Interval Computations
No 4, 1994

Interval Methods and Condition Numbers of Linear Algebraic Systems

Søren Christiansen

The sensitivity of linear systems is often expressed in terms of condition numbers, e.g., the ordinary, the effective, or the local condition number, all of which are computed via singular value decomposition. We express the sensitivity of linear systems by using interval methods, and by means of experiments using three particularly suitable systems we indicate which condition numbers are relevant for creating a relation to the sensitivity using interval methods.

Интервальные методы и числа обусловленности линейных алгебраических систем

С. Христиансен

Чувствительность линейных систем часто выражается в терминах чисел обусловленности, в частности, обыкновенных, эффективных и локальных чисел обусловленности, для получения которых применяется сингулярное разложение матрицы. Мы даем выражение для чувствительности линейной системы интервальными методами и посредством экспериментов с тремя специально подобранными системами показываем, какие из чисел обусловленности более подходят для соотношений чувствительности, использующих интервальную технику.

1 Introduction

When interval methods are used to treat systems of linear algebraic equations, a fundamental feature is that the solution obtained really reflects the sensitivity of the system with respect to perturbations of the given data—both the matrix and the right-hand side. This sensitivity can also be expressed in terms of various condition numbers of the system. The purpose of the present note is to point out some connections between the two subjects. It is done by experiments with three particularly suitable systems.

Condition numbers can be expressed via the Singular Value Decomposition (Section 2), and analogous quantities are mimicked by means of Interval Methods (Section 3). For three particularly suitable systems (Section 4) actual numerical experiments are carried out (Section 5) leading to some conclusions (Section 6) as to the connection between the sensitivity determined by interval methods and the condition numbers.

2 Condition numbers and sensitivity

For a system of linear algebraic equations

$$\underline{\underline{A}}\underline{x} = \underline{b} \tag{2–1}$$

where $\underline{\underline{A}}$ is an $n \times n$-matrix, while $\underline{x}$ and $\underline{b}$ are n-vectors, condition numbers can be expressed via the Singular Value Decomposition (SVD) [1, Section 2.5], which gives

$$\underline{\underline{A}} = \sum_{j=1}^{n} \sigma_j \underline{u}_j \underline{v}_j^T \tag{2–2}$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ are the singular values of $\underline{\underline{A}}$, from which is obtained

$$\underline{b} = \sum_{j=1}^{n} \beta_j \underline{u}_j, \quad \beta_j = \underline{u}_j^T \underline{b}, \qquad \|\underline{b}\| = \left(\sum_{j=1}^{n} \beta_j^2 \right)^{\frac{1}{2}}, \tag{2–3a,b,c}$$

$$\underline{x} = \sum_{j=1}^{n} \xi_j \underline{v}_j, \quad \xi_j = \beta_j / \sigma_j, \quad \|\underline{x}\| = \left(\sum_{j=1}^{n} \xi_j^2 \right)^{\frac{1}{2}}. \tag{2–4a,b,c}$$

Here and subsequently the 2-norm of vectors, the spectral norm, is indicated by $\|\underline{\bullet}\|$, while the subordinate 2-norm of matrices will be indicated by $\|\underline{\underline{\bullet}}\|$.

When $\underline{\underline{A}}$ or $\underline{b}$ are perturbed by $\delta\underline{\underline{A}}$ or $\delta\underline{b}$, the solution $\underline{x}$ is perturbed by $\delta\underline{x}$, where

$$(\underline{\underline{A}} + \delta\underline{\underline{A}})(\underline{x} + \delta\underline{x}) = \underline{b}, \tag{2–5a}$$

$$\underline{\underline{A}}(\underline{x} + \delta\underline{x}) = \underline{b} + \delta\underline{b}. \tag{2–5b}$$

The corresponding sensitivity (for the solution $\underline{x}$) can be expressed in terms of condition numbers

$$\frac{\|\delta\underline{x}\|}{\|\underline{x}\|} \lesssim \kappa_A \frac{\|\delta\underline{\underline{A}}\|}{\|\underline{\underline{A}}\|}, \tag{2–6a}$$

$$\frac{\|\delta\underline{x}\|}{\|\underline{x}\|} \lesssim \kappa_b \frac{\|\delta\underline{b}\|}{\|\underline{b}\|}. \tag{2–6b}$$

In the literature several expressions are given for condition numbers; here three are expressed in terms of quantities from the SVD:

(i) The ordinary condition number [1, p. 80]:

$$\kappa_{ord} = \kappa_{ord}(\underline{\underline{A}}) = \frac{\sigma_1}{\sigma_n}. \tag{2–7a}$$

(ii) The effective condition number [2, p. 965]:

$$\kappa_{\mathit{eff}} = \kappa_{\mathit{eff}}(\underline{\underline{A}}, \underline{b}) = \frac{\|\underline{b}\|}{\sigma_n} \min_k \frac{\sigma_k}{(\beta_k^2 + \beta_{k+1}^2 + \cdots + \beta_n^2)^{\frac{1}{2}}}. \tag{2-7b}$$

(iii) The local condition number [3, p. 936] (or natural condition number [4, p. 164]):

$$\kappa_{loc} = \kappa_{loc}(\underline{\underline{A}}, \underline{b}) = \frac{\|\underline{b}\|}{\sigma_n} \frac{1}{\|\underline{x}\|}. \tag{2-7c}$$

The condition numbers (2–7) (which do not all fit exactly into (2–6), because (2–6) only expresses some principal relations) can be used as follows in (2–6).

