



HAL
open science

Statistical estimates for the conditioning of linear least squares problems

Marc Baboulin, Serge Gratton, Rémi Lacroix, Alan J Laub

► **To cite this version:**

Marc Baboulin, Serge Gratton, Rémi Lacroix, Alan J Laub. Statistical estimates for the conditioning of linear least squares problems. International Conference on Parallel Processing and Applied Mathematics, Sep 2013, Warsaw, Poland. pp.124-133, 10.1007/978-3-642-55224-3_13 . hal-01766923

HAL Id: hal-01766923

<https://hal.science/hal-01766923>

Submitted on 14 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Statistical estimates for the conditioning of linear least squares problems

Marc Baboulin¹, Serge Gratton², Rémi Lacroix¹, and Alan J. Laub⁴

¹ Université Paris-Sud and Inria, France

[marc.baboulin,remi.lacroix]@inria.fr

² ENSEEIHT and CERFACS, France

serge.gratton@enseeiht.fr

³ University of California Los Angeles, USA

laub@ats.ucla.edu

Abstract. In this paper we are interested in computing linear least squares (LLS) condition numbers to measure the numerical sensitivity of an LLS solution to perturbations in data. We propose a statistical estimate for the normwise condition number of an LLS solution where perturbations on data are measured using the Frobenius norm for matrices and the Euclidean norm for vectors. We also explain how condition numbers for the components of an LLS solution can be computed. We present numerical experiments that compare the statistical condition estimates with their corresponding exact values.

Keywords: Linear least squares, condition number, statistical condition estimation, componentwise conditioning.

1 Introduction

We consider the overdetermined linear least squares (LLS) problem

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2, \quad (1)$$

with $A \in \mathbb{R}^{m \times n}$, $m \geq n$ and $b \in \mathbb{R}^m$. We assume throughout this paper that A has full column rank and as a result, Equation (1) has a unique solution $x = A^+b$ where A^+ is the Moore-Penrose pseudoinverse of the matrix A , expressed by $A^+ = (A^T A)^{-1} A^T$. We can find for instance in [7, 13, 19] a comprehensive survey of the methods that can be used for solving efficiently and accurately LLS problems.

The condition number is a measure of the sensitivity of a mapping to perturbations. It was initially defined in [23] as the maximum amplification factor between a small perturbation in the data and the resulting change in the problem solution. Namely, if the solution x of a given problem can be expressed as a function $g(y)$ of a data y , then if g is differentiable (which is the case for many linear algebra problems), the absolute condition number of g at y can be defined as (see e.g. [12])

$$\kappa(y) = \max_{z \neq 0} \frac{\|g'(y) \cdot z\|}{\|z\|}. \quad (2)$$

From this definition, $\kappa(y)$ is a quantity that, for a given perturbation size on the data y , allows us to predict *to first order* the perturbation size on the solution x . Associated with a backward error [26], condition numbers are useful to assess the numerical quality of a computed solution. Indeed numerical algorithms are always subject to errors although their sensitivity to errors may vary. These errors can have various origins like for instance data uncertainty due to instrumental measurements or rounding and truncation errors inherent to finite precision arithmetic.

LLS can be very sensitive to perturbations in data and it is crucial to be able to assess the quality of the solution in practical applications [4]. It was shown in [14] that the 2-norm condition number $\text{cond}(A)$ of the matrix A plays a significant role in LLS sensitivity analysis. It was later proved in [25] that the sensitivity of LLS problems is proportional to $\text{cond}(A)$ when the residual vector is small and to $\text{cond}(A)^2$ otherwise. Then [12] provided a closed formula for the condition number of LLS problems, using the Frobenius norm to measure the perturbations of A . Since then many results on normwise LLS condition numbers have been published (see e.g. [2, 7, 11, 15, 16]).

It was observed in [18] that normwise condition numbers can lead to a loss of information since they consolidate all sensitivity information into a single number. Indeed in some cases this sensitivity can vary significantly among the different solution components (some examples for LLS are presented in [2, 21]). To overcome this issue, it was proposed the notion of “componentwise” condition numbers or condition numbers for the solution components [9]. Note that this approach must be distinguished from the componentwise metric also applied to LLS for instance in [5, 10]. This approach was generalized by the notion of *partial* or *subspace* condition numbers which corresponds to conditioning of $L^T x$ with $L \in \mathbb{R}^{n \times k}$, $k \leq n$, proposed for instance in [2, 6] for least squares and total least squares, or [8] for linear systems. The motivation for computing the conditioning of $L^T x$ can be found for instance in [2, 3] for normwise LLS condition numbers.

