NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. (2008)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/nla.627

Computing the conditioning of the components of a linear least-squares solution

Marc Baboulin[1,2,*,†], Jack Dongarra[2,3,4], Serge Gratton[5,6] and Julien Langou[7]

[1] *Department of Mathematics, University of Coimbra, Coimbra, Portugal*
[2] *Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN, U.S.A.*
[3] *Oak Ridge National Laboratory, Oak Ridge, U.S.A.*
[4] *University of Manchester, Manchester, U.K.*
[5] *Centre National d'Etudes Spatiales, Toulouse, France*
[6] *CERFACS, Toulouse, France*
[7] *Department of Mathematical and Statistical Sciences, University of Colorado at Denver, Denver, CO, U.S.A.*

SUMMARY

In this paper, we address the accuracy of the results for the overdetermined full rank linear least-squares problem. We recall theoretical results obtained in (*SIAM J. Matrix Anal. Appl.* 2007; **29**(2):413–433) on conditioning of the least-squares solution and the components of the solution when the matrix perturbations are measured in Frobenius or spectral norms. Then we define computable estimates for these condition numbers and we interpret them in terms of statistical quantities when the regression matrix and the right-hand side are perturbed. In particular, we show that in the classical linear statistical model, the ratio of the variance of one component of the solution by the variance of the right-hand side is exactly the condition number of this solution component when only perturbations on the right-hand side are considered. We explain how to compute the variance–covariance matrix and the least-squares conditioning using the libraries LAPACK (*LAPACK Users' Guide* (3rd edn). SIAM: Philadelphia, 1999) and ScaLAPACK (*ScaLAPACK Users' Guide*. SIAM: Philadelphia, 1997) and we give the corresponding computational cost. Finally we present a small historical numerical example that was used by Laplace (*Théorie Analytique des Probabilités*. Mme Ve Courcier, 1820; 497–530) for computing the mass of Jupiter and a physical application if the area of space geodesy. Copyright © 2008 John Wiley & Sons, Ltd.

Received 3 December 2007; Revised 13 October 2008; Accepted 13 October 2008

KEY WORDS: linear least squares; statistical linear least squares; parameter estimation; condition number; variance–covariance matrix; LAPACK; ScaLAPACK

*Correspondence to: Marc Baboulin, Department of Mathematics, University of Coimbra, Coimbra, Portugal.
†E-mail: baboulin@mat.uc.pt

1. INTRODUCTION

We consider the linear least-squares problem (LLSP) $\min_{x\in\mathbb{R}^n}\|Ax-b\|_2$, where $b\in\mathbb{R}^m$ and $A\in\mathbb{R}^{m\times n}$ is a matrix of full column rank n.

Our concern comes from the following observation: in many parameter estimation problems, there may be random errors in the observation vector b due to instrumental measurements as well as roundoff errors in the algorithms. The matrix A may be subject to errors in its computation (approximation and/or roundoff errors). In such cases, while the condition number of the matrix A provides some information about the sensitivity of the LLSP to perturbations, a single global conditioning quantity is often not relevant enough since we may have significant disparity between the errors in the solution components. We refer to the last section of the manuscript for illustrative examples. There are several results for analyzing the accuracy of the LLSP by components. For linear systems $Ax=b$ and for LLSP, Chandrasekaran and Ipsen [1] defines the so-called componentwise condition numbers that correspond to amplification factors of the relative errors in solution components due to perturbations in data A or b and explains how to estimate them. For LLSP, Kenney *et al.* [2] proposes to estimate componentwise condition numbers by a statistical method. More recently, Arioli *et al.* [3] developed theoretical results on conditioning of linear functionals of LLSP solutions.

The objective of our paper is to provide computable quantities for the theoretical values given in [3] in order to assess the accuracy of an LLSP solution or some of its components. To achieve this goal, traditional tools for the numerical linear algebra practitioner are condition numbers or backward errors whereas the statistician usually refers to variance or covariance. Our purpose here is to show that these mathematical quantities coming either from numerical analysis or statistics are closely related. We propose formulas for the condition number of the LLSP solution or its components in connection with the linear statistical model when, in addition to the random perturbations of b (that follows a statistical distribution), we can also have non-random perturbations in the matrix A, which can be rounding errors or model errors (e.g. errors coming from linearization, when the initial problem is nonlinear, or simplication in the physics, etc.).

In particular, we will show in Equation (10) that, in the classical linear statistical model, the ratio of the variance of one component of the solution by the variance of the right-hand side is exactly the condition number of this component when perturbations on the right-hand side only are considered. In that sense, we attempt to clarify, similar to [4], the analogy between quantities handled by the linear algebra and the statistical approaches in linear least-squares. Then we define computable estimates for these quantities and explain how they can be computed using the standard libraries LAPACK [5] or ScaLAPACK [6]. As far as we are aware, there is no freeware functionality in Fortran or C that implements the condition number of an LLSP solution or its components and there is no routine in LAPACK or ScaLAPACK for covariance computation. This paper provides us with code fragments, similar to [5, 6]. The resulting LAPACK routine will be in a next release of LAPACK and, since these codes use kernel routines that are also available in ScaLAPACK, they enable us to address also very large computations on parallel computers.

This paper is organized as follows. In Section 2, we recall and exploit some results of practical interest coming from [3]. We also define the condition numbers of an LLSP solution or one component of it. In Section 3, we recall some definitions and results related to the linear statistical model for LLSP, and we interpret the condition numbers in terms of statistical quantities. In Section 4 we provide ways for computing the variance–covariance matrix and LLSP condition numbers using LAPACK (the corresponding ScaLAPACK routines can be used for larger computations).

DOI: 10.1002/nla

In Section 5, we propose two numerical examples that show the relevance of the proposed quantities and their practical computation. The first test case is a historical example from Laplace and the second example is related to gravity field computations. Finally, some concluding remarks are given in Section 6.

Throughout this paper we will use the following notations. We use the Frobenius norm $\|.\|_F$ and the spectral norm $\|.\|_2$ on matrices and the usual Euclidean norm $\|.\|_2$ on vectors. $A^\dagger$ denotes the Moore–Penrose pseudoinverse of A, r denotes the residual vector $b-Ax$, the matrix I is the identity matrix and e_i is the ith canonical vector of $\mathbb{R}^n$.

2. THEORETICAL BACKGROUND FOR LINEAR LEAST-SQUARES CONDITIONING

Following the notations in [3], we consider the function

$$\begin{aligned} g:\mathbb{R}^{m\times n}\times\mathbb{R}^m &\longrightarrow \mathbb{R}^k \\ A,b &\longmapsto g(A,b)=L^{\mathrm{T}}x(A,b)=L^{\mathrm{T}}(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b \end{aligned} \tag{1}$$

where L is an $n\times k$ matrix, with $k\leqslant n$. Since A has full rank n, g is continuously F-differentiable in a neighbourhood of (A,b).