In (2–6a), viz. for perturbations on $\underline{\underline{A}}$: (i) is applicable [5, p. 271], (ii) is applicable for very restricted types of $\delta\underline{\underline{A}}$ [2, p. 967], (iii) does not seem applicable.

In (2–6b), viz. for perturbations on $\underline{b}$: (i) is applicable [5, p. 271], (ii) is applicable [2, p. 965], (iii) is applicable [4, p. 164].

3 Interval methods and sensitivity

The two expressions (2–6), involving the condition numbers, lead us to consider the two quantities

$$k_A := \frac{\|\delta \underline{x}\|/\|\underline{x}\|}{\|\delta \underline{\underline{A}}\|/\|\underline{\underline{A}}\|}, \tag{3–1a}$$

$$k_b := \frac{\|\delta \underline{x}\|/\|\underline{x}\|}{\|\delta \underline{b}\|/\|\underline{b}\|}. \tag{3–1b}$$

For $\underline{\underline{A}}$ and $\underline{b}$ given, it is possible to compute the two quantities (3–1) by choosing $\delta\underline{\underline{A}}$ and $\delta\underline{b}$, respectively. It is further possible to carry out "many" such computations, and thereby trying to determine approximations to the maximal values of the scalar quantities (3–1). But instead of performing "many" experiments using point values, it is the purpose here to use interval methods and to carry out one or two calculations, which shall replace the "many" experiments. The question arises: when (3–1) are computed using interval methods, do we then obtain numbers which resembles the condition numbers obtained by SVD ?

We will in the computation of the quantities (3–1) replace the perturbations $\delta\underline{\underline{A}}$ and $\delta\underline{b}$ by intervals. We indicate intervals by square brackets [ ], and let $[I] := [-\Delta, +\Delta]$ be a (symmetric) scalar interval with midpoint zero and radius Δ. The perturbations on $\underline{\underline{A}}$ and $\underline{b}$ are chosen as follows: On $\underline{\underline{A}}$ is applied one perturbation interval matrix

$$[\delta A] := [I]\underline{\underline{A}} \tag{3–2}$$

which is obtained by multiplying all the point elements of $\underline{\underline{A}}$ by $[I]$, giving a relative perturbation. On $\underline{b}$ is applied two perturbation interval vectors

$$[\delta \underline{b}] := [I]\{11\ldots1\}^T, \tag{3–3a}$$

$$[\delta \underline{b}] := [I]\underline{b} \tag{3–3b}$$

where in (3–3a) all the elements are equal to $[I]$, giving an absolute perturbation, while in (3–3b) all the point elements of $\underline{b}$ are multiplied by $[I]$, giving a relative perturbation. (For $\underline{\underline{A}}$ it is necessary to introduce a relative perturbation in (3–2) in order to avoid the computation of $\|\underline{\underline{A}}\| = \sigma_1$ via SVD for use in (3–7). For $\underline{b}$ there is the freedom to choose between absolute

and relative perturbation, because $\|\underline{b}\|$ is easily determined in (3–8), but the relative perturbation leads to the simplest expression in (3–10).)

Instead of considering (2–1), the following interval systems are formed, with $[\delta\underline{\underline{A}}]$ and $[\delta\underline{b}]$ from (3–2) and (3–3), while $[\underline{\underline{A}}]$ and $[\underline{b}]$ indicates the closest possible intervals enclosing $\underline{\underline{A}}$ and $\underline{b}$,

$$([\underline{\underline{A}}] + [\delta\underline{\underline{A}}])[\underline{x}] = [\underline{b}] \tag{3–4}$$

and

$$\begin{aligned} \underline{\underline{A}}[\underline{x}] &= [\underline{b}] + [\delta\underline{b}], & \text{(3–5a)} \\ [\underline{\underline{A}}][\underline{x}] &= [\underline{b}] + [\delta\underline{b}] & \text{(3–5b)} \end{aligned}$$

with the solution interval vector $[\underline{x}]$. From $[\underline{x}]$ two point vectors are determined

$$\begin{aligned} \check{\underline{x}} &:= \big(\sup([\underline{x}]) + \inf([\underline{x}])\big)/2, & \text{(3–6a)} \\ \delta\hat{\underline{x}} &:= \big(\sup([\underline{x}]) - \inf([\underline{x}])\big)/2 & \text{(3–6b)} \end{aligned}$$

which play the role of $\underline{x}$ and $\delta\underline{x}$, respectively, in (3–1), where the corresponding vector 2-norms are used.

