Journal of Computational and Applied Mathematics 344 (2018) 640–656

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Condition numbers for a linear function of the solution of the linear least squares problem with equality constraints

Huai-An Diao

School of Mathematics and Statistics, Northeast Normal University, No. 5268 Renmin Street, Chang Chun 130024, PR China

ARTICLE INFO

Article history:
Received 25 May 2017
Received in revised form 28 December 2017

MSC:
15A09
15A12
65F35

Keywords:
Linear least squares problem with equality constraints
Linear least squares
Componentwise perturbation
Condition number
Hager–Higham algorithm

ABSTRACT

In this paper, we consider the normwise, mixed and componentwise condition numbers for a linear function $\boldsymbol{L}x$ of the solution x to the linear least squares problem with equality constraints (LSE). The explicit expressions of the normwise, mixed and componentwise condition numbers are derived. Also, we revisit some previous results on the condition numbers of linear least squares problem (LS) and LSE. It is shown that some previous explicit condition number expressions on LS and LSE can be recovered from our new derived condition numbers' formulas. The sharp upper bounds for the derived normwise, mixed and componentwise condition numbers are obtained, which can be estimated efficiently by means of the classical Hager–Higham algorithm for estimating matrix one-norm. Moreover, the proposed condition estimation methods can be incorporated into the generalized QR factorization method for solving LSE. The numerical examples show that when the coefficient matrices of LSE are sparse and badly-scaled, the mixed and componentwise condition numbers can give sharp perturbation bounds, on the other hand normwise condition numbers can severely overestimate the exact relative errors because normwise condition numbers ignore the data sparsity and scaling. However, from the numerical experiments for random LSE problems, if the data is not either sparse or badly scaled, it is more suitable to adopt the normwise condition number to measure the conditioning of LSE since the explicit formula of the normwise condition number is more compact.

1. Introduction

The least squares problem with equality constraints (LSE) has the following form:

$$\text{LSE}: \quad \min_{x\in\mathbb{R}^n} \|Ax-b\|_2 \text{ subject to } Cx=d, \tag{1.1}$$

where $A\in\mathbb{R}^{m\times n}$, $C\in\mathbb{R}^{p\times n}$, $b\in\mathbb{R}^m$, $d\in\mathbb{R}^p$ and $m+p\geq n\geq p$. The rank conditions [1]

$$\operatorname{rank}(C)=p \text{ and } \operatorname{rank}\left(\begin{bmatrix}A\\C\end{bmatrix}\right)=n \tag{1.2}$$

guarantee the existence of the unique solution of LSE [1,2]

$$x=\mathcal{K}b+C_A^{\dagger}d,$$

E-mail address: hadiao@nenu.edu.cn.

https://doi.org/10.1016/j.cam.2018.05.050

where

$$\mathcal{K} = (A\mathcal{P})^\dagger, \quad \mathcal{P} = I_n - C^\dagger C, \quad C_A^\dagger = (I_n - \mathcal{K}A)\, C^\dagger, \tag{1.3}$$

and $B^\dagger$ is the Moore–Penrose inverse of B [1]. Under the rank condition $\operatorname{rank}(C) = p$ the equality constraints $Cx = d$ in (1.1) are consistent, thus LSE (1.1) has solutions. The second rank condition of (1.2) guarantees the uniqueness of the solution to (1.1). On the other hand, the augmented system also defines the unique solution x as follows:

$$\mathcal{A}\mathbf{x} := \begin{bmatrix} \mathbf{0} & \mathbf{0} & C \\ \mathbf{0} & I_m & A \\ C^\top & A^\top & \mathbf{0} \end{bmatrix} \begin{bmatrix} \lambda \\ r \\ x \end{bmatrix} = \begin{bmatrix} d \\ b \\ \mathbf{0} \end{bmatrix} := \mathbf{b}, \tag{1.4}$$

where $A^\top$ is the transpose of A, I_m denotes the $m \times m$ identity matrix, $\mathbf{0}$ is the zeros matrix with conformal dimension, $\lambda \in \mathbb{R}^p$ is a vector of Lagrange multipliers, and r is the residual vector $r = b - Ax$. As stated in [2,3], when the rank condition (1.2) is satisfied, $\mathcal{A}$ is nonsingular and its inverse has the following expression:

$$\mathcal{A}^{-1} = \begin{bmatrix} (AC_A^\dagger)^\top AC_A^\dagger & -(AC_A^\dagger)^\top & (C_A^\dagger)^\top \\ -AC_A^\dagger & I_m - (A\mathcal{P})\mathcal{K} & \mathcal{K}^\top \\ C_A^\dagger & \mathcal{K} & -\left((A\mathcal{P})^\top (A\mathcal{P})\right)^\dagger \end{bmatrix}. \tag{1.5}$$

When $C = \mathbf{0}$ and $d = \mathbf{0}$, LSE is reduced to the classical linear least squares problem (LS) as follows

$$\mathrm{LS}: \quad \min_{x \in \mathbb{R}^n} \|Ax - b\|_2. \tag{1.6}$$

In this case, we know that the rank condition (1.2) becomes $\operatorname{rank}(A) = n$. Thus LS has the unique LS solution $x = A^\dagger b = (A^\top A)^{-1} A^\top b$.

The LSE problem has many applications such as in the analysis of large scale structures [4], and the solution of the inequality constrained least square problem [5] etc. The algorithms and perturbation analysis of LSE can be found in several papers [1–9] and references therein.

Perturbation theory is important in matrix computation, since they can give error bounds for the computed solution. Especially, condition number measures the *worst-case* sensitivity of an input data with respect to *small* perturbations on it; see the recent monograph [10] and references therein. Rice in [11] gave a general theory of condition numbers. Let $\psi : \mathbb{R}^p \to \mathbb{R}^q$ be a mapping, where $\mathbb{R}^p$ and $\mathbb{R}^q$ are the usual p- and q-dimensional Euclidean spaces equipped with some norms, respectively. If ψ is continuous and Fréchet differentiable in the neighborhood of $u_0 \in \mathbb{R}^p$ then, according to [11], the *relative normwise condition number* of ψ at u_0 is given by

$$\operatorname{cond}^\psi(u_0) := \lim_{\varepsilon \to 0} \sup_{\|\Delta u\| \le \varepsilon} \left(\frac{\|\psi(u_0 + \Delta u) - \psi(u_0)\|}{\|\psi(u_0)\|} \Big/ \frac{\|\Delta u\|}{\|u_0\|} \right) = \frac{\|\mathrm{d}\psi(u_0)\| \, \|u_0\|}{\|\psi(u_0)\|}, \tag{1.7}$$

where $\mathrm{d}\psi(u_0)$ is the Fréchet derivative of ψ at u_0. Condition number can tell us the loss of the precision in finite precision computation of a problem. With the backward error of a problem, the relative error of the computed solution can be bounded by the product of condition number and backward error.

When the data is sparse or badly-scaled, componentwise perturbation analysis [12,13] has been proposed to investigate the mixed condition numbers and componentwise condition numbers [14] of the problem in matrix computation. The mixed condition numbers use the componentwise error analysis for the input data, while the normwise error analysis for the output data. On the other hand, the componentwise condition numbers use the componentwise error analysis for both input and output data. Consequently, the perturbation bounds based on the mixed and componentwise condition numbers are more effective and sharper than those based on the normwise condition number when the data is sparse or badly scaled because the normwise condition number defined in (1.7) does not take account of the structure of both input and output data with respect to scaling and/or sparsity.

In some situations, the conditionings of particular components of a solution are different. Thus it is suitable to consider the condition numbers of a linear function of the solution. These type condition numbers had been studied for the LS problem [15,16], the weighted LS problem [17], the total least squares problems [18,19], the indefinite LS problem [20] and the LSE problem [21], etc. In this paper, we will investigate the sensitivity of a linear function $\boldsymbol{L}x$ of the LSE solution x with respect to perturbations on the data A, C, b and d. First, let us introduce the following mapping

$$\begin{aligned} &\Phi : \mathbb{R}^{mn} \times \mathbb{R}^{pn} \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^k \\ &\Phi(\mathrm{vec}(A),\ \mathrm{vec}(C),\ b,\ d) := \boldsymbol{L}\left(\mathcal{K}b + C_A^\dagger d\right), \end{aligned} \tag{1.8}$$

where $\mathrm{vec}(A)$ is a vector obtained by stacking the columns of a matrix A one by one (see [22] for details), and $\boldsymbol{L}$ is an k-by-n, $k \le n$, matrix introduced for the selection of the solution components. For example, when $\boldsymbol{L} = I_n$ $(k = n)$, all the n components of the solution x are equally selected. When $\boldsymbol{L} = e_i^\top$ $(k = 1)$, the ith row of I_n, then only the ith component of the solution is selected. The matrix $\boldsymbol{L}$ is not perturbed in the text.

In this paper the explicit expressions for the normwise, mixed and componentwise condition numbers of the linear function $\boldsymbol{L}x$ of the solution x to LSE (1.1) are considered. We will adopt the definition of condition numbers to derive condition numbers' expressions. Moreover, sharp upper bounds also are deduced, which can be estimated efficiently via the classical Hager–Higham algorithm [23–25]. When the generalized QR factorization (GQR) method [2,6,8] is adopted for solving LSE, the computational complexity of the proposed condition estimations can be significantly reduced through utilizing the already computed matrix decompositions. There have been lots references on the condition numbers analysis for LS [15,16,26,27] and LSE [2,7,21]. We revisit some results on condition numbers for LS and LSE. Our new derived explicit expressions can recover the former expressions under some assumptions. Numerical examples in Section 5 tell us, under some particular situations, the mixed and componentwise condition numbers of the linear function for LSE can be much smaller than the normwise condition numbers. However, when the data is neither sparse or badly scaled, there are little differences between the normwise condition numbers and the mixed/componentwise condition numbers. Thus, it is more suitable to use the normwise condition numbers as the measure for monitoring the conditioning of LSE since the explicit formulas of normwise condition numbers are more compact.

The paper is organized as follows. In Section 2, the normwise, mixed and componentwise condition numbers for LSE are investigated. Also their upper bounds are derived, which can be estimated by the Hager–Higham algorithm [23–25]. We revisit the previous results of the conditioning analysis for LSE and LS in Sections 3 and 4, respectively. We do some numerical examples to test the effectiveness of the proposed condition numbers in Section 5. At end, in Section 6 concluding remarks are drawn and the future research topics are pointed out.

2. Explicit expressions of condition numbers for LSE

In this section we will derive the explicit condition numbers expressions for a linear function of the solution of LSE. Also sharp upper bounds for the normwise, mixed and componentwise condition numbers are obtained. By considering the already computed decomposition of the GQR method [2,6,8] for solving LSE, we can estimate upper bounds efficiently via the Hager–Higham algorithm [23–25].

In the following the normwise condition numbers for LSE are defined. First, we introduce the following product norm [27] to measure the input data $[E, f]$. Let α and β be two positive real numbers. For the data space $\mathbb{R}^{p\times q} \times \mathbb{R}^p$, we use the product norm defined by

$$\|(E, f)\|_{\mathcal{F}} = \sqrt{\alpha^2\|E\|_F^2 + \beta^2\|f\|_2^2}. \tag{2.1}$$

The norm is very flexible since they allow to monitor the perturbations on E and f. For instance, large values of α (resp. β) enable to obtain condition number problems where mainly f (resp. E) is perturbed.

The following lemma concerns with an equivalent expression for a general linear operator's spectral norm, which will be used to derive expressions for the normwise condition number of LSE.

Lemma 1 (*[28, Lemma 2.1]*). *Given matrices $\boldsymbol{L} \in \mathbb{R}^{k\times q}$, $V \in \mathbb{R}^{p\times q}$, $X \in \mathbb{R}^{q\times p}$, $Y \in \mathbb{R}^{q\times q}$ and vectors $s \in \mathbb{R}^q$, $t \in \mathbb{R}^p$, $u \in \mathbb{R}^p$ with two positive real numbers α and β, for the linear operator l defined by*

$$l(V, u) := \boldsymbol{L}\left(-XVs + YV^\top t + Xu\right),$$

its operator spectral norm can be characterized by

$$\|l\|_2 = \sup_{V\neq 0, u\neq 0} \frac{\|l(V,u)\|_2}{\|[V,u]\|_{\mathcal{F}}} = \left\| \boldsymbol{L}\left[-\frac{1}{\beta}\|s\|_2 X, \quad \frac{1}{\alpha}\|t\|_2 Y\right] \begin{bmatrix} c_1 I_m - c_2\dfrac{tt^\top}{\|t\|_2^2} & \dfrac{\beta}{\alpha}\dfrac{ts^\top}{\|t\|_2\|s\|_2} \\ \mathbf{0} & I_n \end{bmatrix} \right\|_2,$$

where $c_1 = \sqrt{\frac{\beta^2}{\alpha^2} + \frac{1}{\|s\|_2^2}}$, $c_2 = c_1 + \frac{1}{\|s\|_2}$. For the linear operator l_M defined by

$$l_M(V) := \boldsymbol{L}\left(-XVs + YV^\top t\right),$$

its operator spectral norm can be characterized by

$$\sup_{V\neq 0} \frac{\|l_M(V)\|_2}{\|V\|_F} = \left\| \boldsymbol{L}\left[-\|s\|_2 X, \quad \|t\|_2 Y\right] \begin{bmatrix} I_m - \dfrac{tt^\top}{\|t\|_2^2} & \dfrac{ts^\top}{\|t\|_2\|s\|_2} \\ \mathbf{0} & I_n \end{bmatrix} \right\|_2. \tag{2.2}$$

Moreover, we have

$$\sup_{V\neq 0} \frac{\|l_M(V)\|_2}{\|V\|_F} \le \sup_{V\neq 0} \frac{\|l_M(V)\|_2}{\|V\|_2} \le \sqrt{2}\sup_{V\neq 0} \frac{\|l_M(V)\|_2}{\|V\|_F}. \tag{2.3}$$

In the following the notation $\frac{a}{b}$ means two vector componentwise division given by

$$\left(\frac{a}{b}\right)_i = \begin{cases} \frac{a_i}{b_i}, & \text{if} \quad b_i \neq 0, \\ 0, & \text{if} \quad b_i = 0, \end{cases} \tag{2.4}$$

where a and b are two conformal dimensional vectors. We define the normwise, mixed and componentwise condition numbers for LSE as follows:

$$\kappa_n := \lim_{\varepsilon\to 0} \sup_{\left\|\left(\begin{bmatrix}\Delta A\\ \Delta C\end{bmatrix}, \begin{bmatrix}\Delta b\\ \Delta d\end{bmatrix}\right)\right\|_{\mathcal{F}} \le \varepsilon} \frac{\|\boldsymbol{L}\Delta x\|_2 \left\|\left(\begin{bmatrix}A\\ C\end{bmatrix}, \begin{bmatrix}b\\ d\end{bmatrix}\right)\right\|_{\mathcal{F}}}{\varepsilon\|\boldsymbol{L}x\|_2}, \tag{2.5}$$

$$\kappa_\infty^{\text{rel}} := \lim_{\varepsilon\to 0} \sup_{\substack{|\Delta A|\le\varepsilon|A|,\ |\Delta C|\le\varepsilon|C| \\ |\Delta b|\le\varepsilon|b|,\ |\Delta d|\le\varepsilon|d|}} \frac{\|\boldsymbol{L}\Delta x\|_\infty}{\varepsilon\|\boldsymbol{L}x\|_\infty},$$

$$\kappa_c := \lim_{\varepsilon\to 0} \sup_{\substack{|\Delta A|\le\varepsilon|A|,\ |\Delta C|\le\varepsilon|C| \\ |\Delta b|\le\varepsilon|b|,\ |\Delta d|\le\varepsilon|d|}} \frac{1}{\varepsilon}\left\|\frac{\boldsymbol{L}\Delta x}{\boldsymbol{L}x}\right\|_\infty,$$

where $|A| = |a_{ij}|, |\Delta A| \le \varepsilon|A|$ should be understood componentwisely, $\|\cdot\|_{\mathcal{F}}$ is the product norm defined by (2.1), and $x+\Delta x$ is the unique optimal solution to

$$\min_{\tilde{x}\in\mathbb{R}^n} \|(A+\Delta A)\tilde{x} - (b+\Delta b)\|_2 \text{ subject to } (C+\Delta C)\tilde{x} = d+\Delta d. \tag{2.6}$$

First, we prove the mapping Φ defined by (1.8) is Fréchet differentiable and its Fréchet derivative is obtained by means of matrix differential calculus [29] in Lemma 3. Before that, we need the following lemma.