Even though condition numbers provide interesting information about the quality of the computed solution, they are expected to be calculated in an acceptable time compared to the cost for the solution itself. Computing the exact (subspace or not) condition number requires $\mathcal{O}(n^3)$ flops when the LLS solution x has been already computed (e.g., using a QR factorization) and can be reused to compute the conditioning [2, 3]. This cost is affordable when compared to the cost for solving the problem ($\mathcal{O}(2mn^2)$ flops when $m \gg n$). However statistical estimates can reduce this cost to $\mathcal{O}(n^2)$ [17, 20]. The theoretical quality of the statistical estimates can be formally measured by the probability to give an estimate in a certain range around the exact value. In this paper we summarize closed formulas for the condition numbers of the LLS solution and of its components, and we propose practical algorithms to compute statistical estimates of these quantities. In particular we derive a new expression for the statistical estimate of the conditioning of x . We also present numerical experiments to compare LLS conditioning with the corresponding statistical estimates.

Notations The notation $\|\cdot\|_2$ applied to a matrix (resp. a vector) refers to the spectral norm (resp. the Euclidean norm) and $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. The matrix I is the identity matrix and e_i is the i th canonical vector. The uniform continuous distribution between a and b is abbreviated $\mathcal{U}(a, b)$ and the normal distribution of mean μ and variance σ^2 is abbreviated $\mathcal{N}(\mu, \sigma^2)$. $\text{cond}(A)$ denotes the 2-norm condition number of a matrix A , defined as $\text{cond}(A) = \|A\|_2 \|A^+\|_2$. The notation $|\cdot|$ applied to a matrix or a vector holds componentwise.

2 Condition estimation for linear least squares

In Section 2.1 we are concerned in calculating the condition number of the LLS solution x and in Section 2.2 we compute or estimate the conditioning of the components of x . We suppose that the LLS problem has already been solved using a QR factorization (the normal equations method is also possible but the condition number is then proportional to $\text{cond}(A)^2$ [7, p. 49]). Then the solution x , the residual $r = b - Ax$, and the factor $R \in \mathbb{R}^{n \times n}$ of the QR factorization of A are readily available (we recall that the Cholesky factor of the normal equations is, in exact arithmetic, equal to R up to some signs). We also make the assumption that both A and b can be perturbed, these perturbations being measured using the weighted product norm $\|(\Delta A, \Delta b)\|_F = \sqrt{\|\Delta A\|_F^2 + \|\Delta b\|_2^2}$ where ΔA and Δb are absolute perturbations of A and b . In addition to providing us with simplified formulas, this product norm has the advantage, mentioned in [15], to be appropriate for estimating the forward error obtained when the LLS problem is solved via normal equations.

2.1 Conditioning of the least squares solution

Exact formula We can obtain from [3] a closed formula for the absolute condition number of the LLS solution as

$$\kappa_{LS} = \|R^{-1}\|_2 \left(\|R^{-1}\|_2^2 \|r\|_2^2 + \|x\|_2^2 + 1 \right)^{\frac{1}{2}}, \quad (3)$$

where x , r and R are exact quantities.

This equation requires mainly to compute the minimum singular value of the matrix A (or R), which can be done using iterative procedures like the inverse power iteration on R , or more expensively with the full SVD of R ($\mathcal{O}(n^3)$ flops). Note that $\|R^{-T}\|_2$ can be approximated by other matrix norms (see [19, p. 293]).

Statistical estimate Similarly to [8] for linear systems, we can estimate the condition number of the LLS solution using the method called *small-sample theory* [20] that provides statistical condition estimates for matrix functions.

