a QR factorization.
2: For $j = 1, 2, \cdots, q$, calculate M_j by (29).
3: compute the absolute condition number $\tilde{\kappa}_{\text{MTLS}}$ by (30) for $\kappa_{\text{MTLS}}(A, b)$.

The sensitivity of componentwise perturbations can be measured by the SCE method similarly. It often leads to a more realistic indication of the accuracy of a computed solution than that from the normwise condition number. Just as shown in [3], the exact value of the condition number for the j-th component of x is computed by

$$\kappa_j^c = \|e_j^T\Big(\gamma B_\lambda^{-1}A^TAB_\lambda^{-1} + \tilde{\sigma}^2\bar{\gamma}B_\lambda^{-1}(I - 2\frac{Cxx^TC}{\bar{\gamma}})B_\lambda^{-1}\Big)e_j\|_2^{1/2}, \tag{31}$$

This algorithm is based on the original idea of SCE [9], which means that we estimate the Fréchet derivative. To see the process more clearly, the readers need to notice that $y_j's$ are just the Fréchet derivative estimates

for the components of the MTLS solution. After generating and normalizing the random elements in the first step, these random elements are overwritten by the componentwise product of (A, b) and these elements. In the main step we need to calculate y_j as

$$y_j = B_\lambda^{-1}(A^T + 2\frac{Cxr^T}{\bar{\gamma}})(\delta b_j - \Delta A_j x) + B_\lambda^{-1}\Delta A_j^T r. \tag{32}$$

At last, the condition vector containing κ_j^c's are computed by

Algorithm 2 SCE-CCE: Subspace condition estimate for Componentwise condition estimate

1: Generate matrices $(\Delta A^{(1)}, \delta b^{(1)}), (\Delta A^{(2)}, \delta b^{(2)}), \cdots, (\Delta A^{(k)}, \delta b^{(k)})$with entries in $\mathcal{N}(0, 1)$. Orthonormalize the following matrix

$$\begin{pmatrix} \Delta\vec{A}^{(1)} & \Delta\vec{A}^{(2)} & \cdots & \Delta\vec{A}^{(k)} \\ \delta b^{(1)} & \delta b^{(2)} & \cdots & \delta b^{(k)} \end{pmatrix}$$

to obtain $(q_1, q_2, \cdots, q_k)$ via modified Gram-Schmidt orthogonalization process. Each q_i can be converted into the desired matrices $(\Delta A^{(i)}, \delta b^{(i)})$ with the unvec operation.
2: Let $p = m(n + 1)$. Approximate ω_p and ω_k by using (28).
3: For $j = 1, 2, \cdots, k$, calculate y_j by (32).
4: compute the absolute condition vector $\bar{\kappa}_{abs}$ by (33).

$$\bar{\kappa}_{abs} = \frac{\omega_k}{\omega_p}\sqrt{|y_1|^2 + |y_2|^2 + \cdots + |y_k|^2}, \tag{33}$$

where the square root and power operation are performed componentwise.

7. Numerical tests

In this section we give numerical examples to check the perturbation bounds, the statistical condition estimates. The following numerical tests are performed via Matlab R2015a with machine precision $\mu = 2.22e-16$ in a laptop with Intel Core (TM)2 Duo CPU by using double precision.

Example 7.1. *Take m = 100, n = 60, n1 = 30, n2 = 30. Choose 0 as the rand seed and use command rand in Matlab to generate a random $m \times n$ matrix A with a uniform distribution on the interval (0,1).*

(1) Choose 1 as the rand seed and generate a random vector b in Matlab. Then $\sigma^2 = 9.88e-1$, $\tilde{\sigma}^2 = 8.38e-1$, $\tilde{\sigma}^2/\tilde{\sigma}_-^2 = 0.85$, where $\tilde{\sigma}_-^2$ is the second smallest singular value of (R_{22}, R_{2b}). The MTLS problem is well conditioned .