Having determined the numerators of (3–1) there remains a problem in the denominators, namely how to choose suitable replacements for $\|\delta\underline{\underline{A}}\|$ and $\|\delta\underline{b}\|$ in terms of the interval matrix $[\delta\underline{\underline{A}}]$ and the interval vector $[\delta\underline{b}]$, because $[\delta\underline{\underline{A}}]$ and $[\delta\underline{b}]$ encompass the zero point matrix and the zero point vector, respectively, which have the norm zero. Therefore one may consider $\|[\delta\underline{\underline{A}}]\| = 0$ and $\|[\delta\underline{b}]\| = 0$ as possibilities, which, when used in (3–1) instead of $\|\delta\underline{\underline{A}}\|$ and $\|\delta\underline{b}\|$, will lead to useless results. But if we pick out the single value Δ from the interval $[I]$, we do not have to consider interval values, and we thereby get for the norm of the perturbation $\delta\underline{\underline{A}}$, cf. (3–2),

$$\|\delta\underline{\underline{A}}\| = \|\Delta\underline{\underline{A}}\| = \Delta\|\underline{\underline{A}}\| \tag{3–7}$$

and for the norm of the two perturbations $\delta\underline{b}$, cf. (3–3),

$$\begin{aligned} \|\delta\underline{b}\| &= \|\underline{\Delta}\|, & \text{(3–8a)} \\ \|\delta\underline{b}\| &= \|\Delta\underline{b}\| = \Delta\|\underline{b}\| & \text{(3–8b)} \end{aligned}$$

where $\underline{\Delta} = \{\Delta\Delta\ldots\Delta\}^T$. The values (3–7) and (3–8) are used in (3–1).

Therefore, instead of k_A and k_b in (3–1), we are now led to consider the two quantities

$$K_A := \frac{\|\delta\underline{\hat{x}}\|/\|\underline{\check{x}}\|}{\Delta}, \tag{3–9}$$

$$K_b := \begin{cases} \dfrac{\|\delta\underline{\hat{x}}\|/\|\underline{\check{x}}\|}{\|\underline{\Delta}\|/\|\underline{b}\|}, & \text{(3–10a)} \\ \dfrac{\|\delta\underline{\hat{x}}\|/\|\underline{\check{x}}\|}{\Delta} & \text{(3–10b)} \end{cases}$$

where all the norms are 2-norms of point vectors introduced above, and derived from results of interval calculations.

The quantities K_A and K_b are assumed to represent intrinsic properties of the linear system, viz., the sensitivity of the solution to (small) perturbations on $\underline{\underline{A}}$ or $\underline{b}$, respectively. They are defined as ideal quantities derived from an ideal interval solution vector $[\underline{x}]$. But in practice the quantities are obtained by actual computations, and it is therefore unavoidable that they also contain a certain contribution, which originates from the actual numerical method of the actual system for performing the interval computations. It is therefore necessary to know how much of the quantities K_A and K_b is due to computational error before final conclusions can be drawn. In order to estimate the magnitude of this unavoidable contribution, the interval solution method in question is applied to linear systems of the form, cf. (3–5),

$$\underline{\underline{A}}[\underline{x}] = [\underline{b}], \tag{3–11a}$$

$$[\underline{\underline{A}}][\underline{x}] = [\underline{b}] \tag{3–11b}$$

with either a point matrix $\underline{\underline{A}}$ or an interval matrix $[\underline{\underline{A}}]$ (which is the closest possible interval enclosing $\underline{\underline{A}}$) and an interval right hand side $[\underline{b}]$ (which is the closest possible interval enclosing $\underline{b}$). These systems are solved by means of the interval system in question, which gives the computed interval solution vector $[\underline{x}]$, from which is again determined $\underline{\check{x}}$ and $\delta\underline{\hat{x}}$, cf. (3–6), and the ratio

$$U := \|\delta\underline{\hat{x}}\|/\|\underline{\check{x}}\|. \tag{3–12}$$

This ratio, which resembles the quantities K_A and K_b, cf. (3–9) and (3–10), should tell about the intrinsic uncertainty of the interval solution method, because no extra external uncertainty is introduced into the linear system.