Let α and β be two positive real numbers. In the present paper, we consider the Euclidean norm for the solution space $\mathbb{R}^k$. For the data space $\mathbb{R}^{m\times n}\times\mathbb{R}^m$, we use the product norms defined by

$$\|(\Delta A,\Delta b)\|_{[F,2]}=\sqrt{\alpha^2\|\Delta A\|_{[F,2]}^2+\beta^2\|\Delta b\|_2^2},\qquad \alpha,\beta>0$$

Following [7], the absolute condition number of g at the point (A,b) using the product norm defined above is given by

$$\kappa_{g,[F,2]}(A,b)=\max_{(\Delta A,\Delta b)}\frac{\|g'(A,b)\cdot(\Delta A,\Delta b)\|_2}{\|(\Delta A,\Delta b)\|_{[F,2]}}$$

where g' denotes the Fréchet derivative of g, i.e. $g'(A,b)$ is the linear operator mapping $\mathbb{R}^{m\times n}\times\mathbb{R}^m$ to $\mathbb{R}^k$ such that

$$\lim_{(\Delta A,\Delta b)\to 0}\frac{\|g(A+\Delta A,b+\Delta b)-g(A,b)-g'(A,b).(\Delta A,\Delta b)\|_2}{\|(\Delta A,\Delta b)\|_{[F,2]}}=0$$

The corresponding relative condition number of g at (A,b) is expressed by

$$\kappa_{g,[F,2]}^{(\mathrm{rel})}(A,b)=\frac{\kappa_{g,F}(A,b)\ \|(A,b)\|_{[F,2]}}{\|g(A,b)\|_2}$$

To address the special cases where only A (resp. b) is perturbed, we also define the quantities

$$\kappa_{g,[F,2]}(A)=\max_{\Delta A}\frac{\left\|\dfrac{\partial g}{\partial A}(A,b)\cdot\Delta A\right\|_2}{\|\Delta A\|_{[F,2]}}$$

DOI: 10.1002/nla

respectively

$$\kappa_{g,2}(b) = \max_{\Delta b} \frac{\left\| \frac{\partial g}{\partial b}(A,b)\cdot \Delta b \right\|_2}{\|\Delta b\|_2}$$

A classical choice for α and β corresponds to the case where perturbations on the data ΔA and Δb are measured relatively to the original data A and b, i.e. $\alpha = 1/\|A\|_{[\mathrm{F},2]}$ and $\beta = 1/\|b\|_2$.

Remark 1

The product norm for the data space is very flexible; the coefficients α and β allow us to monitor the perturbations on A and b. For instance, large values of α (resp. β) enable us to obtain condition number problems where mainly b (resp. A) are perturbed. In particular, we will address the special cases where only b (resp. A) is perturbed by choosing the α and β parameters as $\alpha = +\infty$ and $\beta = 1$ (resp. $\alpha = 1$ and $\beta = +\infty$) since we have

$$\lim_{\alpha\to+\infty} \kappa_{g,[\mathrm{F},2]}(A,b) = \frac{1}{\beta}\kappa_{g,[\mathrm{F},2]}(b) \quad \text{and} \quad \lim_{\beta\to+\infty} \kappa_{g,[\mathrm{F},2]}(A,b) = \frac{1}{\alpha}\kappa_{g,[\mathrm{F},2]}(A)$$

This can be justified as follows:

$$\begin{aligned}\kappa_{g,[\mathrm{F},2]}(A,b) &= \max_{(\Delta A,\Delta b)} \frac{\left\| \frac{\partial g}{\partial A}(A,b)\cdot \Delta A + \frac{\partial g}{\partial b}(A,b)\cdot \Delta b \right\|_2}{\sqrt{\alpha^2 \|\Delta A\|^2_{[\mathrm{F},2]} + \beta^2 \|\Delta b\|_2^2}} \\ &= \max_{(\Delta A,\Delta b)} \frac{\left\| \frac{\partial g}{\partial A}(A,b)\cdot \frac{\Delta A}{\alpha} + \frac{\partial g}{\partial b}(A,b)\cdot \frac{\Delta b}{\beta} \right\|_2}{\sqrt{\|\Delta A\|^2_{[\mathrm{F},2]} + \|\Delta b\|_2^2}}\end{aligned}$$

The above expression represents the operator norm of a linear functional depending continuously on α, and then we get

$$\lim_{\alpha\to+\infty} \kappa_{g,[\mathrm{F},2]}(A,b) = \max_{(\Delta A,\Delta b)} \frac{\left\| \frac{\partial g}{\partial b}(A,b)\cdot \frac{\Delta b}{\beta} \right\|_2}{\sqrt{\|\Delta A\|^2_{[\mathrm{F},2]} + \|\Delta b\|_2^2}} = \max_{\Delta b} \frac{\left\| \frac{\partial g}{\partial b}(A,b)\cdot \frac{\Delta b}{\beta} \right\|_2}{\|\Delta b\|_2} = \frac{1}{\beta}\kappa_{g,[\mathrm{F},2]}(b)$$

The proof is the same for the case where $\beta = +\infty$.

The condition numbers related to $L^{\mathrm{T}} x(A,b)$ are referred to in [3] as partial partial condition numbers (PCN) of the LLSP with respect to the linear operator L.

In this paper, we are interested in computing the PCN for two special cases. The first case is when L is the identity matrix (conditioning of the solution) and the second case is when L is a canonical vector e_i (conditioning of a solution component). We can extract from [3] two theorems that lead to computable quantities in these two special cases.

DOI: 10.1002/nla

Theorem 1
In the general case where ($L \in \mathbb{R}^{n\times k}$), the absolute condition numbers of $g(A,b) = L^{\mathrm{T}}x(A,b)$ in the Frobenius and spectral norms can be, respectively, bounded as follows:

$$\frac{1}{\sqrt{3}}f(A,b) \leqslant \kappa_{g,\mathrm{F}}(A,b) \leqslant f(A,b)$$
$$\frac{1}{\sqrt{3}}f(A,b) \leqslant \kappa_{g,2}(A,b) \leqslant \sqrt{2}f(A,b)$$

where

$$f(A,b) = \left(\|L^{\mathrm{T}}(A^{\mathrm{T}}A)^{-1}\|_2^2 \frac{\|r\|_2^2}{\alpha^2} + \|L^{\mathrm{T}}A^{\dagger}\|_2^2 \left(\frac{\|x\|_2^2}{\alpha^2} + \frac{1}{\beta^2} \right) \right)^{1/2} \tag{2}$$

Theorem 2
In the two particular cases:

1. L is a vector ($L \in \mathbb{R}^n$), or
2. L is the n-by-n identity matrix ($L = I$)

the absolute condition number of $g(A,b) = L^{\mathrm{T}}x(A,b)$ in the Frobenius norm is given by the formula:

$$\kappa_{g,\mathrm{F}}(A,b) = \left(\|L^{\mathrm{T}}(A^{\mathrm{T}}A)^{-1}\|_2^2 \frac{\|r\|_2^2}{\alpha^2} + \|L^{\mathrm{T}}A^{\dagger}\|_2^2 \left(\frac{\|x\|_2^2}{\alpha^2} + \frac{1}{\beta^2} \right) \right)^{1/2}$$

Theorem 2 provides the exact value for the condition number in the Frobenius norm for our two cases of interest ($L = e_i$ and $L = I$). From Theorem 1, we observe that

$$\frac{1}{\sqrt{3}}\kappa_{g,\mathrm{F}}(A,b) \leqslant \kappa_{g,2}(A,b) \leqslant \sqrt{6}\kappa_{g,\mathrm{F}}(A,b) \tag{3}$$

which states that the partial condition number in spectral norm is of the same order of magnitude as the one in Frobenius norm. In the remainder of the paper, the focus is given to the partial condition number in Frobenius norm only.

