Technical Report
RAL-TR-2004-029

Partial condition numbers for linear least squares problems

Mario Arioli, Marc Baboulin, and Serge Gratton

November 15, 2004

Council for the Central Laboratory of the Research Councils

Library and Information Services
CCLRC Rutherford Appleton Laboratory
Chilton Didcot
Oxfordshire OX11 0QX
UK
Tel: +44 (0)1235 445384
Fax: +44(0)1235 446403
Email: library@rl.ac.uk

CCLRC reports are available online at:
http://www.clrc.ac.uk/Activity/ACTIVITY=Publications;SECTION=225;

ISSN 1358-6254

RAL-TR-2004-029

Partial condition numbers for linear least squares problems

Mario Arioli[1] and Marc Baboulin[2] and Serge Gratton[2]

ABSTRACT
We consider here the linear least squares problem $\min_{y \in \mathbb{R}^n} \|Ay - b\|_2$ where $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$ is a matrix of full column rank n and we denote x its solution. We assume that both A and b can be perturbed and that these perturbations are measured using Frobenius norms. In this paper, we are concerned with the condition number of a linear function of x ($L^T x$ where $L \in \mathbb{R}^{n \times k}$) for which we provide an exact formula. This quantity requires the computation of the singular values and the right singular vectors of the matrix A, which can be very expensive in practice. This is why we also propose a statistical method that estimates this condition number by using the exact condition numbers in random orthogonal directions. Provided the triangular R factor of A from $A^T A = R^T R$ is available, this statistical approach enables the computation of a condition estimate in $\mathcal{O}(n^2)$. We also address the question of the numerical reliability of this statistical estimate. In the case where the perturbation of A is measured using the spectral norm, although we do not have a close formula for the condition number, we provide sharp estimates.
Keywords: Linear least squares, normwise condition number, statistical condition estimate, parameter estimation

Keywords: Condition numbers, Least-Squares problems.

AMS(MOS) subject classifications: 65F35, 65F20.

Current reports available by anonymous ftp to ftp.numerical.rl.ac.uk in directory pub/reports.
[1] M.Arioli@rl.ac.uk, Rutherford Appleton Laboratory,
[2] baboulin@cerfacs.fr - gratton@cerfacs.fr, CERFACS, 42 av. Gaspard Coriolis, 31057 Toulouse Cedex, France.

Computational Science and Engineering Department
Atlas Centre
Rutherford Appleton Laboratory
Oxon OX11 0QX

November 11, 2004

Contents

1 Introduction

Perturbation theory has been applied to many fundamental problems of linear algebra such as linear systems, linear least squares, or eigenvalue problems (Björck 1996, Eldén 1980, Higham 2002, Stewart and Sun 1991). In this paper we consider the problem of calculating the quantity $L^T x$, where x is the solution of 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. This estimation is a fundamental problem of parameter estimation in the framework of the Gauss-Markov Model (Rao and Mitra 1971, p. 137). More precisely, we focus here on the evaluation of the sensitivity of $L^T x$ to small perturbations of the matrix A and/or the right-hand side b, where $L \in \mathbb{R}^{n \times k}$ and x is the solution of the LLSP.
The interest for this question stems for instance from parameter estimation where the parameters of the model can often be divided into 2 parts : the variables of physical significance and a set of ancillary variables involved in the models. For example, this situation occurs in the determination of positions using the GPS system, where the 3-D coordinates are the quantities of interest but the statistical model involves other parameters such as clock drift and GPS ambiguities (Kaplan 1996) that are generally estimated during the solution process. It is then crucial to ensure that the solution components of interest can be computed with satisfactory accuracy. The main goal of this paper is to formalize this problem in terms of a condition number and to describe practical methods to compute or estimate this quantity. Note that as far as the sensitivity of a subset of the solution components is concerned, the matrix L is a projection whose columns consist of vectors of the canonical basis of $\mathbb{R}^n$.

The condition number of a map $g : \mathbb{R}^m \mapsto \mathbb{R}^n$ at x_0 measures the sensitivity of $g(x_0)$ to perturbations of x_0. More precisely, suppose that the data space $\mathbb{R}^m$ and the solution space $\mathbb{R}^n$ are equipped respectively with the norms $\|.\|_{\mathcal{D}}$ and $\|.\|_{\mathcal{S}}$, the condition number $K(x_0)$ is defined by

$$K(x_0) = \lim_{\delta \to 0} \sup_{0 < \|x_0 - x\|_{\mathcal{D}} \leq \delta} \frac{\|g(x_0) - g(x)\|_{\mathcal{S}}}{\|x_0 - x\|_{\mathcal{D}}}, \tag{1}$$

whereas the relative condition number is defined by $K^{(rel)}(x_0) = K(x_0)\|x_0\|_{\mathcal{D}} / \|g(x_0)\|_{\mathcal{S}}$. This definition shows that $K(x_0)$ measures an asymptotic sensitivity and that this quantity depends on the chosen norms for the data and solution spaces. If g is a Fréchet-differentiable (F-differentiable) function at x_0, then $K(x_0)$ is the norm of the F-derivative $|||g'(x_0)|||)$ (see (Geurts 1982)), where $|||.|||$ is the operator norm induced by the choice of the norms on the data and solution spaces.

For the full rank LLSP, we have $g(A, b) = (A^T A)^{-1} A^T b$. If we consider the product norm $\|(A, b)\|_F = \sqrt{\|A\|_{\mathrm{F}}^2 + \|b\|_2^2}$ for the data space and $\|x\|_2$ for the solution space, then Gratton (1996) gives an explicit formula for the relative condition number $K^{(rel)}(A, b)$:

$$K^{(rel)}(A, b) = \left\|A^\dagger\right\|_2 \left(\left\|A^\dagger\right\|_2^2 \|r\|_2^2 + \|x\|_2^2 + 1 \right)^{\frac{1}{2}} \frac{\|(A, b)\|_F}{\|x\|_2}$$

where $A^\dagger$ denotes the pseudo inverse of A, $r = b - Ax$ is the residual vector and $\|.\|_{\mathrm{F}}$ and $\|.\|_2$ are respectively the Frobenius and Euclidean norms. But does the value of $K^{(rel)}(A, b)$ give us useful information about the sensitivity of $L^T x$? Can it in some cases overestimate the error in components or on the contrary be too optimistic ?
Let us consider the following example.

$$A = \begin{pmatrix} 1 & 1 & \epsilon^2 \\ \epsilon & 0 & \epsilon^2 \\ 0 & \epsilon & \epsilon^2 \\ \epsilon^2 & \epsilon^2 & 2 \end{pmatrix}, \; x = \begin{pmatrix} \epsilon \\ \epsilon \\ \frac{1}{\epsilon} \end{pmatrix} \text{ and } b = \begin{pmatrix} 3\epsilon \\ \epsilon^2 + \epsilon \\ \epsilon^2 + \epsilon \\ 2\epsilon^3 + \frac{2}{\epsilon} \end{pmatrix},$$

x is here the exact solution of the LLSP $\min_{x \in \mathbb{R}^3} \|Ax - b\|_2$. If we take $\epsilon = 10^{-8}$ then we have $x = (10^{-8}, 10^{-8}, 10^8)^T$ and the solution computed in Matlab using a machine precision $2.22 \cdot 10^{-16}$ is $\tilde{x} =$

$(1.5 \cdot 10^{-8}, 1.5 \cdot 10^{-8}, 10^8)^T$. The LLSP condition number is $K^{(rel)}(A, b) = 2.4 \cdot 10^8$ and the relative errors on the components of x are

$$\frac{|x_1 - \tilde{x}_1|}{|x_1|} = \frac{|x_2 - \tilde{x}_2|}{|x_2|} = 0.5 \text{ and } \frac{|x_3 - \tilde{x}_3|}{|x_3|} = 0.$$