Lemma 2 (*[29, Page 171, Theorem 3]*). *Let T be the set of non-singular real $m \times m$ matrices, and S be an open subset of $\mathbb{R}^{n\times q}$. If the matrix function $F : S \to T$ is k times (continuously) differentiable on S, then so is the matrix function $F^{-1} : S \to T$ defined by $F^{-1}(X) = (F(X))^{-1}$, and*

$$\mathrm{d}F^{-1} = -F^{-1}(\mathrm{d}F)F^{-1}.$$

The Fréchet derivative of Φ involves *Kronecker product*. In the following, we review some basic results on Kronecker product. If $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$, then the Kronecker product $A \otimes B \in \mathbb{R}^{mp\times nq}$ is defined by $A \otimes B = \left[a_{ij}B\right] \in \mathbb{R}^{mp\times nq}$ [22]. The following results can be found in [22]

$$\begin{aligned} &|A\otimes B| = |A|\otimes|B|, \quad \mathrm{vec}(AXB) = (B^\top \otimes A)\mathrm{vec}(X), \\ &\text{for any } A \in \mathbb{R}^{m\times n},\ \Pi(\mathrm{vec}(A)) = \mathrm{vec}(A^\top), \\ &\text{for any vector } y \text{ and matrix } Y,\ \left(y^\top \otimes Y\right)\Pi = Y \otimes y^\top, \end{aligned} \tag{2.7}$$

where $\Pi \in \mathbb{R}^{mn\times mn}$ is the *vec-permutation matrix*; see [22] for details.

Lemma 3. *The function Φ is a continuous mapping on $\mathbb{R}^{mn} \times \mathbb{R}^{pn} \times \mathbb{R}^m \times \mathbb{R}^p$, where Φ is defined by (1.8). In addition, Φ is Fréchet differentiable at (A, C, b, d) and its Fréchet derivative is given by*

$$\begin{aligned} \mathrm{d}\Phi(\mathrm{vec}(A), \mathrm{vec}(C), b, d) = \boldsymbol{L}\Big[&\mathcal{K}(\mathcal{K}^\top \otimes r^\top - x^\top \otimes I_m),\ -C_A^\dagger(x^\top \otimes I_p) - \mathcal{K}[\mathcal{K}^\top \otimes (r^\top A C_A^\dagger)], \\ &\mathcal{K},\ C_A^\dagger\Big]. \end{aligned}$$

Proof. From (1.4), we know that $\mathbf{x} = \mathcal{A}^{-1}\mathbf{b}$. Since $\mathcal{A}$ is invertible, the linear operator Φ defined in (1.8) is continuously Fréchet differentiable in a neighborhood of the data (A, C, b, d) from the theory of the matrix differential calculus [29]. With Lemma 2, we can deduce that

$$\mathrm{d}\mathbf{x} = (\mathrm{d}\mathcal{A}^{-1})\mathbf{b} + \mathcal{A}^{-1}\mathrm{d}\mathbf{b} = -\mathcal{A}^{-1}(\mathrm{d}\mathcal{A})\mathbf{x} + \mathcal{A}^{-1}\mathrm{d}\mathbf{b}.$$

Noting that

$$\mathrm{d}\mathbf{x} = \begin{bmatrix}\mathrm{d}\lambda\\ \mathrm{d}r\\ \mathrm{d}x\end{bmatrix}, \quad \mathrm{d}\mathbf{b} = \begin{bmatrix}\mathrm{d}d\\ \mathrm{d}b\\ \mathbf{0}\end{bmatrix}, \quad \mathrm{d}\mathcal{A} = \begin{bmatrix}\mathbf{0} & \mathbf{0} & \mathrm{d}C\\ \mathbf{0} & \mathbf{0} & \mathrm{d}A\\ \mathrm{d}C^\top & \mathrm{d}A^\top & \mathbf{0}\end{bmatrix},$$

recalling (1.5), and after some algebraic operations we deduce that

$$\begin{aligned} \mathrm{d}x = {}& \mathcal{K}\mathrm{d}b + C_A^\dagger \mathrm{d}d - C_A^\dagger \mathrm{d}Cx - \mathcal{K}\mathrm{d}Ax + ((A\mathcal{P})^\top(A\mathcal{P}))^\dagger(\mathrm{d}A)^\top r \\ &+ ((A\mathcal{P})^\top(A\mathcal{P}))^\dagger(\mathrm{d}C)^\top\lambda. \end{aligned} \tag{2.8}$$

Thus $\mathrm{d}\Phi = \boldsymbol{L}\mathrm{d}x$ since Φ is linear. From (1.4), it can be verified that $C^\top\lambda + A^\top r = 0$. Since C has full row rank ($CC^\dagger = I_p$) it is easy to see that $\lambda = -(AC^\dagger)^\top r$. Substituting the above expression in (2.8), using the equality $(AC^\dagger)^\top r = (AC_A^\dagger)^\top r$ from Lemma 4.2 in [2], and noting $((A\mathcal{P})^\top(A\mathcal{P}))^\dagger = \mathcal{K}\mathcal{K}^\top$, we can derive that

$$\begin{aligned}&\mathrm{d}\Phi(\mathrm{vec}(A),\ \mathrm{vec}(C),\ b,\ d)\cdot\left[\mathrm{vec}(\mathrm{d}A)^\top, \mathrm{vec}(\mathrm{d}C)^\top, \mathrm{d}b^\top, \mathrm{d}d^\top\right]^\top = -\boldsymbol{L}\mathcal{K}\mathrm{d}Ax\\ &\quad + \boldsymbol{L}\mathcal{K}\mathcal{K}^\top(\mathrm{d}A)^\top r - \boldsymbol{L}C_A^\dagger \mathrm{d}Cx - \boldsymbol{L}\mathcal{K}\mathcal{K}^\top(\mathrm{d}C)^\top(AC_A^\dagger)^\top r + \boldsymbol{L}\mathcal{K}\mathrm{d}b + \boldsymbol{L}C_A^\dagger\mathrm{d}d.\end{aligned} \tag{2.9}$$

Applying vec operator to both sides of the above equation and using Kronecker product properties (2.7), we can prove that

$$\begin{aligned}&\mathrm{d}\Phi(\mathrm{vec}(A),\ \mathrm{vec}(C),\ b,\ d)\cdot\left[\mathrm{vec}(\mathrm{d}A)^\top, \mathrm{vec}(\mathrm{d}C)^\top, \mathrm{d}b^\top, \Delta d^\top\right]^\top = -x^\top\otimes(\boldsymbol{L}\mathcal{K})\mathrm{vec}(\mathrm{d}A)\\ &\qquad + r^\top\otimes(\boldsymbol{L}\mathcal{K}\mathcal{K}^\top)\Pi\ \mathrm{vec}(\mathrm{d}A) - x^\top\otimes(\boldsymbol{L}C_A^\dagger)\mathrm{vec}(\mathrm{d}C) - (r^\top AC_A^\dagger)\otimes(\boldsymbol{L}\mathcal{K}\mathcal{K}^\top)\Pi\mathrm{vec}(\mathrm{d}C)\\ &\qquad + \boldsymbol{L}\mathcal{K}\mathrm{d}b + \boldsymbol{L}C_A^\dagger\mathrm{d}d\\ &\quad = -\boldsymbol{L}\mathcal{K}(x^\top\otimes I_m)\mathrm{vec}(\mathrm{d}A) + \boldsymbol{L}\mathcal{K}(\mathcal{K}^\top\otimes r^\top)\mathrm{vec}(\mathrm{d}A) - \boldsymbol{L}C_A^\dagger(x^\top\otimes I_p)\mathrm{vec}(\mathrm{d}C)\\ &\qquad - \boldsymbol{L}\mathcal{K}[\mathcal{K}^\top\otimes(r^\top AC_A^\dagger)]\mathrm{vec}(\mathrm{d}C) + \boldsymbol{L}\mathcal{K}\mathrm{d}b + \boldsymbol{L}C_A^\dagger\mathrm{d}d.\end{aligned}$$

From the above deductions, we complete the proof of this lemma. □

In the following theorem, we will give the explicit expressions for normwise, mixed and componentwise condition numbers of LSE. First, recall the notation $\frac{a}{b}$ is defined by (2.4), where a and b are two conformal dimensional vectors.

Theorem 1. *Let $A \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{p\times n}$, $b \in \mathbb{R}^m$ and $d \in \mathbb{R}^p$, the rank conditions (1.2) be satisfied, and $x = \mathcal{K}b + C_A^\dagger d$ be the solution of LSE (1.1), for normwise, mixed and componentwise condition numbers defined by (2.5), we have*

$$\begin{aligned}\kappa_n &= \frac{\left\|\left(\begin{bmatrix}A\\C\end{bmatrix},\begin{bmatrix}b\\d\end{bmatrix}\right)\right\|_{\mathcal{F}}\cdot\|\boldsymbol{L}\mathcal{N}\|_2}{\|\boldsymbol{L}x\|_2},\\ \kappa_\infty^{\mathrm{rel}} &= \frac{\left\||\boldsymbol{L}\mathcal{M}|\ \mathrm{vec}(|A|) + |\boldsymbol{L}\mathcal{J}|\ \mathrm{vec}(|C|) + |\boldsymbol{L}\mathcal{K}|\ |b| + |\boldsymbol{L}C_A^\dagger|\ |d|\right\|_\infty}{\|\boldsymbol{L}x\|_\infty},\\ \kappa_c &= \left\|\frac{|\boldsymbol{L}\mathcal{M}|\ \mathrm{vec}(|A|) + |\boldsymbol{L}\mathcal{J}|\ \mathrm{vec}(|C|) + |\boldsymbol{L}\mathcal{K}|\ |b| + |\boldsymbol{L}C_A^\dagger|\ |d|}{\boldsymbol{L}x}\right\|_\infty,\end{aligned}$$

where

$$\begin{aligned}&\mathcal{M} = \mathcal{K}(\mathcal{K}^\top\otimes r^\top - x^\top\otimes I_m),\quad \mathcal{J} = C_A^\dagger(x^\top\otimes I_p) + \mathcal{K}[\mathcal{K}^\top\otimes(r^\top AC_A^\dagger)],\\ &\mathcal{N} = \begin{bmatrix}-\frac{1}{\beta}\|x\|_2\mathcal{K} & -\frac{1}{\beta}\|x\|_2C_A^\dagger & \frac{1}{\alpha}\|t\|_2\mathcal{K}\mathcal{K}^\top\end{bmatrix}\begin{bmatrix}c_1I_{m+p} - c_2\dfrac{tt^\top}{\|t\|_2^2} & \dfrac{\beta}{\alpha}\dfrac{tx^\top}{\|x\|_2\|t\|_2}\\ \mathbf{0} & I_n\end{bmatrix},\end{aligned} \tag{2.10}$$

and $t = \begin{bmatrix}r\\-(AC_A^\dagger)^\top r\end{bmatrix}$, $c_1 = \sqrt{\frac{\beta^2}{\alpha^2} + \frac{1}{\|x\|_2^2}}$ and $c_2 = c_1 + \frac{1}{\|x\|_2}$.