Let us denote by $x(A, b)$ the expression of x as a function of the data A and b . Since A has full rank n , $x(A, b)$ is continuously F-differentiable in a neighborhood

of (A, b) . If $x'(A, b)$ is the derivative of this function, then $x'(A, b) \cdot (\Delta A, \Delta b)$ denotes the image of $(\Delta A, \Delta b)$ by the linear function $x'(A, b)$. By Taylor's theorem, the forward error Δx on the solution $x(A, b)$ can be expressed as

$$\Delta x = x'(A, b) \cdot (\Delta A, \Delta b) + \mathcal{O}(\|(\Delta A, \Delta b)\|_F^2). \quad (4)$$

Following the definition given in Equation (2), the condition number of x corresponds to the operator norm of $x'(A, b)$, which is a bound to first order on the sensitivity of x at (A, b) and we have

$$\|\Delta x\|_2 \leq \kappa_{LS} \|(\Delta A, \Delta b)\|_F.$$

We now use [20] to estimate $\|\Delta x\|_2$ by

$$\xi(q) = \frac{\omega_q}{\omega_n} \sqrt{|z_1^T \Delta x|^2 + \dots + |z_q^T \Delta x|^2}, \quad (5)$$

where z_1, \dots, z_q are random orthogonal vectors selected uniformly and randomly from the unit sphere in n dimensions, and ω_q is the Wallis factor defined by

$$\begin{aligned} \omega_1 &= 1, \\ \omega_q &= \frac{1 \cdot 3 \cdot 5 \cdots (q-2)}{2 \cdot 4 \cdot 6 \cdots (q-1)} \text{ for } q \text{ odd,} \\ \omega_q &= \frac{2}{\pi} \frac{2 \cdot 4 \cdot 6 \cdots (q-2)}{1 \cdot 3 \cdot 5 \cdots (q-1)} \text{ for } q \text{ even.} \end{aligned}$$

ω_q can be approximated by $\sqrt{\frac{2}{\pi(q-\frac{1}{2})}}$.

It comes from [20] that if for instance we have $q = 2$, then the probability that $\xi(q)$ lies within a factor α of $\|\Delta x\|_2$ is

$$Pr\left(\frac{\|\Delta x\|_2}{\alpha} \leq \xi(q) \leq \alpha \|\Delta x\|_2\right) \approx 1 - \frac{\pi}{4\alpha^2}. \quad (6)$$

For $\alpha = 10$, we obtain a probability of 99.2%.

For each $i \in \{1, \dots, q\}$, using Equation (2) we have the first-order bound

$$|z_i^T \Delta x| \leq \kappa_i \|(\Delta A, \Delta b)\|_F, \quad (7)$$

where κ_i denotes the condition number of the function $z_i^T x(A, b)$. Then using (5) and (7) we get

$$\xi(q) \leq \frac{\omega_q}{\omega_n} \left(\sum_{i=1}^q \kappa_i^2 \right)^{\frac{1}{2}} \|(\Delta A, \Delta b)\|_F.$$

$\xi(q)$ being an estimate of $\|\Delta x\|_2$, we will use the quantity $\bar{\kappa}_{LS}$ defined by

$$\bar{\kappa}_{LS} = \frac{\omega_q}{\omega_n} \left(\sum_{i=1}^q \kappa_i^2 \right)^{\frac{1}{2}} \quad (8)$$

as an estimate for κ_{LS} .

We point out that $\bar{\kappa}_{LS}$ is a scalar quantity that must be distinguished from the estimate given in [21] which is a vector. Indeed the small-sample theory is used here to derive an estimate of the condition number of x whereas it is used in [21] to derive estimates of the condition numbers of the components of x (see Section 2.2). Now we can derive Algorithm 2.1 that computes $\bar{\kappa}_{LS}$ as expressed in Equation (8) and using the condition numbers of $z_i^T x$. The vectors z_1, \dots, z_q are obtained for instance via a QR factorization of a random matrix $Z \in \mathbb{R}^{n \times q}$. The condition number of $z_i^T x$ can be computed using the expression given in [3]) as

$$\kappa_i = (\|R^{-1}R^{-T}z_i\|_2^2\|r\|_2^2 + \|R^{-T}z_i\|_2^2(\|x\|_2^2 + 1))^{\frac{1}{2}}. \quad (9)$$

The accuracy of the estimate can be tweaked by modifying the number q of considered random samples. The computation of $\bar{\kappa}_{LS}$ requires computing the QR factorization of an $n \times q$ matrix for $\mathcal{O}(nq^2)$ flops. It also involves solving q times two $n \times n$ triangular linear systems, each triangular system being solved in $\mathcal{O}(n^2)$ flops. The resulting computational cost is $\mathcal{O}(2qn^2)$ flops (if $n \gg q$).