(2) Let $b = 1_m$ be an all-1 vector. Then $\sigma^2 = 9.88e-1$, $\tilde{\sigma}^2 = 1.89e-1$, $\tilde{\sigma}^2/\tilde{\sigma}_-^2 = 0.19$. Therefore, the MTLS problem is well conditioned.

b	ε	$\frac{\|\tilde{x}-x\|_2}{\|x\|_2}$	(16)	(20), (25)	$\kappa_{\mathrm{MTLS}}^{rel}(A,b)\frac{\|(\Delta A,\delta b)\|_F}{\|(A,b)\|_F}$	$\frac{\|\tilde{\rho}-\rho\|_2}{\|\rho\|_2}$
rand(m,1)	1e-6	2.609047e-09	9.726393e-8	4.726393e-8	7.412667e-7	1.0203e-09
ones(m,1)	1e-6	4.1806e-11	4.34547e-9	1.34547e-9	3.05349e-8	2.3019e-11

Table 1: Comparisons of forward error and upper bounds for a perturbed MTLS problem and the relative condition number of the MTLS problem

In Table 1, we compare the exact relative error about MTLS solution with the upper bounds (16), (20) with (25) and the above bounds $\kappa_{\mathrm{MTLS}}^{rel}(A,b)\frac{\|(\Delta A,\delta b)\|_F}{\|(A,b)\|_F}$ derived from (18). In actual calculation, we omit the high order term in (25). In addition, we give the exact relative error about smallest sinular value of (A_ε, b).

Example 7.2. *In this example [16] we consider the MTLS problem $Ax \approx b$, where (A, b) is defined by*

$$(A, b) = Y \begin{pmatrix} D \\ 0 \end{pmatrix} Z^T \in R^{m\times(n+1)},$$

where $Y = I_m - 2yy^T$, $y \in R^m$, $Z = I_{n+1} - 2zz^T$, $z \in R^{n+1}$ are random unit vectors, $D = diag(n, n-1, \cdots, 1, 1-\varepsilon_p)$ for a given parameter ε_p.

Throughout this section, we take two samples and the average of the ratios are obtained by 1000 random tests.

We compare the statistical result obtained via Algorithm SCE-NCE with the exact condition number given in (17). Figure 1 shows the performance of SCE on the case where $m = 200$, $n_1 = 40$, $n_2 = 40$. It shows the accuracy of our estimates, and we can find that if $\varepsilon_p = 9.99976032e-1$, *ratio* $= 0.97453$, while $\varepsilon_p = 9.99952397e-5$, *ratio* $= 1.03608$, which confirms the accuracy of the statistical results.

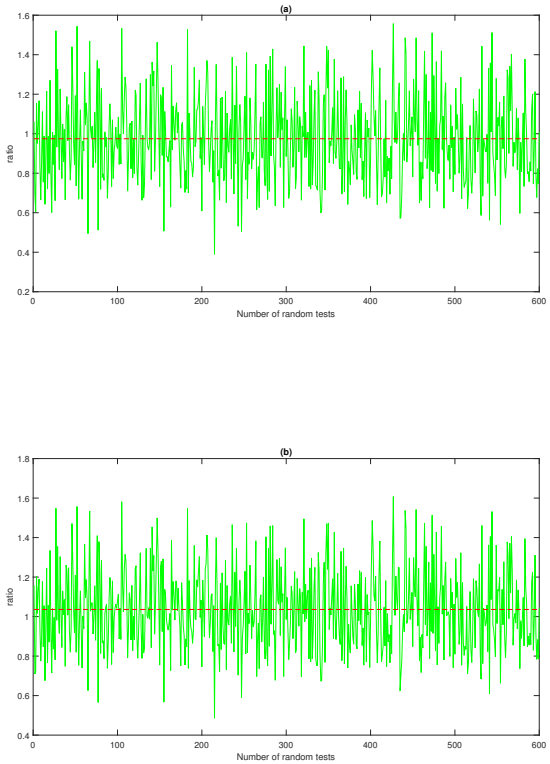

Figure 1: *SCE results $\bar{\kappa}_{mtls}$ compared with the exact condition numbers $\kappa_{MTLS}(A, b)$ of MTLS. Ratio $= \bar{\kappa}_{MTLS}/\kappa_{MTLS}(A, b)$. The tested matrices are of size 200×80 (a): $\varepsilon_p = 9.99976032e-1$ and (b): $\varepsilon_p = 9.99952397e-5$. The horizontal dotted lines stand for the average ratios.*

For componentwise condition estimation, we take $m = 200$, $n_1 = 40$, $n_2 = 40$. in this part. In Figure 2, we plot the ratio for $\varepsilon_p = 9.99976032e-1$ and $\varepsilon_p = 9.99952397e-5$ between the statistical condition estimate via Algorithm SCE-CCE and the exact value computed by (31). It is observed that the ratio is close to 1 for every component of MTLS solution. Therefore the statistical estimate is accurate with high probability in practical computing.