Then, if $L = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$, we expect a large value for the condition number of $L^T x$ because there is a 50% relative error on x_1 and x_2. If now $L = (0, 0, 1)^T$, then we expect that the condition number of $L^T x$ would be close to 1 because $\tilde{x}_3 = x_3$. For these two values of L, the LLSP condition number is far from giving a good idea of the sensitivity of $L^T x$. Note in this case that the perturbations are due to roundoff errors.
Let us now consider a simple example in the framework of parameter estimation where in addition to roundoff errors, random errors are involved. Let $b = \{b_i\}_{i=1,\cdots,10}$ be a series of observed values depending on data $s = \{s_i\}$ where $s_i = 10 + i, i = 1, \cdots, 10$. We determine a 3-degree polynomial that approximates b in the least squares sense, and we suppose that the following relationship holds

$$b = x_1 + x_2 \frac{1}{s} + x_3 \frac{1}{s^2} + x_4 \frac{1}{s^3} \text{ with } x_1 = x_2 = x_3 = x_4 = 1.$$

This corresponds to the LLSP $\min_{x \in \mathbb{R}^4} \|Ax - b\|_2$ where A is the Vandermonde matrix defined by $A_{ij} = \frac{1}{s_i^{j-1}}$. We assume that the perturbation on each b_i is 10^{-8} multiplied by a normally distributed random number. Let $\tilde{x}$ and $\tilde{y}$ be the computed solutions corresponding to two perturbed right-hand sides. Then we obtain the following relative errors on each component:

$$\frac{|\tilde{x}_1 - \tilde{y}_1|}{|\tilde{x}_1|} = 2 \cdot 10^{-7}, \frac{|\tilde{x}_2 - \tilde{y}_2|}{|\tilde{x}_2|} = 6 \cdot 10^{-6}, \frac{|\tilde{x}_3 - \tilde{y}_3|}{|\tilde{x}_3|} = 6 \cdot 10^{-5}, \text{ and } \frac{|\tilde{x}_4 - \tilde{y}_4|}{|\tilde{x}_4|} = 10^{-4}.$$

We have $K^{(rel)}(A, b) = 3.1 \cdot 10^5$. Regarding the disparity between the sensitivity of each component, we need a quantity that evaluates more precisely the sensitivity of each solution component of the LLSP.
The idea of analyzing the accuracy of some solution components in linear algebra is by no means new. For linear systems $Ax = b$, $A \in \mathbb{R}^n$ and for LLSP, Chandrasekaran and Ipsen (1995) define so called componentwise condition numbers that correspond to amplification factors of the relative errors in solution components due to perturbations of data A or b and explains how to estimate them. In our formalism, these quantities are upper bounds of the condition number of $L^T x$ where L is a column of the identity matrix. We also emphasise that the term "componentwise" refers here to the solution components and must be distinguished from the metric used for matrices and for which Wei, Xu, Qiao and Diao (2003) provide a condition number for generalized inversion and linear least squares.
For LLSP, Kenney, Laub and Reese (1998) provide a statistical estimate for componentwise condition numbers due to either relative or structured perturbations. In the case of linear systems, Cao and Petzold (2003) propose a statistical approach, based on the work of Kenney and Laub (1994), that enables to compute the condition number of $L^T x$ in $\mathcal{O}(n^2)$.
Our approach differs from the previous studies in the following aspects:

1. we are interested in the condition of $L^T x$ where L is a general matrix and not only a canonical vector of $\mathbb{R}^n$,
2. we are looking for a condition number based on the Fréchet-derivative, and not only for an upper bound of this quantity.

We present in this paper three ways to obtain information on the condition of $L^T x$. The first one uses an explicit formula based on the singular value decomposition of A. The second is at the same time an upper bound of this condition number and a sharp estimate of it. The third method supplies a statistical estimate. The choice between these three methods will depend on the size of the problem (computational cost) and on the accuracy desired for this quantity.

This paper is organized as follows. In Section 2, we define the notion of a partial condition number and provide a closed formula for it in the general case where $L \in \mathbb{R}^{n \times k}$ and in the particular case when

$L \in \mathbb{R}^n$. Then in Section 3 we establish bounds of the partial condition number in Frobenius as well as in spectral norm, and these bounds can be considered as sharp estimates of it. In Section 4 we describe a statistical method that estimates the partial condition number. In Section 5 we present numerical results in order to compare the statistical estimate and the exact condition number on sample matrices A and p. In Section 6 we give a summary comparing the three ways to compute the condition of $L^T x$ as well as a numerical illustration. Finally some concluding remarks are given in Section 7.
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 $\|.\|_2$ on vectors. I is the identity matrix and e_i is the i-th canonical vector. We also denote by $\mathrm{Im}(A)$ the space spanned by the columns of A and by $\mathrm{Ker}(A)$ the null space of A.

2 The partial condition number of an LLSP

Let L be a $n \times k$ matrix, with $k \leq n$. We consider the function

$$\begin{array}{rcccl} g & : & \mathbb{R}^{m\times n} \times \mathbb{R}^m & \longrightarrow & \mathbb{R}^k \\ & & A, b & \longmapsto & g(A,b) = L^T x(A,b) = L^T (A^T A)^{-1} A^T b. \end{array} \quad (2)$$

Since A has full rank n, g is continuously F-differentiable in a neighborhood of (A,b) and we denote by g' its F-derivative. Let α and β be two positive real numbers. In the present paper we will 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 will use the product norm defined by

$$\|(A,b)\|_{\mathrm{F\ or\ 2}} = \sqrt{\alpha^2 \|A\|^2_{\mathrm{F\ or\ 2}} + \beta^2 \|b\|_2^2},\ \alpha,\beta > 0.$$

This norm is very flexible since it allows to monitor the perturbations on A and b. For instance, large values of α (resp. β) enable to obtain condition number problems where mainly b (resp. A) are perturbed. A more general weighted Frobenius norm $\|(AT, \beta b)\|_{\mathrm{F}}$, where T is a positive diagonal matrix is sometimes chosen. This is for instance the case where Wei, Diao and Qiao (2004) give an explicit expression for the condition number of rank deficient linear least squares using this norm.
According to Geurts (1982), the absolute condition number of g at the point (A,b) is given by:

$$\kappa_{g,\mathrm{F\ or\ 2}}(A,b) = \|g'(A,b)\| = \max_{(\Delta A,\Delta b)} \frac{\|g'(A,b).(\Delta A,\Delta b)\|_2}{\|(\Delta A,\Delta b)\|_{\mathrm{F\ or\ 2}}},$$

and then the relative condition number of g at (A,b) is expressed by

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

Since $L \in \mathbb{R}^{n\times k}$, $k \leq n$, we call the condition number related to $L^T x(A,b)$ a *partial condition number* of the LLSP with respect to the linear operator L. The partial condition number is given by the following theorem.

Theorem 1. *Let $A = U\Sigma V^T$ be the thin singular value decomposition of A (Golub and van Loan 1996) with $\Sigma = diag(\sigma_i)$ and $\sigma_1 \geq \sigma_2 \cdots \geq \sigma_n > 0$. The absolute condition number of $g(A,b) = L^T x(A,b)$ is given by*

$$\kappa_{g,F}(A,b) = \left\|SV^T L\right\|_2$$

where $S \in \mathbb{R}^{n\times n}$ is the diagonal matrix with diagonal elements $S_{ii} = \sigma_i{}^{-1}\sqrt{\frac{\sigma_i{}^{-2}\|r\|_2^2 + \|x\|_2^2}{\alpha^2} + \frac{1}{\beta^2}}$.