For the case $L = I$, the result of Theorem 2 is similar to [8] and [7, p. 92]. The upper bound for $\kappa_{2,\mathrm{F}}(A,b)$ that can be derived from Equation (3) is also the one obtained by Geurts [7] when we consider perturbations in A.

Let us denote by $\kappa_i(A,b)$ the condition number related to the component x_i in Frobenius norm (i.e. $\kappa_i(A,b) = \kappa_{g,\mathrm{F}}(A,b)$ where $g(A,b) = e_i^{\mathrm{T}}x(A,b) = x_i(A,b)$). Then replacing L by e_i in Theorem 2 provides us with an exact expression for computing $\kappa_i(A,b)$, this gives

$$\kappa_i(A,b) = \left(\|e_i^{\mathrm{T}}(A^{\mathrm{T}}A)^{-1}\|_2^2 \frac{\|r\|_2^2}{\alpha^2} + \|e_i^{\mathrm{T}}A^{\dagger}\|_2^2 \left(\frac{\|x\|_2^2}{\alpha^2} + \frac{1}{\beta^2} \right) \right)^{1/2} \tag{4}$$

$\kappa_i(A,b)$ will be referred to as *the condition number of the solution component x_i*.

DOI: 10.1002/nla

Let us denote by $\kappa_{\mathrm{LS}}(A,b)$ the condition number related to the solution x in Frobenius norm (i.e. $\kappa_{\mathrm{LS}}(A,b)=\kappa_{g,\mathrm{F}}(A,b)$ where $g(A,b)=x(A,b)$). Then Theorem 2 provides us with an exact expression for computing $\kappa_{\mathrm{LS}}(A,b)$, that is

$$\kappa_{\mathrm{LS}}(A,b)=\|(A^{\mathrm{T}}A)^{-1}\|_2^{1/2}\left(\frac{\|(A^{\mathrm{T}}A)^{-1}\|_2\|r\|_2^2+\|x\|_2^2}{\alpha^2}+\frac{1}{\beta^2}\right)^{1/2} \tag{5}$$

where we have used the fact that $\|(A^{\mathrm{T}}A)^{-1}\|_2=\|A^{\dagger}\|_2^2$.

$\kappa_{\mathrm{LS}}(A,b)$ will be referred to as *the condition number of the least squares solution*.

Note that Demmel *et al.* [9] define condition numbers for both x and r in order to derive error bounds for x and r but uses infinity norm to measure perturbations.

In this paper, we will also be interested in the special case where only b is perturbed ($\alpha=+\infty$ and $\beta=1$). In this case, we will call $\kappa_i(b)$ the condition number of the solution component x_i, and $\kappa_{\mathrm{LS}}(b)$ the condition number of the least- squares solution. When we restrict the perturbations to be on b, Equation (4) simplifies to

$$\kappa_i(b)=\|e_i^{\mathrm{T}}A^{\dagger}\|_2 \tag{6}$$

and Equation (5) simplifies to

$$\kappa_{\mathrm{LS}}(b)=\|A^{\dagger}\|_2 \tag{7}$$

This latter formula is standard and is in accordance with [10, p. 29].

3. CONDITION NUMBERS AND STATISTICAL QUANTITIES

3.1. Background for the linear statistical model

We consider here the classical linear statistical model

$$b=Ax+\varepsilon,\quad A\in\mathbb{R}^{m\times n},\ \ b\in\mathbb{R}^m,\ \ \mathrm{rank}(A)=n$$

where ε is a vector of random errors having expected value $E(\varepsilon)=0$ and variance–covariance $V(\varepsilon)=\sigma_b^2 I$. In statistical language, the matrix A is referred to as the regression matrix and the unknown vector x is called the vector of regression coefficients.

Following the Gauss–Markov theorem [11], the least-squares estimates $\hat{x}$ is the linear unbiased estimator of x satisfying

$$\|A\hat{x}-b\|_2=\min_{x\in\mathbb{R}^n}\|Ax-b\|_2$$

with minimum variance–covariance equal to

$$C=\sigma_b^2(A^{\mathrm{T}}A)^{-1} \tag{8}$$

Moreover, $1/(m-n)\|b-A\hat{x}\|_2^2$ is an unbiased estimate of σ_b^2. This quantity is sometimes called the mean squared error (MSE).

The diagonal elements c_{ii} of C give the variance of each component $\hat{x}_i$ of the solution. The off-diagonal elements $c_{ij},\ i\neq j$ give the covariance between $\hat{x}_i$ and $\hat{x}_j$.

DOI: 10.1002/nla

We define $\sigma_{\hat{x}_i}$ as the standard deviation of the solution component $\hat{x}_i$ and we have

$$\sigma_{\hat{x}_i} = \sqrt{c_{ii}} \tag{9}$$

In the next section, we will prove that the condition numbers $\kappa_i(A,b)$ and $\kappa_{\mathrm{LS}}(A,b)$ can be related to the statistical quantities $\sigma_{\hat{x}_i}$ and σ_b.

Remark 2

In the linear statistical model, random errors affect exclusively the observation vector b while A is considered as known exactly (however, it is relevant to consider perturbations of A and b in order to model for instance the effect of roundoff errors in the computation). Measurement errors may also affect A. This case is treated by the statistical model referred to as Errors-In-Variables model (see e.g. [10, p. 176]) [12, p. 230], where we have the relation

$$(A+E)x = b+\varepsilon$$

We assume in this model that the rows of $[E,\varepsilon]$ are independently and identically distributed with common zero mean vector and common covariance matrix.

The corresponding linear algebra problem, discussed originally in [13], is called the total least squares (TLS) problem and can be expressed as:

$$\text{find any } x \text{ solving } \hat{A}x = \hat{b}, \text{where } (\hat{A},\hat{b}) \text{ minimizes } \|[\hat{A},\hat{b}]-[A,b]\|_{\mathrm{F}}$$

As mentioned in [12, p. 238], the TLS method enables us to obtain a more accurate solution when entries of A are perturbed under certain conditions. However, comparing the sensitivity of the TLS and the LLS is complicated due to the fact that they correspond to practical situations that are different due to the nature of perturbations. However, Van Huffel and Vandewalle [12] provide upper bounds on errors and show that the sensitivity is related to the size of the residual compared with the problem parameters.