Proof. In view of (1.7) and Lemma 3, for the normwise condition number κ_n, the following equality

$$\kappa_n = \sup_{\substack{\mathrm{d}A\neq 0,\ \mathrm{d}b\neq 0\\ \mathrm{d}C\neq 0,\ \mathrm{d}d\neq 0}} \frac{\left\|\mathrm{d}\Phi\ (\mathrm{vec}(A), \mathrm{vec}(C), b, d)\cdot\left[\mathrm{vec}(\mathrm{d}A)^\top, \mathrm{vec}(\mathrm{d}C)^\top, \mathrm{d}b^\top, \mathrm{d}d^\top\right]^\top\right\|_2}{\left\|\left(\begin{bmatrix}\mathrm{d}A\\ \mathrm{d}C\end{bmatrix},\begin{bmatrix}\mathrm{d}b\\ \mathrm{d}d\end{bmatrix}\right)\right\|_{\mathcal{F}}}\cdot\frac{\left\|\left(\begin{bmatrix}A\\C\end{bmatrix},\begin{bmatrix}b\\d\end{bmatrix}\right)\right\|_{\mathcal{F}}}{\|\boldsymbol{L}x\|_2}$$

holds. From (2.9) we can deduce the following relationships

$$\begin{aligned}&\mathrm{d}\Phi(\mathrm{vec}(A),\ \mathrm{vec}(C),\ b,\ d)\cdot\left[\mathrm{vec}(\mathrm{d}A)^\top, \mathrm{vec}(\mathrm{d}C)^\top, \mathrm{d}b^\top, \mathrm{d}d^\top\right]^\top\\ &= \boldsymbol{L}\left\{-\begin{bmatrix}\mathcal{K} & C_A^\dagger\end{bmatrix}\begin{bmatrix}\mathrm{d}A\\ \mathrm{d}C\end{bmatrix}x + \mathcal{K}\mathcal{K}^\top\begin{bmatrix}\mathrm{d}A\\ \mathrm{d}C\end{bmatrix}^\top\begin{bmatrix}r\\ -(AC_A^\dagger)^\top r\end{bmatrix} + \begin{bmatrix}\mathcal{K} & C_A^\dagger\end{bmatrix}\begin{bmatrix}\mathrm{d}b\\ \mathrm{d}d\end{bmatrix}\right\}.\end{aligned}$$

Thus, identifying $X=\begin{bmatrix}\mathcal{K} & C_A^\dagger\end{bmatrix}$, $Y=\mathcal{K}\mathcal{K}^\top$, $V=\begin{bmatrix}\mathcal{K} & C_A^\dagger\end{bmatrix}$, $u=\begin{bmatrix}\mathrm{d}b\\ \mathrm{d}d\end{bmatrix}$, $s=x$ and $t=\begin{bmatrix}r\\ -(AC_A^\dagger)^\top r\end{bmatrix}$ in Lemma 1, we can derive that

$$\sup_{\substack{\mathrm{d}A\neq 0,\ \mathrm{d}b\neq 0\\ \mathrm{d}C\neq 0,\ \mathrm{d}d\neq 0}} \frac{\|\mathrm{d}\Phi\ (\mathrm{vec}(A),\mathrm{vec}(C),b,d)\cdot(\mathrm{d}A,\mathrm{d}C,\mathrm{d}b,\mathrm{d}d)\|_2}{\left\|\left(\begin{bmatrix}\mathrm{d}A\\ \mathrm{d}C\end{bmatrix},\begin{bmatrix}\mathrm{d}b\\ \mathrm{d}d\end{bmatrix}\right)\right\|_{\mathcal{F}}}$$
$$=\left\|\boldsymbol{L}\begin{bmatrix}-\frac{1}{\beta}\|x\|_2\mathcal{K} & -\frac{1}{\beta}\|x\|_2 C_A^\dagger & \frac{1}{\alpha}\|t\|_2\mathcal{K}\mathcal{K}^\top\end{bmatrix}\begin{bmatrix}c_1 I_{m+p}-c_2\frac{tt^\top}{\|t\|_2^2} & \frac{\beta}{\alpha}\frac{tx^\top}{\|x\|_2\|t\|_2}\\ \mathbf{0} & I_n\end{bmatrix}\right\|_2,$$

then the explicit expression of κ_n has been derived.

For the second part, recalling Δx defined by (2.5), from the definition of Fréchet derivative, we know that

$$\boldsymbol{L}\Delta x=\mathrm{d}\Phi\ (\mathrm{vec}(A),\mathrm{vec}(C),b,d)\left(\mathrm{vec}(\Delta A)^\top,\mathrm{vec}(\Delta C)^\top,\Delta b^\top,\Delta d^\top\right)^\top+\mathcal{O}(\varepsilon^2).$$

Hence, because $|\Delta A|\le\varepsilon|A|$, $|\Delta C|\le\varepsilon|C|$, $|\Delta b|\le\varepsilon|b|$, $|\Delta d|\le\varepsilon|d|$, we can use the monotonicity of infinity norm together with the definition of $\kappa_\infty^{\mathrm{rel}}$ to obtain that

$$\kappa_\infty^{\mathrm{rel}}\le\frac{\left\||\mathrm{d}\Phi\ (\mathrm{vec}(A),\mathrm{vec}(C),b,d)|\left(\mathrm{vec}(|A|)^\top,\mathrm{vec}(|C|)^\top,|b|^\top,|d|^\top\right)^\top\right\|_\infty}{\|\boldsymbol{L}x\|_\infty}. \tag{2.11}$$

For the numerator of the right hand side of the above equation, it is not difficult to see that there exists an index i_0 satisfying that

$$\left\||\mathrm{d}\Phi\ (\mathrm{vec}(A),\mathrm{vec}(C),b,d)|\left(\mathrm{vec}(|A|)^\top,\mathrm{vec}(|C|)^\top,|b|^\top,|d|^\top\right)^\top\right\|_\infty$$
$$=\left\|e_{i_0}^\top\,|\mathrm{d}\Phi\ (\mathrm{vec}(A),\mathrm{vec}(C),b,d)|\left(\mathrm{vec}(|A|)^\top,\mathrm{vec}(|C|)^\top,|b|^\top,|d|^\top\right)^\top\right\|_\infty,$$

where e_{i_0} is i_0th column of the identity matrix. Let Φ_{ij} be the (ij)th entry of the matrix $\mathrm{d}\Phi\ (\mathrm{vec}(A),\mathrm{vec}(C),b,d)\in\mathbb{R}^{k\times(mn+pn+m+p)}$. From the definition of $\kappa_\infty^{\mathrm{rel}}$, it can be verified that the upper bound in (2.11) is attainable at

$$\Delta A_{i_a j_a}=\varepsilon\,\mathrm{sign}(\Phi_{i_0((i_a-1)m+j_a)})|a_{i_a j_a}|,\ \Delta C_{i_c j_c}=\varepsilon\,\mathrm{sign}(\Phi_{i_0(mn+(i_c-1)p+j_c)})|c_{i_c j_c}|,$$
$$\Delta b_{k_b}=\varepsilon\,\mathrm{sign}(\Phi_{i_0(mn+pn+k_b)})|b_{k_b}|,\ \Delta d_{k_d}=\varepsilon\,\mathrm{sign}(\Phi_{i_0(mn+pn+m+k_d)})|b_{k_d}|,$$

where the notation sign is the sign function. Therefore, we prove the expression of $\kappa_\infty^{\mathrm{rel}}$. The proof of the third part is similar to the second part, thus it is omitted. □

As investigated in [7, Section 6], the spectral normwise condition number of LSE when there are no perturbations on b and d can be defined as below:

$$\kappa_{(A,C)}^{(2)}:=\lim_{\varepsilon\to 0}\sup_{\left\|\begin{bmatrix}\Delta A\\ \Delta C\end{bmatrix}\right\|_2\le\varepsilon}\frac{\|\boldsymbol{L}\Delta x\|_2\left\|\begin{bmatrix}A\\ C\end{bmatrix}\right\|_2}{\varepsilon\|\boldsymbol{L}x\|_2},$$

where $x+\Delta x$ the unique solution to the following perturbed LSE:

$$\min_{x+\Delta x\in\mathbb{R}^n}\|(A+\Delta A)(x+\Delta x)-b\|_2 \text{ subject to } (C+\Delta C)(x+\Delta x)=d.$$

However a computable formula for $\kappa_{(A,C)}^{(2)}$ does not exist. On the other hand, the following normwise condition number

$$\kappa_{(A,C)}^{F}:=\lim_{\varepsilon\to 0}\sup_{\left\|\begin{bmatrix}\Delta A\\ \Delta C\end{bmatrix}\right\|_F\le\varepsilon}\frac{\|\boldsymbol{L}\Delta x\|_2\left\|\begin{bmatrix}A\\ C\end{bmatrix}\right\|_2}{\varepsilon\|\boldsymbol{L}x\|_2}$$

can be used to approximate $\kappa_{(A,C)}^{2}$ as done in [7]. From (2.3), it can be verified that

$$\kappa_{(A,C)}^{F}\le\kappa_{(A,C)}^{(2)}\le\sqrt{2}\kappa_{(A,C)}^{F}.$$

Applying (2.2), we can obtain that

$$\kappa_{(A,C)}^{F}=\left\|\begin{bmatrix}A\\ C\end{bmatrix}\right\|_2\left\|\boldsymbol{L}\begin{bmatrix}-\mathcal{K} & -C_A^\dagger & \frac{\|t\|_2}{\|x\|_2}\mathcal{K}\mathcal{K}^\top\end{bmatrix}\begin{bmatrix}I_{m+p}-\frac{tt^\top}{\|t\|_2^2} & \frac{tx^\top}{\|t\|_2\|x\|_2}\\ \mathbf{0} & I_n\end{bmatrix}\right\|_2. \tag{2.12}$$

It was show in the last equation of [7, Page 1196] that

$$\kappa^{F,1}_{(A,C)\mapsto x} := \lim_{\varepsilon\to 0} \sup_{\left\|\begin{bmatrix}\Delta A\\ \Delta C\end{bmatrix}\right\|_F \le \varepsilon} \frac{\|\Delta x\|_2/\|x\|_2}{\left\|\begin{bmatrix}\Delta A\\ \Delta C\end{bmatrix}\right\|_F \Big/ \left\|\begin{bmatrix}A\\ C\end{bmatrix}\right\|_2}$$

$$= \left\|\begin{bmatrix}A\\ C\end{bmatrix}\right\|_2 \left\| \begin{bmatrix} -\mathcal{K} & -C_A^{\dagger} & \frac{\|y\|_2}{\|x\|_2}\mathcal{K}\mathcal{K}^{\top}\end{bmatrix} \begin{bmatrix} I_{m+p} - \frac{yy^{\top}}{\|y\|_2^2} & \frac{yx^{\top}}{\|y\|_2\|x\|_2} \\ \mathbf{0} & I_n \end{bmatrix} \right\|_2, \; y = \begin{bmatrix} r \\ -(AC^{\dagger})^{\top} r\end{bmatrix}. \tag{2.13}$$

Recalling $(AC^{\dagger})^{\top}r = (AC_A^{\dagger})^{\top}r$ from Lemma 4.2 in [2], it yields

$$y = \begin{bmatrix} r \\ -(AC_A^{\dagger})^{\top} r\end{bmatrix} = t,$$

where t is defined in (2.10). Thus when $\boldsymbol{L} = I_n$, the formula (2.13) of $\kappa^{F,1}_{(A,C)\mapsto x}$ is identical to the one (2.12) of $\kappa^F_{(A,C)}$.

In the reminder of this section, we will derive the upper bounds for κ_n, $\kappa_\infty^{\rm rel}$ and κ_c, which can be estimated efficiently by the Hager–Higham algorithm [23–25] when we adopt the GQR method [2,6,8,30] to solve LSE, because the already computed decompositions can be used to reduce the computational cost of the proposed condition estimation method.

The GQR [2,6,8,30] is an efficient and stable method to solve LSE. We first review GQR method. Let $A \in \mathbb{R}^{m\times n}$ and $C \in \mathbb{R}^{p\times n}$ with $m+p \ge n \ge p$. The generalized QR factorization was introduced by Hammarling [6] and Paige [8], which further was analyzed by Anderson et al. [30]. There are orthogonal matrices $Q \in \mathbb{R}^{n\times n}$ and $U \in \mathbb{R}^{m\times m}$ such that

$$U^{\top}AQ = \begin{array}{c} \\ m-n+p \\ n-p \end{array} \!\!\begin{array}{c} \begin{array}{cc} p & n-p \end{array} \\ \begin{pmatrix} L_{11} & \mathbf{0} \\ L_{21} & L_{22} \end{pmatrix} \end{array}, \quad CQ = \begin{array}{c} \begin{array}{cc} p & n-p \end{array} \\ \begin{pmatrix} S & \mathbf{0} \end{pmatrix} \end{array}, \tag{2.14}$$

where $L_{11} \in \mathbb{R}^{(m-n+p)\times p}$, $L_{21} \in \mathbb{R}^{(n-p)\times p}$, $L_{22} \in \mathbb{R}^{(n-p)\times(n-p)}$ and $S \in \mathbb{R}^{p\times p}$. Moreover, $L_{22} \in \mathbb{R}^{(n-p)\times(n-p)}$ and $S \in \mathbb{R}^{p\times p}$ are lower triangular. If rank condition (1.2) holds, then L_{22}, S are nonsingular [2, Theorem 2.1]. The generalized QR factorization method for solving LSE can be summarized as follows. Let $y_1 \in \mathbb{R}^p$ be the solution of the triangular system $Sy_1 = d$ and y_2 be the solution to the triangular system $L_{22}y_2 = v_2 - L_{21}y_1$, where

$$v = U^{\top}b = \begin{array}{c} m-n+p \\ n-p \end{array} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix},$$

where $v_1 \in \mathbb{R}^{m-n+p}$ and $v_2 \in \mathbb{R}^{n-p}$. Then the solution x to LSE can be computed by $x = Qy$, where $y = [y_1^{\top}, y_2^{\top}]^{\top}$. Thus when implementing the GQR method, the decomposition (2.14) has already been computed, which can be utilized to devise the method based on the Hager–Higham algorithm to estimate the upper bounds for κ_n, $\kappa_\infty^{\rm rel}$ and κ_c. Thus the computational cost of condition estimation methods can be reduced.

In the following, we will give upper bounds for κ_n, $\kappa_\infty^{\rm rel}$ and κ_c, which can be estimated efficiently by the Hager–Higham algorithm [23–25]. The proposed condition estimation method can be incorporated into the GQR method for solving LSE.

Corollary 1. *With the notations before, denoting*

$$\kappa_n^{\rm U} = \frac{\left(\frac{\|x\|_2}{\beta}\|\mathcal{K}\|_2 + \frac{\|x\|_2}{\beta}\left\|C_A^{\dagger}\right\|_2 \|A\|_F + \frac{\|r\|_2\left(1+\left\|AC_A^{\dagger}\right\|_2\right)}{\alpha}\|\mathcal{K}\|_2^2\right)}{\|\boldsymbol{L}x\|_2} \times \|\boldsymbol{L}\|_2 \left\|\left(\begin{bmatrix}A\\C\end{bmatrix}, \begin{bmatrix}b\\d\end{bmatrix}\right)\right\|_{\mathcal{F}} \sqrt{\max\left\{1, \frac{\beta^2}{\alpha^2} + \frac{1}{\|x\|_2^2}\right\} + \frac{2\beta}{\alpha}},$$

$$\kappa_\infty^{\rm U} = \frac{\left\|\boldsymbol{L}\mathcal{K}D_{|A|\,|x|}\right\|_\infty + \left\|\boldsymbol{L}\mathcal{K}\mathcal{K}^{\top}D_{|A^{\top}|\,|r|}\right\|_\infty + \left\|\boldsymbol{L}C_A^{\dagger}D_{|C|\,|x|}\right\|_\infty}{\|\boldsymbol{L}x\|_\infty} + \frac{\left\|\boldsymbol{L}\mathcal{K}\mathcal{K}^{\top}D_{|C^{\top}|\,|(AC_A^{\dagger})^{\top}r|}\right\|_\infty + \|\boldsymbol{L}\mathcal{K}D_b\|_\infty + \left\|\boldsymbol{L}C_A^{\dagger}D_d\right\|_\infty}{\|\boldsymbol{L}x\|_\infty},$$

$$\kappa_c^{\rm U} = \left\|D_{\boldsymbol{L}x}^{-1}\boldsymbol{L}\mathcal{K}D_{|A|\,|x|}\right\|_\infty + \left\|D_{\boldsymbol{L}x}^{-1}\boldsymbol{L}\mathcal{K}\mathcal{K}^{\top}D_{|A^{\top}|\,|r|}\right\|_\infty + \left\|D_{\boldsymbol{L}x}^{-1}\boldsymbol{L}C_A^{\dagger}D_{|C|\,|x|}\right\|_\infty + \left\|D_{\boldsymbol{L}x}^{-1}\boldsymbol{L}\mathcal{K}\mathcal{K}^{\top}D_{|C^{\top}|\,|(AC_A^{\dagger})^{\top}r|}\right\|_\infty + \left\|D_{\boldsymbol{L}x}^{-1}\boldsymbol{L}\mathcal{K}D_b\right\|_\infty + \left\|D_{\boldsymbol{L}x}^{-1}\boldsymbol{L}C_A^{\dagger}D_d\right\|_\infty,$$

where $D_v = {\rm diag}(v)$ is a diagonal matrix with the ith diagonal element being v_i, we have

$$\kappa_n \le \kappa_n^{\rm U}, \quad \kappa_\infty^{\rm rel} \le \kappa_\infty^{\rm U}, \quad \kappa_c \le \kappa_c^{\rm U}.$$