4 Systems of equations

The quantities K_A, K_b, and U are to be compared with the condition numbers κ_{ord}, κ_{eff}, and κ_{loc}. This is done by experiments with three particularly suitable systems, each one having a matrix which depends strongly on a real parameter; the matrices are introduced below (Subsection 4.1, 4.2, and 4.3). The condition number κ_{ord} is determined solely from the matrix $\underline{\underline{A}}$; contrary to this the condition numbers κ_{eff} and κ_{loc} also depend on the actual right-hand side $\underline{b}$. We therefore have to introduce some right-hand side vectors, and the following turns out to be of interest:

$$\underline{b}_\alpha = \{1 \ldots 1\}^T, \tag{4–1a}$$

$$\underline{b}_\beta = \{+1 \; -1 \; +1 \; \ldots \; -1\}^T, \tag{4–1b}$$

$$\underline{b}_\gamma = \{0 \ldots 0 \, \frac{1}{2} \, 1 \ldots 1 \, \frac{1}{2}\}^T. \tag{4–1c}$$

The vector $\underline{b}_\gamma$ is only used when n is even in which case the number of zero's and the number of one's are equal.

The right-hand sides (4–1) and the three matrices given below, (4–2), (4–5), and (4–8), are all easy to construct. The matrices all depend strongly on a real parameter d, i.e., $\underline{\underline{A}} = \underline{\underline{A}}(d)$, and it turns out that they have interesting features concerning how the various condition numbers depend on d and n, for various right-hand sides; the matrices may all become bad-conditioned by a suitable choice of d. Thus it is possible to draw conclusions based on a qualitative behaviour of the varions condition numbers, as functions of d or n, in order to find which condition numbers are of relevance. A quantitative comparison of condition numbers for fixed values of d and n will not be informative; it may even lead to careless conclusions.

4.1 Matrix I

The matrix stems from a numerical solution of a linear boundary integral equation of the first kind with logarithmic kernel, where the boundary is a smooth, closed curve. Several physical problems can be formulated in terms of such equations as mentioned in [6]. When the curve is described parametrically as $z = z(t)$, $0 \le t \le 1$, where z is a complex variable, and when the points $z_j = z(t_j)$, with $z'_j = z'(t_j)$, $t_j = j/n$, $j = 1, 2, \ldots, n$, are

picked out, the elements of the matrix $\underline{\underline{A}}$ are

$$a_{jk} = \left\{ \begin{array}{ll} -\frac{2}{n} \ln|z_j - z_k|; & j \neq k \\ -\frac{2}{n} \ln\left|\frac{z'_j}{2\pi n}\right|; & j = k \end{array} \right\}. \tag{4–2}$$

The formula (4–2) is given in [7], and is derived from [8] and also [6]. As boundary curve is here used an ellipse with semiaxes a and b, with the axis-ratio $\varepsilon = b/a$, while $d = (a+b)/2$, $d > 0$, gives the size of the ellipse

$$z(t) = \frac{2d}{1+\varepsilon}\big(\cos(2\pi t) + i\varepsilon \sin(2\pi t)\big); \quad 0 \leq t \leq 1. \tag{4–3}$$

The matrix (4–2) is seen to be real and symmetric. If the matrix is derived from the curve (4–3) with $\varepsilon = 1$ it is furthermore circulant, and has the eigenvalues [7]

$$\lambda_k = -2\ln(d); \quad k = 0, \tag{4–4a}$$

$$\frac{1}{|k|} < \lambda_k \leq \frac{2\ln(2)}{|k|}; \quad k \neq 0, \quad -\frac{n}{2} < k \leq \frac{n}{2}. \tag{4–4b}$$

It turns out that $\underline{\underline{A}}$ is singular for d very near 1, and that it tends to be singular for $d \to 0$ or $d \to \infty$. For $\varepsilon = 1$ we see from (4–4) that the matrix $\underline{\underline{A}}$ is singular for $d = 1$, exactly. Just by choosing d suitably it is possible to get a matrix as bad conditioned as wanted, in that κ_{ord} will tend to infinity for $d \to 0$, $d \to 1$, or $d \to \infty$. For this matrix and with the right-hand sides (4–1) interesting results are available from a detailed investigation [7]: In many cases κ_{eff} and κ_{loc} are identical, but it is possible to device such a $\underline{b}$ that κ_{eff} and κ_{loc} are distinct, both as functions of d and also as functions of n. In particular for $\underline{b} = \underline{b}_\gamma$, $\kappa_{\mathit{eff}} = O(n)$, $\kappa_{loc} = O(n^{\frac{1}{2}})$ (this result is due to Professor Jukka Saranen, Oulu, Finland). It is thereby possible to distinguish between these two condition numbers in a qualitative manner.