The TLS method involves an SVD computation and the computational cost is higher than that of a classical LLS (about $2mn^2+12n^3$ as mentioned in [14, p. 598], to be compared with the approximately $2mn^2$ flops required for LLS solved via Householder QR factorization).

3.2. Perturbation on b only

Using Formula (8), the variance c_{ii} of the solution component $\hat{x}_i$ can be expressed as

$$c_{ii} = e_i^{\mathrm{T}} C e_i = \sigma_b^2 e_i^{\mathrm{T}} (A^{\mathrm{T}}A)^{-1} e_i$$

We note that $(A^{\mathrm{T}}A)^{-1} = A^{\dagger}A^{\dagger\mathrm{T}}$ so that

$$c_{ii} = \sigma_b^2 e_i^{\mathrm{T}} (A^{\dagger}A^{\dagger\mathrm{T}}) e_i = \sigma_b^2 \|e_i^{\mathrm{T}} A^{\dagger}\|_2^2$$

Using Equation (9), we get

$$\sigma_{\hat{x}_i} = \sqrt{c_{ii}} = \sigma_b \|e_i^{\mathrm{T}} A^{\dagger}\|_2$$

Finally from Equation (6), we get

$$\sigma_{\hat{x}_i} = \sigma_b \kappa_i(b) \tag{10}$$

DOI: 10.1002/nla

Equation (10) shows that the condition number $\kappa_i(b)$ relates linearly to the standard deviation of σ_b with the standard deviation of $\sigma_{\hat{x}_i}$.

Now if we consider the constant vector ℓ of size n, we have (see [11])

$$\text{variance}(\ell^{\mathrm{T}}\hat{x}) = \ell^{\mathrm{T}} C \ell$$

Since C is symmetric, we can write

$$\max_{\|\ell\|_2=1} \text{variance}(\ell^{\mathrm{T}}\hat{x}) = \|C\|_2$$

Using the fact that $\|C\|_2 = \sigma_b^2 \|(A^{\mathrm{T}}A)^{-1}\|_2 = \sigma_b^2 \|A^\dagger\|_2^2$, and Equation (7), we get

$$\max_{\|\ell\|_2=1} \text{variance}(\ell^{\mathrm{T}}\hat{x}) = \sigma_b^2 \kappa_{\mathrm{LS}}(b)^2$$

or, if we call $\sigma(\ell^{\mathrm{T}}\hat{x})$ the standard deviation of $\ell^{\mathrm{T}}\hat{x}$,

$$\max_{\|\ell\|_2=1} \sigma(\ell^{\mathrm{T}}\hat{x}) = \sigma_b \kappa_{\mathrm{LS}}(b)$$

Note that $\sigma_b = \max_{\|\ell\|_2=1} \sigma(\ell^{\mathrm{T}}\varepsilon)$ since $V(\varepsilon) = \sigma_b^2 I$.

Remark 3

Matlab has a routine LSCOV that computes the quantities $\sqrt{c_{ii}}$ in a vector STDX and the squared error MSE using the syntax [X,STDX,MSE]=LSCOV(A,B).

Then the condition numbers $\kappa_i(b)$ can be computed by the matlab expression STDX/sqrt(MSE).

3.3. *Perturbation on A and b*

We now provide the expression of the condition number given in Equation (4) and in Equation (5) in terms of statistical quantities.

Observing the following relations:

$$C_i = \sigma_b^2 e_i^{\mathrm{T}} (A^{\mathrm{T}}A)^{-1} \quad \text{and} \quad c_{ii} = \sigma_b^2 \|e_i^{\mathrm{T}} A^\dagger\|_2^2$$

where C_i is the ith column of the variance–covariance matrix, the condition number of x_i given in Equation (4) can expressed as

$$\kappa_i(A,b) = \frac{1}{\sigma_b}\left(\frac{\|C_i\|_2^2}{\sigma_b^2}\frac{\|r\|_2^2}{\alpha^2} + c_{ii}\left(\frac{\|x\|_2^2}{\alpha^2} + \frac{1}{\beta^2}\right)\right)^{1/2}$$

The quantity σ_b^2 will often be estimated by $1/(m-n)\,\|r\|_2^2$ in which case the expression can be simplified

$$\kappa_i(A,b) = \frac{1}{\sigma_b}\left(\frac{m-n}{\alpha^2}\|C_i\|_2^2 + c_{ii}\left(\frac{\|x\|_2^2}{\alpha^2} + \frac{1}{\beta^2}\right)\right)^{1/2} \tag{11}$$

In the standard case where perturbations on the data are measured relatively to the original data A and b, i.e. when $\alpha = 1/\|A\|_{\mathrm{F}}$ and $\beta = 1/\|b\|_2$, we get

$$\kappa_i(A,b) = \frac{1}{\sigma_b}((m-n)\,\|C_i\|_2^2\,\|A\|_{\mathrm{F}}^2 + c_{ii}(\|A\|_{\mathrm{F}}^2\,\|x\|_2^2 + \|b\|_2^2))^{1/2}$$

DOI: 10.1002/nla

Table I. Condition number expressions for the full rank LLSP.

Source	Data	Formula	Status
$\kappa_{\mathrm{LS}}(A,b)$ (Equation (13), [3, 8])	$\sqrt{\frac{\|\delta A\|_F^2}{\|A\|_F^2}+\frac{\|\delta b\|_2^2}{\|b\|_2^2}}$	$\frac{\|C\|_2^{1/2}}{\sigma_b}(\|A\|_F^2((m-n)\,\|C\|_2+\|x\|_2^2)+\|b\|_2^2)^{1/2}$	Exact
Geurts [7]	$\frac{\|\delta A\|_F}{\|A\|_F}$	$\frac{\|A\|_F\|C\|_2^{1/2}}{\sigma_b}((m-n)\,\|C\|_2+\|x\|_2^2)^{1/2}$	Exact
Björck [10]	$\frac{\|\delta A\|_2}{\|A\|_2}$	$\frac{\|A\|_2\|C\|_2^{1/2}}{\sigma_b}(\sqrt{m-n}\,\|C\|_2^{1/2}+\|x\|_2)$	Estimate
Grcar [15]	$\max\left\{\frac{\|\delta A\|_{[F,2]}}{\|A\|_{[F,2]}},\frac{\|\delta b\|_2}{\|b\|_2}\right\}$	$\frac{\|C\|_2^{1/2}}{\sigma_b}(\|A\|_{[F,2]}(\sqrt{m-n}\,\|C\|_2^{1/2}+\|x\|_2)+\|b\|_2)$	Estimate

From Equation (5), we obtain

$$\kappa_{\mathrm{LS}}(A,b)=\frac{\|C\|_2^{1/2}}{\sigma_b}\left(\frac{\|C\|_2\|r\|_2^2}{\alpha^2\sigma_b^2}+\frac{\|x\|_2^2}{\alpha^2}+\frac{1}{\beta^2}\right)^{1/2}$$

The quantity σ_b^2 will often be estimated by $1/(m-n)\,\|r\|_2^2$ in which case the expression can be simplified

$$\kappa_{\mathrm{LS}}(A,b)=\frac{\|C\|_2^{1/2}}{\sigma_b}\left(\frac{m-n}{\alpha^2}\|C\|_2+\frac{\|x\|_2^2}{\alpha^2}+\frac{1}{\beta^2}\right)^{1/2} \tag{12}$$

In the case where $\alpha=1/\|A\|_F$ and $\beta=1/\|b\|_2$, we get

$$\kappa_{\mathrm{LS}}(A,b)=\frac{\|C\|_2^{1/2}}{\sigma_b}(\|A\|_F^2((m-n)\,\|C\|_2+\|x\|_2^2)+\|b\|_2^2)^{1/2} \tag{13}$$

Note in Equations (11) and (12) the dependence in, respectively, $\|C_i\|_2$ and $\|C\|_2$ when the variance σ_b is large (i.e. the residual is large).