Proof. Recalling the expression of κ_n given by Theorem 1, which involves the spectral norm of $\boldsymbol{L}\mathcal{N}$ and $\mathcal{N}$ is defined in (2.10), it can be deduced that

$$\begin{bmatrix} c_1 I_{m+p} - c_2 \frac{tt^\top}{\|t\|_2^2} & \frac{\beta}{\alpha}\frac{tx^\top}{\|x\|_2\|t\|_2} \\ \mathbf{0} & I_n \end{bmatrix} \cdot \begin{bmatrix} c_1 I_{m+p} - c_2 \frac{tt^\top}{\|t\|_2^2} & \mathbf{0} \\ \frac{\beta}{\alpha}\frac{xt^\top}{\|x\|_2\|t\|_2} & I_n \end{bmatrix} = \begin{bmatrix} (\frac{\beta^2}{\alpha^2} + \frac{1}{\|x\|_2^2}) I_{m+p} & \frac{\beta}{\alpha}\frac{tx^\top}{\|t\|_2\|x\|_2} \\ \frac{\beta}{\alpha}\frac{xt^\top}{\|t\|_2\|x\|_2} & I_n \end{bmatrix}$$

$$= \begin{bmatrix} (\frac{\beta^2}{\alpha^2} + \frac{1}{\|x\|_2^2}) I_{m+p} & \mathbf{0} \\ \mathbf{0} & I_n \end{bmatrix} + \frac{\beta}{\alpha} \begin{bmatrix} \mathbf{0} & \frac{tx^\top}{\|t\|_2\|x\|_2} \\ \frac{xt^\top}{\|t\|_2\|x\|_2} & \mathbf{0} \end{bmatrix} := \mathcal{A}_1 + \frac{\beta}{\alpha}\left(\mathbf{ab}^\top + \mathbf{ba}^\top\right), \tag{2.15}$$

where $\mathbf{a} = [t^\top/\|t\|_2 \quad \mathbf{0}]^\top$ and $\mathbf{b} = [\mathbf{0} \quad x^\top/\|x\|_2]^\top$, thus from Weyl theorem [31], we have

$$\left\| \begin{bmatrix} c_1 I_{m+p} - c_2 \frac{tt^\top}{\|t\|_2^2} & \frac{\beta}{\alpha}\frac{tx^\top}{\|x\|_2\|t\|_2} \\ \mathbf{0} & I_n \end{bmatrix} \right\|_2 \le \sqrt{\lambda_{\max}(\mathcal{A}_1) + \frac{\beta}{\alpha}\lambda_{\max}\left(\mathbf{ab}^\top + \mathbf{ba}^\top\right)}$$
$$\le \sqrt{\max\left\{1, \frac{\beta^2}{\alpha^2} + \frac{1}{\|x\|_2^2}\right\} + \frac{2\beta}{\alpha}}, \tag{2.16}$$

where $\lambda_{\max}(\mathcal{B})$ is the maximum eigenvalue of a semi-definite matrix $\mathcal{B}$. Combining (2.16) with the following fact

$$\left\| \left[-\frac{1}{\beta}\|x\|_2\mathcal{K} \quad -\frac{1}{\beta}\|x\|_2 C_A^\dagger \quad \frac{1}{\alpha}\|t\|_2\mathcal{K}\mathcal{K}^\top \right] \right\|_2 \le \frac{\|x\|_2}{\beta}\|\mathcal{K}\|_2 + \frac{\|x\|_2}{\beta}\left\|C_A^\dagger\right\|_2 \|A\|_F + \frac{\|t\|_2}{\alpha}\|\mathcal{K}\|_2^2 .$$

we can prove the first statement of this corollary.

Noting that for any matrix $B \in \mathbb{R}^{p\times q}$ and diagonal matrix $D_v \in \mathbb{R}^{q\times q}$, we have

$$\|BD_v\|_\infty = \|\,|BD_v|\,\|_\infty = \|\,|B|\,|D_v|\,\|_\infty = \|\,|B|\,|D_v|\mathbf{e}\,\|_\infty = \|\,|B|\,|v|\|_\infty . \tag{2.17}$$

where $\mathbf{e} = [1, \ldots, 1]^\top \in \mathbb{R}^q$. Using the Kronecker product property (2.7) and triangle inequality, we have

$$\begin{aligned}
&\left\| |\boldsymbol{L}\mathcal{M}|\operatorname{vec}(|A|) + |\boldsymbol{L}\mathcal{J}|\operatorname{vec}(|C|) + |\boldsymbol{L}\mathcal{K}|\,|b| + |\boldsymbol{L}C_A^\dagger|\,|d| \right\|_\infty \\
&\le \|\,|\boldsymbol{L}\mathcal{M}|\operatorname{vec}(|A|)\|_\infty + \|\,|\boldsymbol{L}\mathcal{J}|\operatorname{vec}(|C|)\|_\infty + \|\,|\boldsymbol{L}\mathcal{K}|\,|b|\|_\infty + \left\||\boldsymbol{L}C_A^\dagger|\,|d|\right\|_\infty \\
&\le \left\| |\boldsymbol{L}\mathcal{K}(\mathcal{K}^\top \otimes r^\top)|\operatorname{vec}(|A|)\right\|_\infty + \left\| |\boldsymbol{L}\mathcal{K}(x^\top \otimes I_m)|\operatorname{vec}(|A|)\right\|_\infty \\
&\quad + \left\| \left|\boldsymbol{L}\mathcal{K}[\mathcal{K}^\top \otimes (r^\top A C_A^\dagger)]\right| \operatorname{vec}(|C|)\right\|_\infty + \left\| \left|\boldsymbol{L}C_A^\dagger(x^\top \otimes I_p)\right| \operatorname{vec}(|C|)\right\|_\infty \\
&\quad + \|\,|\boldsymbol{L}\mathcal{K}|\,|b|\|_\infty + \left\||\boldsymbol{L}C_A^\dagger|\,|d|\right\|_\infty \\
&\le \left\| |\boldsymbol{L}\mathcal{K}\mathcal{K}^\top|(|r^\top| \otimes I_n)\Pi\operatorname{vec}(|A|)\right\|_\infty + \left\| |\boldsymbol{L}\mathcal{K}|(|x^\top| \otimes I_m)\operatorname{vec}(|A|)\right\|_\infty \\
&\quad + \left\| \left|\boldsymbol{L}\mathcal{K}\mathcal{K}^\top\right| [|(r^\top A C_A^\dagger)| \otimes I_n]\Pi\operatorname{vec}(|C|)\right\|_\infty + \left\| \left|\boldsymbol{L}C_A^\dagger\right| (|x^\top| \otimes I_p)\operatorname{vec}(|C|)\right\|_\infty \\
&\quad + \|\,|\boldsymbol{L}\mathcal{K}|\,|b|\|_\infty + \left\||\boldsymbol{L}C_A^\dagger|\,|d|\right\|_\infty \\
&= \|\,|\boldsymbol{L}\mathcal{K}|\,|A|\,|x|\|_\infty + \left\||\boldsymbol{L}\mathcal{K}\mathcal{K}^\top|\,|A^\top|\,|r|\right\|_\infty + \left\||\boldsymbol{L}C_A^\dagger|\,|C|\,|x|\right\|_\infty \\
&\quad + \left\||\boldsymbol{L}\mathcal{K}\mathcal{K}^\top|\,|C^\top|\,|(AC_A^\dagger)^\top r|\right\|_\infty + \|\,|\boldsymbol{L}\mathcal{K}|\,|b|\|_\infty + \left\||\boldsymbol{L}C_A^\dagger|\,|d|\right\|_\infty \\
&= \left\|\boldsymbol{L}\mathcal{K}D_{|A|\,|x|}\right\|_\infty + \left\|\boldsymbol{L}\mathcal{K}\mathcal{K}^\top D_{|A^\top|\,|r|}\right\|_\infty + \left\|\boldsymbol{L}C_A^\dagger D_{|C|\,|x|}\right\|_\infty \\
&\quad + \left\|\boldsymbol{L}\mathcal{K}\mathcal{K}^\top D_{|C^\top|\,|(AC_A^\dagger)^\top r|}\right\|_\infty + \|\boldsymbol{L}\mathcal{K}D_b\|_\infty + \left\|\boldsymbol{L}C_A^\dagger D_d\right\|_\infty ,
\end{aligned}$$

where in the last equality we use (2.17). Thus we can prove the upper bound for $\kappa_\infty^{\mathrm{rel}}$ based the above inductions. Similarly, we can derive the upper bound for κ_c. □

Remark 1. From some examples in Section 5, the upper bounds $\kappa_\infty^{\mathrm{U}}$ and κ_c^{U} are attainable, thus they are sharp. On the other hand, when $\alpha = \beta = 1$ and $\boldsymbol{L} = I_n$, the upper bound for κ_n can be simplified to

$$\kappa_n^{\mathrm{U}} = \left(\|\mathcal{K}\|_2 + \left\|C_A^\dagger\right\|_2 \|A\|_F + \frac{\|r\|_2\left(1 + \left\|AC_A^\dagger\right\|_2\right)}{\|x\|_2}\|\mathcal{K}\|_2^2 \right) \times \left\| \left(\begin{bmatrix} A \\ C \end{bmatrix}, \begin{bmatrix} b \\ d \end{bmatrix} \right) \right\|_{\mathcal{F}} \sqrt{3 + \frac{1}{\|x\|_2^2}}.$$

Then the above bound indicates that if the residual r is small or zero, the sensitivity is governed by $\|\mathcal{K}\|_2$ and $\left\|C_A^{\dagger}\right\|_2$, otherwise by $\|\mathcal{K}\|_2^2$.

If the factorization (2.14) is computed, the following expressions can be verified:

$$\begin{aligned}
\mathcal{K} &= Q\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & L_{22}^{-1}\end{bmatrix}U^{\top}, \quad \mathcal{K}\mathcal{K}^{\top} = Q\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & L_{22}^{-1}L_{22}^{-\top}\end{bmatrix}Q^{\top},\\
C_A^{\dagger} &= Q\begin{bmatrix} I_p\\ -L_{22}^{-1}L_{21}\end{bmatrix}S^{-1}, \quad (AC_A^{\dagger})^{\top}r = S^{-\top}L_{11}^{\top}(v_1 - L_{11}y_1),\\
r &= U\begin{bmatrix} v_1 - L_{11}y_1\\ \mathbf{0}\end{bmatrix}, \quad \mathcal{K}r = \mathbf{0}.
\end{aligned} \tag{2.18}$$

Thus, for the each term in κ_n^{U}, $\kappa_{\infty}^{\mathrm{U}}$ and κ_c^{U}, we can use the classical condition estimation method [23–25] to estimate them. This method is an efficient method for estimating one-norm of a matrix $\mathcal{B}$, which involves a sequence of matrix–vector multiplications $\mathcal{B}v$ and $\mathcal{B}^{\top}v$. By taking account of the decompositions (2.18), the matrix–vector multiplications during estimating each term of κ_n^{U}, $\kappa_{\infty}^{\mathrm{U}}$ and κ_c^{U} can be computed through solving some triangular linear system with different right hands. Thus the computational complexity of the algorithms to estimate κ_n^{U}, $\kappa_{\infty}^{\mathrm{U}}$ and κ_c^{U} can be reduced significantly compared with the GQR method. The detailed descriptions of the Hager–Higham algorithm [23–25] to estimate κ_n^{U}, $\kappa_{\infty}^{\mathrm{U}}$ and κ_c^{U} are omitted.

3. Revisiting previous results on normwise condition numbers for LSE

The normwise condition number for LSE has been investigated in [2,21], respectively. In this section, we will prove that κ_n in Theorem 1 is identical to κ_2 given by (3.2) [21] under some reasonable assumptions. Also the relationship between κ_1 of [2] and κ_2 is investigated.