Algorithm 2.1 Statistical condition estimation for linear least squares solution

Require: $q \geq 1$, the number of samples

Generate q vectors $z_1, z_2, \dots, z_q \in \mathbb{R}^n$ with entries in $\mathcal{U}(0, 1)$

Orthonormalize the vectors z_i using a QR factorization

for $j = 1$ to q **do**

Compute $\kappa_j = (\|R^{-1}R^{-T}z_j\|_2^2\|r\|_2^2 + \|R^{-T}z_j\|_2^2(\|x\|_2^2 + 1))^{\frac{1}{2}}$

end for

Compute $\bar{\kappa}_{LS} = \frac{\omega_q}{\omega_n} \sqrt{\sum_{j=1}^q \kappa_j^2}$ with $\omega_q = \sqrt{\frac{2}{\pi(q-\frac{1}{2})}}$

2.2 Componentwise condition estimates

In this section, we focus on calculating the condition number for each component of the LLS solution x . The first one is based on the results from [3] and enables us to compute the exact value of the condition numbers for the i th component of x . The other is a statistical estimate from [21].

Exact formula By considering in Equation (9) the special case where $z_i = e_i$, we can express in Equation (10) the condition number of the component $x_i = e_i^T x$ and then calculate a vector $\kappa_{CW} \in \mathbb{R}^n$ with components κ_i being the exact condition number for the i th component expressed by

$$\kappa_i = (\|R^{-1}R^{-T}e_i\|_2^2\|r\|_2^2 + \|R^{-T}e_i\|_2^2(\|x\|_2^2 + 1))^{\frac{1}{2}}. \quad (10)$$

The computation of one κ_i requires two triangular solves ($R^T y = e_i$ and $Rz = y$) corresponding to $2n^2$ flops. When we want to compute all κ_i , it is more efficient to solve $RY = I$ and then compute YY^T , which requires about $2n^3/3$ flops.

Statistical condition estimate We can find in [21] three different algorithms to compute statistical componentwise condition estimation for LLS problems. Algorithm 2.2 corresponds to the algorithm that uses unstructured perturbations and it can be compared with the exact value given in Equation (10). Algorithm 2.2 computes a vector $\bar{\kappa}_{CW} = (\bar{\kappa}_1, \dots, \bar{\kappa}_n)^T$ containing the statistical estimate of each κ_i . Depending on the needed accuracy for the statistical estimation, the number of random perturbations $q \geq 1$ applied to the input data in Algorithm 2.2 can be adjusted. This algorithm involves two $n \times n$ triangular solves with q right-hand sides, which requires about qn^2 flops.

Algorithm 2.2 Componentwise statistical condition estimate for linear least squares

Require: $q \geq 1$, the number of perturbations of input data

for $j = 1$ to q **do**

 Generate $S_j \in \mathbb{R}^{n \times n}$, $g_j \in \mathbb{R}^n$ and $h_j \in \mathbb{R}^n$ with entries in $\mathcal{N}(0, 1)$

 Compute $u_j = R^{-1}(g_j - S_j x + \|Ax - b\|_2 R^{-T} h_j)$

end for

Let $p = m(n + 1)$ and compute vector $\bar{\kappa}_{CW} = \frac{\sum_{j=1}^q |u_j|}{q\omega_p\sqrt{p}}$ with $\omega_q = \sqrt{\frac{2}{\pi(q-\frac{1}{2})}}$

3 Numerical experiments

In the following experiments, random LLS problems are generated using the method given in [22] for generating LLS test problems with known solution x and residual norm. Random problems are obtained using the quantities m , n , ρ , l such that $A \in \mathbb{R}^{m \times n}$, $\|r\|_2 = \rho$ and $\text{cond}(A) = n^l$. The matrix A is generated using

$$A = Y \begin{pmatrix} D \\ 0 \end{pmatrix} Z^T, \quad Y = I - 2yy^T, \quad Z = I - 2zz^T$$

where $y \in \mathbb{R}^m$ and $z \in \mathbb{R}^n$ are random unit vectors and $D = n^{-l} \text{diag}(n^l, (n-1)^l, (n-2)^l, \dots, 1)$. We have $x = (1, 2^2, \dots, n^2)^T$, the residual vector is given by $r = Y \begin{pmatrix} 0 \\ v \end{pmatrix}$ where $v \in \mathbb{R}^{m-n}$ is a random vector of norm ρ and the right-hand side is given by $b = Y \begin{pmatrix} DZx \\ v \end{pmatrix}$. In Section 3.1, we will consider LLS problems of size $m \times n$ with $m = 9984$ and $n = 2496$. All the experiments were performed using the library LAPACK 3.2 [1] from Netlib.