Proof. Let $\Delta A \in \mathbb{R}^{m\times n}$ and $\Delta b \in \mathbb{R}^m$. Using the chain rules of composition of derivatives, we get

$$g'(A,b).(\Delta A,\Delta b) = L^T(A^TA)^{-1}\Delta A^T(b - A(A^TA)^{-1}A^Tb) - L^T(A^TA)^{-1}A^T\Delta A(A^TA)^{-1}A^Tb + L^TA^\dagger\Delta b$$

i.e

$$g'(A,b).(\Delta A,\Delta b) = L^T(A^TA)^{-1}\Delta A^T r - L^T A^\dagger \Delta A x + L^T A^\dagger \Delta b. \quad (3)$$

We write $\Delta A = \Delta A_1 + \Delta A_2$ by defining $\Delta A_1 = AA^\dagger \Delta A$ (projection of ΔA on $\mathrm{Im}(A)$) and $\Delta A_2 = (I - AA^\dagger)\Delta A$ (projection of ΔA on $\mathrm{Im}(A)^\perp$). Since $r \in \mathrm{Im}(A)^\perp$ and $A^\dagger \Delta A_2 = 0$, we obtain

$$g'(A,b).(\Delta A, \Delta b) = L^T(A^TA)^{-1}\Delta A_2^T r - L^T A^\dagger \Delta A_1 x + L^T A^\dagger \Delta b. \tag{4}$$

We now prove that $\kappa_{g,F}(A,b) \leq \left\|SV^TL\right\|_2$. Let u_i and v_i be the i-th column of respectively U and V. From $A^\dagger = V\Sigma^{-1}U^T$, we get $AA^\dagger = UU^T = \sum_{i=1}^n u_i u_i^T$ and since $\sum_{i=1}^n v_i v_i^T = I$, we have $\Delta A_1 = \sum_{i=1}^n u_i u_i^T \Delta A$ and $\Delta A_2 = (I - AA^\dagger)\Delta A \sum_{i=1}^n v_i v_i^T$. Moreover, still using the thin SVD of A and $A^\dagger$, it follows that

$$(A^TA)^{-1}v_i = \frac{v_i}{\sigma_i^2} \text{ and } A^\dagger u_i = v\Sigma^{-1}e_i. \tag{5}$$

Thus (4) becomes

$$\begin{aligned} g'(A,b).(\Delta A, \Delta b) &= \sum_{i=1}^n L^T v_i [v_i^T \Delta A^T (I - AA^\dagger)\frac{r}{\sigma_i^2} - u_i^T \Delta A \frac{x}{\sigma_i} + u_i^T \frac{\Delta b}{\sigma_i}] \\ &= L^T \sum_{i=1}^n v_i y_i, \end{aligned}$$

where we set $y_i = v_i^T \Delta A^T (I - AA^\dagger)\frac{r}{\sigma_i^2} - u_i^T \Delta A \frac{x}{\sigma_i} + u_i^T \frac{\Delta b}{\sigma_i} \in \mathbb{R}$.
Thus if $Y = (y_1, y_2, \cdots, y_n)^T$, we get $\|g'(A,b).(\Delta A, \Delta b)\|_2 = \left\|L^TVY\right\|_2$ and then

$$\|g'(A,b).(\Delta A, \Delta b)\|_2 = \left\|L^TVSS^{-1}Y\right\|_2 \leq \left\|SV^TL\right\|_2 \left\|S^{-1}Y\right\|_2.$$

We denote by $w_i = \frac{v_i^T \Delta A^T (I - AA^\dagger) r}{S_{ii}\sigma_i^2} - \frac{u_i^T \Delta A x}{S_{ii}\sigma_i} + \frac{u_i^T \Delta b}{S_{ii}\sigma_i}$ the i-th component of $S^{-1}Y$. Then we have

$$\begin{aligned} |w_i| &\leq \alpha \left\|v_i^T \Delta A^T (I - AA^\dagger)^T\right\|_2 \frac{\|r\|_2}{\alpha S_{ii}\sigma_i^2} + \alpha \left\|u_i^T \Delta A\right\|_2 \frac{\|x\|_2}{\alpha S_{ii}\sigma_i} + \frac{\beta |u_i^T \Delta b|}{\beta S_{ii}\sigma_i} \\ &\leq (\frac{\|r\|_2^2}{\alpha^2 S_{ii}^2 \sigma_i^4} + \frac{\|x\|_2^2}{\alpha^2 S_{ii}^2\sigma_i^2} + \frac{1}{\beta^2 S_{ii}^2 \sigma_i^2})^{\frac{1}{2}} (\alpha^2 \left\|(I - AA^\dagger)\Delta A v_i\right\|_2^2 + \alpha^2 \left\|u_i^T \Delta A\right\|_2^2 + \beta^2 |u_i^T \Delta b|^2)^{\frac{1}{2}} \\ &= \frac{S_{ii}}{S_{ii}} (\alpha^2 \left\|(I - AA^\dagger)\Delta A v_i\right\|_2^2 + \alpha^2 \left\|u_i^T \Delta A\right\|_2^2 + \beta^2 |u_i^T \Delta b|^2)^{\frac{1}{2}}. \end{aligned}$$

Hence

$$\begin{aligned} \left\|S^{-1}Y\right\|_2^2 &\leq \sum_{i=1}^n \alpha^2 \left\|(I - AA^\dagger)\Delta A v_i\right\|_2^2 + \alpha^2 \left\|u_i^T \Delta A\right\|_2^2 + \beta^2 |u_i^T \Delta b|^2 \\ &= \alpha^2 \left\|(I - AA^\dagger)\Delta A V\right\|_F^2 + \alpha^2 \left\|U^T \Delta A\right\|_F^2 + \beta^2 \left\|U^T \Delta b\right\|_2^2 \\ &= \alpha^2 \left\|(I - AA^\dagger)\Delta A\right\|_F^2 + \alpha^2 \left\|U^T \Delta A\right\|_F^2 + \beta^2 \left\|U^T \Delta b\right\|_2^2. \end{aligned}$$

Since $\left\|U^T \Delta A\right\|_F = \left\|UU^T \Delta A\right\|_F = \left\|AA^\dagger \Delta A\right\|_F$ and $\left\|U^T \Delta b\right\|_2 = \left\|UU^T \Delta b\right\|_2 \leq \|\Delta b\|_2$, we get

$$\left\|S^{-1}Y\right\|_2^2 \leq \alpha^2 \|\Delta A_1\|_F^2 + \alpha^2 \|\Delta A_2\|_F^2 + \beta^2 \|\Delta b\|_2^2.$$

From $\|\Delta A\|_F^2 = \|\Delta A_1\|_F^2 + \|\Delta A_2\|_F^2$, we get $\left\|S^{-1}Y\right\|_2^2 \leq \|(\Delta A, \Delta b)\|_F^2$ and thus

$$\|g'(A,b).(\Delta A, \Delta b)\|_2 \leq \left\|SV^TL\right\|_2 \|(\Delta A, \Delta b)\|_F.$$

So we have shown that $\left\|SV^TL\right\|_2$ is an upper bound for $\kappa_{g,F}(A,b)$.
We now prove that this upper bound can be reached, that is, that $\left\|SV^TL\right\|_2 = \frac{\left\|g'(A,b).(\Delta A, \Delta b)\right\|_2}{\|(\Delta A, \Delta b)\|_F}$ holds for some $(\Delta A, \Delta b) \in \mathbb{R}^{m\times n} \times \mathbb{R}^m$.
Let $(\Delta A, \Delta b)$ be expressed as

$$(\Delta A, \Delta b) = (\Delta A_2 + \Delta A_1, \Delta b) = (\sum_{i=1}^n \frac{\alpha_i}{\alpha} \frac{r}{\|r\|_2} v_i^T + \sum_{i=1}^n \frac{\beta_i}{\alpha} u_i \frac{x^T}{\|x\|_2}, \sum_{i=1}^n \frac{\gamma_i}{\beta} u_i).$$

