On Mixed and Componentwise Condition Numbers for Moore-Penrose Inverse and Linear Least Squares Problems [a]

Yimin WEI

Department of Mathematics
Fudan University
Shanghai, 200433
People's Republic of China
E-mail: ymwei@fudan.edu.cn

[a] Joint with F. Cucker and H. Diao

Outline

- Introduction
- Preliminaries
- Moore-Penrose inverse
- Linear least squares problems
- Full row rank and underdetermined system
- Numerical examples

Introduction

General considerations

A general theory of condition numbers was first given by Rice in 1966.

Let $\phi : \mathbb{R}^s \to \mathbb{R}^t$ be a mapping, $\mathbb{R}^s$ and $\mathbb{R}^t$ are the s- and t-dimensional Euclidean spaces equipped with some norms.

If ϕ is continuous and Fréchet differentiable in the neighborhood of $a_0 \in \mathbb{R}^s$ then, according to Rice, the *relative normwise condition number* of a_0 is given by

$$\begin{aligned} \mathrm{cond}(a_0) : &= \lim_{\varepsilon \to 0} \sup_{\|\Delta a\| \le \varepsilon} \left(\frac{\|\phi(a_0 + \Delta a) - \phi(a_0)\|}{\|\phi(a_0)\|} \Big/ \frac{\|\Delta a\|}{\|a_0\|} \right) \\ &= \frac{\|\phi'(a_0)\| \|a_0\|}{\|\phi(a_0)\|}, \end{aligned}$$

where $\phi'(a_0)$ is the Fréchet derivative of ϕ at a_0.

A drawback of condition numbers is that they ignore the structure of both input and output data with respect to scaling and/or sparsity.

To tackle this drawback, another approach known as *componentwise analysis*, has been increasingly considered.

Two different kinds of condition number were studied. Firstly, those measuring the errors in the output using norms and *the* input perturbations componentwise. Secondly, those measuring both the error in the output and the perturbation in the input componentwise. The resulting condition numbers are called *mixed* and *componentwise*, respectively, by Gohberg and Koltracht in 1993.

By their very nature, condition numbers are defined as limits of suprema. Therefore, their definition does not suggest a way to compute them from the input data. To do so, equivalent explicit expressions are sought or, alternatively, easy to compute and sufficiently sharp upper bounds. This has been extensively done for many problems in linear algebra, mostly for normwise condition numbers.

Main Definitions and Results

To define mixed and componentwise condition numbers the following form of “distance” function will be useful.

Let $a, b \in \mathbb{R}^n$. We denote by a/b the element in $\mathbb{R}^n$ whose ith component is a_i/b_i if $b_i \neq 0$ and 0 otherwise. Then

$$d(a, b) = \left\| \frac{a - b}{b} \right\|_\infty = \max_{\substack{i=1,\ldots,n \\ b_i \neq 0}} \left\{ \frac{|a_i - b_i|}{|b_i|} \right\}.$$

Note that

$$d(a, b) = \min\{\nu \geq 0 \mid |a_i - b_i| \leq \nu |b_i| \text{ for } i = 1, \ldots, n\}.$$

Also, if $b = 0$ then $d(a, b) = 0$. We can extend the function d to matrices in an obvious manner. We introduce a notation allowing us to do so smoothly. For a matrix $A \in \mathbb{R}^{m \times n}$ we define $\mathsf{vec}(A) \in \mathbb{R}^{mn}$ by $\mathsf{vec}(A) = [a_1^{\mathrm{T}}, \ldots, a_n^{\mathrm{T}}]^{\mathrm{T}}$, where $A = [a_1, \ldots, a_n]$ with $a_i \in \mathbb{R}^m$, $i = 1, \ldots, n$. Then

$$d(A, B) = d(\mathsf{vec}(A), \mathsf{vec}(B)).$$

Note that vec is a homeomorphism between $\mathbb{R}^{m\times n}$ and $\mathbb{R}^{mn}$. In addition, it transforms norms in the sense that, for all $A \in \mathbb{R}^{m\times n}$,

$$\|\mathsf{vec}(A)\|_2 = \|A\|_F \qquad \text{and} \qquad \|\mathsf{vec}(A)\|_\infty = \|A\|_{\max} \tag{1}$$

where $\| \ \|_F$ is the Frobenius norm given by

$$\|A\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2\right)^{1/2}$$

and $\| \ \|_{\max}$ is the max norm given by

$$\|A\|_{\max} = \max_{i,j} |A_{ij}|.$$

Let $\| \ \|_\alpha$ be a norm in $\mathbb{R}^p$. Denote $B_\alpha(a,\varepsilon) = \{x \in \mathbb{R}^p \mid \|x - a\|_\alpha \leq \varepsilon\}$ and $B^0(a,\varepsilon) = \{x \mid d(x,a) \leq \varepsilon\}$.

For a partial function $F : \mathbb{R}^p \to \mathbb{R}^q$, denote by $\mathsf{Dom}(f)$ its domain of definition.

Definition 2.1 *Let* $F : \mathbb{R}^p \to \mathbb{R}^q$ *be a continuous mapping defined on an open set* $\mathsf{Dom}(F) \subset \mathbb{R}^p$ *such that* $0 \notin \mathsf{Dom}(F)$*. Let* $a \in \mathsf{Dom}(F)$ *such that* $F(a) \neq 0$*.*

(i) *Let* $\| \ \|_\alpha$ *and* $\| \ \|_\beta$ *be norms in* $\mathbb{R}^p$ *and* $\mathbb{R}^q$ *respectively. The* normwise condition number *of* F *at* a *(with respect to the norms* $\| \ \|_\alpha$ *and* $\| \ \|_\beta$*) is defined by*

$$\kappa(F, a) = \lim_{\varepsilon \to 0} \sup_{\substack{x \in B_\alpha(a,\varepsilon) \\ x \neq a}} \frac{\|F(x) - F(a)\|_\beta}{\|x - a\|_\alpha} \frac{\|a\|_\alpha}{\|F(a)\|_\beta}.$$

(ii) *The* mixed condition number *of* F *at* a *is defined by*

$$m(F, a) = \lim_{\varepsilon \to 0} \sup_{\substack{x \in B^0(a,\varepsilon) \\ x \neq a}} \frac{\|F(x) - F(a)\|_\infty}{\|F(a)\|_\infty} \frac{1}{d(x, a)}.$$

(iii) *Suppose* $F(a) = (f_1(a), \ldots, f_q(a))$ *is such that* $f_j(a) \neq 0$ *for* $j = 1, \ldots, q$*. Then the* componentwise condition number *of* F *at* a *is*

$$c(F, a) = \lim_{\varepsilon \to 0} \sup_{\substack{x \in B^0(a,\varepsilon) \\ x \neq a}} \frac{d(F(x), F(a))}{d(x, a)}.$$

In this talk we consider several condition numbers for the *Moore-Penrose inverse* of $A \in \mathbb{R}^{m\times n}$. This is the unique $n \times m$ matrix $A^\dagger$ satisfying the following four matrix equations

$$AA^\dagger A = A, \quad A^\dagger AA^\dagger = A^\dagger, \quad (AA^\dagger)^{\mathrm{T}} = AA^\dagger, \quad (A^\dagger A)^{\mathrm{T}} = A^\dagger A.$$

For a real matrix M, M^{T} denotes its transpose matrix.

Firstly consider the normwise condition number for both the 2-norm and the Frobenius norm in $\mathbb{R}^{m\times n}$. Definition 2.1 yields,

$$\kappa_2^\dagger(A) := \lim_{\varepsilon\to 0} \sup_{\|\Delta A\|_2 \le \varepsilon} \frac{\|(A+\Delta A)^\dagger - A^\dagger\|_2}{\|\Delta A\|_2} \frac{\|A\|_2}{\|A^\dagger\|_2}$$

and

$$\kappa_F^\dagger(A) := \lim_{\varepsilon\to 0} \sup_{\|\Delta A\|_F \le \varepsilon} \frac{\|(A+\Delta A)^\dagger - A^\dagger\|_F}{\|\Delta A\|_F} \frac{\|A\|_F}{\|A^\dagger\|_F}.$$

We are also interested in the mixed and componentwise condition numbers for the Moore-Penrose inverse.

In these cases, identifying $\mathbb{R}^{m\times n}$ with $\mathbb{R}^{mn}$ via vec and using (1), Definition 2.1 yields

$$m_\dagger(A) := \lim_{\varepsilon\to 0} \sup_{\|\Delta A/A\|_{\max}\le\varepsilon} \frac{\|(A+\Delta A)^\dagger - A^\dagger\|_{\max}}{\|A^\dagger\|_{\max}} \frac{1}{\|\Delta A/A\|_{\max}}$$

and

$$c_\dagger(A) := \lim_{\varepsilon\to 0} \sup_{\|\Delta A/A\|_{\max}\le\varepsilon} \frac{1}{\|\Delta A/A\|_{\max}} \left\| \frac{(A+\Delta A)^\dagger - A^\dagger}{A^\dagger} \right\|_{\max}.$$

$\frac{B}{A}$ is an entrywise division defined by $\frac{B}{A} := \mathsf{vec}^{-1}(\mathsf{vec}(B)/\mathsf{vec}(A))$.

In a similar way, one defines, given a full column rank matrix A and a vector b, the condition numbers $\kappa_2^{\mathsf{ls}}(A,b)$, $\kappa_F^{\mathsf{ls}}(A,b)$, $m^{\mathsf{ls}}(A,b)$, and $c^{\mathsf{ls}}(A,b)$ for the computation of the solution x of the least squares (LS) problem

$$\min_{v\in\mathbb{R}^n} \|Av-b\|_2$$

and the condition numbers $m^{\mathsf{res}}(A,b)$, and $c^{\mathsf{res}}(A,b)$ for the computation of the residue $\|Ax-b\|$.