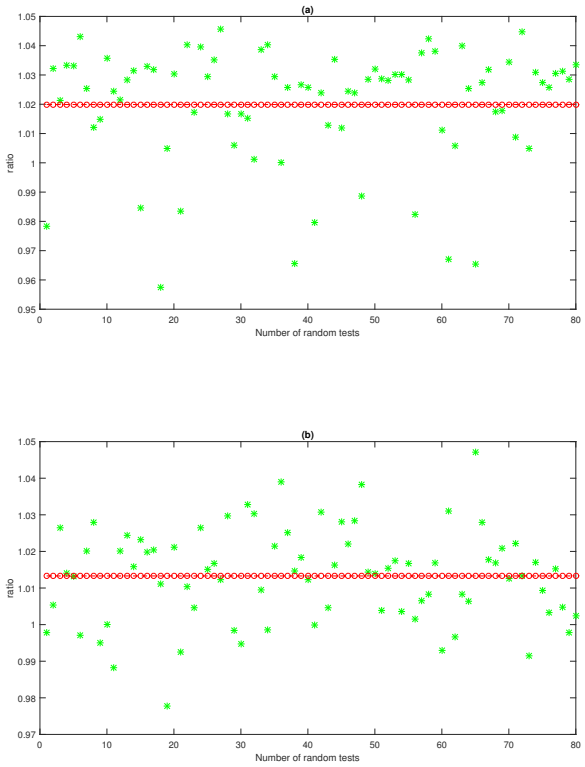

Figure 2: *SCE results* $\bar{\kappa}_{abs}$ *compared with the exact condition numbers* κ_j^c *of MTLS. Ratio* $= \bar{\kappa}_{abs}/\kappa_j^c$. *The tested matrices are of size* 200×80 *(a):* $\varepsilon_p = 9.99976032e-1$ *and (b):* $\varepsilon_p = 9.99952397e-5$. *The horizontal dotted lines stand for the average ratios.*

8. Conclusions

In this paper, we mainly present the perturbation analysis of the mixed least squares-total least squares (MTLS). In the analysis of the first order perturbation, we first provide an upper bound that one can establish the normwise condition number formulas for the MTLS problem from it. For easy estimation, we show a lower bound for the normwise condition number which is proved to be optimal. In order to overcome the problems encountered in calculating the normwise condition number, we give an upper bound for computing more effectively and nonstandard and unusual perturbation bounds for the MTLS problem. Both of the two types of the perturbation bounds can enjoy storage and computational advantages, and demonstrate the superiority of them by the numerical examples. For getting more insight into the sensitivity of the MTLS technique with respect to perturbations in all data, we analyze the corrections applied by MTLS to the data in $Ax \approx b$ to make the set compatible. We also deduce the assumptions about the underlying pertubation model. Thus indicates how closely the data A, b fit the so-called general errors-in-variables model. They all reveal that the same parameters ($\tilde{\sigma}$ and C) mainly determine the correspondences and differences between both techniques. On how to estimate the conditioning of the MTLS problem more effectively, we propose statistical algorithms by taking advantage of the superiority of SCE techniques. The SCE results are compared with the exact values in numerical experiments.

References

[1] Q. H. Liu, M. H. Wang, *On the weighting method for mixed least squares-total least squares problems*, Numer. Linear. Algebra Appl., 2017, 24(5): e2094.

[2] F. Cucker, H. Diao, Y. Wei, *On mixed and componentwise condition numbers for Moore-Penrose inverse and linear least squares problems*, Math. Comp., 76(2007)947–963.