4.2 Matrix II

From [9; p. 10] we are led to derive the tridiagonal, non-symmetric matrix

$$\underline{\underline{A}} = \begin{bmatrix} 1 & d & 0 & \cdot & \cdot & \cdot \\ \varepsilon d & 1 & d & 0 & \cdot & \cdot \\ 0 & \varepsilon d & 1 & d & 0 & \cdot \\ \cdot & 0 & \varepsilon d & 1 & d & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 & \varepsilon d & 1 \end{bmatrix}_{n \times n} \tag{4–5}$$

where d and ε are real; for $\varepsilon = 1$ the matrix is symmetric. The matrix has the eigenvalues

$$\lambda_k = 1 + 2d\sqrt{\varepsilon}\cos\frac{k\,\pi}{n+1}; \quad k = 1, 2, \ldots, n \tag{4–6}$$

which result for $\varepsilon = 1$ coincides with [9; p. 10]. A special case of the matrix, viz. $\varepsilon = 1$ and $d = -1/2$, appears when desribing small movements of identical particles placed equidistantly on an eleatic string.

From (4–6) we see:

- For $\varepsilon \leq 0$: $\lambda_k \neq 0$ for all d and k, and $\underline{\underline{A}}$ is non-singular, but may become ill-conditioned for $n \to \infty$.

- For $\varepsilon \geq 0$:

  (1) When $0 \leq |d| < \tilde{d} := 1/(2\sqrt{\varepsilon})$: $\lambda_k \neq 0$ for k, and $\underline{\underline{A}}$ is non-singular, but may become ill-conditioned for $n \to \infty$, but κ_{ord} will behave smoothly for $n \to \infty$.

  (2) When $\varepsilon = 1$ (and $0 \leq |d| < 1/2$):

  $$\kappa_{ord} = \max_k |\lambda_k| / \min_k |\lambda_k| \to (1 + 2|d|)/(1 - 2\|d\|) \quad \text{for } n \to \infty. \tag{4–7}$$

  (3) When $\tilde{d} < |d|$: $\lambda_k = 0$ for n or $n-1$ values of d, opposite in pairs, so that $\underline{\underline{A}}$ is singular for these values of d; for $n \to \infty$ (and a fixed value of d) κ_{ord} will behave non-smoothy because the matrix will become (nearly) singular for various values of n.

4.3 Matrix III

From [9; p. 77, Example 4.20] we are led to derive the pentadiagonal, non-symmetric matrix

$$\underline{\underline{A}} = \begin{bmatrix} -1 & d & 1 & 0 & \dots & & & \\ \varepsilon d & 0 & d & 1 & 0 & \dots & & \\ 1 & \varepsilon d & 0 & d & 1 & 0 & \dots & \\ 0 & 1 & \varepsilon d & 0 & d & 1 & 0 & \dots \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ & \dots & 0 & 1 & \varepsilon d & 0 & d & 1 \\ & & \dots & 0 & 1 & \varepsilon d & 0 & d \\ & & & \dots & 0 & 1 & \varepsilon d & -1 \end{bmatrix}_{n \times n} \tag{4–8}$$

where d and ε are real; for $\varepsilon = 1$ the matrix is symmetric and has the eigenvalues [9; p. 77, Example 4.20]

$$\lambda_k = 2 \left(2 \cos^2 \frac{k\,\pi}{n+1} - 1 - d \cos \frac{k\,\pi}{n+1} \right); \quad k = 1, 2, \dots, n. \tag{4–9}$$

From (4–9) we see that $\lambda_k = 0$ for n or $n-1$ values of d, opposite in pairs, so that $\underline{\underline{A}}$ is singular for these values of d; for $n \to \infty$ (for a fixed value of d) κ_{ord} will behave non-smoothly because the matrix will become (nearly) singular for various values of n.

5 Numerical experiments

We will try to draw some conclusions from experiments which we shall carry out with the systems of linear algebraic equations introduced in Section 4, namely by combining the three different sets of matrices, (4–2), (4–5), and (4–8), with the three different right-hand sides, (4–1). On the one hand we compute the three condition numbers, κ_{ord}, κ_{eff}, and κ_{loc}, (2–7), via a singular value decomposition (SVD) by means of point value methods applied to the system (2–1). On the other hand we apply interval value methods to the systems (3–4), (3–5), or (3–11), with the perturbations (3–7) or (3–8), with $\Delta = 10^{-6}$, in order to obtain K_A, (3–9), K_b, (3–10), and U, (3–12). The experiments, which are all carried out with point value methods and interval value methods in pairs, are described in the following

three subsections, which correspond to the three subsections of Section 4 for the three sets of matrices.