Preliminaries

Kronecker products

$A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{p\times q}$, the *Kronecker product* $A \otimes B \in \mathbb{R}^{mp\times nq}$ is

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{bmatrix}.$$

Note that we consider vectors $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$ as matrices in $\mathbb{R}^{m\times 1}$ and $\mathbb{R}^{n\times 1}$.

In this case, we have

$$u \otimes v = \mathsf{vec}(uv^{\mathrm{T}}) \in \mathbb{R}^{mn}.$$

The following results can be found

$$\begin{aligned}
(A+B)\otimes(C+D) &= A\otimes C + B\otimes C + A\otimes D + B\otimes D,\\
(A\otimes C)(B\otimes D) &= (AB)\otimes(CD),\\
A\otimes(B\otimes C) &= (A\otimes B)\otimes C,\\
(A\otimes B)^{\mathrm{T}} &= A^{\mathrm{T}}\otimes B^{\mathrm{T}},\\
\|A\otimes B\| &= \|A\|\|B\|,\\
|A\otimes B| &= |A|\otimes|B|,\\
\mathsf{vec}(AXB) &= (B^{\mathrm{T}}\otimes A)\mathsf{vec}(X),
\end{aligned}$$

where $|A| = (|A_{ij}|)$, A_{ij} is the (i,j)-th entry of A, and

$\|A\|$ denotes either $\|\ \|_2$ or $\|\ \|_F$.

It is proven that there exists a matrix $\Pi \in \mathbb{R}^{mn\times mn}$ such that, for all $A \in \mathbb{R}^{m\times n}$,

$$\Pi(\mathsf{vec}(A)) = \mathsf{vec}(A^{\mathrm{T}}). \tag{2}$$

The matrix Π is called the *vec-permutation matrix*. Here Π can be represented explicitly

$$\Pi = \sum_{i=1}^{n} \sum_{j=1}^{m} E_{ij}(m \times n) \otimes E_{ji}(n \times m). \tag{3}$$

Here $E_{ij}(m \times n) = e_i^{(m)}(e_j^{(n)})^{\mathrm{T}} \in \mathbb{R}^{m\times n}$ denotes the (i, j)-th elementary matrix and $e_i^{(m)}$ is the vector $\left[0, \ldots, 0, 1, 0, \ldots, 0\right]^{\mathrm{T}} \in \mathbb{R}^m$, the 1 in the i-th component.

Also it is proved that for any vector $y \in \mathbb{R}^p$ and matrix $Y \in \mathbb{R}^{p\times q}$,

$$\left(y^{\mathrm{T}} \otimes Y\right) \Pi = Y \otimes y^{\mathrm{T}}. \tag{4}$$

Singular Value Decomposition and Moore-Penrose Inverse

Assume that $A \in \mathbb{R}^{m\times n}$ with $\mathsf{rank}(A) = n$. The singular value decomposition (SVD) of A is given by

$$A = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^{\mathrm{T}}, \tag{5}$$

where $U = (u_1, \ldots, u_m) \in \mathbb{R}^{m\times m}, V = (v_1, \ldots, v_n) \in \mathbb{R}^{n\times n}, \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n), \sigma_1 \geqslant \ldots \geqslant \sigma_n > 0$. Express the Moore-Penrose inverse of A by

$$A^{\dagger} = V \left[\Sigma^{-1}, 0\right] U^{\mathrm{T}} = \left(A^{\mathrm{T}}A\right)^{-1} A^{\mathrm{T}},$$

and

$$A^{\dagger}A = I_n, \quad I_m - AA^{\dagger} = U \begin{bmatrix} 0 & 0 \\ 0 & I_{m-n} \end{bmatrix} U^{\mathrm{T}}, \quad A^{\dagger}A^{\dagger\,\mathrm{T}} = (A^{\mathrm{T}}A)^{-1},$$

where I_n is the identity matrix of order n.

$AA^\dagger, (I_m - AA^\dagger)$ are the matrices of the orthogonal projections of $\mathbb{R}^m$ onto $\mathcal{R}(A)$ and $\mathcal{N}(A^\dagger) = \mathcal{N}(A^{\mathrm{T}})$ respectively, where $\mathcal{R}(A)$ is the range space of A and $\mathcal{N}(A^\dagger)$ is the null space of $A^\dagger$.

Proposition 1 *Let* $U = [u_1, u_2, \ldots, u_m] \in \mathbb{R}^{m\times m}$ *and* $V = [v_1, v_2, \ldots, v_n] \in \mathbb{R}^{n\times n}$. *Then*

$$\mathcal{R}(A) = \mathsf{span}\{u_1, \ldots, u_n\}, \quad \textit{and} \quad \mathcal{N}(A^\dagger) = \mathsf{span}\{u_{n+1}, \ldots, u_m\},$$

$$\mathsf{span}\{u_1, \ldots, u_n\} = \left\{ u \in \mathbb{R}^m \mid u = \sum_{i=1}^{n} \alpha_i u_i,\ \alpha_i \in \mathbb{R},\ i = 1, 2, \ldots, n \right\}.$$

In addition,

$$A^\dagger u_n = \|A^\dagger\|_2 v_n, \qquad v_n^{\mathrm{T}} A^\dagger = \|A^\dagger\|_2 u_n^{\mathrm{T}}, \qquad \|A^\dagger\|_2 = \frac{1}{\sigma_n},$$

$$A = \sum_{i=1}^{n} \sigma_i u_i v_i^{\mathrm{T}}, \quad \textit{and} \quad A^\dagger = \sum_{i=1}^{n} \frac{1}{\sigma_i} v_i u_i^{\mathrm{T}}.$$

Condition numbers and differentiability

The following lemma gives expressions for the normwise, mixed and componentwise condition numbers for differentiable functions. In its statement, and in all what follows, if $a \in \mathbb{R}^p$ we denote by $\mathsf{Dg}(a)$ the $p \times p$ diagonal matrix with $a_1, \ldots, a_p$ in the diagonal.

Lemma 3.1 [6] *Let $F : \mathbb{R}^p \to \mathbb{R}^q$ be as in Definition 2.1 and $a \in \mathsf{Dom}(F)$ be such that F is Fréchet differentiable at a. Then,*

(a) $\kappa(F, a) = \dfrac{\|DF(a)\|_{\alpha\beta}\|a\|_\alpha}{\|F(a)\|_\beta}$.

(b) $m(F, a) = \dfrac{\|DF(a)\mathsf{Dg}(a)\|_\infty}{\|F(a)\|_\infty}$.

(c) $c(F, a) = \|\mathsf{Dg}(F(a))^{-1}DF(a)\mathsf{Dg}(a)\|_\infty$. □

To use the lemma above for the Moore-Penrose inverse (of full-rank matrices) we introduce some notation. Consider the sets

$$S = \{G \in \mathbb{R}^{m\times n} \mid \mathsf{rank}(G) = n\}, V = \{g \in \mathbb{R}^{mn} \mid g = \mathsf{vec}(G), G \in S\}.$$

Note that the set S is open in $\mathbb{R}^{m\times n}$ since its complement is the union of the sets $\det(G_s) = 0$ where G_s runs over all $n \times n$ submatrices of G. Moreover, since vec is a homeomorphism between $\mathbb{R}^{m\times n}$ and $\mathbb{R}^{mn}$, it follows that V is open as well.

Now define the mapping $\Phi : S \to \mathbb{R}^{n\times m}$ given by $\Phi(G) = G^\dagger$. Also, define $\phi : V \to \mathbb{R}^{mn}$ by $\phi(\mathsf{vec}(G)) = \mathsf{vec}(\Phi(G))$. By definition (and taking $\|\ \|_\beta$ to be the 2-norm in $\mathbb{R}^m$ and $\|\ \|_\alpha$ to be the operator norm with respect to the 2-norm in both $\mathbb{R}^n$ and $\mathbb{R}^m$) we have,

$$\kappa_2^\dagger(A) = \kappa(\Phi; A), \qquad \kappa_F^\dagger(A) = \kappa(\phi; \mathsf{vec}(A)),$$

as well as

$$m_\dagger(A) = m(\phi; \mathsf{vec}(A)), \qquad \text{and} \qquad c_\dagger(A) = c(\phi; \mathsf{vec}(A)).$$

To make use of the above we would like to have explicit expressions for the derivatives $D\Phi$ and $D\phi$. Lemma 3.3 below exhibits such expressions. Its proof uses the following well-known result.

Lemma 3.2 *Let $A \in \mathbb{R}^{m\times n}$ and suppose $\{A_k\}$ is a sequence of $m \times n$ matrices satisfying $\lim_{k\to\infty} A_k = A$. A necessary and sufficient condition for $\lim_{k\to\infty} A_k^\dagger = A^\dagger$ is*

$$\mathsf{rank}(A_k) = \mathsf{rank}(A)$$

for sufficiently large k.

Lemma 3.3 *Both Φ and ϕ are continuous mappings and ϕ is Fréchet differentiable at a for all $a \in V$. If $a \in V$ and $A \in S$ are such that $a = \mathsf{vec}(A)$ then*

$$D\phi(a)(e) = \left[-\left(A^{\dagger^{\mathrm{T}}} \otimes A^\dagger\right) + \left((I - AA^\dagger) \otimes (A^{\mathrm{T}}A)^{-1}\right)\Pi\right] e,$$

$$D\Phi(A)(E) = -A^\dagger E A^\dagger + \left(A^{\mathrm{T}}A\right)^{-1} E^{\mathrm{T}}(I - AA^\dagger),$$

where $e \in \mathbb{R}^{mn}$ and $E \in \mathbb{R}^{m\times n}$.