[3] M. Baboulin, S. Gratton, *A contribution to the conditioning of the total least-squares problem,* SIAM J. Matrix Anal. Appl., 32(2011)685–699.
[4] G. H. Golub, L. Van, F. Charles, *An analysis of the total least squares problem,* SIAM J. Numer. Anal., 17(1980)883–893.
[5] G. H. Golub, L. Van, F. Charles, *Matrix Computations,* Johns Hopkins Studies in the Mathematical Sciences, 4-th, Johns Hopkins University Press, Baltimore, MD, 2013.
[6] S. Gratton, *On the condition number of linear least squares problems in a weighted Frobenius norm,* BIT, 36(1996)523–530.
[7] G. W. Stewart, *A second order perturbation expansion for small singular values,* Linear Algebra Appl., 56(1984)231–235.
[8] B. Y. Li, Z. X. Jia, *Some results on condition numbers of the scaled total least squares problem,* Linear Algebra Appl., 435(2011)674–686.
[9] C. S. Kenney, A. J. Laub, *Small-sample statistical condition estimates for general matrix functions,* SIAM J. Sci. Comput., 15(1994)36–61.
[10] M. Baboulin, S. Gratton, R. Lacroix, A. Laub, *Efficient computation of condition estimates for linear least squares problems,* To appear in the Proceedings of the 10th International Conference on Parallel Processing and Applied Mathematics, PPAM 2013(09/2013).
[11] C. S. Kenney, A. J. Laub, M. S. Reese, *Statistical condition estimation for linear least squares,* SIAM J. Matrix Anal. Appl., 19(1998)906–923.
[12] X. G. Liu, *Solvability and perturbation analysis of the total least squares problem,* Acta Math. Appl. Sinica, 19(1996)254–262.
[13] S. Van Huffel, J. Vandewalle, *The total least squares problem, frontiers in applied mathematics,* SIAM, Philadelphia, PA, 1991.
[14] M. S. Wei, *The analysis for the total least squares problem with more than one solution,* SIAM J. Matrix Anal. Appl., 13(1992)746–763.
[15] H. A. Diao, X. H. Shi, Y. M. Wei, *Effective condition numbers and small sample statistical condition estimation for the generalized Sylvester equation,* Sci. China Math., 56(2013)967–982.
[16] P. P. Xie, H. Xiang, Y. M. Wei, *A contribution to perturbation analysis for total least squares problems,* Numer Algor., 2(2017)381–395.
[17] B. Zheng, Z. Yang, *Perturbation analysis for mixed least squaresCtotal least squares problems,* Numer Linear Algebra Appl., 2019, 26(4): e2239.
[18] A. Laub, J. Xia, *Applications of statistical condition estimation to the solution of linear systems,* Numer. Linear Algebra Appl., 15(2008)489–513.
[19] G. H. Golub , C. F. Loan Van, *An analysis of the total least squares problem,* SIAM J. Matrix Anal A., 17 (1980)883–893.
[20] S. Van Huffel, J. Vandewalle, *The total least-squares problem: computational aspects and analysis,* SIAM Philadelphia, PA: Frontier in Applied Mathematics, 1991.
[21] M. S. Wei, *Algebraic relations between the total least squares and least squares problems with more than one solution,* Numer Math., 62(1992)123–148.
[22] S. Van Huffel , H. Y. Zha, *Restricted total least squares problem: formulation, algorithm, and properties,* SIAM J. Matrix Anal A., 12(1991)292–309.
[23] S. Van Huffel, J. Vandewalle, *Analysis and properties of the generalized total least squares problem $AX \approx B$ when some or all columns in A are subject to error,* SIAM J. Matrix Anal A., 10(1989)294–315.
[24] C. C. Paige, M. S. Wei, *Analysis of the generalized total least squares problem $AX \approx B$ when some columns of A are free of error,* Numer Math., 65(1993)177–202.
[25] S. J, Yan, J. Y. Fan, *The solution set of the mixed LS-TLS problem,* Int. J. Comput. Math., 77(2001)545–561.
[26] G. Stewart, *Perturbation bounds for the QR factorization of a matrix,* SIAM J. Numer. Anal., 14(1977)509-518.
[27] J. Ding, A. H. Zhou, *Eigenvalues of rank-one updated matrices with some applications,* Appl. Math. Lett., 20 (2007)1223–1226.
[28] T. R. Yang, M. Lin, *Rayleigh quotient iteration for a total least squares filter in robot navigation,* Chapter Signal Analysis and Prediction in Applied and Numerical Harmonic Analysis Birkhuser. Boston: MA, 1998.