Note also that the matlab routine LSCOV mentioned in Remark 3 can compute the whole covariance matrix and enable us to obtain the condition numbers expressed in Equations (11) and (12).

It could be also interesting to compare the expression of $\kappa_{\mathrm{LS}}(A,b)$ given by Equation (13) with the existing formulas from the literature. Table I gives the condition number of the LLSP solution that is proposed by other authors. For sake of consistency, we modified these formulas in such a way that the condition number is expressed as a function of the statistical quantities $\|C\|_2$ and σ_b. We observe in Table I that the exact formulas are obtained when using the Frobenius norm on matrices and that, in spectral norm, we have estimates. The inequality (3) shows that Equations (11) and (12) give very sharp estimates of the corresponding condition number when perturbations of A are measured in spectral norm (within a factor $\sqrt{6}$). Note also in Table I that

DOI: 10.1002/nla

the metric chosen for measuring perturbations on data are different and that in two cases, only perturbations on A are considered.

4. COMPUTATION WITH LAPACK

Section 2 provides us with formulas to compute the condition numbers κ_i and κ_{LS}. As explained in Section 3, those quantities are intimately interrelated with the entries of the variance–covariance matrix. The goal of this section is to present practical methods and codes to compute those quantities efficiently with LAPACK and ScaLAPACK. The assumption made is that the LLSP has already been solved with either the normal equations method or a QR factorization approach. Therefore the solution vector $\hat{x}$, the norm of the residual $\|\hat{r}\|_2$, and the R-factor R of the QR factorization of A are readily available (we recall that the Cholesky factor of the normal equations is the R-factor of the QR factorization up to some signs). In the example codes, we have used the LAPACK routine DGELS that solves the LLSP using QR factorization of A. Note that it is possible to have a more accurate solution using extra-precise iterative refinement [9]. We will use the fact that $1/(m-n)\,\|b-A\hat{x}\|_2^2$ is an unbiased estimate of σ_b^2. We wish to compute the following quantities related to the variance–covariance matrix C

- the ith column $C_i = \sigma_b^2 (A^{\mathrm{T}}A)^{-1} e_i$,
- the ith diagonal element $c_{ii} = \sigma_b^2 \|e_i^{\mathrm{T}} A^{\dagger}\|_2^2$,
- the whole matrix C.

We note that the quantities C_i, c_{ii}, and C are of interest for statisticians. The NAG routine F04YAF [16] is indeed an example of tool to compute these three quantities.

For the two first quantities of interest, we note that

$$\|e_i^{\mathrm{T}} A^{\dagger}\|_2^2 = \|R^{-\mathrm{T}} e_i\|_2^2 \quad \text{and} \quad (A^{\mathrm{T}}A)^{-1} e_i = R^{-1}(R^{-T} e_i)$$

4.1. Computation of the ith column C_i

C_i can be computed with two n-by-n triangular solves

$$R^{\mathrm{T}} y = e_i \quad \text{and} \quad Rz = y$$

The ith column of C can be computed by the following code fragment.

Code 1:

```
CALL DGELS( ’N’, M, N, 1, A, LDA, B, LDB, WORK, LWORK, INFO )
RESNORM = DNRM2( (M−N), B(N+1), 1)
SIGMA2 = RESNORM**2/DBLE(M−N)
E(1:N) = 0.D0
E(I) = 1.D0
CALL DTRSV( ’U’, ’T’, ’N’, N−I+1, A(I,I), LDA, E(I), 1)
CALL DTRSV( ’U’, ’N’, ’N’, N, A, LDA, E, 1)
CALL DSCAL( N, SIGMA2, E, 1)
```

This requires about $2n^2$ flops (in addition to the cost of solving the LLSP using DGELS).

c_{ii} can be computed by one n-by-n triangular solve and taking the square of the norm of the solution, which involves about $(n-i+1)^2$ flops. It is important to note that the larger i, the less expensive to obtain c_{ii}. In particular if $i=n$ then only one operation is needed: $c_{nn}=R_{nn}^{-2}$. This suggests that a correct ordering of the variables can save some computation.

4.2. *Computation of the* i*th diagonal element* c_{ii}

From $c_{ii}=\sigma_b^2\|e_i^{\mathrm{T}}R^{-1}\|_2^2$, it comes that each c_{ii} corresponds to the norm of the ith row of R^{-1}. Then the diagonal elements of C can be computed by the following code fragment.

Code 2:

```
CALL DGELS( ’N’, M, N, 1, A, LDA, B, LDB, WORK, LWORK, INFO)
RESNORM = DNRM2((M−N), B(N+1), 1)
SIGMA2 = RESNORM**2/DBLE(M-N)
CALL DTRTRI( ’U’, ’N’, N, A, LDA, INFO)
DO I=1,N
    CDIAG(I) = DNRM2( N−I+1, A(I,I), LDA)
    CDIAG(I) = SIGMA2 * CDIAG(I)**2
END DO
```

This requires about $n^3/3$ flops (plus the cost of DGELS).

4.3. *Computation of the whole matrix* C

In order to compute explicit all the coefficients of the matrix C, one can use the routine DPOTRI, which computes the inverse of a matrix from its Cholesky factorization. First the routine computes the inverse of R using DTRTRI and then performs the triangular matrix–matrix multiply $R^{-1}R^{-\mathrm{T}}$ by DLAUUM. This requires about $2n^3/3$ flops. We can also compute the variance–covariance matrix without inverting R using for instance the algorithm given in [10, p. 119], but the computational cost remains $2n^3/3$ (plus the cost of DGELS).

We can obtain the upper triangular part of C by the following code fragment.

Code 3:

```
CALL DGELS( ’N’, M, N, 1, A, LDA, B, LDB, WORK, LWORK, INFO)
RESNORM = DNRM2((M−N), B(N+1), 1)
SIGMA2 = RESNORM**2/DBLE(M−N)
CALL DPOTRI( ’U’, N, A, LDA, INFO)
CALL DLASCL( ’U’, 0, 0, N, N, 1.D0, SIGMA2, N, N, A, LDA, INFO)
```

4.4. *Condition numbers computation*

We give in Table II the LAPACK routines used for computing the condition numbers of an LLSP solution or its components and the corresponding number of floating-point operations per second. Since the LAPACK routines involved in the covariance and/or LLSP condition numbers have their equivalent in the parallel library ScaLAPACK, then this table is also available when using ScaLAPACK. This enables us to easily compute these quantities for larger LLSP.