Moore Penrose inverse

Explicit expressions for the condition numbers for the Moore-Penrose inverse computation.

Theorem 4.1 *Let* $A \in \mathbb{R}^{m \times n}$ *be such that* $\mathsf{rank}(A) = n$. *Then*

(a) $$\kappa_2^\dagger(A) = \frac{\|D\Phi(A)\|_2 \, \|A\|_2}{\|A^\dagger\|_2}$$

(b) $$\kappa_F^\dagger(A) = \frac{\left\| \left({A^\dagger}^{\mathrm{T}} \otimes A^\dagger \right) - ((I - AA^\dagger) \otimes (A^{\mathrm{T}}A)^{-1}) \, \Pi \right\|_2 \|A\|_F}{\|A^\dagger\|_F}$$

$$= \frac{\|A^\dagger\|_2^2 \|A\|_F}{\|A^\dagger\|_F}$$

(c) $$m_\dagger(A) = \frac{\left\| \left| \left[\left({A^\dagger}^{\mathrm{T}} \otimes A^\dagger \right) - \left((I - AA^\dagger) \otimes (A^{\mathrm{T}}A)^{-1} \right) \Pi \right] \right| \mathsf{vec}(|A|) \right\|_\infty}{\|\mathsf{vec}(A^\dagger)\|_\infty}$$

(d) $$c_{\dagger}(A) = \left\| \frac{\left|\left[\left(A^{\dagger\mathrm{T}} \otimes A^{\dagger}\right) - \left((I - AA^{\dagger}) \otimes (A^{\mathrm{T}}A)^{-1}\right)\Pi\right]\right| \mathsf{vec}(|A|)}{\mathsf{vec}(A^{\dagger})} \right\|_{\infty}$$

Remark 1 *Theorem 4.1 provides explicit expressions for* $\kappa_F^{\dagger}(A)$*,* $m_{\dagger}(A)$*, and* $c_{\dagger}(A)$ *but not for* $\kappa_2^{\dagger}(A)$ *due to the occurrence of the factor* $\|D\Phi(A)\|_2$*. The most explicit expression is*

$$\|D\Phi(A)\|_2 = \max_{\|E\|_2=1} \left\| A^{\dagger}EA^{\dagger} - \left(A^{\mathrm{T}}A\right)^{-1} E^{\mathrm{T}}(I - AA^{\dagger}) \right\|_2$$

which easily follows from Lemma 3.3. Yet, we will give sufficiently tight upper and lower bounds for $\kappa_2^{\dagger}(A)$ *in Corollary 4.3 below.*

Lemma 4.2 *Let* $A, \Delta A \in \mathbb{R}^{m\times n}$ *be such that* $\mathsf{rank}(A) = \mathsf{rank}(A + \Delta A) = n$*. Then*

$$\|(A + \Delta A)^{\dagger} - A^{\dagger}\|_F \ \leq \ \|A^{\dagger}\|_2 \|(A + \Delta A)^{\dagger}\|_2 \|\Delta A\|_F,$$

$$\|(A + \Delta A)^{\dagger} - A^{\dagger}\|_2 \ \leq \ \sqrt{2}\|A^{\dagger}\|_2 \|(A + \Delta A)^{\dagger}\|_2 \|\Delta A\|_2.$$

Theorem 4.1 gives explicit expressions for the condition numbers $\kappa_F^\dagger(A)$, $m_\dagger(A)$, and $c_\dagger(A)$. While these expressions are sharp, the one for $\kappa_2^\dagger(A)$ may not be easy to compute by its dependance on the derivative $D\Phi(A)$, and those for $m_\dagger(A)$ and $c_\dagger(A)$ may not be so by their dependance on the (large) matrix Π and the need to compute Kronecker products. The next corollary gives easier to compute upper bounds for these three condition numbers. It also gives a lower bound for $\kappa_2^\dagger(A)$.

Corollary 4.3 *In the hypothesis of Theorem* 4.1 *we have*

(a) $\|A\|_2\|A^\dagger\|_2 \leq \kappa_2^\dagger(A) \leq \sqrt{2}\|A\|_2\|A^\dagger\|_2,$

(b) $$m^\dagger(A) \leq \frac{\left\||A^\dagger||A||A^\dagger| + |(A^{\mathrm{T}}A)^{-1}||A^{\mathrm{T}}||I - AA^\dagger|\right\|_{\max}}{\|A^\dagger\|_{\max}},$$

(c) $$c^\dagger(A) \leq \left\|\frac{|A^\dagger||A||A^\dagger| + |(A^{\mathrm{T}}A)^{-1}||A^{\mathrm{T}}||I - AA^\dagger|}{A^\dagger}\right\|_{\max}.$$

Linear Least Squares Problems

We consider linear least squares problems (LS)

$$\min_{v\in\mathbb{R}^n} \|Av - b\|_2, \tag{6}$$

where $A \in \mathbb{R}^{m\times n}$, $\mathsf{rank}(A) = n$, and $b \in \mathbb{R}^m$. Since A, as a linear map, is injective there is a unique minimizer x for (6). This minimizer satisfies

$$A^{\mathrm{T}}Ax = A^{\mathrm{T}}b, \tag{7}$$

and therefore,

$$x = A^{\dagger}b = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b.$$

Let x be as above, $\Delta b \in \mathbb{R}^m$, and $\Delta A \in \mathbb{R}^{m\times n}$ such that $\mathsf{rank}(A + \Delta A) = n$. Consider the problem

$$\min_{w\in\mathbb{R}^n} \|(A + \Delta A)w - (b + \Delta b)\|_2 \tag{8}$$

Then there is a unique minimizer y and letting $\Delta x := y - x$ we have

$$\Delta x = (A + \Delta A)^{\dagger}(b + \Delta b) - x.$$

The normwise, mixed and componentwise condition numbers for LS are defined as follows. Let $\Delta A_1 = AA^{\dagger}\Delta A$ and $\Delta A_2 = (I - AA^{\dagger})\Delta A$. Then

$$\begin{aligned}
\kappa_2^{\mathsf{ls}}(A,b) &:= \lim_{\varepsilon\to 0} \sup_{\substack{\sqrt{\|\Delta A_1\|_2^2 + \|\Delta A_2\|_2^2} \le \varepsilon \|A\|_2 \\ \|\Delta b\|_2 \le \varepsilon \|b\|_2}} \frac{\|\Delta x\|_2}{\varepsilon \|x\|_2}, \\
\kappa_F^{\mathsf{ls}}(A,b) &:= \lim_{\varepsilon\to 0} \sup_{\substack{\|\Delta A\|_F \le \varepsilon \|A\|_F \\ \|\Delta b\|_2 \le \varepsilon \|b\|_2}} \frac{\|\Delta x\|_2}{\varepsilon \|x\|_2}, \\
m^{\mathsf{ls}}(A,b) &:= \lim_{\varepsilon\to 0} \sup_{\substack{|\Delta A| \le \varepsilon |A| \\ |\Delta b| \le \varepsilon |b|}} \frac{\|\Delta x\|_\infty}{\varepsilon \|x\|_\infty}, \\
c^{\mathsf{ls}}(A,b) &:= \lim_{\varepsilon\to 0} \sup_{\substack{|\Delta A| \le \varepsilon |A| \\ |\Delta b| \le \varepsilon |b|}} \frac{1}{\varepsilon} \left\| \frac{\Delta x}{x} \right\|_\infty .
\end{aligned}$$

Just as in the previous section, to comfortably make use of Lemma 3.1, we define the mappings $\Psi : S \times \mathbb{R}^m \to \mathbb{R}^n$ by

$$\Psi : (G, f) := G^\dagger f$$

and $\psi : V \times \mathbb{R}^m \to \mathbb{R}^n$ by

$$\psi(g, f) := (\mathsf{vec}^{-1} g)^\dagger f.$$

For the normwise condition numbers of the mapping Ψ, we consider two norms defined on $\mathbb{R}^{m\times n} \times \mathbb{R}^m$. These norms depend on the pair (A, b) and are respectively given by

$$\|(G, f)\|_{\mathrm{M}} = \max\left\{\frac{1}{\|A\|_2}\sqrt{\|G_1\|_2^2 + \|G_2\|_2^2}, \frac{1}{\|b\|_2}\|f\|_2\right\}, \quad (9)$$

$$\|(G, f)\|_{\mathrm{Fro}} = \max\left\{\frac{1}{\|A\|_F}\|G\|_F, \frac{1}{\|b\|_2}\|f\|_2\right\},$$

where $(G, f) \in \mathbb{R}^{m\times n} \times \mathbb{R}^m$, $G_1 = AA^\dagger G$ and $G_2 = (I - AA^\dagger)G$.