Cox and Higham [2] defined the *relative* normwise condition number for LSE as follows

$$\mathrm{cond}(A, C, b, d) := \lim_{\epsilon_1\to 0}\sup \frac{\|\Delta x\|_2}{\epsilon_1\,\|x\|_2},$$

where $x + \Delta x$ is the unique solution to the perturbed LSE (2.6), $\epsilon_1 = \min\{\epsilon : \|\Delta A\|_F \leqslant \epsilon\|A\|_F,\ \|\Delta b\|_2 \leqslant \epsilon\|b\|_2,\ \|\Delta C\|_F \leqslant \epsilon\|C\|_F,\ \|\Delta d\|_2 \leqslant \epsilon\|d\|_2\}$ and proved that

$$\frac{\|\Delta x\|_2}{\|x\|_2} \le \kappa_1\epsilon_1 + \mathcal{O}(\epsilon_1^2),$$

where

$$\begin{aligned}
\kappa_1 = \Big(&\|C_A^{\dagger}\|_2\|d\|_2 + \|\mathcal{K}\|_2\|b\|_2 + \left\|x^{\top}\otimes C_A^{\dagger} + \left[(r^{\top}AC_A^{\dagger})\otimes\left(\mathcal{K}\mathcal{K}^{\top}\right)\right]\Pi\right\|_2\|C\|_F\\
&\left\|-x^{\top}\otimes\mathcal{K} + \left[r^{\top}\otimes(\mathcal{K}\mathcal{K}^{\top})\right]\Pi\right\|_2\|A\|_F\Big)/\|x\|_2,
\end{aligned} \tag{3.1}$$

and the following relationship holds

$$\mathrm{cond}(A, C, b, d) \le \kappa_1 \le 4\,\mathrm{cond}(A, C, b, d).$$

Li and Wang [21] defined the *absolute* normwise condition number for a linear function of the solution x to LSE as follows:

$$\kappa_2^{\mathrm{abs}} := \lim_{\epsilon_2\to 0}\sup\frac{\|\boldsymbol{L}\Delta x\|_2}{\epsilon_2} = \|M_1\|_2 = \|K\|_2^{1/2},$$

where $x + \Delta x$ is the unique solution to the perturbed LSE (2.6),

$$\epsilon_2 := \sqrt{\alpha_A^2\|\Delta A\|_F^2 + \alpha_C^2\|\Delta C\|_F^2 + \alpha_b^2\|\Delta b\|_2^2 + \alpha_d^2\|\Delta d\|_2^2}$$

with $\alpha_A > 0$, $\alpha_C > 0$, $\alpha_b > 0$ and $\alpha_d > 0$, and

$$\begin{aligned}
M_1 &= \left[\frac{\left[r^{\top}\otimes(\boldsymbol{L}\mathcal{K}\mathcal{K}^{\top})\right]\Pi - x^{\top}\otimes(\boldsymbol{L}\mathcal{K})}{\alpha_A}\quad -\frac{x^{\top}\otimes C_A^{\dagger} + \left[(r^{\top}AC_A^{\dagger})\otimes\left(\mathcal{K}\mathcal{K}^{\top}\right)\right]\Pi}{\alpha_B}\quad \frac{\boldsymbol{L}\mathcal{K}}{\alpha_b}\quad \frac{\boldsymbol{L}C_A^{\dagger}}{\alpha_d}\right],\\
K &= \left(\frac{\|r\|_2^2}{\alpha_A^2} + \frac{\|r^{\top}AC_A^{\dagger}\|_2^2}{\alpha_C^2}\right)\boldsymbol{L}\left(\left((A\mathcal{P})^{\top}A\mathcal{P}\right)^{\dagger}\right)^2\boldsymbol{L}^{\top}\\
&\quad + \left(\frac{\|x\|_2^2}{\alpha_C^2} + \frac{1}{\alpha_d^2}\right)\boldsymbol{L}C_A^{\dagger}(C_A^{\dagger})^{\top}\boldsymbol{L}^{\top} + \frac{1}{\alpha_C^2}\boldsymbol{L}\left((A\mathcal{P})^{\top}A\mathcal{P}\right)^{\dagger}xr^{\top}AC_A^{\dagger}(C_A^{\dagger})^{\top}\boldsymbol{L}^{\top}\\
&\quad + \frac{1}{\alpha_C^2}\boldsymbol{L}C_A^{\dagger}(C_A^{\dagger})^{\top}A^{\top}rx^{\top}\left((A\mathcal{P})^{\top}A\mathcal{P}\right)^{\dagger}\boldsymbol{L}^{\top} + \left(\frac{\|x\|_2^2}{\alpha_A^2} + \frac{1}{\alpha_b^2}\right)\boldsymbol{L}\left(\left((A\mathcal{P})^{\top}A\mathcal{P}\right)^{\dagger}\right)\boldsymbol{L}^{\top}.
\end{aligned}$$

So the explicit expression for the *relative* normwise condition number for a linear function of the solution x to LSE was given by

$$\begin{aligned}\kappa_2 &:= \lim_{\epsilon_2\to 0}\sup \frac{\|\boldsymbol{L}\Delta x\|_2}{\epsilon_2\,\|\boldsymbol{L}x\|_2}\cdot\sqrt{\alpha_A^2\|A\|_F^2+\alpha_C^2\|C\|_F^2+\alpha_b^2\|b\|_2^2+\alpha_d^2\|d\|_2^2}\\ &=\frac{\|K\|_2^{1/2}}{\|\boldsymbol{L}x\|_2}\cdot\sqrt{\alpha_A^2\|A\|_F^2+\alpha_C^2\|C\|_F^2+\alpha_b^2\|b\|_2^2+\alpha_d^2\|d\|_2^2}.\end{aligned}\tag{3.2}$$

Noting that $\left((A\mathcal{P})^\top A\mathcal{P}\right)^\dagger=\mathcal{K}\mathcal{K}^\top$, the matrix K can be rewritten as

$$\begin{aligned}K &= \left(\frac{\|r\|_2^2}{\alpha_A^2}+\frac{\|r^\top AC_A^\dagger\|_2^2}{\alpha_C^2}\right)\boldsymbol{L}\left(\mathcal{K}\mathcal{K}^\top\right)^2\boldsymbol{L}^\top+\left(\frac{\|x\|_2^2}{\alpha_A^2}+\frac{1}{\alpha_b^2}\right)\boldsymbol{L}\mathcal{K}\mathcal{K}^\top\boldsymbol{L}^\top\\ &\quad+\left(\frac{\|x\|_2^2}{\alpha_C^2}+\frac{1}{\alpha_d^2}\right)\boldsymbol{L}C_A^\dagger(C_A^\dagger)^\top\boldsymbol{L}^\top+\frac{1}{\alpha_C^2}\boldsymbol{L}\mathcal{K}\mathcal{K}^\top xr^\top AC_A^\dagger(C_A^\dagger)^\top\boldsymbol{L}^\top\\ &\quad+\frac{1}{\alpha_C^2}\boldsymbol{L}C_A^\dagger(C_A^\dagger)^\top A^\top rx^\top\mathcal{K}\mathcal{K}^\top\boldsymbol{L}^\top.\end{aligned}\tag{3.3}$$

In view of the definitions of κ_n given by (2.5) and k_2 given by (3.2), if we choose $\alpha_A=\alpha_C=\alpha$ and $\alpha_b=\alpha_d=\beta$, it is not difficult to see that the definition of κ_n is equivalent to the definition of κ_2. In the proposition below, we will prove that they are identical from their expressions.

Proposition 1. *Let the expressions of κ_n and κ_2 be given by Theorem 1 and (3.2), respectively, then when $\alpha_A=\alpha_C=\alpha$ and $\alpha_b=\alpha_d=\beta$ hold, we have $\kappa_n=\kappa_2$.*

Proof. From (2.15), and recalling that $\mathcal{N}$ is given by (2.10), it can be verified that

$$\begin{aligned}\mathcal{N}\mathcal{N}^\top &= \begin{bmatrix}-\frac{1}{\beta}\|x\|_2\mathcal{K} & -\frac{1}{\beta}\|x\|_2C_A^\dagger & \frac{1}{\alpha}\|t\|_2\mathcal{K}\mathcal{K}^\top\end{bmatrix}\begin{bmatrix}(\frac{\beta^2}{\alpha^2}+\frac{1}{\|x\|_2^2})I_{m+p} & \frac{\beta}{\alpha}\frac{tx^\top}{\|t\|_2\|x\|_2}\\ \frac{\beta}{\alpha}\frac{xt^\top}{\|t\|_2\|x\|_2} & I_n\end{bmatrix}\begin{bmatrix}-\frac{1}{\beta}\|x\|_2\mathcal{K}^\top\\ -\frac{1}{\beta}\|x\|_2(C_A^\dagger)^\top\\ \frac{1}{\alpha}\|t\|_2\mathcal{K}\mathcal{K}^\top\end{bmatrix}\\ &=\left(\frac{1}{\beta^2}+\frac{1}{\alpha^2}\|x\|_2^2\right)\left(\mathcal{K}\mathcal{K}^\top+C_A^\dagger(C_A^\dagger)^\top\right)+\frac{1}{\alpha^2}\left(\|r\|_2^2+\|r^\top AC_A^\dagger\|_2^2\right)\left(\mathcal{K}\mathcal{K}^\top\right)^2\\ &\quad-\frac{1}{\alpha^2}\begin{bmatrix}\mathcal{K} & C_A^\dagger\end{bmatrix}\begin{bmatrix}r\\ -(AC_A^\dagger)^\top r\end{bmatrix}x^\top\mathcal{K}\mathcal{K}^\top-\frac{1}{\alpha^2}\mathcal{K}\mathcal{K}^\top x\begin{bmatrix}r^\top & -r^\top AC_A^\dagger\end{bmatrix}\begin{bmatrix}\mathcal{K}^\top\\ (C_A^\dagger)^\top\end{bmatrix}\\ &=\left(\frac{1}{\beta^2}+\frac{\|x\|_2^2}{\alpha^2}\right)\left(\mathcal{K}\mathcal{K}^\top+C_A^\dagger(C_A^\dagger)^\top\right)+\frac{\|r\|_2^2+\|r^\top AC_A^\dagger\|_2^2}{\alpha^2}\left(\mathcal{K}\mathcal{K}^\top\right)^2\\ &\quad+\frac{1}{\alpha^2}C_A^\dagger(AC_A^\dagger)^\top rx^\top\mathcal{K}\mathcal{K}^\top+\frac{1}{\alpha^2}\mathcal{K}\mathcal{K}^\top xr^\top AC_A^\dagger(C_A^\dagger)^\top,\end{aligned}\tag{3.4}$$

where in the last equality we use the fact $\mathcal{K}r=\mathbf{0}$ from (2.18). Thus comparing (3.4) with (3.3), we can prove that

$$\boldsymbol{L}\mathcal{N}\mathcal{N}^\top\boldsymbol{L}^\top=K$$

whenever $\alpha_A=\alpha_C=\alpha$ and $\alpha_b=\alpha_d=\beta$. Using the fact that $\|K\|_2=\|\boldsymbol{L}\mathcal{N}\|_2^2$ and considering the expressions of κ_n and κ_2, we complete the proof. □

Li and Wang in [21, Remark 2.1] derived that

$$\|M_1\|_2\le\gamma_{\rm up}\le 4\,\|M_1\|_2,$$

where

$$\begin{aligned}\gamma_{\rm up} &= \|C_A^\dagger\|_2+\|\mathcal{K}\|_2+\left\|x^\top\otimes C_A^\dagger+\left[(r^\top AC_A^\dagger)\otimes\left(\mathcal{K}\mathcal{K}^\top\right)\right]\Pi\right\|_2\\ &\quad+\left\|\left[r^\top\otimes(\mathcal{K}\mathcal{K}^\top)\right]\Pi-x^\top\otimes\mathcal{K}\right\|_2,\end{aligned}$$

which is related to κ_1 given by (3.1). However, the relationship between *relative* normwise condition numbers κ_1 and κ_2 is not investigated. In Proposition 2, we will prove that there are no big differences between κ_1 and κ_2 from their explicit expressions (3.1) and (3.2) for the common choices $\boldsymbol{L}=I_n$, and $\alpha_A=\alpha_C=\alpha_b=\alpha_d=1$ of κ_2.

Proposition 2. *With the notations above, when $\boldsymbol{L} = I_n$, and $\alpha_A = \alpha_C = \alpha_b = \alpha_d = 1$, we have*

$$\kappa_1 \le 2\,\kappa_2 \le 4\,\kappa_1.$$

Proof. Using Cauchy–Schwarz inequality and $\|\mathcal{B}\|_2 \le \|\mathcal{A}\|_2$ where $\mathcal{B}$ is a submatrix of $\mathcal{A}$, we have

$$\begin{aligned}
&\|C_A^{\dagger}\|_2\|d\|_2 + \|\mathcal{K}\|_2\|b\|_2 + \left\|x^\top \otimes C_A^{\dagger} + \left[(r^\top A C_A^{\dagger}) \otimes \left(\mathcal{K}\mathcal{K}^\top\right)\right]\Pi\right\|_2 \|C\|_F \\
&\quad + \left\|-x^\top \otimes \mathcal{K} + \left[r^\top \otimes (\mathcal{K}\mathcal{K}^\top)\right]\Pi\right\|_2 \|A\|_F \\
&\le \sqrt{\|C_A^{\dagger}\|_2^2 + \|\mathcal{K}\|_2^2 + \left\|x^\top \otimes C_A^{\dagger} + \left[(r^\top A C_A^{\dagger}) \otimes \left(\mathcal{K}\mathcal{K}^\top\right)\right]\Pi\right\|_2^2 + \left\|\left[r^\top \otimes (\mathcal{K}\mathcal{K}^\top)\right]\Pi - x^\top \otimes \mathcal{K}\right\|_2^2} \\
&\quad \cdot \sqrt{\|d\|_2^2 + \|b\|_2^2 + \|C\|_F^2 + \|A\|_F^2} \\
&\le 2\,\|M_1\|_2 \cdot \sqrt{\|d\|_2^2 + \|b\|_2^2 + \|C\|_F^2 + \|A\|_F^2}
\end{aligned}$$

whenever $\boldsymbol{L} = I_n$, and $\alpha_A = \alpha_C = \alpha_b = \alpha_d = 1$. Thus from the expression of κ_1 and κ_2, we prove that $\kappa_1 \le 2\,\kappa_2$.

On the other hand, for a given matrix $\mathcal{A}$ which is partitioned as $\mathcal{A} = \begin{bmatrix}\mathcal{A}_1 & \mathcal{A}_2 & \mathcal{A}_3 & \mathcal{A}_4\end{bmatrix}$, using the fact $\|\mathcal{A}\|_2 \le \sum_{i=1}^4 \|\mathcal{A}_i\|_2$, it is not difficult to prove that

$$\begin{aligned}
\|K\|_2^{1/2} &= \|M_1\|_2 \le \|C_A^{\dagger}\|_2 + \|\mathcal{K}\|_2 + \left\|x^\top \otimes C_A^{\dagger} + \left[(r^\top A C_A^{\dagger}) \otimes \left(\mathcal{K}\mathcal{K}^\top\right)\right]\Pi\right\|_2 \\
&\quad + \left\|\left[r^\top \otimes (\mathcal{K}\mathcal{K}^\top)\right]\Pi - x^\top \otimes \mathcal{K}\right\|_2 \\
&= \|C_A^{\dagger}\|_2\|d\|_2 \frac{1}{\|d\|_2} + \left\|x^\top \otimes C_A^{\dagger} + \left[(r^\top A C_A^{\dagger}) \otimes \left(\mathcal{K}\mathcal{K}^\top\right)\right]\Pi\right\|_2 \|C\|_F \frac{1}{\|C\|_F} \\
&\quad + \|\mathcal{K}\|_2\|b\|_2 \frac{1}{\|b\|_2} + \left\|\left[r^\top \otimes (\mathcal{K}\mathcal{K}^\top)\right]\Pi - x^\top \otimes \mathcal{K}\right\|_2 \|A\|_F \frac{1}{\|d\|_F} \\
&\le \Bigg(\|C_A^{\dagger}\|_2\|d\|_2 + \|\mathcal{K}\|_2\|b\|_2 + \left\|x^\top \otimes C_A^{\dagger} + \left[(r^\top A C_A^{\dagger}) \otimes \left(\mathcal{K}\mathcal{K}^\top\right)\right]\Pi\right\|_2 \|C\|_F \\
&\quad \left\|-x^\top \otimes \mathcal{K} + \left[r^\top \otimes (\mathcal{K}\mathcal{K}^\top)\right]\Pi\right\|_2 \|A\|_F\Bigg) \max\left\{\frac{1}{\|A\|_F}, \frac{1}{\|C\|_F}, \frac{1}{\|b\|_2}, \frac{1}{\|d\|_2}\right\}.
\end{aligned}$$

From the above inequality, and the expressions of κ_1 and κ_2, we finish the proof. $\square$

4. Revisiting previous results on condition numbers for LS

In this section, we will give the explicit expressions of mixed and componentwise condition numbers for a linear function of the solution to LS in Proposition 3, which can recover the corresponding ones given in [26] when $\boldsymbol{L} = I_n$ and is identical to the counterparts of [16]. We will derive the former explicit formulas of the normwise condition numbers of LS [27] and [15] from the explicit expression of κ_n when $C = \mathbf{0}$ and $d = \mathbf{0}$.