Then it follows from (4) and (5) that

$$\begin{aligned} g'(A,b).(\Delta A,\Delta b) &= L^T(A^TA)^{-1}\sum_{i=1}^{n}\frac{\alpha_i}{\alpha}\|r\|_2 v_i^T - L^TA^\dagger\sum_{i=1}^{n}\frac{\beta_i}{\alpha}u_i\|x\|_2 + L^TA^\dagger\sum_{i=1}^{n}\frac{\gamma_i}{\beta}u_i \\ &= L^T\sum_{i=1}^{n}\frac{\alpha_i}{\alpha\sigma_i^2}v_i\|r\|_2 - L^T\sum_{i=1}^{n}\frac{\beta_i}{\alpha\sigma_i}v_i\|x\|_2 + L^T\sum_{i=1}^{n}\frac{\gamma_i}{\beta\sigma_i}u_i \\ &= \sum_{i=1}^{n}L^Tv_i(\frac{\alpha_i}{\alpha\sigma_i^2}\|r\|_2 - \frac{\beta_i}{\alpha\sigma_i}\|x\|_2 + \frac{\beta\gamma_i}{\sigma_i}). \end{aligned}$$

Thus by denoting $\xi_i = [L^Tv_i\frac{\|r\|_2}{\alpha\sigma_i^2}, -L^Tv_i\frac{\|x\|_2}{\alpha\sigma_i}, \frac{L^Tv_i}{\beta\sigma_i}] \in \mathbb{R}^{1\times 3}$ and $\Gamma = [\xi_1,\ldots,\xi_n] \in \mathbb{R}^{1\times 3n}$, and $X = (\alpha_1,\beta_1,\gamma_1,\cdots,\alpha_n,\beta_n,\gamma_n)^T \in \mathbb{R}^{3n\times 1}$ we get

$$g'(A,b).(\Delta A,\Delta b) = \Gamma X. \tag{6}$$

Since ${u_i}^Tr = 0$, we have $\mathrm{trace}(\Delta A_1^T\Delta A_2) = 0$, and $\|\Delta A\|_F^2 = \|\Delta A_1\|_F^2 + \|\Delta A_2\|_F^2$. Then, using the fact that

$$\|(\Delta A,\Delta b)\|_F^2 = \sum_{i=1}^{n}\alpha_i^2 + \sum_{i=1}^{n}\beta_i^2 + \sum_{i=1}^{n}\gamma_i^2 = \|X\|_2^2,$$

Equation (6) yields

$$\frac{\|g'(A,b).(\Delta A,\Delta b)\|_2}{\|(\alpha\Delta A,\beta\Delta b)\|_F} = \frac{\|\Gamma X\|_2}{\|X\|_2}.$$

We know that $\|\Gamma\|_2 = \max_X \frac{\|\Gamma X\|_2}{\|X\|_2}$ is reached for some $X = (\alpha_1,\beta_1,\gamma_1,\cdots,\alpha_n,\beta_n,\gamma_n)^T$. Then for the $(\Delta A,\Delta b)$ corresponding to this X, we have $\frac{\left\|g'(A,b).(\Delta A,\Delta b)\right\|_2}{\|(\Delta A,\Delta b)\|_F} = \|\Gamma\|_2$.
Since in addition $\Gamma = [L^Tv_1[\frac{\|r\|_2}{\alpha\sigma_1^2}, -\frac{\|x\|_2}{\alpha\sigma_1}, \frac{1}{\beta\sigma_1}],\cdots,L^Tv_n[\frac{\|r\|_2}{\alpha\sigma_n^2}, -\frac{\|x\|_2}{\alpha\sigma_n}, \frac{1}{\beta\sigma_n}]]$, we get

$$\begin{aligned} \Gamma\Gamma^T &= L^Tv_1(\frac{\|r\|_2^2}{\alpha^2\sigma_1^4} + \frac{\|x\|_2^2}{\alpha^2\sigma_1^2} + \frac{1}{\beta^2\sigma_1^2})v_1^TL + \cdots + L^Tv_n(\frac{\|r\|_2^2}{\alpha^2\sigma_n^4} + \frac{\|x\|_2^2}{\alpha^2\sigma_n^2} + \frac{1}{\beta^2\sigma_n^2})v_n^TL \\ &= L^Tv_1S_{11}^2v_1^TL + \cdots + L^Tv_nS_{nn}^2v_n^TL \\ &= (L^TVS)(SV^TL). \end{aligned}$$

Hence

$$\|\Gamma\|_2 = \sqrt{\|\Gamma\Gamma^T\|_2} = \left\|SV^TL\right\|_2$$

and $\alpha_1,\beta_1,\gamma_1,\cdots,\alpha_n,\beta_n,\gamma_n$ are such that $\frac{\left\|g'(A,b).(\Delta A,\Delta b)\right\|_2}{\|(\alpha\Delta A,\beta\Delta b)\|_F} = \left\|SV^TL\right\|_2$.
Thus $\left\|SV^TL\right\|_2 \le \kappa_{g,F}(A,b)$, which concludes the proof. $\square$

Let l_j be the j-th column of L, $j = 1,\cdots,k$. From

$$SV^TL = \begin{pmatrix} S_{11}{v_1}^T \\ \vdots \\ S_{nn}{v_n}^T \end{pmatrix}(l_1,\cdots,l_k) = \begin{pmatrix} S_{11}{v_1}^Tl_1 & \cdots & S_{11}{v_1}^Tl_k \\ \vdots & & \vdots \\ S_{nn}{v_n}^Tl_1 & \cdots & S_{nn}{v_n}^Tl_k \end{pmatrix},$$

it follows that $\left\|SV^TL\right\|_2$ is large when there exists at least one large S_{ii} and a l_j such that ${v_i}^Tl_j \neq 0$. In particular, the condition number of $L^Tx(A,b)$ is large when A has small singular values and L has components in the corresponding right singular vectors or when $\|r\|_2$ is large.
Let us study the particular case where L is a vector i.e when g is a scalar derived function.

Corollary 1. *In the particular case when L is a vector ($L \in \mathbb{R}^n$), the absolute condition number of $g(A,b) = L^T x(A,b)$ is given by*

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

Proof. By replacing $(A^TA)^{-1} = V\Sigma^{-2}V^T$ and $A^\dagger = V\Sigma^{-1}U^T$ in the expression of $K = (\left\|L^T(A^TA)^{-1}\right\|_2^2 \|r\|_2^2 + \left\|L^TA^\dagger\right\|_2^2 (\|x\|_2^2 + 1))^{\frac{1}{2}}$ we get

$$\begin{aligned} K^2 &= \left\|L^TV\Sigma^{-2}V^T\right\|_2^2 \frac{\|r\|_2^2}{\alpha^2} + \left\|L^TV\Sigma^{-1}U^T\right\|_2^2 (\frac{\|x\|_2^2}{\alpha^2} + \frac{1}{\beta^2}) \\ &= \left\|L^TV\Sigma^{-2}\right\|_2^2 \frac{\|r\|_2^2}{\alpha^2} + \left\|L^TV\Sigma^{-1}\right\|_2^2 (\frac{\|x\|_2^2}{\alpha^2} + \frac{1}{\beta^2}) \\ &= \left\|\Sigma^{-2}V^TL\right\|_2^2 \frac{\|r\|_2^2}{\alpha^2} + \left\|\Sigma^{-1}V^TL\right\|_2^2 (\frac{\|x\|_2^2}{\alpha^2} + \frac{1}{\beta^2}). \end{aligned}$$