Lemma 5.1 *We have*

$$\kappa_2^{\mathsf{ls}}(A,b) = \kappa_{\mathrm{M}}(\Psi; A, b), \quad \kappa_F^{\mathsf{ls}}(A,b) = \kappa_{\mathrm{Fro}}(\Psi; A, b),$$

and

$$m^{\mathsf{ls}}(A,b) = m(\psi; a, b), \quad c^{\mathsf{ls}}(A,b) = c(\psi; a, b).$$

Lemma 5.2 *The set $V \times \mathbb{R}^m$ is open and ψ is a continuous mapping on $V \times \mathbb{R}^m$. In addition, ψ is Fréchet differentiable at (A,b) and $D\psi(A,b)$ is given by*

$$\begin{aligned} D\Psi(A,b)(G,f) &= -A^{\dagger}Gx + (A^{\mathrm{T}}A)^{-1}G^{\mathrm{T}}r + A^{\dagger}f, \\ D\psi(A,b)(G,f) &= \left[-\left(x^{\mathrm{T}} \otimes A^{\dagger}\right) + (A^{\mathrm{T}}A)^{-1} \otimes r^{\mathrm{T}}, A^{\dagger}\right] \begin{bmatrix} \mathsf{vec}(G) \\ f \end{bmatrix}. \end{aligned}$$

where $r = b - Ax$.

We next combine Lemmas 3.1 and 5.2 to get expressions for normwise, mixed and componentwise condition number of LS.

Theorem 5.3 *Let* $A \in \mathbb{R}^{m\times n}$, $\mathsf{rank}(A) = n$, *and* $b \in \mathbb{R}^m$. *We have*

$$\kappa_2^{\mathsf{ls}}(A,b) = \|A\|_2\|A^\dagger\|_2\sqrt{1 + \frac{\|A^\dagger\|_2^2\|r\|_2^2}{\|x\|_2^2} + \frac{\|A^\dagger\|_2\|b\|_2}{\|x\|_2}},$$

$$\kappa_F^{\mathsf{ls}}(A,b) = \|A\|_F\|A^\dagger\|_2\sqrt{1 + \frac{\|A^\dagger\|_2^2\|r\|_2^2}{\|x\|_2^2} + \frac{\|A^\dagger\|_2\|b\|_2}{\|x\|_2}},$$

$$m^{\mathsf{ls}}(A,b) = \frac{\left\| |[(x^{\mathrm{T}} \otimes A^\dagger) - (A^{\mathrm{T}}A)^{-1} \otimes r^{\mathrm{T}}]| \,\mathsf{vec}(|A|) + |A^\dagger||b| \right\|_\infty}{\|x\|_\infty},$$

$$c^{\mathsf{ls}}(A,b) = \left\| \frac{|(x^{\mathrm{T}} \otimes A^\dagger) - (A^{\mathrm{T}}A)^{-1} \otimes r^{\mathrm{T}}| \,\mathsf{vec}(|A|) + |A^\dagger||b|}{x} \right\|_\infty.$$

Furthermore, if $r = 0$ *(i.e., for consistent linear systems* $Ax = b$*)*

we have

$$\begin{aligned}
\kappa_2^{\mathsf{ls}}(A,b) &= \|A\|_2\|A^\dagger\|_2 + \frac{\|A^\dagger\|_2\|b\|_2}{\|x\|_2},\\
\kappa_F^{\mathsf{ls}}(A,b) &= \|A\|_F\|A^\dagger\|_2 + \frac{\|A^\dagger\|_2\|b\|_2}{\|x\|_2},\\
m^{\mathsf{ls}}(A,b) &= \frac{\||A^\dagger||A||x| + |A^\dagger||b|\|_\infty}{\|x\|_\infty},\\
c^{\mathsf{ls}}(A,b) &= \left\|\frac{|A^\dagger||A||x| + |A^\dagger||b|}{x}\right\|_\infty .
\end{aligned}$$

Remark 2 *When $n = m$ the consistent case of Theorem 5.3 recovers the known expressions [11] for normwise, mixed and componentwise condition numbers for nonsingular linear systems.*

Corollary 5.4 *We have the following bounds*

$$
\begin{aligned}
m^{\mathsf{ls}}(A,b) &\le \frac{\| |A^\dagger||A||x| + |(A^{\mathrm{T}}A)^{-1}||A^{\mathrm{T}}||r| + |A^\dagger||b| \|_\infty}{\|x\|_\infty},\\
c^{\mathsf{ls}}(A,b) &\le \left\| \frac{|A^\dagger||A||x| + |(A^{\mathrm{T}}A)^{-1}||A^{\mathrm{T}}||r| + |A^\dagger||b|}{|x|} \right\|_\infty .
\end{aligned}
$$

Condition numbers bound the *worst-case* sensitivity of an input data only to *small* perturbations. If ε is the size of the perturbation, a term $\mathcal{O}(\varepsilon^2)$ is neglected and therefore, the bound only holds for ε small enough. One says that condition numbers are *first order* bounds for these sensitivities. Occasionaly, one is interested in bounds for *unrestricted* perturbations. The following result exhibits such unrestricted perturbation bounds for LS.

Theorem 5.5 *Let* $A, \Delta A \in \mathbb{R}^{m\times n}$ *satisfying* $\mathsf{rank}(A) = \mathsf{rank}(A+\Delta A) = n$. *Let* $\Delta b \in \mathbb{R}^m$ *and* x, y *be the solutions of* (6) *and* (8) *respectively. If for some* $E \in \mathbb{R}^{m\times n}$ *and some* $f \in \mathbb{R}^m$ *we have* $|\Delta A| \le \varepsilon E$ *and* $|\Delta b| \le \varepsilon f$ *then*

$$\frac{\|y-x\|_\infty}{\|x\|_\infty} \le \varepsilon \frac{\| |[(y^{\mathrm{T}} \otimes A^\dagger) - (A^{\mathrm{T}}A)^{-1} \otimes s^{\mathrm{T}}]| \,\mathsf{vec}(E) + |A^\dagger| f \|_\infty}{\|x\|_\infty}, \tag{10}$$

$$\frac{\|s-r\|_\infty}{\|r\|_\infty} \le \varepsilon \frac{\left\| \left| \left[y^{\mathrm{T}} \otimes (I - AA^\dagger) + {A^\dagger}^{\mathrm{T}} \otimes s^{\mathrm{T}} \right] \right| \mathsf{vec}(E) + |I - AA^\dagger|\, f \right\|_\infty}{\|r\|_\infty},$$

where $s = b + \Delta b - (A + \Delta A)y$.

The next corollary gives (easier to compute, no occurrences of $\otimes$) upper bounds for the residual vector mixed and componentwise condition numbers.

Corollary 5.6 *Let* $A \in \mathbb{R}^{m\times n}$ *satisfy* $\mathsf{rank}(A) = n$. *Let* $b \in \mathbb{R}^m$, x *be the solution of* (6) *and* $r = Ax - b$. *Then*

$$
\begin{aligned}
m^{\mathsf{res}}(A,b) &\le \mathtt{m}^{\mathsf{upper}}_{\mathsf{res}}(\mathtt{A},\mathtt{b}) \\
&:= \frac{\left\| |I - AA^{\dagger}|\,|A||x| + \left|AA^{\dagger}{A^{\dagger}}^{\mathrm{T}}\right|\,|A|^{\mathrm{T}}|r| + |I - AA^{\dagger}|\,|b| \right\|_{\infty}}{\|r\|_{\infty}}, \\
c^{\mathsf{res}}(A,b) &\le = \mathsf{c}^{\mathsf{upper}}_{\mathsf{res}}(\mathtt{A},\mathtt{b}) \\
&:= \left\| \frac{|I - AA^{\dagger}|\,|A||x| + \left|AA^{\dagger}{A^{\dagger}}^{\mathrm{T}}\right|\,|A|^{\mathrm{T}}|r| + |I - AA^{\dagger}|\,|b|}{r} \right\|_{\infty}.
\end{aligned}
$$

Full row rank and underdetermined systems

Suppose $A \in \mathbb{R}^{m \times n}$ with $\mathsf{rank}(A) = m$. Then the Moore-Penrose inverse of A can be written as

$$A^\dagger = A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}.$$

The next result exhibits expressions for the condition numbers of A (for the Moore-Penrose inverse). The proof follows the same lines as that of Theorem 4.1.

Theorem 6.1 *Let $A \in \mathbb{R}^{m\times n}$ be such that $\mathsf{rank}(A) = m$. Then*

$$\begin{aligned}
\|A\|_2\|A^\dagger\|_2 &\le \kappa_2^\dagger(A) = \frac{\max\limits_{\|E\|_2=1} \left\|A^\dagger E A^\dagger - (I - A^\dagger A)E^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}\right\|_2 \|A\|_2}{\|A^\dagger\|_2}\\
&\le \sqrt{2}\|A\|_2\|A^\dagger\|_2,\\
\kappa_F^\dagger(A) &= \frac{\left\|\left({A^\dagger}^{\mathrm{T}} \otimes A^\dagger\right) - ((AA^{\mathrm{T}})^{-1} \otimes (I - A^\dagger A))\,\Pi\right\|_2 \|A\|_F}{\|A^\dagger\|_F}\\
&= \frac{\|A^\dagger\|_2^2\|A\|_F}{\|A^\dagger\|_F},\\
m^\dagger(A) &= \frac{\left\|\left|\left[\left({A^\dagger}^{\mathrm{T}} \otimes A^\dagger\right) - \left((AA^{\mathrm{T}})^{-1} \otimes (I - A^\dagger A)\right)\Pi\right]\right| \mathsf{vec}(|A|)\right\|_\infty}{\|\mathsf{vec}(A^\dagger)\|_\infty};\\
c^\dagger(A) &= \left\|\frac{\left|\left[\left({A^\dagger}^{\mathrm{T}} \otimes A^\dagger\right) - \left((AA^{\mathrm{T}})^{-1} \otimes (I - A^\dagger A)\right)\Pi\right]\right| \mathsf{vec}(|A|)}{\mathsf{vec}(A^\dagger)}\right\|_\infty
\end{aligned}$$