5.1 Experiments with matrix I

We here use an IBM 3090 mainframe to treat the system with the matrix (4–2) and (4–3) (where we use $\varepsilon = 1$ for simplicity), and the right-hand sides (4–1). For the point value computations we here use FORTRAN [10], double precision, and find the SVD by means of [11]; for the interval value computations we here use the language ACRITH–XSC [12, 13] , double precision (which is convenient and seemingly accurate, but unfortunately only available on some IBM mainframes). For the perturbation $\delta\underline{b}$ we here use the absolute one (3–3a) which leads to the quantity K_b in the less simple form (3–10a). Because ACRITH–XSC can solve linear algebraic equations with point value matrix (and interval value right-hand side) we solve the equations as follows: Equation (3–4), for the determination of K_A, using DILIN0 [13, p. C–10] for interval value matrices; equations (3–5a) and (3–11a), for the determination of K_b and U, respectively, using DLIN0 [13, p. C–9] for point value matrices. From the various interval value solutions the three ratios (3–9), (3–10a), and (3–12) are computed using [14]. Below we give some examples where the various computed quantities are shown graphically as function of d or n, using the convenient and flexible plotting system AMFPLOT [15]. The choice of the parameters used is guided by the results of the investigation [7] presented in Section 4.1.

Example I–1. With $\underline{b} = \underline{b}_\gamma$, (4–1c), and n fixed, we have [7] for $d \to 1$ that $\kappa_{ord} \to \infty$ while $\kappa_{\mathit{eff}} = \mathrm{O}(1)$ and $\kappa_{loc} = \mathrm{O}(1)$. With $n = 12$ the quantities κ_{ord}, κ_{eff}, κ_{loc} and K_A, K_b are computed as functions of d for $0 < d \leq 2.0$ and shown in Figure I–1. We notice that K_A resembles κ_{ord}, while K_b resembles κ_{eff} or κ_{loc}, but that it is not possible to decide clearly between κ_{eff} or κ_{loc}. In the next Example an experiment is carried out, which makes it possible to decide.

Example I–2. With $\underline{b} = \underline{b}_\gamma$, (4–1c), and $d > 1$ fixed, we have [3] for $n \to \infty$ that $\kappa_{ord} = \mathrm{O}(n)$ and $\kappa_{\mathit{eff}} = \mathrm{O}(n)$ while $\kappa_{loc} = \mathrm{O}(n^{\frac{1}{2}})$. With $d = 2.0$ the quantities κ_{ord}, κ_{eff}, κ_{loc} and K_A, K_b are computed as functions of n for $4 \leq n \leq 40$ and shown in Figure I–2. We notice that K_A resembles κ_{ord} (as seen in Example I–1), while K_b resembles κ_{loc} and not κ_{eff}.

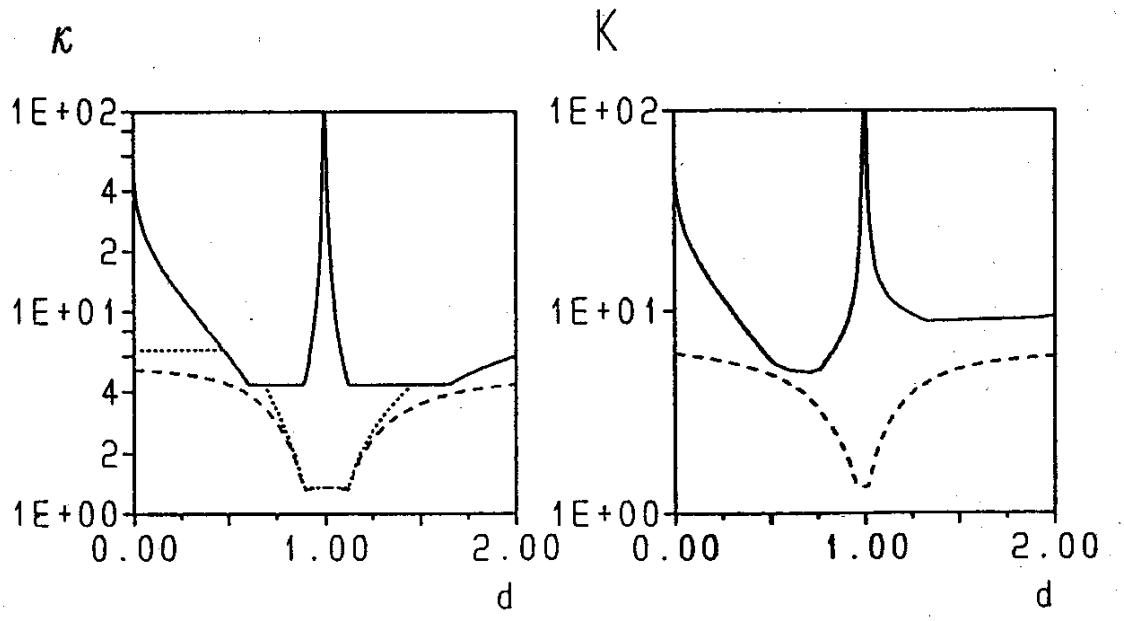

Figure I–1: See Example I–1. As functions of d are shown: left picture: κ_{ord} (———), κ_{eff} $(\cdot\cdot\cdot\cdot)$, κ_{loc} $(---)$; right picture: K_A (———), K_b $(---)$.