By writing $(z_1, \cdots, z_n)^T$ the vector $V^TL \in \mathbb{R}^n$ we have

$$\begin{aligned} K^2 &= \sum_{i=1}^n \frac{z_i^2}{\sigma_i^4} \frac{\|r\|_2^2}{\alpha^2} + \sum_{i=1}^n \frac{z_i^2}{\sigma_i^2} (\frac{\|x\|_2^2}{\alpha^2} + \frac{1}{\beta^2}) \\ &= \sum_{i=1}^n \frac{z_i^2}{\sigma_i^2} (\sigma_i^{-2} \frac{\|r\|_2^2 + \|x\|_2^2}{\alpha^2} + \frac{1}{\beta^2}) \\ &= \sum_{i=1}^n S_{ii}^2 z_i^2 \\ &= \left\|SV^TL\right\|_2^2, \end{aligned}$$

and Theorem 1 gives the result. □

Remark 1. In the general case where L is an $n \times k$ matrix, the computation of $\kappa_{g,F}(A,b)$ via the exact formula given in Theorem 1 requires the computation of the singular values and the right singular vectors of A, which might be expensive in practice since its involves $2mn^2$ operations if we use a R-SVD algorithm and if $m \gg n$ (see Golub and van Loan, 1996, p. 254). If the LLSP is solved using a direct method, the R factor of the QR decomposition of A (or equivalently in exact arithmetic, the Cholesky factor of A^TA) might be available. Since the right singular vectors of A are also those of R, the condition number can be computed in about $12n^3$ flops using the Golub-Reinsch SVDmethod (Golub and van Loan, 1996, p. 254). Using R is even more interesting when $L \in \mathbb{R}^n$, since from

$$\left\|L^TA^+\right\|_2 = \left\|R^{-T}L\right\|_2 \text{ and } \left\|L^T(A^TA)^{-1}\right\|_2 = \left\|R^{-1}(R^{-T}L)\right\|_2, \tag{7}$$

it follows that the computation of $\kappa_{g,F}(A,b)$ can be done by solving two successive n-by-n triangular systems which involves about $2n^2$ flops.

3 Sharp estimate of the partial condition number in Frobenius and spectral norms

In many cases, obtaining a lower and/or an upper bound of $\kappa_{g,F}(A,b)$ is satisfactory when these bounds are tight enough and significantly cheaper to compute than the exact formula. Moreover, many applications use conditions numbers expressed in spectral norm. We give in the following Theorem sharp bounds for the partial condition numbers in the Frobenius and spectral norms.

Theorem 2. *The absolute condition numbers of $g(A,b) = L^T x(A,b)$ in the Frobenius and spectral norms can be respectively bounded as follows*

$$\frac{f(A,b)}{\sqrt{3}} \leq \kappa_{g,F}(A,b) \leq f(A,b)$$

$$\frac{f(A,b)}{\sqrt{3}} \leq \kappa_{g,2}(A,b) \leq \sqrt{2} f(A,b)$$

where

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

Proof. Let v_n and w_n be the right singular vectors corresponding to the largest singular values of respectively $L^T(A^TA)^{-1}$ and $L^TA^\dagger$. Let $(\Delta A, \Delta b)$ be expressed as

$$(\Delta A, \Delta b) = (a_1 \frac{r}{\|r\|_2} v_n^T + a_2 w_m \frac{x^T}{\|x\|_2}, a_3 w_m).$$

By replacing this value of $(\Delta A, \Delta b)$ in (3) we get

$$\begin{aligned} g'(A,b).(\Delta A, \Delta b) &= a_1 \left\| L^T (A^T A)^{-1} \right\|_2 \|r\|_2 - a_2 \left\| L^T A^\dagger \right\|_2 \|x\|_2 \\ &+ a_2 L^T (A^T A)^{-1} \frac{x}{\|x\|_2} w_m^T r - a_1 L^T A^\dagger \frac{r}{\|r\|_2} v_n^T x + a_3 \left\| L^T A^\dagger \right\|_2. \end{aligned}$$

Since $r \in \text{ Im}(A)^\perp$ we have $A^\dagger r = 0$. Moreover we have $w_m \in \text{ Ker}(L^T A^\dagger)^\perp$ and thus $w_m \in \text{ Im}(A^{+T} L)$ and can be written $w_m = A^{+T} L \delta$. Then $w_m^T r = \delta^T L^T A^\dagger r = 0$. It follows that

$$g'(A,b).(\Delta A, \Delta b) = a_1 \left\| L^T (A^T A)^{-1} \right\|_2 \|r\|_2 - a_2 \left\| L^T A^\dagger \right\|_2 \|x\|_2 + a_3 \left\| L^T A^\dagger \right\|_2$$

and

$$\|g'(A,b).(\Delta A, \Delta b)\|_2 = \left\| \left[\frac{\left\| L^T (A^T A)^{-1} \right\|_2 \|r\|_2}{\alpha}, -\frac{\left\| L^T A^\dagger \right\|_2 \|x\|_2}{\alpha}, \frac{\left\| L^T A^\dagger \right\|_2}{\beta} \right] \begin{bmatrix} \alpha a_1 \\ \alpha a_2 \\ \beta a_3 \end{bmatrix} \right\|_2.$$

On the other hand, we have $\|(\Delta A, \Delta b)\|_{\text{F or 2}} = \sqrt{\alpha^2 \|\Delta A\|_{\text{F or 2}}^2 + \beta^2 \|\Delta b\|_2^2}$. From the well-known inequality $\forall (c,d) \in \mathbb{R}_+^2, \frac{c+d}{\sqrt{2}} \leq \sqrt{c^2 + d^2} \leq c + d$ and since $\|\Delta A\|_{\text{F or 2}} \leq |a_1| + |a_2|$, we get

$$\begin{aligned} \|(\Delta A, \Delta b)\|_{\text{F or 2}} &\leq \alpha \|\Delta A\|_{\text{F or 2}} + \beta \|\Delta b\|_2 \\ &\leq \alpha |a_1| + \alpha |a_2| + \beta |a_3| \\ &\leq \sqrt{3} \sqrt{\alpha^2 a_1^2 + \alpha^2 a_2^2 + \beta^2 a_3^2}. \end{aligned}$$

Then

$$\frac{\|g'(A,b).(\Delta A, \Delta b)\|_2}{\|(\Delta A, \Delta b)\|_{\text{F or 2}}} \geq \frac{1}{\sqrt{3}} \frac{\left\| \left[\frac{\left\| L^T (A^T A)^{-1} \right\|_2 \|r\|_2}{\alpha}, -\frac{\left\| L^T A^\dagger \right\|_2 \|x\|_2}{\alpha}, \frac{\left\| L^T A^\dagger \right\|_2}{\beta} \right] \begin{bmatrix} \alpha a_1 \\ \alpha a_2 \\ \beta a_3 \end{bmatrix} \right\|_2}{\sqrt{\alpha^2 a_1^2 + \alpha^2 a_2^2 + \beta^2 a_3^2}}.$$

Let us denote by $R(a_1, a_2, a_3)$ the right-hand side of the above inequality. We know that there exists (a_1, a_2, a_3) that maximizes $R(a_1, a_2, a_3)$ and this maximum is equal to

$$\left\| \left[\frac{\left\| L^T (A^T A)^{-1} \right\|_2 \|r\|_2}{\alpha}, -\frac{\left\| L^T A^\dagger \right\|_2 \|x\|_2}{\alpha}, \frac{\left\| L^T A^\dagger \right\|_2}{\beta} \right] \right\|_2$$

i.e $f(A,b)$. Thus $\kappa_{g,\mathrm{F\ or\ 2}}(A,b) \geq \frac{1}{\sqrt{2}} f(A,b)$.