Corollary 6.2 *Let* $A \in \mathbb{R}^{m\times n}$ *be such that* $\mathsf{rank}(A) = m$. *Then*

$$
\begin{aligned}
m^{\dagger}(A) &\le \frac{\left\| |A^{\dagger}||A||A^{\dagger}| + |I - A^{\dagger}A||A^{\mathrm{T}}||(AA^{\mathrm{T}})^{-1}| \right\|_{\max}}{\|A^{\dagger}\|_{\max}}, \\
c^{\dagger}(A) &\le \left\| \frac{|A^{\dagger}||A||A^{\dagger}| + |I - A^{\dagger}A||A^{\mathrm{T}}||(AA^{\mathrm{T}})^{-1}|}{A^{\dagger}} \right\|_{\max}.
\end{aligned}
$$

For underdetermined systems

$$Av = b,$$

where $A \in \mathbb{R}^{m\times n}$ with $\mathsf{rank}(A) = m$ and $b \in \mathbb{R}^{m}$, the set of solutions is an affine subspace of $\mathbb{R}^{n}$ with the dimension of $\mathcal{N}(A)$. It contains a unique point x minimizing the 2-norm. It is well known that this solution is $x = A^{\dagger}b$. Consider the problem of, giving A and b, find x. This problem induces condition numbers $m^{\mathsf{min}}(A, b)$ and $c^{\mathsf{min}}(A, b)$. The following result exhibits expressions for these condition numbers.

Theorem 6.3 *Let $A \in \mathbb{R}^{m\times n}$ with* $\mathsf{rank}(A) = m$ *and* $b \in \mathbb{R}^m$*. Then*

$$m^{\mathsf{min}}(A,b) = \frac{\| |[(x^{\mathrm{T}} \otimes A^{\dagger}) - (I - A^{\dagger}A) \otimes (x^{\mathrm{T}}A^{\dagger})]| \,\mathsf{vec}(|A|) + |A^{\dagger}||b| \|_{\infty}}{\|x\|_{\infty}},$$

$$c^{\mathsf{min}}(A,b) = \left\| \frac{|(x^{\mathrm{T}} \otimes A^{\dagger}) - (I - A^{\dagger}A) \otimes (x^{\mathrm{T}}A^{\dagger})| \,\mathsf{vec}(|A|) + |A^{\dagger}||b|}{x} \right\|_{\infty}.$$

Corollary 6.4 *Let $A \in \mathbb{R}^{m\times n}$ with* $\mathsf{rank}(A) = m$ *and $b \in \mathbb{R}^m$. Then*

$$\begin{aligned} m^{\mathsf{min}}(A,b) &\le \mathtt{m}^{\mathsf{upper}}_{\mathsf{min}}(\mathtt{A},\mathtt{b}) \\ &:= \frac{\left\| |A^{\dagger}||A||x| + |I - A^{\dagger}A||A^{\mathrm{T}}||{A^{\dagger}}^{\mathrm{T}}x| + |A^{\dagger}||b| \right\|_{\infty}}{\|x\|_{\infty}}, \\ c^{\mathsf{min}}(A,b) &\le \mathsf{c}^{\mathsf{upper}}_{\mathsf{min}}(\mathtt{A},\mathtt{b}) \\ &:= \left\| \frac{|A^{\dagger}||A||x| + |I - A^{\dagger}A||A^{\mathrm{T}}||{A^{\dagger}}^{\mathrm{T}}x| + |A^{\dagger}||b|}{x} \right\|_{\infty}. \end{aligned}$$

Numerical Examples and Comparisons with Previous Work

A brief description of some previous work

Probably the first mixed perturbation analysis was done by Skeel [18]. He performed a mixed perturbation analysis for nonsingular linear systems of equations and a mixed error analysis for Gaussian elimination. For nonsingular linear systems $Ax = b$, where $A \in \mathbb{R}^{n\times n}$ Skeel defined the (mixed) condition number as

$$\kappa_\infty(A, b) := \lim_{\varepsilon\to 0} \sup_{\substack{|\Delta A|\le\varepsilon|A| \\ |\Delta b|\le\varepsilon|b|}} \frac{\|\Delta x\|_\infty}{\varepsilon\|x\|_\infty}.$$

He then showed (see also [11, 12]) that

$$\kappa_\infty(A, b) = \frac{\||A^{-1}||A||x| + |A^{-1}||b|\|_\infty}{\|x\|_\infty}.$$

One can get the following relationship

$$\mathrm{cond}_\infty(A,b) \le \kappa_\infty(A,b) \le 2\mathrm{cond}_\infty(A,b),$$

where

$$\mathrm{cond}_\infty(A,b) := \frac{\||A^{-1}||A||x|\|_\infty}{\|x\|_\infty},$$

(also introduced in [18]) and

$$\mathrm{cond}_\infty(A) := \left\||A^{-1}||A|\right\|_\infty \le \kappa_\infty(A) := \|A\|_\infty \|A^{-1}\|_\infty. \qquad (11)$$

A remarkable feature of $\mathrm{cond}_\infty(A)$ is that it is invariant under row scaling (i.e., $\mathrm{cond}_\infty(A) = \mathrm{cond}_\infty(DA)$ for all non-singular diagonal matrix D).

Skeel's condition number is of mixed type. It is defined using componentwise perturbations on the input data and infinity norm in the solution. In [17], Rohn introduced a new relative condition number measuring both perturbation in the input data and error in the componentwise. For the (i,j) entry of matrix inversion and i-th component of $Ax = b$ Rohn defined

$$
\begin{aligned}
c_{ij}(A) &:= \lim_{\varepsilon\to 0} \sup_{|\Delta A|\le\varepsilon|A|} \frac{|(A+\Delta A)^{-1} - A^{-1}|_{ij}}{\varepsilon|A^{-1}|_{ij}}, \\
c_i(A,b) &:= \lim_{\varepsilon\to 0} \sup_{\substack{|\Delta A|\le\varepsilon|A| \\ |\Delta b|\le\varepsilon|b|}} \frac{|(A+\Delta A)^{-1}(b+\Delta b) - A^{-1}b|_i}{\varepsilon|A^{-1}b|_i},
\end{aligned}
$$

and showed that

$$
c_{ij}(A) = \frac{(|A^{-1}||A||A^{-1}|)_{ij}}{|A^{-1}|_{ij}}, \quad c_i(A,b) = \frac{(|A^{-1}||A||x| + |A^{-1}||b|)_i}{|x|_i}.
$$

They were Gohberg and Koltracht [6] who named Skeel's condition number *mixed* to distinguish it from componentwise condition numbers. They also gave explicit expressions for both mixed and componentwise condition numbers, always for square systems of linear equations.

The work of Skeel was soon extended to rectangular systems and matrices. Perturbation theory for rectangular matrices and linear least squares problems already existed for the normwise case (cf. [19, 23]).

Björck in [3] extended (11) to consistent systems $Ax = b$ with $A \in \mathbb{R}^{m\times n}$ and $\mathsf{rank}(A) = n$ (i.e., such that, for the solution x, $r := Ax - b = 0$). He gave the following first order upper bound

$$\|\Delta x\|_\infty \le 2\varepsilon \left(\big\| |A^\dagger||A||x| \big\|_\infty \right) + \mathcal{O}(\varepsilon^2) \le 2\mathrm{cond}_\infty^\dagger(A)\|x\|_\infty + \mathcal{O}(\varepsilon^2),$$

where $\mathrm{cond}_\infty^\dagger(A) = \| |A^\dagger||A| \|_\infty$. Here the perturbation satisfies

$$|\Delta A| \le \varepsilon |A|, \quad |\Delta b| \le \varepsilon |b|.$$

Geurts [5] first gave an expression for the normwise condition number for full column rank linear least squares problems, with respect to the Frobenius norm, when there is only perturbation on A (without perturbations on b). He proved that

$$\begin{aligned} \kappa_F^G(A,b) &:= \lim_{\varepsilon\to 0} \sup_{\|\Delta A\|_F \le \varepsilon \|A\|_F} \frac{\|\Delta x\|_2}{\varepsilon \|x\|_2} \\ &= \|A\|_F \|A^\dagger\|_2 \left(\frac{\|A^\dagger\|_2^2 \|b - Ax\|_2^2}{\|x\|_2^2} + 1 \right)^{\frac{1}{2}} . \end{aligned} \qquad (12)$$

Gratton [9] considered perturbations on both A and b and gave an expression for the normwise condition number, with respect to a "weighted" Frobenius norm of the pair (A, b), for full column rank linear least squares problems. For $\alpha, \beta > 0$ he showed that

$$\begin{aligned}\kappa_{F(\alpha,\beta)}(A, b) &:= \lim_{\varepsilon\to 0} \sup_{\|[\alpha\Delta A,\beta\Delta b]\|_F\le\varepsilon\|[\alpha A,\beta b]\|_F} \frac{\|\Delta x\|_2}{\varepsilon\|x\|_2} \\ &= \frac{\|A^\dagger\|_2 \left\|\left[\alpha A,\ \ \beta b\right]\right\|_F}{\|x\|_2}\sqrt{\frac{\|x\|_2^2 + \|A^\dagger\|_2^2\|r\|_2^2}{\alpha^2} + \frac{1}{\beta^2}}.\end{aligned}$$

The authors claim they considered this weighted Frobenius norm due to its flexibility. Taking large values of α amounts to perturb b only.