Table II. Computation of least-squares conditioning with (Sca)LAPACK.

Condition number	Linear algebra operation	(Sca)LAPACK routines	Flops count
$\kappa_i(A,b)$	$R^{\mathrm{T}}y=e_i$ and $Rz=y$	2 calls to (P)DTRSV	$2n^2$
all $\kappa_i(A,b),\ i=1,n$	$RY=I$ and compute YY^{T}	(P)DPOTRI	$2n^3/3$
all $\kappa_i(b),\ i=1,n$	invert R	(P)DTRTRI	$n^3/3$
$\kappa_{\mathrm{LS}}(A,b)$	estimate $\|R^{-1}\|_{1 \text{ or } \infty}$	(P)DTRCON	$\mathcal{O}(n^2)$
	compute $\|R^{-1}\|_{\mathrm{F}}$	(P)DTRTRI	$n^3/3$

Remark 4

The cost for computing all the $\kappa_i(A,b)$ or estimating $\kappa_{\mathrm{LS}}(A,b)$ is always $\mathcal{O}(n^3)$. When $m \gg n$, this cost is affordable if we compare it to the cost of the least-squares solution using Householder QR factorization ($2mn^2-2n^3/3$) or the normal equations ($mn^2+n^3/3$).

Remark 5

For estimating $\kappa_{\mathrm{LS}}(A,b)$, we need to have an estimate of $\|A^{\dagger}\|_2$ i.e. $\|R^{-1}\|_2$. The computation of $\|R^{-1}\|_2$ requires to compute the minimum singular value of the matrix A (or R). One way is to compute the full SVD of A (or R) which requires $\mathcal{O}(n^3)$ flops. As an alternative, $\|R^{-1}\|_2$ can be approximated using other matrix norms. For instance, $\|R^{-1}\|_1$ or $\|R^{-1}\|_\infty$ can be estimated using Higham modification [17, p. 293] of Hager's [18] method as it is implemented in the LAPACK routine DTRCON. The cost is $\mathcal{O}(n^2)$.

It is also interesting to evaluate $\|R^{-1}\|_2$ by considering $\|R^{-1}\|_{\mathrm{F}}$ since we have $\|R^{-1}\|_{\mathrm{F}}^2 = \|R^{-\mathrm{T}}\|_{\mathrm{F}}^2 = \mathrm{tr}(R^{-1}R^{-T}) = 1/\sigma_b^2\,\mathrm{tr}(C)$, where $\mathrm{tr}(C)$ denotes the trace of the matrix C, i.e. $\sum_{i=1}^{n} c_{ii}$.

When only b is perturbed ($\alpha=+\infty$ and $\beta=1$), then we get $\kappa_{\mathrm{LS}}(b) \simeq \sqrt{\mathrm{tr}(C)}/\sigma_b$.

This result relates to [19, p. 167], where $\mathrm{tr}(C)$ measures the squared effect on the LLSP solution x to small changes in b.

5. NUMERICAL EXPERIMENTS

5.1. *Laplace's computation of the mass of Jupiter and assessment of the validity of its results*

In [20], Laplace computes the mass of Jupiter, Saturn and Uranus and provides the variances associated with those variables in order to assess the quality of the results. The data come from the French astronomer Bouvart in the form of the normal equations given in below.

$$
\begin{aligned}
795\,938 z_0 - 12\,729\,398 z_1 + 6788.2 z_2 - 1959.0 z_3 + 696.13 z_4 + 2602 z_5 &= 7212.600\\
-12\,729\,398 z_0 + 424\,865\,729 z_1 - 153\,106.5 z_2 - 39\,749.1 z_3 - 5459 z_4 + 5722 z_5 &= -738\,297.800\\
6788.2 z_0 - 153\,106.5 z_1 + 71.8720 z_2 - 3.2252 z_3 + 1.2484 z_4 + 1.3371 z_5 &= 237.782\\
-1959.0 z_0 - 39\,749.1 z_1 - 3.2252 z_2 + 57.1911 z_3 + 3.6213 z_4 + 1.1128 z_5 &= -40.335 \qquad (14)
\end{aligned}
$$

$$696.13z_0-5459z_1+1.2484z_2+3.6213z_3+21.543z_4+46.310z_5=-343.455$$

$$2602z_0+5722z_1+1.3371z_2+1.1128z_3+46.310z_4+129z_5=-1002.900$$

For computing the mass of Jupiter, we know that Bouvart performed $m=129$ observations and there are $n=6$ variables in the system. The residual of the solution $\|b-A\hat{x}\|_2^2$ is also given by Bouvart and is 31 096. Of the six unknowns, Laplace only seeks one, the second variable z_1. The mass of Jupiter in term of the mass of the Sun is given by z_1 and the formula:

$$\text{mass of Jupiter}=\frac{1+z_1}{1067.09}$$

It turns out that the first variable z_0 represents the mass of Uranus through the formula

$$\text{mass of Uranus}=\frac{1+z_0}{19\,504}$$

If we solve the system (14), we obtain the solution vector

$$0.08954\quad -0.00304\quad -11.53658\quad -0.51492\quad 5.19460\quad -11.18638$$

From z_1, we can compute the mass of Jupiter as a fraction of the mass of the Sun and we obtain 1070. This value is rather accurate since the correct value according to NASA is 1048. From z_0, we can compute the mass of Uranus as a fraction of the mass of the Sun and we obtain 17 918. This value is inaccurate since the correct value according to NASA is 22 992.

Laplace has computed the variance of z_0 and z_1 to assess the fact that z_1 was probably correct and z_0 probably inaccurate. To compute those variances, Laplace first performed a Cholesky factorization from right to left of the system (14) then, since the variables were correctly ordered, the number of operations involved in the computation of the variances of z_0 and z_1 was minimized. The variance–covariance matrix for Laplace's system is:

$$\begin{pmatrix} 0.005245 & -0.000004 & -0.499200 & 0.137212 & 0.235241 & -0.186069 \\ \cdot & 0.000004 & 0.009873 & 0.003302 & 0.002779 & -0.001235 \\ \cdot & \cdot & 71.466023 & -5.441882 & -16.672689 & 14.922752 \\ \cdot & \cdot & \cdot & 10.860492 & 5.418506 & -4.896579 \\ \cdot & \cdot & \cdot & \cdot & 66.088476 & -28.467391 \\ \cdot & \cdot & \cdot & \cdot & \cdot & 15.874809 \end{pmatrix}$$

Our computation gives us that the variance for the mass of Jupiter is 4.383233×10^{-6}. For reference, Laplace in 1820 computed 4.383209×10^{-6}. (We deduce the variance from Laplace's value 5.0778624. To get what we now call the variance, one needs to compute the quantity: $1/(2*10*5.0778624)*m/(m-n)$.)