Gratton [27] introduced and derived the normwise absolute condition number for LS as follows:

$$\kappa_{\mathrm{LS}}^{G} := \lim_{\varepsilon \to 0} \sup_{\|(\Delta A,\, \Delta b)\|_{\mathcal{F}} \le \varepsilon} \frac{\|\Delta x\|_2}{\varepsilon} = \|A^{\dagger}\|_2 \sqrt{\frac{1}{\beta^2} + \frac{\|x\|_2^2 + \|r\|_2^2\,\|A^{\dagger}\|_2^2}{\alpha^2}}, \tag{4.1}$$

where $x + \Delta x$ is the unique solution to the perturbed LS:

$$\min_{x+\Delta x} \|(A+\Delta A)(x+\Delta x) - (b+\Delta b)\|_2\,.$$

Later, Arioli et al. [15] defined the partial normwise absolute condition number for LS as follows:

$$\kappa_{\mathrm{LS},\boldsymbol{L}} := \lim_{\varepsilon \to 0} \sup_{\|(\Delta A,\, \Delta b)\|_{\mathcal{F}} \le \varepsilon} \frac{\|\boldsymbol{L}\Delta x\|_2}{\varepsilon} = \left\|S V^\top L^\top\right\|_2, \tag{4.2}$$

where $A = U\Sigma V^\top$ is the thin SVD of A, $\Sigma = \mathrm{diag}(\sigma_i)$ with $\sigma_1 \ge \sigma_2 \ge \cdots \sigma_n > 0$, and

$$S = \mathrm{diag}\left(\sigma_i^{-1}\sqrt{\frac{\sigma_i^{-2}\|r\|_2^2 + \|x\|_2^2}{\alpha^2} + \frac{1}{\beta^2}}\right).$$

Cucker et al. [26] studied the mixed and componentwise condition numbers for LS as below:

$$m(A,b) := \lim_{\varepsilon \to 0} \sup_{\substack{|\Delta A| \le \varepsilon |A|,\\ |\Delta b| \le \varepsilon |b|}} \frac{\|\Delta x\|_\infty}{\varepsilon \|x\|_\infty} = \frac{\left\|\left|\left((A^\top A)^{-1} \otimes r^\top - x^\top \otimes A^{\dagger}\right)\right| \mathrm{vec}(|A|) + |A^{\dagger}|\,|b|\right\|_\infty}{\|x\|_\infty},$$

$$c(A,b) := \lim_{\varepsilon\to 0} \sup_{\substack{|\Delta A|\le\varepsilon|A|\\ |\Delta b|\le\varepsilon|b|}} \frac{1}{\varepsilon}\left\|\frac{\Delta x}{x}\right\|_\infty = \left\|\frac{\left|\left((A^\top A)^{-1}\otimes r^\top - x^\top\otimes A^\dagger\right)\right| \mathrm{vec}(|A|) + |A^\dagger|\,|b|}{x}\right\|_\infty . \tag{4.3}$$

In [16], Baboulin and Gratton considered the mixed and componentwise condition numbers for a linear function of the solution to LS and their explicit expressions were given in [16, Equations (3.3) and (3.7)] as below

$$\kappa^{\mathrm{abs}}_{\mathrm{LS},\infty} = \lim_{\varepsilon\to 0} \sup_{\substack{|\Delta A|\le\varepsilon|A|,\\ |\Delta b|\le\varepsilon|b|}} \frac{\|\boldsymbol{L}\Delta x\|_\infty}{\varepsilon} = \left\|\left|\left((\boldsymbol{L}(A^\top A)^{-1})\otimes r^\top - x^\top\otimes(\boldsymbol{L}A^\dagger)\right)\right|\mathrm{vec}(|A|) + |\boldsymbol{L}A^\dagger|\,|b|\right\|_\infty ,$$

$$\kappa^{G}_{\mathrm{LS},c} = \lim_{\varepsilon\to 0} \sup_{\substack{|\Delta A|\le\varepsilon|A|\\ |\Delta b|\le\varepsilon|b|}} \frac{1}{\varepsilon}\left\|\frac{\boldsymbol{L}\Delta x}{\boldsymbol{L}x}\right\|_\infty = \left\|\frac{\left|\left((\boldsymbol{L}(A^\top A)^{-1})\otimes r^\top - x^\top\otimes(\boldsymbol{L}A^\dagger)\right)\right|\mathrm{vec}(|A|) + |\boldsymbol{L}A^\dagger|\,|b|}{\boldsymbol{L}x}\right\|_\infty ,$$

When $C = \mathbf{0}$ and $d = \mathbf{0}$, LSE (1.1) is reduced to LS (1.6). Since the rank condition (1.2) guarantees $\mathrm{rank}(A) = n$, it is easy to verify that

$$\mathcal{P} = I_n, \quad \mathcal{K} = A^\dagger, \quad C_A^\dagger = \mathbf{0}, \tag{4.4}$$

where $\mathcal{P}$ and $\mathcal{K}$ are defined in (1.3). With the property $A^\dagger(A^\dagger)^\top = (A^\top A)^{-1}$ for the full column rank matrix A, the matrices $\mathcal{M}$ and $\mathcal{J}$ given by (2.10) are simplified to

$$\mathcal{M} = A^\dagger((A^\dagger)^\top\otimes r^\top - x^\top\otimes I_m) = (A^\top A)^{-1}\otimes r^\top - x^\top\otimes A^\dagger, \quad \mathcal{J} = \mathbf{0}.$$

Thus from Theorem 1, we have the following proposition.

Proposition 3. *The mixed and componentwise condition numbers for a linear function of the solution to LS can be defined and characterized by*

$$\kappa^{\mathrm{rel}}_{\mathrm{LS},\infty} = \lim_{\varepsilon\to 0} \sup_{\substack{|\Delta A|\le\varepsilon|A|,\\ |\Delta b|\le\varepsilon|b|}} \frac{\|\boldsymbol{L}\Delta x\|_\infty}{\varepsilon\|\boldsymbol{L}x\|_\infty} = \frac{\left\|\left|\boldsymbol{L}\left((A^\top A)^{-1}\otimes r^\top - x^\top\otimes A^\dagger\right)\right|\mathrm{vec}(|A|) + |\boldsymbol{L}A^\dagger|\,|b|\right\|_\infty}{\|\boldsymbol{L}x\|_\infty},$$

$$\kappa_{\mathrm{LS},c} = \lim_{\varepsilon\to 0} \sup_{\substack{|\Delta A|\le\varepsilon|A|\\ |\Delta b|\le\varepsilon|b|}} \frac{1}{\varepsilon}\left\|\frac{\boldsymbol{L}\Delta x}{\boldsymbol{L}x}\right\|_\infty = \left\|\frac{\left|\boldsymbol{L}\left((A^\top A)^{-1}\otimes r^\top - x^\top\otimes A^\dagger\right)\right|\mathrm{vec}(|A|) + |\boldsymbol{L}A^\dagger|\,|b|}{\boldsymbol{L}x}\right\|_\infty ,$$

which can recover the corresponding ones given by (4.3) *when we take* $\boldsymbol{L} = I_n$.

Remark 2. From the Kronecker product property (2.7), it is easy to see that $\kappa^{\mathrm{rel}}_{\mathrm{LS},\infty} = \kappa^{\mathrm{abs}}_{\mathrm{LS},\infty}/\|\boldsymbol{L}x\|_\infty$ and $\kappa_{\mathrm{LS},c} = \kappa^{G}_{\mathrm{LS},c}$. Here we use a different methodology to deduce the expressions of $\kappa^{\mathrm{rel}}_{\mathrm{LS},\infty}$ and $\kappa_{\mathrm{LS},c}$ while Baboulin and Gratton [16] adopted the dual techniques to derive the corresponding expressions of $\kappa^{\mathrm{abs}}_{\mathrm{LS},\infty}$ and $\kappa^{G}_{\mathrm{LS},c}$.

In Proposition 4, we will derive the explicit formulas (4.1) and (4.2) of κ^G_{LS} and $\kappa_{\mathrm{LS},\boldsymbol{L}}$ directly from the explicit expression κ_n given by Theorem 1 when we let $C = \mathbf{0}$ and $d = \mathbf{0}$. We should remark that the following proof of deriving the explicit formulas (4.1) and (4.2) is different and simpler compared with the former deductions for κ^G_{LS} and $\kappa_{\mathrm{LS},\boldsymbol{L}}$ in [15,27], respectively.

Proposition 4. *The expressions of* κ^G_{LS} *and* $\kappa_{\mathrm{LS},\boldsymbol{L}}$, *which are given by* (4.1) *and* (4.2) *respectively, can be derived from the expression of* κ_n *given by* Theorem 1.

Proof. Plugging (4.4), $C = \mathbf{0}$ and $d = \mathbf{0}$ into (3.4), we have

$$\mathcal{N}\mathcal{N}^\top = \left(\frac{1}{\beta^2} + \frac{\|x\|_2^2}{\alpha^2}\right)A^\dagger(A^\dagger)^\top + \frac{\|r\|_2^2}{\alpha^2}\left(A^\dagger(A^\dagger)^\top\right)^2, \tag{4.5}$$

Thus it can be check that

$$(\sigma_1(\mathcal{N}))^2 = \left(\frac{1}{\beta^2} + \frac{\|x\|_2^2}{\alpha^2}\right)(\sigma_1(A^\dagger))^2 + \frac{\|r\|_2^2}{\alpha^2}(\sigma_1(A^\dagger))^4, \tag{4.6}$$

where $\sigma_i(\mathcal{A})$ is the ith largest singular values of a matrix $\mathcal{A}$. Comparing the definitions of κ_n given by (2.5) and κ^G_{LS} given by (4.1), from Theorem 1 and (4.6), it is easy to see that

$$\kappa^G_{\mathrm{LS}} = \frac{\|\boldsymbol{L}x\|_2}{\left\|\left(\begin{bmatrix}A\\C\end{bmatrix}, \begin{bmatrix}b\\d\end{bmatrix}\right)\right\|_{\mathcal{F}}}\kappa_n = \|\boldsymbol{L}\mathcal{N}\|_2 = \sqrt{\left(\frac{1}{\beta^2} + \frac{\|x\|_2^2}{\alpha^2}\right)(\sigma_1(A^\dagger))^2 + \frac{\|r\|_2^2}{\alpha^2}(\sigma_1(A^\dagger))^4}$$

$$= \|A^\dagger\|_2\sqrt{\frac{1}{\beta^2} + \frac{\|x\|_2^2 + \|r\|_2^2\,\|A^\dagger\|_2^2}{\alpha^2}}$$

whenever $\boldsymbol{L} = I_n$, $C = \mathbf{0}$ and $d = \mathbf{0}$. Using the thin SVD of A, from (4.5), we have

$$\boldsymbol{L}\mathcal{N}\mathcal{N}^\top\boldsymbol{L}^\top = \boldsymbol{L}V\left(\left(\frac{1}{\beta^2} + \frac{\|x\|_2^2}{\alpha^2}\right)\Sigma^{-2} + \frac{\|r\|_2^2}{\alpha^2}\Sigma^{-4}\right)V^\top\boldsymbol{L}^\top = \boldsymbol{L}VS^2V^\top\boldsymbol{L}^\top.$$

Thus we know that $\|\boldsymbol{L}\mathcal{N}\|_2 = \left\|SV^\top\boldsymbol{L}^\top\right\|_2$, and in view of the definitions of κ_n given by (2.5) and $\kappa_{\mathrm{LS},\boldsymbol{L}}$ given by (4.2), we can prove that

$$\kappa_{\mathrm{LS},\boldsymbol{L}} = \frac{\|\boldsymbol{L}x\|_2}{\left\|\left(\begin{bmatrix}A\\C\end{bmatrix},\begin{bmatrix}b\\d\end{bmatrix}\right)\right\|_{\mathcal{F}}}\kappa_n = \|\boldsymbol{L}\mathcal{N}\|_2 = \left\|SV^\top\boldsymbol{L}^\top\right\|_2,$$

which completes the proof of this proposition. □

5. Numerical examples

In this section we will do some numerical examples. In order to check the effectiveness of proposed condition numbers, we will compare the exact relative errors of the solution with the linear asymptotic perturbation bounds based on the product of condition numbers and the pre-generated perturbation magnitude. Thus in the following, we will first construct the data for the LSE problem, then generate perturbations on the data, and finally compute the unperturbed and perturbed solutions corresponding to the unperturbed and perturbed data, respectively. Since the GQR method [2,6,8,30] is proved to be numerically stable [2], in this section we adopt the GQR method to compute the solutions to LSE. All the computations are carried out using Matlab 8.1 with the machine precision $\mu = 2.2 \times 10^{-16}$.

Example 1. Let v be a 4×1 vector with $v_4 = 1/\eta$ where η is a small positive number, and other components are set to 1. We construct the data A, C, b and d as follows:

$$A = \begin{bmatrix}1&0&0&0\\0&0&0&0\\0&1&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&\delta&0\\0&0&0&0\\0&0&0&\delta\end{bmatrix} \in \mathbb{R}^{9\times4},\ C = \begin{bmatrix}0&1&0&0\\1&0&0&0\end{bmatrix} \in \mathbb{R}^{2\times4},\ b = A\cdot v + 10^{-5}\cdot b_2,\ d = \begin{bmatrix}1\\1\end{bmatrix},$$

where b_2 is a unitary vector satisfying $A^\top b_2 = \mathbf{0}$ and δ is a small positive number. Obviously, the matrix A has many zero components and is bad scaled because of the appearance of δ. Thus, it is reasonable to measure the error on the input data by using componentwise perturbation analysis instead of the normwise perturbation analysis. Also it can be verified that the rank conditions (1.2) are satisfied. For the perturbations, we generate them as

$$\begin{aligned}&\Delta A = 10^{-\ell}\cdot\Delta A_1 \odot A,\ \Delta C = 10^{-\ell}\cdot\Delta C_1 \odot C,\\&\Delta b = 10^{-\ell}\cdot\Delta b_1 \odot b,\ \Delta d = 10^{-\ell}\cdot\Delta d_1 \odot d,\end{aligned} \tag{5.1}$$

where $\ell \in \mathbb{N}$ describes the perturbation magnitude, each component of $\Delta A_1 \in \mathbb{R}^{m\times n}$, $\Delta C_1 \in \mathbb{R}^{p\times n}$, $\Delta b_1 \in \mathbb{R}^m$ and $\Delta d_1 \in \mathbb{R}^p$ with $m = 9$, $n = 4$ and $p = 2$ is uniformly distributed in the interval $(-1, 1)$, and $\odot$ denotes the componentwise multiplication of two conformal dimensional matrices. When the perturbations are small enough, we denote the unique solution by $\tilde{x}$ of the perturbed LSE problem (2.6). We use the GQR method [2] to compute the solution x and the perturbed solution $\tilde{x}$ separately. Usually the solution x has badly scaled components, for example, the last component of x is of order $1/\eta$ while other components of x are of $\mathcal{O}(1)$.