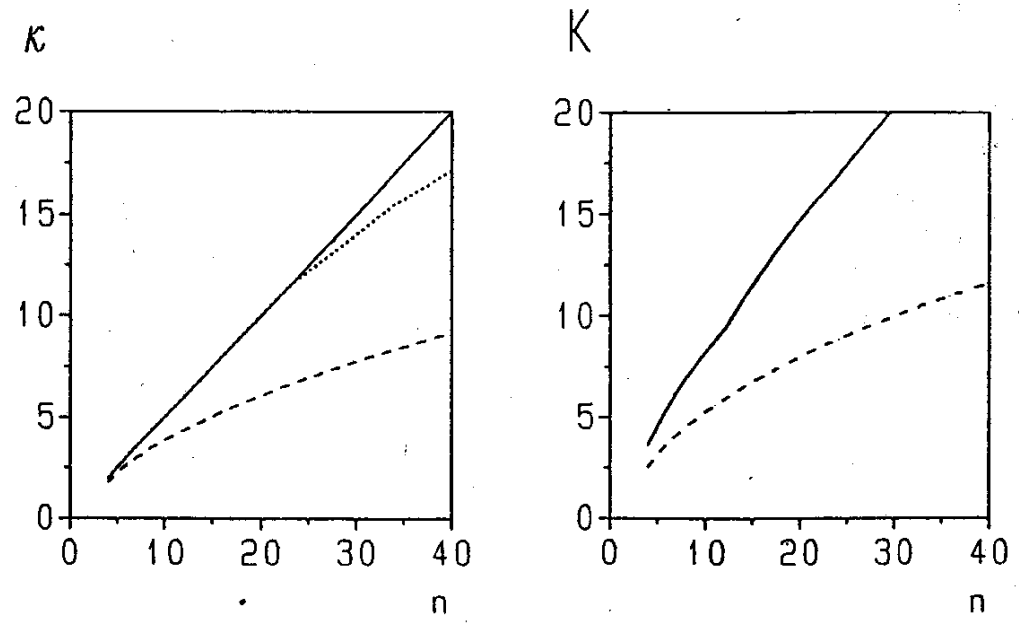

Figure I–2: See Example I–2. As functions of n are shown: left picture: κ_{ord} (———), κ_{eff} $(\cdot\cdot\cdot\cdot)$, κ_{loc} $(---)$; right picture: K_A (———), K_b $(---)$.

Example I–3. With $\underline{b} = \underline{b}_\alpha$, (4–1a), and n fixed, we have [7] that $\kappa_{eff} = \kappa_{loc}$ and for $d \to 1$ that $\kappa_{ord} \to \infty$ while $\kappa_{eff} = \kappa_{loc} \to 1$. With $n = 12$ the quantities κ_{ord}, κ_{eff}, κ_{loc} and K_A, K_b are computed as functions of d for $0 < d \leq 2.0$ and shown in Figure I–3. We notice that K_A resembles κ_{ord} (as seen in Example I–2), while K_b resembles $\kappa_{eff} = \kappa_{loc}$.