3.1 Accuracy of statistical estimates

Conditioning of LLS solution In this section we compare the statistical estimate $\bar{\kappa}_{LS}$ obtained via Algorithm 2.1 with the exact condition number κ_{LS}

computed using Equation (3). In our experiments, the statistical estimate is computed using two samples ($q = 2$). For seven different values for $\text{cond}(A) = n^l$ (l ranging from 0 to 3, $n = 2496$) and several values of $\|r\|_2$, we report in Table 1 the ratio $\bar{\kappa}_{LS}/\kappa_{LS}$, which is the average of the ratios obtained for 100 random problems.

Table 1: Ratio between statistical and exact condition numbers ($q = 2$)

$\text{cond}(A)$	n^0	$n^{\frac{1}{2}}$	n^1	$n^{\frac{3}{2}}$	n^2	$n^{\frac{5}{2}}$	n^3
$\ r\ _2 = 10^{-10}$	57.68	3.32	1.46	1.19	1.10	1.03	1.07
$\ r\ _2 = 10^{-5}$	57.68	3.33	1.45	1.18	1.07	1.09	1.05
$\ r\ _2 = 1$	57.68	3.36	1.45	1.19	1.19	1.05	1.15
$\ r\ _2 = 10^5$	57.68	3.33	1.24	1.04	1.05	1.05	1.02
$\ r\ _2 = 10^{10}$	57.68	1.44	1.07	1.09	1.00	1.01	1.07

The results in Table 1 show the relevance of the statistical estimate presented in Section 2.1. For $l \geq \frac{1}{2}$ the averaged estimated values never differ from the exact value by more than one order of magnitude. We observe that when l tends to 0 (i.e., $\text{cond}(A)$ gets close to 1) the estimate becomes less accurate. This can be explained by the fact that the statistical estimate $\bar{\kappa}_{LS}$ is based on evaluating the Frobenius norm of the Jacobian matrix [17]. Actually some additional experiments showed that $\bar{\kappa}_{LS}/\kappa_{LS}$ evolves exactly like $\|R^{-1}\|_F^2 / \|R^{-1}\|_2^2$. In this particular LLS problem we have

$$\begin{aligned} \|R^{-1}\|_F^2 / \|R^{-1}\|_2^2 &= (1 + (n/(n-1))^{2l} + (n/(n-2))^{2l} + \dots + n^{2l}) / n^{2l} \\ &= \sum_{k=1}^n \frac{1}{k^{2l}}. \end{aligned}$$

Then when l tends towards 0, $\|R^{-1}\|_F / \|R^{-1}\|_2 \sim \sqrt{n}$, whereas this ratio gets closer to 1 when l increases. This is consistent with the well-known inequality $1 \leq \|R^{-1}\|_F / \|R^{-1}\|_2 \leq \sqrt{n}$. Note that the accuracy of the statistical estimate does not vary with the residual norm.

Componentwise condition estimation Figure 1 depicts the conditioning for all LLS solution components, computed as $\kappa_i/|x_i|$ where κ_i is obtained using Equation (10). Figures 1(a) and 1(b) correspond to random LLS problems with respectively $\text{cond}(A) = 2.5 \cdot 10^3$ and $\text{cond}(A) = 2.5 \cdot 10^9$. These figures show the interest of the componentwise approach since the sensitivity to perturbations of each solution component varies significantly (from 10^2 to 10^8 for $\text{cond}(A) = 2.5 \cdot 10^3$, and from 10^7 to 10^{16} for $\text{cond}(A) = 2.5 \cdot 10^9$). The normalized condition number of the solution computed using Equation (3) is $\kappa_{LS}/\|x\|_2 = 2.5 \cdot 10^3$ for

$\text{cond}(A) = 2.5 \cdot 10^3$ and $\kappa_{LS}/\|x\|_2 = 4.5 \cdot 10^{10}$ for $\text{cond}(A) = 2.5 \cdot 10^9$, which in both cases greatly overestimates or underestimates the conditioning of some components. Note that the LLS sensitivity is here well measured by $\text{cond}(A)$ since $\|r\|_2$ is small compared to $\|A\|_2$ and $\|x\|_2$, as expected from [25] (otherwise it would be measured by $\text{cond}(A)^2$).