For the $\boldsymbol{L}$ matrix in our condition numbers, we choose

$$\boldsymbol{L}_0 = I_4,\quad \boldsymbol{L}_1 = \begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&0&1\end{bmatrix} \in \mathbb{R}^{3\times4},\quad \boldsymbol{L}_2 = \begin{bmatrix}0&0&0&1\end{bmatrix} \in \mathbb{R}^{1\times4}.$$

Thus, corresponding to the above three matrices, the whole x, the subvector $[x_1,\ x_2,\ x_3]^\top$, and the component x_4 are selected respectively.

We measure the normwise, mixed and componentwise relative errors in $\boldsymbol{L}x$ defined by

$$r_2^{\mathrm{rel}} = \frac{\|\boldsymbol{L}\tilde{x} - \boldsymbol{L}x\|_2}{\|\boldsymbol{L}x\|_2},\quad r_\infty^{\mathrm{rel}} = \frac{\|\boldsymbol{L}\tilde{x} - \boldsymbol{L}x\|_\infty}{\|\boldsymbol{L}x\|_\infty},\quad r_c^{\mathrm{rel}} = \left\|\frac{\boldsymbol{L}\tilde{x} - \boldsymbol{L}x}{\boldsymbol{L}x}\right\|_\infty.$$

Table 1
Comparison of condition numbers with the corresponding relative errors for Example 1.

L	r_2^{rel}	κ_1	κ_n	κ_n^{U}	r_∞^{rel}	$\kappa_\infty^{\mathrm{rel}}$	$\kappa_\infty^{\mathrm{U}}$	r_c^{rel}	κ_c	κ_c^{U}
$\ell=6$,	$\eta=10^{-3}$,	$\delta=10^{-3}$								
I_4	7.91e−07	1.42e+03	3.00e+03	5.20e+03	7.91e−07	2.00e+00	2.00e+00	7.91e−07	2.00e+00	4.00e+00
L_1	1.79e−07		1.73e+06	3.00e+06	2.97e−07	2.00e+00	4.00e+00	2.97e−07	2.00e+00	4.00e+00
L_2	7.91e−07		3.00e+03	5.20e+03	7.91e−07	2.00e+00	2.00e+00	7.91e−07	2.00e+00	2.00e+00
$\ell=6$,	$\eta=10^{-3}$,	$\delta=10^{-6}$								
I_4	6.18e−07	1.42e+06	2.83e+06	5.00e+06	6.18e−07	2.00e+00	2.00e+00	6.18e−07	2.00e+00	4.00e+00
L_1	4.57e−08		1.63e+09	2.89e+09	6.48e−08	2.00e+00	4.00e+00	6.48e−08	2.00e+00	4.00e+00
L_2	6.18e−07		2.83e+06	5.00e+06	6.18e−07	2.00e+00	2.00e+00	6.18e−07	2.00e+00	2.00e+00
$\ell=12$,	$\eta=10^{-3}$,	$\delta=10^{-3}$								
I_4	1.46e−13	1.42e+03	3.00e+03	5.20e+03	1.46e−13	2.00e+00	2.00e+00	7.54e−13	2.00e+00	4.00e+00
L_1	4.44e−13		1.73e+06	3.00e+06	7.54e−13	2.00e+00	4.00e+00	7.54e−13	2.00e+00	4.00e+00
L_2	1.46e−13		3.00e+03	5.20e+03	1.46e−13	2.00e+00	2.00e+00	1.46e−13	2.00e+00	2.00e+00
$\ell=12$,	$\eta=10^{-3}$,	$\delta=10^{-6}$								
I_4	3.78e−13	1.42e+06	2.83e+06	5.00e+06	3.78e−13	2.00e+00	2.00e+00	8.66e−13	2.00e+00	4.00e+00
L_1	6.98e−13		1.63e+09	2.89e+09	8.66e−13	2.00e+00	4.00e+00	8.66e−13	2.00e+00	4.00e+00
L_2	3.78e−13		2.83e+06	5.00e+06	3.78e−13	2.00e+00	2.00e+00	3.78e−13	2.00e+00	2.00e+00

And in the rest of this section we always consider the case $\alpha = \beta = 1$ for the normwise condition number κ_n. From the definition of condition numbers, the following quantities

$$10^{-\ell} \cdot \kappa_1, \text{ and } 10^{-\ell} \cdot \kappa_n \tag{5.2}$$

are the linear normwise asymptotic perturbation bounds. On the other hand,

$$10^{-\ell} \cdot \kappa_\infty^{\mathrm{rel}}, \text{ and } 10^{-\ell} \cdot \kappa_c, \tag{5.3}$$

are the linear mixed and componentwise asymptotic perturbation bounds, respectively.

In Table 1, we do numerical experiments for different choices of ℓ, η and δ. It can be observed that when δ varies from 10^{-3} to 10^{-6}, LSE tends to be more ill-conditioned with respect to the decreasing of δ in the sense of normwise perturbation analysis, while the mixed and componentwise condition numbers are always $\mathcal{O}(1)$. Thus for these particular LSE problems, they are less sensitive to the componentwise perturbations. When the perturbation magnitudes $10^{-\ell}$ decrease, the corresponding relative errors also decrease as shown in the first column of Table 1. The linear normwise, mixed and componentwise asymptotic perturbation bounds defined by (5.2) and (5.3) can always bound the exact relative errors, and we should point out the linear normwise asymptotic perturbation bounds are pessimistic since the normwise condition numbers ignore the scaling and sparsity structure of the data. On the other hand, linear mixed and componentwise asymptotic perturbation bounds $10^{-\ell} \cdot \kappa_\infty^{\mathrm{rel}}$ and $10^{-\ell} \cdot \kappa_c$ are greater than the exact relative errors r_∞^{rel} and r_c^{rel} at most of one hundredfold, respectively. Most of them are only tenfold of the corresponding relative errors. The derived upper bounds $\kappa_\infty^{\mathrm{U}}$ and κ_c^{U} can be equal to the exact mixed and componentwise condition numbers for some cases, for example when $\boldsymbol{L} = \boldsymbol{L}_2$. Thus the proposed upper bounds $\kappa_\infty^{\mathrm{U}}$ and κ_c^{U} are sharp. The normwise upper bounds κ_n^{U} always have the same order of the corresponding κ_n, which means that κ_n^{U} is also effective. Also we should point out that when $\boldsymbol{L} = \boldsymbol{L}_1$, κ_n can be much larger than κ_1. The reason of the above phenomenon is that only the last component of x has the order of $1/\eta$ while other components are $\mathcal{O}(1)$. From our experiments, when $\boldsymbol{L} = \boldsymbol{L}_1$, the values of the numerator of κ_n are approximately equal to values of the numerator of κ_1 for all cases. On the other hand, the denominator of κ_n is about $\mathcal{O}(1)$ and the denominator of κ_n is about $\mathcal{O}(1/\eta)$ because $\boldsymbol{L}_1 x$ does not select the last component of x, thus the ratios of κ_1 and κ_n are about $\mathcal{O}(1/\eta)$ when $\boldsymbol{L} = \boldsymbol{L}_1$. For other choices of $\boldsymbol{L}$, the values of κ_n have the same order of the values of κ_1 and $\kappa_n > \kappa_1$ always holds which coincide with Proposition 2.

Example 2. In this example, we test random LSE problems and always choose $\boldsymbol{L} = I_n$. Let the matrix A, given κ_A, be generated as $A = QDU$, where $D \in \mathbb{R}^{m\times n}$ is a diagonal matrix with decreasing diagonal values geometrically distributed between 1 and κ_A, and $Q \in \mathbb{R}^{m\times m}$ and $U \in \mathbb{R}^{n\times n}$ are random orthogonal matrices generated by the MATLAB built-in function `gallery`(‘`qumlt`’, n). Furthermore, A is normalized such that $\|A\|_2 = 1$. The matrix $B \in \mathbb{R}^{p\times n}$, given its condition number κ_B, is formed by using MATLAB routine $B =$ `gallery`(‘`randsvd`’, $[p, n]$, κ_B) with $\|B\|_2 = 1$ and its singular values are geometrically distributed between 1 and $1/\kappa_B$. We construct the random vectors b and d which are satisfied with the standard Gaussian distribution for LSE. For all the experiments, we choose $m = 50$, $n = 20$ and $p = 15$. The perturbations on the data are generated by (5.1). For each generated data and corresponding perturbed data, we compute the solutions via the GQR method.

Table 2 indicates that the conditioning of LSE is more dominated by the quantity κ_B for these constructed examples. When κ_B increase from 10^2 to 10^8, the values of $\log(\kappa_1)$, $\log(\kappa_n)$, $\log(\kappa_\infty^{\mathrm{rel}})$ and $\log(\kappa_c)$ also increase. Meanwhile, the relative

Table 2
Comparison of condition numbers with the corresponding relative errors for Example 2.

κ_A	r_2^{rel}	κ_1	κ_n	κ_n^{U}	r_∞^{rel}	$\kappa_\infty^{\mathrm{rel}}$	$\kappa_\infty^{\mathrm{U}}$	r_c^{rel}	κ_c	κ_c^{U}
$\ell = 6$,	$\kappa_B = 10^2$									
10	2.58e−06	1.80e+02	4.18e+02	3.19e+03	2.71e−06	1.17e+02	1.85e+02	3.72e−05	1.42e+03	2.08e+03
10^3	3.28e−06	2.91e+02	5.17e+02	2.36e+03	3.98e−06	1.10e+02	1.76e+02	3.62e−05	1.65e+03	2.53e+03
10^5	6.10e−06	2.79e+02	4.41e+02	3.02e+03	5.71e−06	1.26e+02	2.76e+02	3.53e−05	6.43e+03	8.31e+03
10^7	6.92e−06	1.55e+03	2.44e+03	8.34e+03	8.83e−06	1.00e+03	1.31e+03	3.32e−05	3.04e+04	3.93e+04
$\ell = 6$,	$\kappa_B = 10^4$									
10	2.39e−04	1.19e+04	3.39e+04	2.25e+05	1.83e−04	6.07e+03	8.83e+03	9.57e−03	4.78e+05	5.45e+05
10^3	4.92e−04	2.13e+04	4.82e+04	2.34e+05	5.43e−04	1.29e+04	1.70e+04	9.70e−03	4.51e+05	5.70e+05
10^5	4.08e−04	1.55e+04	3.24e+04	1.37e+05	3.85e−04	7.81e+03	1.22e+04	3.89e−03	5.28e+04	8.45e+04
10^7	7.63e−04	2.60e+04	5.11e+04	1.86e+05	6.00e−04	1.12e+04	1.75e+04	1.24e−02	2.26e+05	3.36e+05
$\ell = 6$,	$\kappa_B = 10^6$									
10	2.28e−02	1.55e+06	4.79e+06	2.96e+07	2.69e−02	6.10e+05	1.24e+06	8.07e−02	5.41e+06	8.65e+06
10^3	1.94e−02	1.92e+06	4.63e+06	2.11e+07	2.12e−02	9.92e+05	1.82e+06	3.32e−01	3.38e+07	4.20e+07
10^5	1.93e−02	1.53e+06	3.38e+06	1.36e+07	2.76e−02	8.72e+05	1.38e+06	2.49e−01	1.17e+07	1.72e+07
10^7	2.03e−02	6.49e+06	1.37e+07	3.46e+07	1.46e−02	2.07e+06	2.55e+06	8.95e−01	1.15e+08	1.36e+08
$\ell = 6$,	$\kappa_B = 10^8$									
10	7.91e−01	1.86e+08	5.60e+08	2.61e+09	6.25e−01	1.12e+08	1.46e+08	3.72e+00	9.75e+08	1.14e+09
10^3	9.95e−01	1.41e+08	3.62e+08	1.77e+09	8.35e−01	6.10e+07	1.18e+08	4.48e+00	1.68e+09	2.09e+09
10^5	8.68e−01	2.37e+08	5.36e+08	1.73e+09	8.64e−01	1.16e+08	1.37e+08	3.06e+00	7.15e+08	8.67e+08
10^7	8.53e−01	2.55e+08	5.51e+08	1.62e+09	9.09e−01	7.26e+07	1.10e+08	3.19e+00	8.41e+08	1.07e+09

normwise, mixed and componentwise errors r_2^{rel}, r_∞^{rel} and r_c^{rel} increase. Even for $\kappa_B = 10^8$, the small componentwise perturbation magnitude 10^{-6} on the input data can result in $\mathcal{O}(10^{-1})$ perturbations for normwise/mixed relative errors (r_n^{rel} and r_∞^{rel}) and $\mathcal{O}(1)$ for componentwise relative errors r_c^{rel}. Thus under this case, LSE problem is very ill-conditioned, which is justified by the condition number numerical values from the last four rows of Table 2. However, it can been seen that the exact relative errors are always smaller than the corresponding linear normwise, mixed and componentwise asymptotic perturbation bounds (5.2) and (5.3), which means that the condition numbers are effective. The upper bounds κ_n^{U}, $\kappa_\infty^{\mathrm{U}}$ and κ_c^{U} are always greater than the corresponding normwise, mixed and componentwise condition numbers. Moreover, the upper bounds of condition numbers are at most tenfold of the corresponding exact condition numbers. Also, the normwise condition numbers κ_1 and κ_n are greater than the mixed condition numbers $\kappa_\infty^{\mathrm{rel}}$ for all cases, while the differences between them are marginal. At last, we should conclude that when the data is not either sparse or badly scaled, it is more suitable to adopt the normwise condition numbers κ_1 and κ_n to measure the conditioning of LSE since they have more compact formulas.

In the rest of this section, we do some numerical experiments for LSE from piecewise-polynomial data fitting problem [32, Chapter 16].