From the variance–covariance matrix, one can assess that the computation of the mass of Jupiter (second variable) is extremely reliable while the computation of the mass of Uranus (first variable) is not. For more details, we recommend to read [21].

DOI: 10.1002/nla

5.2. *Gravity field computation*

A classical example of parameter estimation problem is the computation of the Earth's gravity field coefficients. More specifically, we estimate the parameters of the gravitational potential that can be expressed in spherical coordinates (r, θ, λ) by [22]

$$V(r,\theta,\lambda) = \frac{GM}{R}\sum_{\ell=0}^{\ell_{\max}}\left(\frac{R}{r}\right)^{\ell+1}\sum_{m=0}^{\ell}\overline{P}_{\ell m}(\cos\theta)[\overline{C}_{\ell m}\cos m\lambda + \overline{S}_{\ell m}\sin m\lambda] \tag{15}$$

where G is the gravitational constant, M is the Earth's mass, R is the Earth's reference radius, the $\overline{P}_{\ell m}$ represent the fully normalized Legendre functions of degree ℓ and order m and $\overline{C}_{\ell m}$, $\overline{S}_{\ell m}$ are the corresponding normalized harmonic coefficients. The objective here is to compute the harmonic coefficients $\overline{C}_{\ell m}$ and $\overline{S}_{\ell m}$ most accurately as possible. The number of unknown parameters is expressed by $n = (\ell_{\max}+1)^2$. These coefficients are computed by solving a LLSP that may involve millions of observations and tens of thousands of variables. More details about the physical problem and the solution methods can be found in [23]. The data used in the following experiments were provided by CNES[‡] and they correspond to 10 days of observations using GRACE[§] measurements (about 166 000 observations). We compute the spherical harmonic coefficients $\overline{C}_{\ell m}$ and $\overline{S}_{\ell m}$ up to a degree $\ell_{\max} = 50$; except the coefficients $\overline{C}_{11}, \overline{S}_{11}, \overline{C}_{00}, \overline{C}_{10}$ that are *a priori* known. Then we have $n = 2597$ unknowns in the corresponding least-squares problems (note that the GRACE satellite enables us to compute a gravity field model up to degree 150). The problem is solved using the normal equations method and we have the Cholesky decomposition $A^{\mathrm{T}}A = U^{\mathrm{T}}U$.

We compute the relative condition numbers of each coefficient x_i using the formula

$$\kappa_i^{(\mathrm{rel})}(b) = \|e_i^{\mathrm{T}}U^{-1}\|_2\,\|b\|_2/|x_i|$$

and the following code fragment, derived from Code 2, in which the array D contains the normal equations $A^{\mathrm{T}}A$ and the vector X contains the right-hand side $A^{\mathrm{T}}b$.

```
CALL DPOSV( 'U', N, 1, D, LDD, X, LDX, INFO)
CALL DTRTRI( 'U', 'N', N, D, LDD, INFO)
DO I=1,N
   KAPPA(I) = DNRM2( N−I+1, D(I,I), LDD) * BNORM/ABS(X(I))
END DO
```

Figure 1 represents the relative condition numbers of all the n coefficients. We observe the disparity between the condition numbers (between 10^2 and 10^8). To be able to give a physical interpretation, we need first to sort the coefficients by degrees and orders as given in the development of $V(r, \theta, \lambda)$ in Expression (15).

In Figure 2, we plot the condition numbers of the coefficients $\overline{C}_{\ell m}$ as a function of the degrees and orders (the curve with the $\overline{S}_{\ell m}$ is similar). We notice that for a given order, the condition number increases with the degree and that, for a given degree, the variation of the sensitivity with the order is less significant.

[‡] Centre National d'Etudes Spatiales, Toulouse, France.

[§] Gravity Recovery and Climate Experiment, NASA, launched March 2002.

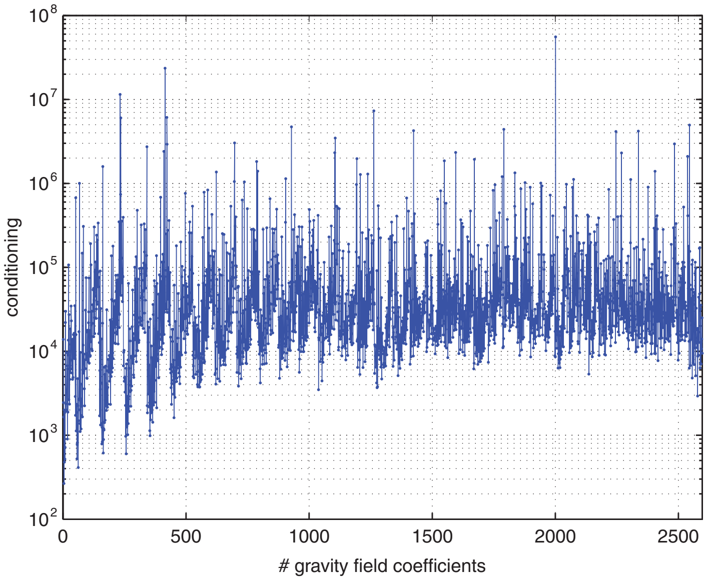

Figure 1. Amplitude of the relative condition numbers for the gravity field coefficients.

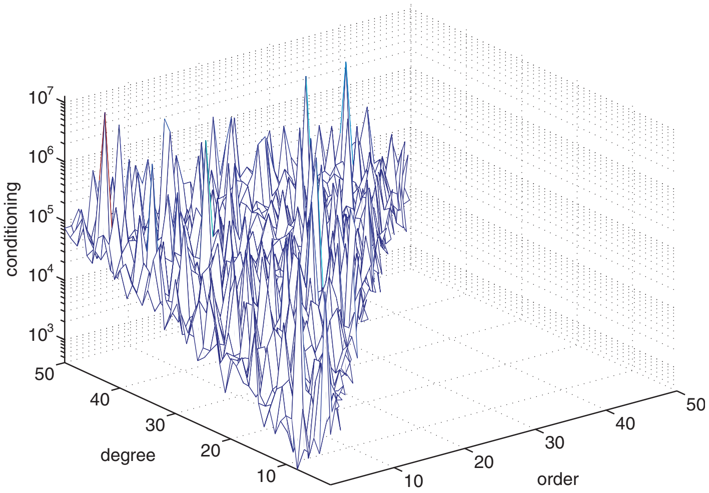

Figure 2. Conditioning of spherical harmonic coefficients $\overline{C}_{\ell m}$ $(2\leqslant\ell\leqslant 50, 1\leqslant m\leqslant 50)$.