Malyshev [14] defined a normwise condition number for full column rank LS problems when there is only perturbation on A as

$$\kappa_{F,2}(A) := \lim_{\varepsilon\to 0} \sup_{\|\Delta A\|_F\le\varepsilon} \left(\frac{\|\Delta x\|_2}{\|x\|_2}\Big/\frac{\|\Delta A\|_F}{\|A\|_2}\right)$$

and proved that

$$\kappa_{F,2}(A) = \|A\|_2 \|A^\dagger\|_2 \sqrt{1 + \left(\|A^\dagger\|_2 \frac{\|r\|_2}{\|x\|_2} \right)^2}.$$

Recently Grcar [10] gave an optimal backward error analysis for full column rank linear least squares problems and introduced another approach to obtain expressions for condition numbers. He used the optimal backward error to define the condition number $\chi_F^{\mathsf{LS,rel}}(A)$ and obtained, in the normwise case and for the Frobenius norm, an expression for $\chi_F^{\mathsf{LS,rel}}(A)$ similar to that by Geurts (12) above. More precisely, assume $\mathsf{rank}(A) = n$ and let x be the solution of

$$\min_u \|b - Au\|_2,$$

and $x + \Delta x$ be that of the same problem for the perturbed matrix $A + \Delta A$. Then the optimal backward error is defined as

$$\mu_2^{\mathsf{LS}}(x + \Delta x) = \min_{x+\Delta x = \underset{u}{\operatorname{argmin}} \|b-(A+E)u\|_2} \|E\|_2.$$

The quantity $\mu_F^{\mathsf{LS}}(x+\Delta x)$ is defined similarly but taking the Frobenius norm instead of the spectral one. Grcar related this optimal backward error to conditioning by proving that

$$\begin{aligned}
\chi_2^{\mathsf{LS,rel}}(A) &:= \limsup_{\|\Delta x\|_2\to 0} \frac{\|\Delta x\|_2}{\|\Delta A\|_2}\cdot\frac{\|A\|_2}{\|x\|_2} = \limsup_{\|\Delta x\|_2\to 0} \frac{\|\Delta x\|_2}{\mu_2^{\mathrm{LS}}(x+\Delta x)}\cdot\frac{\|A\|_2}{\|x\|_2},\\
\chi_F^{\mathsf{LS,rel}}(A) &:= \limsup_{\|\Delta x\|_2\to 0} \frac{\|\Delta x\|_2}{\|\Delta A\|_F}\cdot\frac{\|A\|_F}{\|x\|_2} = \limsup_{\|\Delta x\|_2\to 0} \frac{\|\Delta x\|_2}{\mu_F^{\mathrm{LS}}(x+\Delta x)}\cdot\frac{\|A\|_F}{\|x\|_2}.
\end{aligned}$$

He then proved [10, Theorem 5.1]

$$\begin{aligned}
\chi_F^{\mathsf{LS,rel}}(A) &= \|A\|_F\|A^\dagger\|_2\left(1+\frac{\|A^\dagger|_2^2\|r\|_2^2}{\|x\|_2^2}\right)^{1/2},\\
\left(1+\frac{\|A^\dagger\|_2^2\|r\|_2^2}{\|x\|_2^2}\right)^{1/2}\|A\|_2\|A^\dagger\|_2 &\le \chi_2^{\mathsf{LS,rel}}(A)\\
&\le \left(1+\frac{\|A^\dagger\|_2\|r\|_2}{\|x\|_2}\right)\|A\|_2\|A^\dagger\|_2.
\end{aligned}$$

A mixed perturbation analysis for full column rank linear least squares problems first appears in [3] and variations of it appear in [1]. The first order perturbation bound shown in [3] is

$$\|\Delta x\|_\infty \leq \varepsilon \left(\left\| |A^\dagger||A||x| + |A^\dagger||b| \right\|_\infty + \left\| \left| \left(A^{\mathrm{T}}A\right)^{-1} \right| |A^{\mathrm{T}}| \, |r| \right\|_\infty \right) + \mathcal{O}(\varepsilon^2).$$

Numerical Experiments

In this subsection we report the results of some numerical experiments and we use them to compare the bounds obtained in this paper with those obtained in previous works. All the computations were carried out using MATLAB 7.0 with machine precision $\epsilon \approx 2.2 \times 10^{-16}$.

(1) Björck [3] derived the following first order mixed[1] perturbation bound for LS,

$$\|\Delta x\|_\infty \leq \varepsilon \left(\left\| |A^\dagger||A||x| + |A^\dagger||b| \right\|_\infty + \left\| \left| \left(A^{\mathrm{T}}A\right)^{-1} \right| |A^{\mathrm{T}}| \, |r| \right\|_\infty \right) + \mathcal{O}(\varepsilon^2).$$

[1] In [3] it is called "componentwise" but in this paper we follow the terminology introduced in [6].

Recall, here the perturbation satisfies

$$|\Delta A| \le \varepsilon |A|, \quad |\Delta b| \le \varepsilon |b|.$$

From Theorem 5.3 we can also deduce a first order mixed perturbation bound for LS,

$$\begin{aligned} \|\Delta x\|_\infty \ \le \ & \varepsilon \left\| \left| \left[-\left(x^{\mathrm{T}} \otimes A^\dagger\right) + \left(r^{\mathrm{T}} \otimes (A^{\mathrm{T}}A)^{-1}\right) \Pi \right] \right| \mathsf{vec}(|A|) + |A^\dagger||b| \right\|_\infty \\ + \ & \mathcal{O}(\varepsilon^2). \end{aligned}$$

Let

$$\mu_{\mathrm{old}} = \left\| |A^\dagger||A||x| + |A^\dagger||b| \right\|_\infty + \left\| \left| \left(A^{\mathrm{T}}A\right)^{-1} \right| \left|A^{\mathrm{T}}\right| |r| \right\|_\infty ,$$

and

$$\mu_{\mathrm{new}} = \left\| \left| \left[-\left(x^{\mathrm{T}} \otimes A^\dagger\right) + \left(r^{\mathrm{T}} \otimes (A^{\mathrm{T}}A)^{-1}\right) \Pi \right] \right| \mathsf{vec}(|A|) + |A^\dagger||b| \right\|_\infty .$$

By Corollary 5.4 and using the triangular inequality, we have that

$$\mu_{\mathrm{new}} \le \mu_{\mathrm{old}}.$$

The following example shows that μ_{new} can be substantially smaller than μ_{old}.

Example 1 *Let* $A = \begin{bmatrix} 2.2288 & -0.5756 \\ -0.2108 & 0.0557 \\ -2.1716 & 0.5622 \end{bmatrix}$ *and* $b = \begin{bmatrix} 0.001 \\ 0 \\ 0 \end{bmatrix}$. *Then*

$$A^{\dagger} = \begin{bmatrix} 82.0630 & 125.0857 & 71.6261 \\ 316.4566 & 483.7978 & 277.8458 \end{bmatrix}, \quad x = \begin{bmatrix} 0.0808 \\ 0.3115 \end{bmatrix},$$

$$r = 1.0\mathit{e\text{-}003} \times \begin{bmatrix} 0.2577 \\ -0.3222 \\ 0.2958 \end{bmatrix}$$

and

$$\mu_{\text{old}} = 494.1189, \quad \mu_{\text{new}} = 65.1780, \quad \frac{\mu_{\text{old}} - \mu_{\text{new}}}{\mu_{\text{old}}} = 0.8681.$$

(2) Example 1 shows that μ_{new} can be substantially smaller than μ_{old}. A natural question is how much smaller is “in general.” A possible answer to this question could be obtained by considering the average of both μ_{old} and μ_{new} for random pairs (A, b).

The following table does so. We considered pairs (A, b) with A a $m \times n$ random matrix whose entries are independent random variables normally distributed (mean zero, variance one) and b a n-dimensional random vector with the same distribution. Each row in the table corresponds to a pair (m, n). Averages are over a sample of 1000 pairs (A, b).

(3) N. Higham obtained unrestricted mixed bounds for the solution and residue of full-column rank LS problems.

Theorem 7.1 [12, Theorem 19.2] *Let $A \in \mathbb{R}^{m\times n}$, $m \geq n$ and $A + \Delta A$ be of full rank. Let $x, y \in \mathbb{R}^n$ be the solutions of* (6) *and* (8), *respectively and let $r = Ax - b$, $s = (A + \Delta A)y - (b + \Delta b)$. Then, for any monotonic norm $\|\ \|$,*

$$\frac{\|y-x\|}{\|x\|} \leq \varepsilon \frac{\||A^\dagger|\,(|b|+|A||y|)\| + \||(A^{\mathrm{T}}A)^{-1}|\,|A^{\mathrm{T}}||s|\|}{\|x\|},$$

$$\frac{\|r-s\|}{\|r\|} \leq \varepsilon \frac{\||I-AA^\dagger|\,(|b|+|A||y|)\| + \left\||A^{\dagger^{\mathrm{T}}}|A^{\mathrm{T}}||s|\right\|}{\|x\|},$$

These bounds are approximately attainable.