Example 3. Given N points $(\mathbf{x}_i, \mathbf{y}_i)$ in the plane, we are seeking to find a piecewise-polynomial function $\widehat{f}(\mathbf{x})$ fitting the above set of the points, where

$$\widehat{f}(\mathbf{x}) = \begin{cases} f_1(\mathbf{x}), & \mathbf{x} \le \mathbf{a}, \\ f_2(\mathbf{x}), & \mathbf{x} > \mathbf{a} \end{cases}$$

with $\mathbf{a}$ given, and $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ polynomials of degree three or less,

$$f_1(\mathbf{x}) = x_1 + x_2\mathbf{x} + x_3\mathbf{x}^2 + x_4\mathbf{x}^3, \quad f_2(\mathbf{x}) = x_5 + x_6\mathbf{x} + x_7\mathbf{x}^2 + x_8\mathbf{x}^3.$$

The conditions that $f_1(\mathbf{a}) = f_2(\mathbf{a})$ and $f_1^{'}(\mathbf{a}) = f_2^{\prime}(\mathbf{a})$ are imposed, so that $\widehat{f}(\mathbf{x})$ is continuous and has a continuous first derivative at $\mathbf{x} = \mathbf{a}$. Suppose the N data $(\mathbf{x}_i, \mathbf{y}_i)$ are numbered so that $\mathbf{x}_1, \ldots, \mathbf{x}_M \le \mathbf{a}$ and $\mathbf{x}_{M+1}, \ldots, \mathbf{x}_N > \mathbf{a}$. The sum of squares of the prediction errors is

$$\sum_{i=1}^{M} \left(x_1 + x_2\mathbf{x}_i + x_3\mathbf{x}_i^2 + x_4\mathbf{x}_i^3 - \mathbf{y}_i\right)^2 + \sum_{i=M+1}^{N} \left(x_5 + x_6\mathbf{x}_i + x_7\mathbf{x}_i^2 + x_8\mathbf{x}_i^3 - \mathbf{y}_i\right)^2.$$

The conditions $f_1(\mathbf{a}) - f_2(\mathbf{a}) = 0$ and $f_1^{'}(\mathbf{a}) - f_2^{\prime}(\mathbf{a}) = 0$ are two linear equations

$$\begin{aligned} x_1 + x_2\mathbf{a} + x_3\mathbf{a}^2 + x_4\mathbf{a}^3 - (x_5 + x_6\mathbf{a} + x_7\mathbf{a}^2 + x_8\mathbf{a}^3) &= 0, \\ x_2 + 2x_3\mathbf{a} + 3x_4\mathbf{a}^2 - (x_6 + 2x_7\mathbf{a} + 3x_8\mathbf{a}^2) &= 0. \end{aligned}$$

Table 3
Comparison of condition numbers with the corresponding relative errors for Example 3.

$\mathbf{a}$	r_2^{rel}	κ_1	κ_n	κ_n^{U}	r_∞^{rel}	$\kappa_\infty^{\mathrm{rel}}$	$\kappa_\infty^{\mathrm{U}}$	r_c^{rel}	κ_c	κ_c^{U}
$\ell = 6$	$M = 100$	$N = 200$								
0.1	8.21e−05	5.56e+05	1.22e+06	3.09e+06	8.45e−05	1.35e+03	1.41e+03	8.67e−05	1.35e+03	1.41e+03
0.3	7.96e−06	6.68e+03	1.01e+04	8.05e+04	1.09e−05	6.93e+02	8.19e+02	1.83e−05	1.32e+03	1.50e+03
0.5	1.44e−05	1.75e+03	2.16e+03	8.74e+04	1.40e−05	8.53e+02	1.09e+03	1.77e−04	9.33e+03	1.07e+04
0.7	2.35e−04	1.04e+04	1.91e+04	4.02e+05	2.35e−04	5.62e+03	6.54e+03	2.61e−04	5.88e+03	7.42e+03
0.9	5.23e−01	7.22e+07	1.24e+08	2.90e+08	5.43e−01	1.77e+07	1.78e+07	1.35e+00	4.40e+07	4.43e+07
$\ell = 6$	$M = 200$	$N = 400$								
0.1	2.44e−04	2.43e+06	9.71e+06	2.23e+07	2.89e−04	4.33e+03	4.49e+03	2.89e−04	4.33e+03	4.49e+03
0.3	1.32e−05	6.80e+03	2.74e+04	2.25e+05	1.42e−05	4.96e+02	5.70e+02	5.28e−05	1.35e+03	1.47e+03
0.5	1.73e−05	2.07e+03	2.34e+03	1.69e+05	1.87e−05	8.85e+02	1.26e+03	8.05e−05	5.48e+03	6.74e+03
0.7	2.30e−05	2.46e+04	8.18e+04	8.43e+05	2.38e−05	5.82e+03	9.20e+03	2.44e−05	6.45e+03	1.01e+04
0.9	4.45e−03	3.10e+06	9.80e+06	2.46e+07	4.58e−03	5.84e+05	7.95e+05	4.99e−03	6.37e+05	8.67e+05
$\ell = 6$	$M = 300$	$N = 600$								
0.1	1.43e−05	8.56e+05	3.38e+06	8.43e+06	1.52e−05	1.48e+03	1.56e+03	1.52e−05	1.48e+03	1.56e+03
0.3	2.52e−06	4.85e+03	1.10e+04	1.98e+05	1.91e−06	4.19e+02	5.01e+02	6.68e−06	5.85e+02	7.12e+02
0.5	1.55e−05	2.59e+03	5.81e+03	3.64e+05	1.92e−05	9.88e+02	1.42e+03	2.27e−05	1.09e+03	1.56e+03
0.7	2.27e−04	1.85e+04	4.82e+04	9.24e+05	2.39e−04	5.16e+03	7.99e+03	2.69e−04	5.67e+03	8.71e+03
0.9	8.05e−02	4.60e+06	1.98e+07	4.62e+07	7.87e−02	1.14e+06	1.16e+06	8.76e−02	1.27e+06	1.29e+06

The coefficients $x = [x_1, \ldots, x_8]^\top$ that minimize the sum of squares of the prediction errors, subject to the continuity constraints, can be determined by solving LSE (1.1), where

$$A = \begin{bmatrix} 1 & \mathbf{x}_1 & \mathbf{x}_1^2 & \mathbf{x}_1^3 & 0 & 0 & 0 & 0 \\ 1 & \mathbf{x}_2 & \mathbf{x}_2^2 & \mathbf{x}_2^3 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \mathbf{x}_M & \mathbf{x}_M^2 & \mathbf{x}_M^3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \mathbf{x}_{M+1} & \mathbf{x}_{M+1}^2 & \mathbf{x}_{M+1}^3 \\ 0 & 0 & 0 & 0 & 1 & \mathbf{x}_{M+2} & \mathbf{x}_{M+2}^2 & \mathbf{x}_{M+2}^3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1 & \mathbf{x}_N & \mathbf{x}_N^2 & \mathbf{x}_N^3 \end{bmatrix}, \quad b = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_{M+1} \\ \vdots \\ \mathbf{y}_N \end{bmatrix},$$

$$C = \begin{bmatrix} 1 & \mathbf{a} & \mathbf{a}^2 & \mathbf{a}^3 & -1 & -\mathbf{a} & -\mathbf{a}^2 & -\mathbf{a}^3 \\ 0 & 1 & 2\mathbf{a} & 3\mathbf{a}^2 & 0 & -1 & -2\mathbf{a} & -3\mathbf{a}^2 \end{bmatrix}, \quad d = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

We randomly sample $\mathbf{x}_i \in [0,\ 1]$. For a randomly generated piecewise-polynomial function $\widehat{f}(\mathbf{x})$ with a predetermined $\mathbf{a}$, we compute the corresponding function value $\mathbf{y}_i = \widehat{f}(\mathbf{x}_i)$. As in the previous examples, we add random perturbations on the data via (5.1). The exact and perturbed solutions x and $x + \Delta x$ are computed by the GQR method [2]. In the following numerical experiments, we always choose $\boldsymbol{L} = I_n$ and compute the exact relative errors with the corresponding condition numbers for different choices of $\mathbf{a}$, M and N. The numerical results are displayed in Table 3.

From Table 3, we observe that when $\mathbf{a} = 0.1$ and $\mathbf{a} = 0.9$, the corresponding LSE problems have large normwise condition numbers while the mixed and componentwise condition numbers only are large for $\mathbf{a} = 0.9$. There are big differences between normwise condition numbers and the counter parts of mixed/componentwise condition number when $\mathbf{a} = 0.1$ for different choices of M and N. In Table 3, all of mixed/componentwise condition numbers are smaller than the corresponding normwise condition numbers for different cases of $\mathbf{a}$, M and N. Also it can be seen that when $\mathbf{a} = 0.9$ the exact relative errors are large which are consistent with the corresponding bounds given by condition numbers. Moreover, the upper bounds of condition numbers can bound the related exact condition numbers. The numerical results show that our proposed condition numbers and their corresponding upper bounds are effective.

6. Concluding remarks

In this paper we studied the perturbation analysis for the least squares problem with equality constraints. Condition number expressions for the linear function of the LSE solution were derived. Moreover, sharp upper bounds for normwise, mixed and componentwise condition numbers could be estimated efficiently by the Hager–Higham algorithm [23–25] via taking account of the already computed decompositions of matrices when the generalized QR factorization method [2,6,8,30] is adopted to solve LSE. On the other hand, some previous explicit condition number expressions on LS and LSE could be recovered from our new derived condition numbers' formulas. Numerical examples validated the effectiveness of the proposed condition numbers.

Acknowledgments

The author would like to thank the referees for their constructive comments, which led to improvements of our manuscript. This work was supported by the Fundamental Research Funds for the Central Universities under the grant 2412017FZ007 and 2412018JC014.

References

[1] Å. Björck, Numerical Methods for Least Squares Problems, SIAM, Philadelphia, 1996.
[2] Anthony J. Cox, Nicholas J. Higham, Accuracy and stability of the null space method for solving the equality constrained least squares problem, BIT 39 (1) (1999) 34–50.
[3] Lars Eldén, Perturbation theory for the least squares problem with linear equality constraints, SIAM J. Numer. Anal. 17 (3) (1980) 338–350.
[4] J.L. Barlow, N.K. Nichols, R.J. Plemmons, Iterative methods for equality-constrained least squares problems, SIAM J. Sci. Stat. Comput. 9 (5) (1988) 892–906.
[5] Charles L. Lawson, Richard J. Hanson, Solving Least Squares Problems, in: Classics in Applied Mathematics, vol. 15, SIAM, Philadelphia, PA, 1995. Revised reprint of the 1974 original.
[6] S. Hammarling, The numerical solution of the general Gauss-Markov linear model, in: Mathematics in Signal Processing (Bath, 1985), in: Inst. Math. Appl. Conf. Ser. New Ser., vol. 12, Oxford Univ. Press, New York, 1987, pp. 441–456.
[7] A.N. Malyshev, A unified theory of conditioning for linear least squares and Tikhonov regularization solutions, SIAM J. Matrix Anal. Appl. 24 (4) (2003) 1186–1196.
[8] C.C. Paige, Some aspects of generalized QR factorizations, in: Reliable Numerical Computation, in: Oxford Sci. Publ., Oxford Univ. Press, New York, 1990, pp. 73–91.
[9] Mu Sheng Wei, Perturbation theory for the rank-deficient equality constrained least squares problem, SIAM J. Numer. Anal. 29 (5) (1992) 1462–1481.
[10] Peter Bürgisser, Felipe Cucker, Condition: The Geometry of Numerical Algorithms, Springer, Heidelberg, 2013.
[11] John R. Rice, A theory of condition, SIAM J. Numer. Anal. 3 (1966) 287–310.
[12] Robert D. Skeel, Scaling for numerical stability in Gaussian elimination, J. ACM 26 (3) (1979) 494–526.
[13] J. Rohn, New condition numbers for matrices and linear systems, Computing 41 (1–2) (1989) 167–169.
[14] I. Gohberg, I. Koltracht, Mixed, componentwise, and structured condition numbers, SIAM J. Matrix Anal. Appl. 14 (3) (1993) 688–704.
[15] Mario Arioli, Marc Baboulin, Serge Gratton, A partial condition number for linear least squares problems, SIAM J. Matrix Anal. Appl. 29 (2) (2007) 413–433.
[16] Marc Baboulin, Serge Gratton, Using dual techniques to derive componentwise and mixed condition numbers for a linear function of a linear least squares solution, BIT 49 (1) (2009) 3–19.
[17] Huai-An Diao, Liming Liang, Sanzheng Qiao, A condition analysis of the weighted linear least squares problem using dual norms, Linear Multilinear Algebra 66 (6) (2018) 1085–1103.
[18] Marc Baboulin, Serge Gratton, A contribution to the conditioning of the total least-squares problem, SIAM J. Matrix Anal. Appl. 32 (3) (2011) 685–699.
[19] Huai-An Diao, Yang Sun, Mixed and componentwise condition numbers for a linear function of the solution of the total least squares problem, Linear Algebra Appl. 544 (2018) 1–29.
[20] Huai-An Diao, Tong-Yu Zhou, Backward error and condition number analysis for the indefinite linear least squares problem, Int. J. Comput. Math. (2018). http://dx.doi.org/10.1080/00207160.2018.1467007.
[21] Hanyu Li, Shaoxin Wang, Partial condition number for the equality constrained linear least squares problem, Calcolo (2017). http://dx.doi.org/10.1007/s10092-017-0221-8.
[22] Alexander Graham, Kronecker Products and Matrix Calculus: With Applications, in: Ellis Horwood Series in Mathematics and its Applications, Ellis Horwood Ltd. Chichester; Halsted Press [John Wiley & Sons, Inc.], New York, 1981.
[23] William W. Hager, Condition estimates, SIAM J. Sci. Stat. Comput. 5 (2) (1984) 311–316.
[24] Nicholas J. Higham, FORTRAN codes for estimating the one-norm of a real or complex matrix, with applications to condition estimation, ACM Trans. Math. Softw. 14 (4) (1988) 381–396.
[25] Nicholas J. Higham, Experience with a matrix norm estimator, SIAM J. Sci. Stat. Comput. 11 (4) (1990) 804–809.
[26] Felipe Cucker, Huaian Diao, Yimin Wei, On mixed and componentwise condition numbers for Moore-Penrose inverse and linear least squares problems, Math. Comp. 76 (258) (2007) 947–963.
[27] S. Gratton, On the condition number of linear least squares problems in a weighted frobenius norm, BIT Numer. Math. 36 (3) (1996) 523–530.
[28] Huai-An Diao, On condition numbers for least squares with quadric inequality constraint, Comput. Math. Appl. 73 (4) (2017) 616–627.
[29] Jan R. Magnus, Heinz Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, John Wiley & Sons, Ltd., Chichester, 2007.
[30] E. Anderson, Zhaojun. Bai, J. Dongarra, Generalized QR factorization and its applications, Linear Algebra Appl. 162/164 (1992) 243–271. Directions in matrix theory (Auburn, AL, 1990).
[31] Gene H. Golub, Charles F. Van Loan, Matrix Computations, fourth ed., in: Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, 2013.
[32] Stephen Boyd, Lieven Vandenberghe, Introduction to Applied Linear Algebra-Vectors, Matrices, and Least Squares. https://web.stanford.edu/~boyd/vmls/vmls.pdf, 2017.