Let us now establish the upper bound for $\kappa_{g,F}(A,b)$. From (4) we get

$$\begin{aligned} \|g'(A,b).(\Delta A,\Delta b)\|_2 &\leq \left\|L^T(A^TA)^{-1}\right\|_2 \|\Delta A_2\|_2 \|r\|_2 + \left\|L^TA^\dagger\right\|_2 \|\Delta A_1\|_2 \|x\|_2 + \left\|L^TA^\dagger\right\|_2 \|\Delta b\|_2 \\ &= YX, \end{aligned}$$

where

$$Y = \left(\frac{\left\|L^T(A^TA)^{-1}\right\|_2 \|r\|_2}{\alpha}, \frac{\left\|L^TA^\dagger\right\|_2 \|x\|_2}{\alpha}, \frac{\left\|L^TA^\dagger\right\|_2}{\beta} \right)$$

and

$$X = (\alpha \|\Delta A_2\|_2, \alpha \|\Delta A_1\|_2, \beta \|\Delta b\|_2)^T.$$

Thus $\|g'(A,b).(\Delta A,\Delta b)\|_2 \leq \|Y\|_2 \|X\|_2$, with

$$\|X\|_2^2 = \alpha^2 \|\Delta A_1\|_2^2 + \alpha^2 \|\Delta A_2\|_2^2 + \beta^2 \|\Delta b\|_2 \leq \alpha^2 \|\Delta A_1\|_F^2 + \alpha^2 \|\Delta A_2\|_F^2 + \beta^2 \|\Delta b\|_2 .$$

Then, since $\|\Delta A\|_F^2 = \|\Delta A_1\|_F^2 + \|\Delta A_2\|_F^2$, we have $\|X\|_2 \leq \|(\Delta A,\Delta b)\|$ and it implies that $\kappa_{g,F}(A,b) \leq \|Y\|_2$ i.e $\kappa_{g,F}(A,b) \leq f(A,b)$.
For the upper bound for $\kappa_{g,2}(A,b)$, we get from (3) that

$$\begin{aligned} \|g'(A,b).(\Delta A,\Delta b)\|_2 &\leq (\left\|L^T(A^TA)^{-1}\right\|_2 \|r\|_2 + \left\|L^TA^\dagger\right\|_2 \|x\|_2) \|\Delta A\|_2 + \left\|L^TA^\dagger\right\|_2 \|\Delta b\|_2 \\ &= Y'X', \end{aligned}$$

where $Y' = \left(\frac{\left\|L^T(A^TA)^{-1}\right\|_2 \|r\|_2 + \left\|L^TA^\dagger\right\|_2 \|x\|_2}{\alpha}, \frac{\left\|L^TA^\dagger\right\|_2}{\beta} \right)$ and $X' = (\alpha \|\Delta A\|_2, \beta \|\Delta b\|_2)^T$. Since $\|X'\|_2 = \|(\Delta A,\Delta b)\|_2$ we have $\kappa_{g,F}(A,b) \leq \|Y'\|_2$ and from $\|Y'\|_2 \leq \sqrt{2}\|Y\|_2$ and obtain $\kappa_{g,2}(A,b) \leq \sqrt{2} f(A,b)$ which concludes the proof. $\square$

Theorem 2 shows that $f(A,b)$ can be considered as a very sharp estimate of the partial condition number expressed either in Frobenius or spectral norm. Indeed, it lies within a factor $\sqrt{3}$ of $\kappa_{g,F}(A,b)$ or $\kappa_{g,2}(A,b)$. Moreover (7) shows that if the R factor of A is available, $f(A,b)$ can be computed by solving two n-by-n triangular systems with k right-hand sides and thus the computational cost is $2kn^2$.

Remark 2. We can check on the following example that $\kappa_{g,F}(A,b)$ is not equal to $f(A,b)$. Let us consider

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \ L = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } b = \begin{pmatrix} 2/\sqrt{2} \\ 1/\sqrt{2} \\ 1 \end{pmatrix}.$$

We have

$$x = (1/\sqrt{2}, 1/\sqrt{2})^T \text{ and } \|x\|_2 = \|r\|_2 = 1,$$

and we get

$$\kappa_{g,F}(A,b) = \frac{\sqrt{45}}{4} \ < \ f(A,b) = \frac{\sqrt{13}}{2}.$$

Remark 3. Using the definition of the condition number and of the product norms $\|(A,b)\|_{\mathrm{F\ or\ 2}} = \sqrt{\alpha^2 \|A\|_{\mathrm{F\ or\ 2}}^2 + \beta^2 \|b\|_2^2}$, tight estimates for the partial condition number for perturbations of A only (resp. b only) can be obtained by taking $\alpha > 0$ and $\beta = +\infty$ (resp. $\beta > 0$ and $\alpha = +\infty$) in Theorem 2.

4 Statistical estimation of the partial condition number

Let $z_1, z_2, \cdots, z_q$ be q orthonormal vectors uniformly and randomly selected in the unit sphere S_{k-1} in k dimensions ($q \leq k$) and let us denote $g_i(A,b) = (Lz_i)^T x(A,b)$. Since $Lz_i \in \mathbb{R}^n$, the absolute condition number of g_i can be computed via the exact formula given in Corollary 1 i.e

$$\kappa_{g_i,F}(A,b) = \left(\left\| (Lz_i)^T (A^T A)^{-1} \right\|_2^2 \frac{\|r\|_2^2}{\alpha^2} + \left\| (Lz_i)^T A^\dagger \right\|_2^2 \left(\frac{\|x\|_2^2}{\alpha^2} + \frac{1}{\beta^2} \right) \right)^{\frac{1}{2}}. \tag{8}$$

We define the random variable $\phi(q)$ by

$$\phi(q) = (\frac{k}{q} \sum_{i=1}^{q} \kappa_{g_i,F}(A,b)^2)^{\frac{1}{2}}.$$

Let the operator $E(.)$ denote the expected value. The following proposition shows that the root mean squared of $\phi(q)$, defined by $R(\phi(q)) = \sqrt{E(\phi(q)^2)}$ can be considered as an estimate for the condition number of $g(A,b) = L^T x(A,b)$.

Proposition 1. *The absolute condition number can be bounded as follows:*

$$\frac{R(\phi(q))}{\sqrt{k}} \leq \kappa_{g,F}(A,b) \leq R(\phi(q)). \tag{9}$$

Proof. Let vec be the operator that stacks the column of a matrix into a long vector and M be the k-by-$m(n+1)$ matrix such that $vec(g'(A,b).(\Delta A, \Delta b)) = M \begin{pmatrix} vec(\alpha \Delta A) \\ vec(\beta \Delta b) \end{pmatrix}$. Then we have:

$$\begin{aligned} \kappa_{g,F}(A,b) &= \max_{(\Delta A, \Delta b)} \frac{\|g'(A,b).(\Delta A, \Delta b)\|_2}{\|(\Delta A, \Delta b)\|_F} = \max_{(\Delta A, \Delta b)} \frac{\|vec(g'(A,b).(\Delta A, \Delta b))\|_2}{\left\| \begin{pmatrix} vec(\alpha \Delta A) \\ vec(\beta \Delta b) \end{pmatrix} \right\|_2} \\ &= \max_{z \in \mathbb{R}^{m(n+1)}, z \neq 0} \frac{\|M \ z\|_2}{\|z\|_2} = \|M\|_2 = \left\|M^T\right\|_2. \end{aligned}$$

Gudmundsson, Kenney and Laub (1995) prove that $\frac{k}{q} \left\|M^T Z\right\|_F^2$ as an estimator of the Frobenius norm of the $m(n+1)$-by-k matrix M^T where $Z = [z_1, z_2, \cdots, z_q]$ is a k-by-q random matrix with orthonormal columns and show that $E(\frac{k}{q} \left\|M^T Z\right\|_F^2) = \left\|M^T\right\|_F^2$. From

$$\begin{aligned} \left\|M^T Z\right\|_F^2 &= \left\|Z^T M\right\|_F^2 \\ &= \left\| \begin{pmatrix} z_1^T M \\ \vdots \\ z_q^T M \end{pmatrix} \right\|_F^2 \end{aligned}$$

we get, since $z_i^T M$ is a vector,

$$\begin{aligned} \left\|M^T Z\right\|_F^2 &= \sum_{i=1}^{q} \left\|z_i^T M\right\|_2^2 \\ &= \sum_{i=1}^{q} \|g_i'(A,b)\|^2. \end{aligned}$$

Eventually we obtain

$$\|M\|_F^2 = E(\frac{k}{q}\sum_{i=1}^{q}\kappa_{g_i,F}(A,b)^2) = E(\phi(q)^2).$$

Moreover $M^T \in \mathbb{R}^{m(n+1)\times k}$ and Equation 9 follows from the well-known inequality

$$\frac{\|M^T\|_F}{\sqrt{k}} \le \|M^T\|_2 \le \|M^T\|_F.$$

Then we will consider $\phi(q)\frac{\|(A,b)\|_F}{\|L^T\tilde{x}\|_2}$ as an estimator of $\kappa_{g,F}^{(rel)}(A,b)$.