The following example demonstrates that our bound (10) is sharper

m	n	$\mu_{\text{old}}^{\text{av}}$	$\mu_{\text{new}}^{\text{av}}$	$(\mu_{\text{old}}^{\text{av}} - \mu_{\text{new}}^{\text{av}})/\mu_{\text{old}}^{\text{av}}$
4	3	19.6395	15.5837	0.2065
8	6	16.0568	13.0890	0.1848
12	9	15.6835	12.5166	0.2019
16	12	18.2260	14.4856	0.2052
20	15	19.5803	15.6184	0.2023
40	30	24.2459	19.2853	0.2046
50	33	27.1138	21.6387	0.2019
80	52	30.2019	24.0522	0.2036
100	57	23.0314	18.3845	0.2018
100	70	42.4747	33.7707	0.2049
200	100	21.8254	17.3910	0.2032
200	150	70.4169	55.6292	0.2100
250	150	35.8237	28.4127	0.2069
300	150	25.1857	20.0293	0.2047
400	200	28.0440	22.2526	0.2065
500	250	30.5930	24.2659	0.2068
600	200	16.6290	13.2902	0.2008
800	200	12.7852	10.2728	0.1965

than the bound in Theorem 7.1. It also shows that the first order bound μ_{new} for $m^{\mathsf{ls}}(A, b)$ is attainable.

Example 2 *Let* $\varepsilon = 10^{-4}$ *and*

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \Delta A = \varepsilon \begin{bmatrix} 0 & -1 & 0 \\ -1 & 1 & -1 \\ 0 & -1 & 1 \\ 1 & -1 & 0 \end{bmatrix}, \quad \Delta b = \varepsilon \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Then $|\Delta A| = \varepsilon A$ *and* $|\Delta b| = \varepsilon b$. *We can compute the solution*

$$x = A^{\dagger} b = \begin{bmatrix} -0.5 \\ 0.75 \\ -0.5 \end{bmatrix},$$

and

$$y = (A + \Delta A)^{\dagger}(b + \Delta b) = \begin{bmatrix} -0.50025005499999 \\ 0.75032507250149 \\ -0.50025005499999 \end{bmatrix}.$$

α	3.250725014936062 e-004
β	3.249999999999998 e-004
η_{old}	1.887415685321830 e+000
η_{new}	3.250725014935217 e-004
γ_{old}	4.249999999999997 e-004
γ_{new}	3.249999999999998 e-004

Now we denote

$$\begin{aligned} \alpha &= \|y-x\|_\infty, \quad \beta = \left\|-A^\dagger \Delta A x + \left(A^{\mathrm{T}}A\right)^{-1} r + A^\dagger \Delta b\right\|_\infty, \\ \eta_{\text{old}} &= \varepsilon\, \left\||A^\dagger|\,(|b|+|A||y|)\right\| + \left\||(A^{\mathrm{T}}A)^{-1}|\,|A^{\mathrm{T}}||s|\right\|_\infty, \\ \eta_{\text{new}} &= \varepsilon\, \left\|\left|\left[\left(y^{\mathrm{T}} \otimes A^\dagger\right) - \left(s^{\mathrm{T}} \otimes (A^{\mathrm{T}}A)^{-1}\right)\Pi\right]\right| \mathsf{vec}(|A|) + |A^\dagger||b|\right\|_\infty, \\ \gamma_{\text{old}} &= \varepsilon\, \mu_{\text{old}}, \\ \gamma_{\text{new}} &= \varepsilon\, \mu_{\text{new}}, \end{aligned}$$

and compare the results in the following table. We see that

$\alpha - \gamma_{\text{new}} = 7.250149360645691 \times \varepsilon^2, \alpha \approx \eta_{\text{new}}, \alpha < \eta_{\text{old}}, \beta < \gamma_{\text{old}}, \beta \approx \gamma_{\text{new}}.$

Then, for the perturbation $(\Delta A, \Delta b)$, *the first order upper bound*

μ_{new} *is attainable (since* $\beta \approx \gamma_{\text{new}}$*). This means our bound is sharper than the one shown in [3]. We also note that the bound* η_{old} *shown in Theorem 7.1 is larger than* η_{new}*. Actually,*

$$\eta_{\text{old}}/\varepsilon = 1.887415685321829e{+}004, \eta_{\text{new}}/\varepsilon = 3.250725014935217e{+}000.$$

(4) Wedin [23] proved the following normwise perturbation result for full column rank LS (see also [12, Theorem 19.1]) when the perturbation $(\Delta A, \Delta b)$ satisfies $\|\Delta A\|_2 \le \varepsilon \|A\|_2$, $\|\Delta b\|_2 \le \varepsilon \|b\|_2$ and $\varepsilon \|A\|_2 \|A^\dagger\|_2 < 1$,

$$\begin{aligned} \frac{\|\Delta x\|_2}{\|x\|_2} &\le \frac{\varepsilon \|A\|_2 \|A^\dagger\|_2}{1 - \varepsilon \|A\|_2 \|A^\dagger\|_2} \left(2 + (\|A\|_2 \|A^\dagger\|_2 + 1) \frac{\|r\|_2}{\|A\|_2 \|x\|_2} \right) \\ &= \varepsilon U_{\text{W}} + \mathcal{O}(\varepsilon^2). \end{aligned}$$

Here $U_{\text{W}} = \|A\|_2 \|A^\dagger\|_2 \left(2 + (1 + \|A\|_2 \|A^\dagger\|_2) \frac{\|r\|_2}{\|A\|_2 \|x\|_2} \right)$.

Taking $\alpha = \beta = 1$ in Gratton's result (13) we obtain the bound

$$\kappa_{F(1,1)}(A, b) = \frac{\|A^\dagger\|_2 \left\| \begin{bmatrix} A & b \end{bmatrix} \right\|_F}{\|x\|_2} \sqrt{\|x\|_2^2 + \|A^\dagger\|_2^2 \|r\|_2^2 + 1}.$$

U_{W}	$\kappa_{F(1,1)}(A,b)$	$\kappa_2^{\mathsf{ls}}(A,b)$	$\kappa_F^{\mathsf{ls}}(A,b)$
1.0613e+006	1.0225e+006	6.7065e+005	7.6924e+005

$m^{\mathsf{ls}}(A,b)$	$c^{\mathsf{ls}}(A,b)$
5.2061e+003	1.0495e+004

We compare our normwise, mixed and componentwise condition numbers with $\kappa_{F(1,1)}(A,b)$ and the first order bound U_{W} in the following table. Here we take $m = 25$, $n = 10$ and A the $m \times n$ Vandermonde matrix whose (i,j)-th element is given by $\Big((j-1)/(n-1)\Big)^{i-1}$. The vector b was randomly generated.

In this table we can see that, for this particular example, mixed and componentwise condition numbers are smaller than the normwise condition number, and the new normwise $\kappa_2^{\mathsf{ls}}(A,b)$, $\kappa_F^{\mathsf{ls}}(A,b)$ are smaller than the previous normwise bounds derived by Wedin and Gratton for the relative error of the solution x.

We also used the pair (A,b) above to compare the upper bounds in Corollary 5.4 with $m^{\mathsf{ls}}(A,b)$ and $c^{\mathsf{ls}}(A,b)$ as well as with Wedin and

$m^{\mathsf{ls}}(A,b)$	$m_{\mathsf{ls}}^{\mathsf{upper}}(A,b)$	$c^{\mathsf{ls}}(A,b)$	$c_{\mathsf{ls}}^{\mathsf{upper}}(A,b)$
5.2061e+003	1.0788e+004	1.0495e+004	1.0788e+004

$m^{\mathsf{res}}(A,b)$	$m_{\mathsf{res}}^{\mathsf{upper}}(A,b)$	$c^{\mathsf{res}}(A,b)$	$c_{\mathsf{res}}^{\mathsf{upper}}(A,b)$
2.9734e+003	3.5654e+003	6.0366e+004	2.0788e+004

Gratton bounds above.

We can see that the bounds $m_{\mathsf{ls}}^{\mathsf{upper}}(A,b)$ and $c_{\mathsf{ls}}^{\mathsf{upper}}(A,b)$ are not too far away from the quantities they bound and that they are sharper than Wedin and Gratton bounds.

Similarly, the next table shows that, always for the pair (A,b) above, the bounds $m_{\mathsf{res}}^{\mathsf{upper}}(A,b)$ and $c_{\mathsf{res}}^{\mathsf{upper}}(A,b)$ are, again, not too far away from the quantities they bound.

Let $\varepsilon = 10^{-8}$ and (E,f) be random (the distribution of each entry being uniform on the interval $(-1,1)$). Then let $\Delta A_{ij} = \varepsilon E_{ij} A_{ij}$ and $\Delta b_i = \varepsilon f_i b_i$. Note that $|\Delta A| \leq \varepsilon |A|$ and $|\Delta b| \leq \varepsilon |b|$. Let $x + \Delta x = (A + \Delta A)^{\dagger}(b + \Delta b)$ be the solution of the perturbed problem.

γ_1	1.2198e-005	$\varepsilon\, U_{\mathrm{W}}$	0.0106
γ_1	1.2198e-005	$\varepsilon\, \kappa_{F(1,1)}(A,b)$	0.0102
γ_1	1.2198e-005	$\varepsilon\, \kappa_2^{\mathsf{ls}}(A,b)$	0.0067
γ_1	1.2198e-005	$\varepsilon\, \kappa_F^{\mathsf{ls}}(A,b))$	0.0077
γ_2	1.2157e-005	$\varepsilon\, m^{\mathsf{ls}}(A,b)$	5.2061 e-005
γ_3	2.5315e-005	$\varepsilon\, c^{\mathsf{ls}}(A,b)$	1.0495 e-004
γ_4	1.4204e-006	$\varepsilon\, m^{\mathsf{res}}(A,b)$	2.9734e-005
γ_5	8.8357e-006	$\varepsilon\, c^{\mathsf{res}}(A,b)$	6.0366e-004

Denote

$$\gamma_1 = \frac{\|\Delta x\|_2}{\|x\|_2}, \gamma_2 = \frac{\|\Delta x\|_\infty}{\|x\|_\infty}, \gamma_3 = \left\|\frac{\Delta x}{x}\right\|_\infty, \gamma_4 = \frac{\|\Delta r\|_\infty}{\|r\|_\infty}, \gamma_5 = \left\|\frac{\Delta r}{r}\right\|_\infty.$$

Then (performing this experiment once) we obtained

Again, we see that mixed and componentwise perturbation bounds are tighter than normwise perturbation bounds.