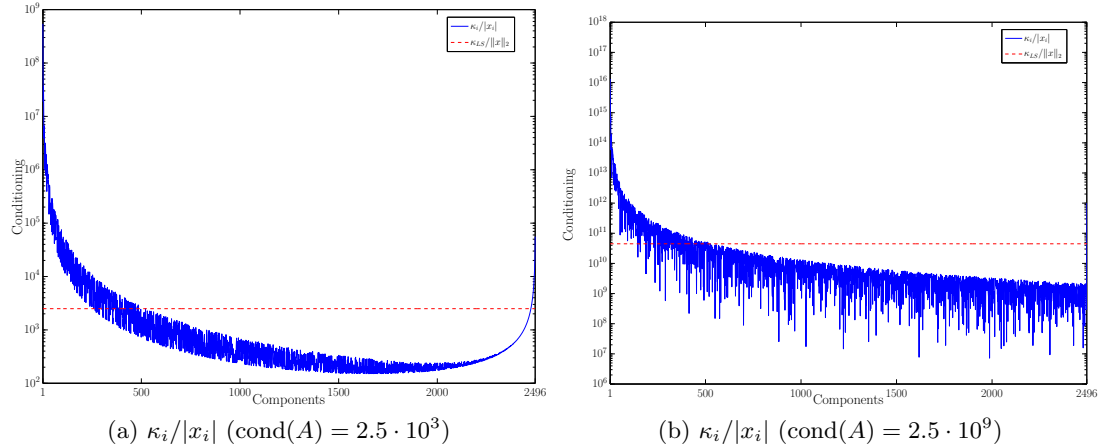


Fig. 1: Componentwise condition numbers of LLS (problem size 9984×2496)

In Figure 2 we represent for each solution component, the ratio between the statistical condition estimate computed via Algorithm 2.2, considering two samples ($q = 2$), and the exact value computed using Equation (10). The ratio is computed as an average on 100 random problems. We observe that this ratio is lower than 1.2 for the case $\text{cond}(A) = 2.5 \cdot 10^3$ (Figure 2 (a)) and close to 1 for the case $\text{cond}(A) = 2.5 \cdot 10^9$ (Figure 2 (b)), which also confirms that, similarly to $\bar{\kappa}_{LS}$ in Section 3.1, the statistical condition estimate is more accurate for larger values of $\text{cond}(A)$.

4 Conclusion

We illustrated how condition numbers of a full column rank LLS problem can be easily computed using exact formulas or statistical estimates at an affordable flop count. Numerical experiments on random LLS problems showed that the statistical estimates provide good accuracy by using only 2 random orthogonal vectors. Subsequently to this work, new routines will be proposed in the public domain libraries LAPACK and MAGMA [24] to compute exact values and statistical estimates for LLS conditioning.