□

The root mean squared of $\phi(q)$ is an upper bound of $\kappa_g(A,b)$, and estimates $\kappa_{g,F}(A,b)$ within a factor $\sqrt{k}$. Proposition 1 leads to computing the condition number of each $g_i(A,b), i = 1,\cdots,q$. From Remark 1, it follows that the computational cost of each $\kappa_{g_i,F}(A,b)$ is $2n^2$ (if the R factor of the QR decompsition of A is available). Hence, for a given sample of vectors z_i, $i = 1,\ldots,q$, computing $\phi(q)$ requires about $2qn^2$ flops.
However, Proposition 1 is mostly of theoretical interest, since it relies on the computation of the root mean squared of a random variable, without providing a practical method to obtain it. The next proposition, by the use of the small sample estimate theory developed by Gudmundsson et al. (1995) answers this question by showing that the evaluation of $\phi(q)$ using only one sample of q vectors $z_1, z_2, \cdots, z_q$ in the unit sphere may provide an acceptable estimate.

Proposition 2. *Using the conjecture by Gudmundsson et al. (1995, p. 781), we have the following result: if $\alpha > 10$, then*

$$Pr\left(\frac{\phi(q)}{\alpha\sqrt{k}} \le \kappa_{g,F}(A,b) \le \alpha\phi(q)\right) \ge 1 - \alpha^{-q}.$$

This probability approaches 1 very fast as q increases For $\alpha = 11$ and $q = 3$ the probability for $\phi(q)$ to estimate $\kappa_{g,F}(A,b)$ within a factor $11\sqrt{k}$ is 99.9%.

Proof. We define as in the proof of Proposition 1 the matrix M as the matrix related to the *vec* operation representing the linear operator $g(A,b)$. From (Gudmundsson et al. 1995, p. 781 and 783) we get

$$Pr\left(\frac{\|M\|_F}{\alpha} \le \phi(q) \le \alpha\|M\|_F\right) \ge 1 - \alpha^{-m}, \tag{10}$$

then the result follows from inequality $\frac{\|M^T\|_F}{\sqrt{k}} \le \|M^T\|_2 \le \|M^T\|_F$. □

We see from this proposition that it may not be necessary to estimate the root mean squared of $\phi(q)$ using sophisticated algorithms. Indeed only one sample of $\phi(q)$ obtained for $q = 3$ provides an estimate of $\kappa_{g,F}(A,b)$ within a factor $\alpha\sqrt{k}$.

Remark 4. If $k = 1$ then $Z = 1$ and the problem is reduced to computing $\kappa_{g_1}(A,b)$ In this case, $\phi(1)$ is exactly the partial condition number of $L^T x(A,b)$.

Remark 5. Concerning the computation of the statistical estimate in the presence of roundoff-errors, the numerical reliability of the statistical estimate relies on an accurate computation of the $\kappa_{g_i,F}(A,b)$ for a given z_i. Let A be a 17-by-13 Vandermonde matrix, b a random vector and $L \in \mathbb{R}^n$ the right singular vector v_n
Using the Mathematica software performing computations in exact arithmetic, we obtained $\kappa_{g,F}^{(rel)}(A,b) \approx 5\cdot 10^8$. If the triangular factor R form $A^TA = R^TR$ is obtained by the QR decomposition of A, we get $\kappa_{g,F}^{(rel)}(A,b) \approx 5\cdot 10^8$. If R is computed via a classical Cholesky factorization, we get $\kappa_{g,F}(A,b)^{(rel)} \approx 10^{10}$. Corollary 1 and Remark 1 show that the computation of $\kappa_{g,F}(A,b)^{(rel)}$ involves linear systems of the kind $A^TAx = d$, which differs from the usual normal equation in their right-hand side. Our observation that for this kind of ill-conditioned systems, a QR factorization is more accurate than a Cholesky factorization is in agreement with the results of Frayssé, Gratton and Toumazou (2000).

5 Numerical experiments

All experiments were performed in Matlab 6.5 using a machine precision $2.22 \cdot 10^{-16}$.

5.1 Examples

For the examples of Section 1, we compute the partial condition number using the formula given in Theorem 1.
In the first example we have

$$A = \begin{pmatrix} 1 & 1 & \epsilon^2 \\ \epsilon & 0 & \epsilon^2 \\ 0 & \epsilon & \epsilon^2 \\ \epsilon^2 & \epsilon^2 & 2 \end{pmatrix}$$

and we assume that only A is perturbed. If we consider the values for L that are $\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $L = (0,0,1)^T$ then we obtain partial condition numbers $\kappa_{g,F}^{(rel)}(A)$ that are respectively 10^{24} and 1.22, as expected since there is 50% relative error on x_1 and x_2 and there is no error on x_3.

In the second example where A is the $10-by-4$ Vandermonde matrix defined by $A_{ij} = \frac{1}{(10+i)^{j-1}}$ and only b is perturbed, the partial condition numbers $\kappa_{g,F}^{(rel)}(b)$ with respect to each component x_1, x_2, x_3, x_4 are respectively $4.5 \cdot 10^2, 2 \cdot 10^4, 3 \cdot 10^5, 1.4 \cdot 10^6$ which is consistent with the error variation given in Section 1 for each component.

5.2 Average behaviour of the statistical estimate

We compare here the statistical estimate described in the previous section to the partial condition number obtained via the exact formula given in Theorem 1. We suppose that only A is perturbed and then the partial condition number can be expressed as $\kappa_{g,F}^{(rel)}(A)$. We use the method described by Paige and Saunders (1982) in order to construct test problems $[A, x, r, b] = P(m, n, n_r, l)$ with

$$A = Y \begin{pmatrix} D \\ 0 \end{pmatrix} Z^T \in \mathbb{R}^{m \times n}, Y = I - 2yy^T, Z = I - 2zz^T,$$

where $y \in \mathbb{R}^m$ and $z \in \mathbb{R}^n$ are random unit vectors and $D = n^{-l} diag(n^l, (n-1)^l, \cdots, 1)$.
$x = (1, 2^2, \cdots, n^2)^T$ is given and $r = Y \begin{pmatrix} 0 \\ c \end{pmatrix} \in \mathbb{R}^m$ is computed with $c \in \mathbb{R}^{m-n}$ random vector of norm n_r. The right-hand side is $b = Y \begin{pmatrix} DZx \\ c \end{pmatrix}$. By construction, the condition number of A and D is n^l.

In our experiments, we consider the matrices

$$A = \begin{pmatrix} A_1 & E' \\ E & A_2 \end{pmatrix} \text{ and } L = \begin{pmatrix} I \\ 0 \end{pmatrix},$$

where $A_1 \in \mathbb{R}^{m_1 \times n_1}$, $A_2 \in \mathbb{R}^{m_2 \times n_2}$, $L \in \mathbb{R}^{n \times n_1}$, $m_1 + m_2 = m$, $n_1 + n_2 = n$, and E and E' contain the same element e_p which defines the coupling between A_1 and A_2.
A_1 and A_2 are randomly generated using respectively $P(m_1, n_1, n_{r_1}, l_1)$ and $P(m_2, n_2, n_{r_2}, l_2)$.
For each sample matrix, we compute in Matlab:

1. the partial condition number $\kappa_{g,F}^{(rel)}(A)$ using the exact formula given in Theorem 1 and based on the singular value decomposition of A,
2. the statistical estimate $\phi(3)$ using three random orthogonal vectors and computing each $\kappa_{g_i,F}(A,b)$, $i=1,2,3$ with the R factor of the QR decomposition of A.