(5) In the sequel we consider underdetermined linear systems $Ax = b$ with A full row rank. In this case, Golub and Van Loan [7, Theorem 5.7.1] proved the following normwise perturbation result.

Theorem 7.2 *Let $A \in \mathbb{R}^{m\times n}$ and $0 \neq b \in \mathbb{R}^m$. Suppose that* $\mathsf{rank}(A) = m \leq n$ *and that $\Delta A \in \mathbb{R}^{m\times n}$ and $\Delta b \in \mathbb{R}^m$ satisfy*

$$\varepsilon = \max\{\|\Delta A\|_2/\|A\|_2,\ \|\Delta b\|_2/\|b\|_2\} < \sigma_m(A).$$

If x and $x + \Delta x$ are the minimum norm solutions of $Ax = b$ and $(A + \Delta A)y = b + \Delta b$, respectively, then

$$\frac{\|\Delta x\|_2}{\|x\|_2} \leq 3\kappa_2(A)\varepsilon + \mathcal{O}(\varepsilon^2), \tag{13}$$

where $\kappa_2(A) = \|A\|_2\|A^\dagger\|_2$.

Demmel and Higham got the following mixed (but, in contrast with our exposition, with respect to the 2-norm) perturbation result.

Theorem 7.3 [4, Theorem 2.1] *Let $A \in \mathbb{R}^{m\times n}$ and $0 \neq b \in \mathbb{R}^m$. Suppose that* $\mathsf{rank}(A) = m \leq n$ *and that*

$$|\Delta A| \leq \varepsilon E, \quad |\Delta b| \leq \varepsilon f,$$

where $E \geq 0, f \geq 0$ and $\varepsilon\|E\|_2\|A^\dagger\|_2 < 1$. If x and y are the minimum norm solutions to $Ax = b$ and $(A + \Delta A)y = b + \Delta b$

respectively, then

$$\frac{\|\Delta x\|_2}{\|x\|_2} \le \frac{\varepsilon}{\|x\|_2}\left(\left\||I - A^{\dagger}A|E^{\mathrm{T}}|A^{\dagger^{\mathrm{T}}}x|\right\|_2 + \left\||A^{\dagger}|\,(f + E|x|)\right\|_2\right) + \mathcal{O}(\varepsilon^2). \tag{14}$$

When $E = |A|$ *and* $f = |b|$,

$$\frac{\|\Delta x\|_2}{\|x\|_2} \le 3\mathrm{cond}_2(A)\varepsilon + \mathcal{O}(\varepsilon^2), \tag{15}$$

where $\mathrm{cond}_2(A) = \||A||A^{\dagger}|\|_2$.

From Theorem 6.3, we obtain the first order perturbation bounds

$$\begin{aligned} \frac{\|\Delta x\|_\infty}{\|x\|_\infty} &\le \ \varepsilon m^{\mathsf{min}}(A, b) + \mathcal{O}(\varepsilon^2), \\ \left\|\frac{\Delta x}{x}\right\|_\infty &\le \ \varepsilon c^{\mathsf{min}}(A, b) + \mathcal{O}(\varepsilon^2). \end{aligned}$$

The following table compare these bounds with $3\kappa_2(A)$ and $3\mathrm{cond}_2(A)$ (note, however, that $3\mathrm{cond}_2(A)$ and $m^{\mathsf{min}}(A, b)$ bound $\frac{\|\Delta x\|}{\varepsilon\|x\|}$ for diferent

$3\kappa_2(A)$	$3\mathrm{cond}_2(A)$	$m^{\mathsf{min}}(A,b)$	$c^{\mathsf{min}}(A,b)$
6.0797e+004	5.0055 e+004	8.9969e+003	1.5667 e+004

$m^{\mathsf{min}}(A,b)$	$m^{\mathsf{upper}}_{\mathsf{min}}(A,b)$	$c^{\mathsf{min}}(A,b)$	$c^{\mathsf{upper}}_{\mathsf{min}}(A,b)$
8.9969e+003	1.2449e+004	1.5667 e+004	1.6588e+004

norms). Here we took $m = 7$, $n = 10$ and A the $m \times n$ Vandermonde matrix whose (i, j)-th element is given by $\Big((j-1)/(n-1)\Big)^{i-1}$. The vector b was randomly generated.

We see again that mixed and componentwise condition numbers are smaller than normwise condition numbers and that the bound $m^{\mathsf{min}}(A,b)$ is sharper than $3\mathrm{cond}_2(A)$. Also, we can compare the upper bounds in Corollary 6.4 both with $m^{\mathsf{min}}(A,b)$ and $c^{\mathsf{min}}(A,b)$ and with $3\kappa_2(A)$ and $3\mathrm{cond}_2(A)$.

We find again that $m^{\mathsf{upper}}_{\mathsf{min}}(A,b)$ and $c^{\mathsf{upper}}_{\mathsf{min}}(A,b)$ are not too far away from the quantities they bound and that they are sharper than the bounds in Theorems 7.2 and 7.3.

γ_1	2.3592e-005	$\varepsilon\, 3\kappa_2(A)$	6.0797 e-004
γ_2	3.3334e-005	$\varepsilon\, 3\mathrm{cond}_2(A)$	5.0055 e-004
γ_2	3.3334e-005	$\varepsilon\, m^{\mathsf{min}}(A,b)$	8.9969 e-005
γ_3	5.8047e-005	$\varepsilon\, c^{\mathsf{min}}(A,b)$	1.5667 e-004

Let $(\Delta A, \Delta b)$ and Δx be as in the end of (4) above. Denote

$$\gamma_1 = \frac{\|\Delta x\|_2}{\|x\|_2}, \quad \gamma_2 = \frac{\|\Delta x\|_\infty}{\|x\|_\infty}, \quad \gamma_3 = \left\|\frac{\Delta x}{x}\right\|_\infty .$$

Comparing these quantities with their bounds we obtain and verify, again, that mixed and componentwise give tighter bounds on relative errors.

References

[1] M. Arioli, I.S. Duff and P.P.M. de Rijk, *An augmented system approach to sparse least-squares problems*, Numer. Math. 55(1989), pp.667-684.

[2] A. Ben-Israel and T.N.E. Greville, *Generalized Inverses: Theory and Applications*, 2nd Edition, Springer Verlag, New York, 2003.

[3] Å. Björck, *Component-wise perturbation analysis and error bounds for linear least squares solutions*, BIT, 31(1991), pp.238-244.

[4] J. Demmel and N. Higham, *Improved error bounds for underdetermined system solvers*, SIAM J. Matrix Anal. Appl., 14 (1993), pp.1-14.

[5] A.J. Geurts, *A Contribution to the theory of condition*, Numer. Math., 39(1982), pp. 85-96.

[6] I. Gohberg and I. Koltracht, *Mixed, componentwise, and structured condition numbers*, SIAM J. Matrix Anal. Appl., 14(1993), pp. 688-704.

[7] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 3rd Edition, John Hopkins University Press, Baltimore, MD, 1996.

[8] A. Graham, *Kronecker Products and Matrix Calculus with Application*, Wiley, New York, 1981.

[9] S. Gratton, *On the condition number of linear least squares problems in a weighted Frobenius norm*, BIT, 36(1996), no.3, pp. 523-530.

[10] J.F. Grcar, *Optimal sensitivity analysis of linear least squares*, Lawrence Berkeley National Laboratory, Report LBNL-52434, 2003.

[11] N.J. Higham, *A survey of componentwise perturbation theory in numerical linear algebra,* Proceedings of Symposia in Applied Mathematics, Vol.48, 1994, pp. 49-77.

[12] N.J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd edition, SIAM, Philadelphia, 2002.

[13] R.A. Horn and C.R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

[14] A.N. Malyshev, *A unified theory of conditioning for linear least squares and Tikhonov regularization solutions*, SIAM J. Matrix Anal. Appl., 24(2003), no.4, pp. 1186-1196.

[15] C.D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia, 2000.

[16] J.R. Rice, *A theory of condition,* SIAM J. Numer. Anal., 3(1966), pp. 217-232.

[17] J.Rohn, *New condition numbers for matrices and linear systems,* Computing, 41(1989), pp. 167-169.

[18] R.D. Skeel, *Scaling for numerical stability in Gaussian elimination*, J. Assoc. Comput. Mach., 26(1979), No.3, pp.817-526.

[19] G.W. Stewart, *On the perturbation of pseudo-inverses, projections and linear least sqaures problems*, SIAM Rev., 19 (1977), pp.634-662.

[20] G.W. Stewart and J.-G. Sun, *Matrix Perturbation Theory*, Academic Press, New York, 1990.

[21] C.F. Van Loan, *The ubiquitous Kronecker product*, J. Comput. Appl. Math., 123(2000), pp. 85-100.

[22] G. Wang, Y. Wei and S. Qiao, *Generalized Inverses: Theory and Computations*, Science Press, Beijing/New York, 2004.

[23] P.Å. Wedin, *Perturbation theory for pseudo-inverses*, BIT, 13(1973), pp.217-232.