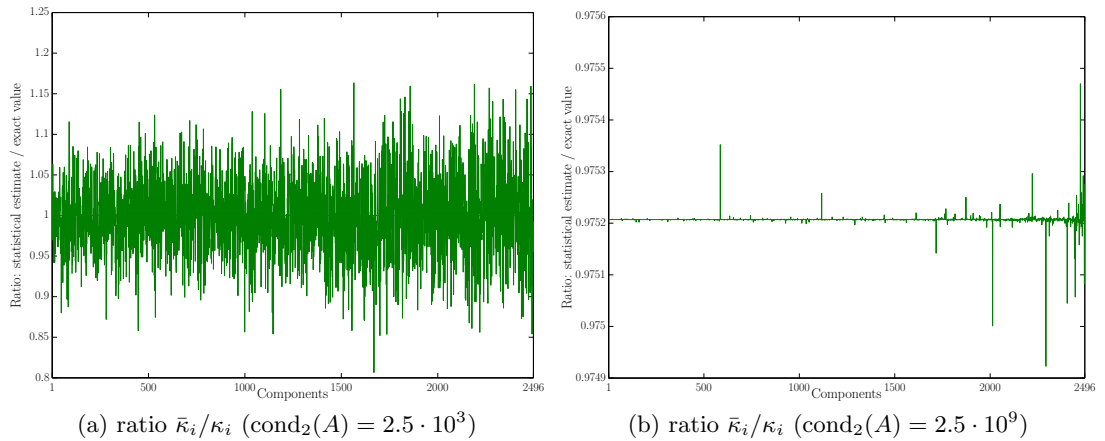


Fig. 2: Comparison between componentwise exact and statistical condition numbers

References

1. E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide*. SIAM, Philadelphia, 1999. Third edition.
2. M. Arioli, M. Baboulin, and S. Gratton. *A partial condition number for linear least squares problems*. SIAM J. Matrix Anal. and Appl., 29(2):413–433, 2007.
3. M. Baboulin, J. Dongarra, S. Gratton, and J. Langou. *Computing the conditioning of the components of a linear least squares solution*. Numerical Linear Algebra with Applications, 16(7):517–533, 2009.
4. M. Baboulin, L. Giraud, S. Gratton, and J. Langou. *Parallel tools for solving incremental dense least squares problems. Application to space geodesy*. Journal of Algorithms and Computational Technology, 3(1):177–133, 2009.
5. M. Baboulin and S. Gratton. *Using dual techniques to derive componentwise and mixed condition numbers for a linear functional of a linear least squares solution*. BIT, 49(1):3–19, 2009.
6. M. Baboulin and S. Gratton. *A contribution to the conditioning of the total least squares problem*. SIAM J. Matrix Anal. and Appl., 32(3):685–699, 2011.
7. Å. Björck. *Numerical methods for least squares problems*. SIAM, Philadelphia, 1996.
8. Y. Cao and L. Petzold. *A subspace error estimate for linear systems*. SIAM J. Matrix Anal. and Appl., 24:787–801, 2003.
9. S. Chandrasekaran and I. C. F Ipsen. *On the sensitivity of solution components in linear systems of equations*. Numerical Linear Algebra with Applications, 2:271–286, 1995.
10. F. Cucker, H. Diao, and Y. Wei. *On mixed and componentwise condition numbers for Moore-Penrose inverse and linear least squares problems*. Mathematics of Computation, 76(258):947–963, 2007.
11. L. Eldén. *Perturbation theory for the least squares problem with linear equality constraints*. SIAM J. Numerical Analysis, 17:338–350, 1980.

12. A. J. Geurts. *A contribution to the theory of condition*. Numerische Mathematik, 39:85–96, 1982.
13. G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 1996. Third edition.
14. G. H. Golub and J. H. Wilkinson. *Note on the iterative refinement of least squares solution*. Numerische Mathematik, 9(2):139–148, 1966.
15. S. Gratton. *On the condition number of linear least squares problems in a weighted Frobenius norm*. BIT, 36:523–530, 1996.
16. J. F. Grcar. *Adjoint formulas for condition numbers applied to linear and indefinite least squares*. Lawrence Berkeley National Laboratory Technical Report, LBNL-55221, 2004.
17. T. Gudmundsson, C. S. Kenney, and A. J. Laub. *Small-sample statistical estimates for matrix norms*. SIAM J. Matrix Anal. and Appl., 16(3):776–792, 1995.
18. N. J. Higham. *A survey of componentwise perturbation theory in numerical linear algebra*. Mathematics of Computation 1943–1993: A Half Century of Computational Mathematics (W. Gautschi, ed.), Proceedings of Symposia in Applied Mathematics, American Mathematical Society, Providence, 48:49–77, 1994.
19. N. J. Higham. *Accuracy and Stability of Numerical Algorithms*. SIAM, Philadelphia, 2002. Second edition.
20. C. S. Kenney and A. J. Laub. *Small-sample statistical condition estimates for general matrix functions*. SIAM J. Sci. Comput., 15(1):36–61, 1994.
21. C. S. Kenney, A. J. Laub, and M. S. Reese. *Statistical condition estimation for linear least squares*. SIAM J. Matrix Anal. and Appl., 19(4):906–923, 1998.
22. C. C. Paige and M. A. Saunders. *LSQR: An algorithm for sparse linear equations and sparse least squares*. ACM Trans. Math. Softw., 8(1):43–71, 1982.
23. J. Rice. *A theory of condition*. SIAM J. Numerical Analysis, 3:287–310, 1966.
24. S. Tomov, J. Dongarra, and M. Baboulin. *Towards dense linear algebra for hybrid GPU accelerated manycore systems*. Parallel Computing, 36(5&6):232–240, 2010.
25. P.-Å. Wedin. *Perturbation theory for pseudo-inverses*. BIT, 13:217–232, 1973.
26. J. H. Wilkinson. *Rounding errors in algebraic processes*. Her Majesty's Stationery Office, London, vol. 32, 1963.