We can also study the effect of regularization on the conditioning. The physicists use in general a Kaula [24] regularization technique that consists of adding to $A^{\mathrm{T}}A$ a diagonal matrix $D=\operatorname{diag}(0,\ldots,0,\delta,\ldots,\delta)$, where δ is a constant that is proportional to $10^{-5}/\ell_{\max}^2$ and the nonzero terms in D correspond to the variables that need to be regularized. An example of the effect of Kaula regularization is shown in Figure 3 where we consider the coefficients of order 0 also called

DOI: 10.1002/nla

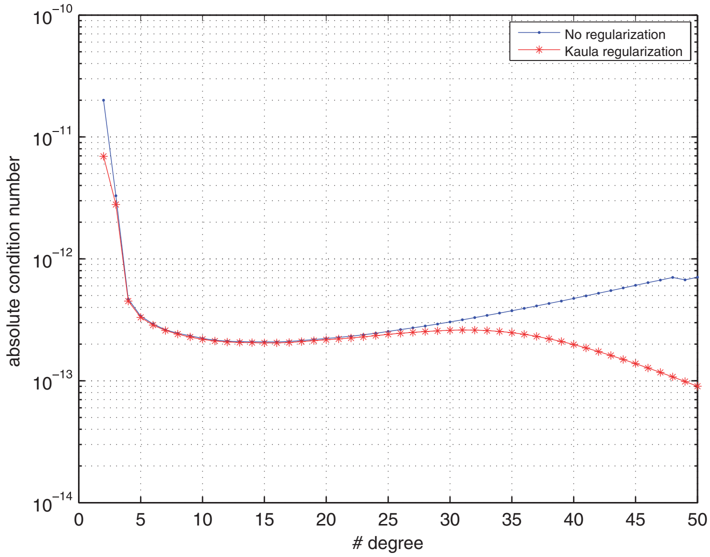

Figure 3. Effect of regularization on zonal coefficients $\overline{C}_{\ell 0}$ $(2\leqslant\ell\leqslant 50)$.

zonal coefficients. We compute here the absolute condition numbers of these coefficients using the formula $\kappa_i(b)=\|e_i^{\mathrm{T}}U^{-1}\|_2$. Note that the $\kappa_i(b)$ are much lower that 1. This is not surprising because typically in our application $\|b\|_2\sim 10^5/$ and $|x_i|\sim 10^{-12}$, which would make the associated relative condition numbers greater than 1. We observe that the regularization is effective on coefficients of highest degree that are in general more sensitive to perturbations.

6. CONCLUSION

To assess the accuracy of a linear least-squares solution, the practitioner of numerical linear algebra uses generally quantities like condition numbers or backward errors when the statistician is more interested in covariance analysis. In this paper we proposed quantities that talk to both communities and that can assess the quality of the solution of a least-squares problem or one of its components. We provided practical ways to compute these quantities using (Sca)LAPACK and we experimented with these computations on practical examples including a real physical application in the area of space geodesy. When there are measurement errors in matrices, it could be more appropriate to use the TLS method. This will be studied in a future work.

REFERENCES

1. Chandrasekaran S, Ipsen ICF. On the sensitivity of solution components in linear systems of equations. *SIAM Journal on Matrix Analysis and Applications* 1995; **16**(1):93–112.
2. Kenney CS, Laub AJ, Reese MS. Statistical condition estimation for linear least squares. *SIAM Journal on Matrix Analysis and Applications* 1998; **19**(4):906–923.
3. Arioli M, Baboulin M, Gratton S. A partial condition number for linear least-squares problems. *SIAM Journal on Matrix Analysis and Applications* 2007; **29**(2):413–433.

DOI: 10.1002/nla

4. Higham NJ, Stewart GW. Numerical linear algebra in statistical computing. In *The State of the Art in Numerical Analysis*, Iserles A, Powell MJD (eds). Oxford University Press: Oxford, 1987; 41–57.
5. Anderson E, Bai Z, Bischof C, Blackford S, Demmel J, Dongarra J, Du Croz J, Greenbaum A, Hammarling S, McKenney A, Sorensen D. *LAPACK Users' Guide* (3rd edn). SIAM: Philadelphia, 1999.
6. Blackford LS, Choi J, Cleary A, D'Azevedo E, Demmel J, Dhillon I, Dongarra J, Hammarling S, Henry G, Petitet A, Stanley K, Walker D, Whaley RC. *ScaLAPACK Users' Guide*. SIAM: Philadelphia, 1997.
7. Geurts AJ. A contribution to the theory of condition. *Numerische Mathematik* 1982; **39**:85–96.
8. Gratton S. On the condition number of linear least squares problems in a weighted Frobenius norm. *BIT Numerical Mathematics* 1996; **36**(3):523–530.
9. Demmel J, Hida Y, Li XS, Riedy EJ. Extra-precise iterative refinement for overdetermined least squares problems. *Technical Report EECS-2007-77*, UC Berkeley, Also LAPACK Working Note 188, 2007.
10. Björck Å. *Numerical Methods for Least Squares Problems*. SIAM: Philadelphia, 1966.
11. Zelen M. Linear estimation and related topics. In *Survey of Numerical Analysis*, Todd J (ed.). McGraw-Hill: New York, 1962; 558–584.
12. Van Huffel S, Vandewalle J. *The Total Least Squares Problem. Computational Aspects and Analysis*. SIAM: Philadelphia, 1991.
13. Golub GH, van Loan CF. An analysis of the total least squares problem. *SIAM Journal on Numerical Analysis* 1980; **17**:883–893.
14. Golub GH, van Loan CF. *Matrix Computations* (3rd edn). The Johns Hopkins University Press: Baltimore, MD, 1996.
15. Grcar JF. Adjoint formulas for condition numbers applied to linear and indefinite least squares. *Technical Report LBNL-55221*, Lawrence Berkeley National Laboratory, 2004.
16. The Numerical Algorithms Group. *NAG Library Manual*, *Mark 21*, NAG, 2006.
17. Higham NJ. *Accuracy and Stability of Numerical Algorithms* (2nd edn). SIAM: Philadelphia, 2002.
18. Hager WW. Condition estimates. *SIAM Journal on Statistical and Scientific Computing* 1984; **5**(2):311–316.
19. Farebrother RW. *Linear Least Squares Computations*. Marcel Dekker: New York, 1988.
20. Laplace PS. Premier supplément. Sur l'application du calcul des probabilités à la philosophie naturelle. *Théorie Analytique des Probabilités*. Mme Ve Courcier, 1820; 497–530.
21. Langou J. Review of 'théorie analytique des probabilités. premier supplément. sur l'application du calcul des probabilités à la philosophie naturelle' from Laplace PS. *Technical Report*, CU Denver, 2007.
22. Balmino G, Cazenave A, Comolet-Tirman A, Husson JC, Lefebvre M. *Cours de géodésie dynamique et spatiale*. Ecole Nationale Supérieure de Techniques Avancées, 1982.
23. Baboulin M. Solving large dense linear least squares problems on parallel distributed computers. application to the earth's gravity field computation. *Ph.D. Thesis*, Institut National Polytechnique de Toulouse, 2006.
24. Kaula WM. *Theory of Satellite Geodesy*. Blaisdell Press: Waltham, MA, 1966.

DOI: 10.1002/nla