These data are then compared by computing the ratio $\gamma = \frac{\phi(3)}{\kappa_{g,F}^{(rel)}(A)}$.
Table 1 contains the mean $\overline{\gamma}$ and the standard deviation s of γ obtained on 1000 random matrices with $m_1 = 12, n_1 = 10, m_2 = 17, n_2 = 13$ by varying the condition numbers ${n_1}^{l_1}$ and ${n_2}^{l_2}$ of respectively A_1 and A_2 and the coupling coefficient e_p. The residual norms are set to $n_{r_1} = n_{r_2} = 1$. In all cases, $\overline{\gamma}$ is close to 1 and s is about 0.3. The statistical estimate $\phi(3)$ lies within a factor 1.22 of $\kappa_{g,F}^{(rel)}(A)$ which is very accurate in condition number estimation. We notice that in two cases, $\phi(3)$ is lower than 1. This is possible because Proposition 1 shows that $E(\phi(3)^2)$ is an upper bound of $\kappa_{g,F}(A)^2$ but not necessarily $\phi(3)^2$.

condition		$e_p = 10^{-5}$		$e_p = 1$		$e_p = 10^5$	
l_1	l_2	$\overline{\gamma}$	s	$\overline{\gamma}$	s	$\overline{\gamma}$	s
1	1	1.22	$2.28 \cdot 10^{-1}$	1.15	$2.99 \cdot 10^{-1}$	1.07	$3.60 \cdot 10^{-1}$
1	8	1.02	$3.19 \cdot 10^{-1}$	1.22	$3.05 \cdot 10^{-1}$	1.21	$3.35 \cdot 10^{-1}$
8	1	$9 \cdot 10^{-1}$	$3 \cdot 10^{-1}$	1.13	$3 \cdot 10^{-1}$	1.06	$3.45 \cdot 10^{-1}$
8	8	$9.23 \cdot 10^{-1}$	$2.89 \cdot 10^{-1}$	1.22	$2.95 \cdot 10^{-1}$	1.18	$3.33 \cdot 10^{-1}$

Table 1: Ratio between statistical and exact condition number of $L^T x$.

6 Estimates vs exact formula

We assume that the R factor of the QR decomposition of A is known. We gather in Table 2 the results obtained in this paper in terms of accuracy and flops counts for the estimation of the partial condition number for the LLSP. Table 3 shows the actual estimates and flops counts in the particular situation where

$\kappa_{g,F}(A,b)$	flops	accuracy
exact formula $n \ll m$	$12n^3$	exact
sharp estimate $f(A,b)$ $k \ll n$	$2kn^2$	$\frac{f(A,b)}{\sqrt{2}} \leq \kappa_{g,F}(A,b) \leq f(A,b)$
stat. estimate $\phi(q)$ $q \ll k$	$2qn^2$	$\frac{\phi(q)}{\alpha\sqrt{k}} \leq \kappa_{g,F}(A,b) \leq \alpha\phi(q)$ $Pr \geq 1 - \alpha^{-q}$ for $\alpha > 10$

Table 2: Comparison between exact formula and estimates for $\kappa_{g,F}(A,b)$

$$m = 1500,\ n = 1000,\ k = 50,$$

$$A_1 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix},\ L_1 = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix},$$

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & I_{n-2} \\ 0 & 0 \end{pmatrix} \text{ and } b = \frac{1}{\sqrt{2}}(2, 1, \cdots, 1)^T, \; L = \begin{pmatrix} L_1 & 0 \\ 0 & I_{k-2} \\ 0 & 0 \end{pmatrix}.$$

We see here that the statistical estimates may provide information on the condition number using a very small amount of floating point operations compared to the two other methods.

$\kappa_{g,F}^{(rel)}(A,b)$	$f(A,b)\frac{\lVert(A,b)\rVert_F}{\lVert L^T\tilde{x}\rVert_2}$	$\phi(q)\frac{\lVert(A,b)\rVert_F}{\lVert L^T\tilde{x}\rVert_2}$
$2.09 \cdot 10^2$	$2.18 \cdot 10^2$	$11.44 \cdot 10^2$
12 Gflops	100 Mflops	6 Mflops

Table 3: Flops and accuracy : exact formula vs estimates

7 Conclusion

We have shown the relevance of the partial condition number shown for test cases from parameter estimation. This partial condition number evaluates the sensitivity of $L^T x$ where x is the solution of a (LLSP) when A and/or b are perturbed. It can be computed via a close formula, a sharp estimate or a statistical estimate. The quantity to compute depends on the size of the LLSP (computational cost) and on the needed accuracy.

References

Björck, Å. (1996), *Numerical Methods for Least Squares Problems*, SIAM.

Cao, Y. and Petzold, L. (2003), 'A subspace error estimate for linear systems', *SIAM Matrix Analysis and Applications* **24**, 787–801.

Chandrasekaran, S. and Ipsen, I. C. F. (1995), 'On the sensitivity of solution components in linear systems of equations', *Numerical Linear Algebra with Applications* **2**, 271–286.

Eldén, L. (1980), 'Perturbation theory for the least squares problem with linear equality constraints', *SIAM J. Numer. Anal.* **17**(3), 338–350.

Frayssé, V., Gratton, S. and Toumazou, V. (2000), 'Structured backward error and condition number for linear systems of the type $A^*Ax = b$', *BIT* **40**, 74–83.

Geurts, A. J. (1982), 'A contribution to the theory of condition', *Numerische Mathematik* **39**, 85–96.

Golub, G. and van Loan, C. (1996), *Matrix Computations*, The Johns Hopkins University Press. Third edition.

Gratton, S. (1996), 'On the condition number of linear least squares problems in a weighted frobenius norm', *BIT* **36**, 523–530.

Gudmundsson, T., Kenney, C. S. and Laub, A. J. (1995), 'Small-sample statistical estimates for matrix norms', *SIAM Matrix Analysis and Applications* **16**, 776–792.

Higham, N. (2002), *Accuracy and Stability of Numerical Algorithms*, SIAM. Second edition.

Kaplan, E. D. (1996), *Understanding GPS : Principles and Applications*, Artech House Publishers, Boston.

Kenney, C. S. and Laub, A. J. (1994), 'Small-sample statistical condition estimates for general matrix functions', *SIAM J. Sci. Comput.* **15**, 36–61.

Kenney, C. S., Laub, A. J. and Reese, M. S. (1998), 'Statistical condition estimation for linear least squares', *SIAM Matrix Analysis and Applications* **19**, 906–923.

Paige, C. C. and Saunders, M. A. (1982), 'LSQR: An algorithm for sparse linear equations and sparse least squares', *ACM Trans. Math. Software* **8**, 43–71.

Rao, C. R. and Mitra, S. K. (1971), *Generalized Inverse of Matrices and Its Applications*, Wiley, New York.

Stewart, G. W. and Sun, J. (1991), *Matrix Perturbation Theory*, Academic Press, New York.

Wei, Y., Diao, H. and Qiao, S. (2004), Condition number for weighted linear least squares problem and its condition number, Technical Report CAS 04-02-SQ, Department of Computing and Software, McMaster University, Hamilton, Ontario, Canada.

Wei, Y., Xu, W., Qiao, S. and Diao, H. (2003), Componentwise condition numbers for generalized matrix inversion and linear least squares, Technical Report CAS 03-12-SQ, Department of Computing and Software, McMaster University, Hamilton, Ontario, Canada.