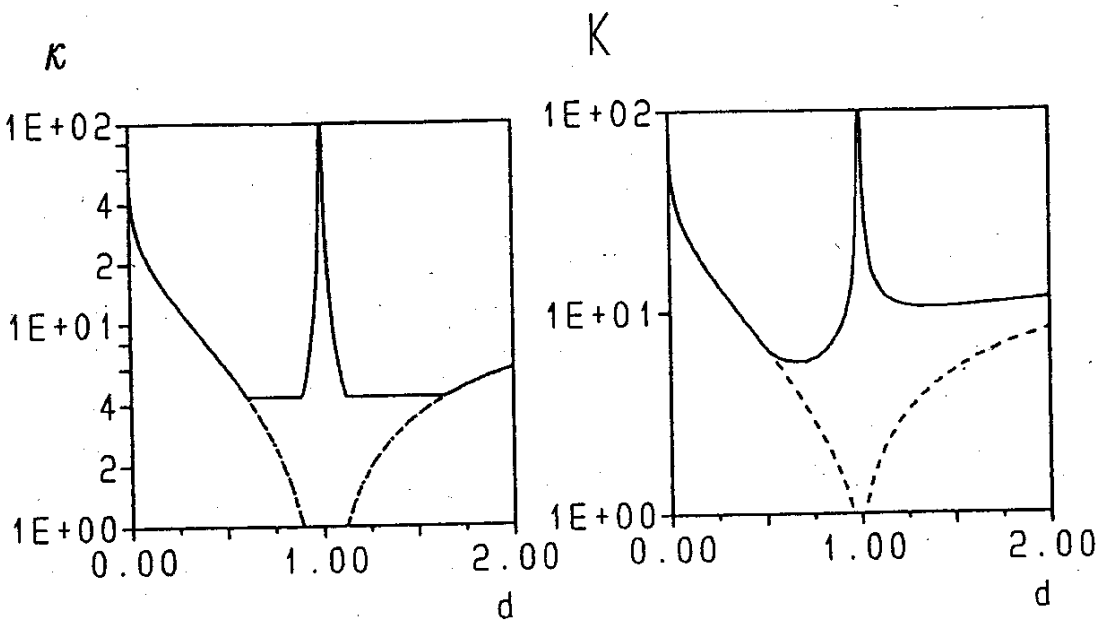

Figure I–3: See Example I–3. As functions of d are shown: left picture: κ_{ord} (———), κ_{eff} and κ_{loc} (coinciding) $(- \cdot -)$; right picture: K_A (———), K_b $(- - -)$.

Example I–4. With $\underline{b} = \underline{b}_\alpha$, (4–1a), and n fixed, we have [7] that $\kappa_{eff} = \kappa_{loc}$ and for $d \to 1$ that $\kappa_{ord} \to \infty$ and $\kappa_{eff} = \kappa_{loc} \to \infty$. With $n = 12$ the quantities κ_{ord}, κ_{eff}, κ_{loc} and K_A, K_b are computed as functions of d for $0 < d \leq 2.0$ and shown in Figure I–4. We notice that K_A resembles κ_{ord}, while K_b resembles $\kappa_{eff} = \kappa_{loc}$.

Example I–5. With $\underline{b} = \underline{b}_\alpha$, (4–1a), and n fixed, we have [7] that $\kappa_{eff} = \kappa_{loc}$ and from (4–4a) that $\underline{x} = \Big(-1/\big(2\ln(d)\big)\Big)\ \{1\,1\ldots1\}^T$. With $n = 12$ the quantities κ_{ord}, κ_{eff}, κ_{loc}, and U are computed as functions of d for $0 < d \leq 2.0$ and shown in Figure I–5. We notice that U is extremely small ($\approx 10^{-16}$). It may seem as if U behave erratic, but that a closer look reveals that U is proportional to $\kappa_{eff} = \kappa_{loc}$, but the proportionality factor turns out to depend on d. This is because $\|\delta\underline{\hat{x}}\|$ is piecewise constant near

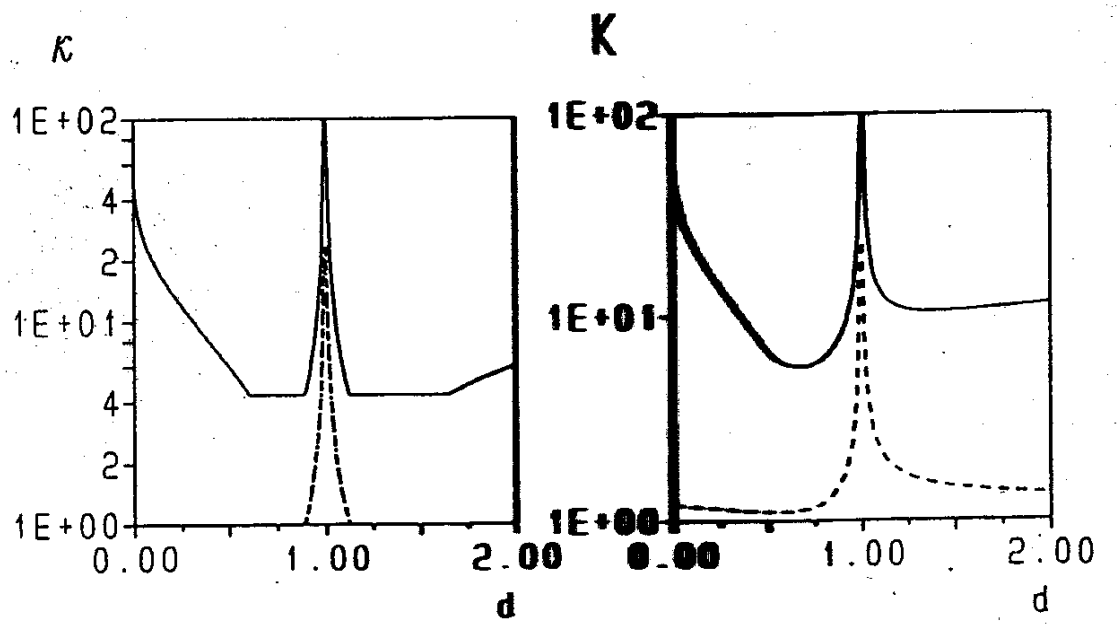

Figure I–4: See Example I–4. As functions of d are shown: left picture: κ_{ord} (———), κ_{eff} and κ_{loc} (coinciding) $(- \cdot -)$; right picture: K_A (———), K_b $(- - -)$.

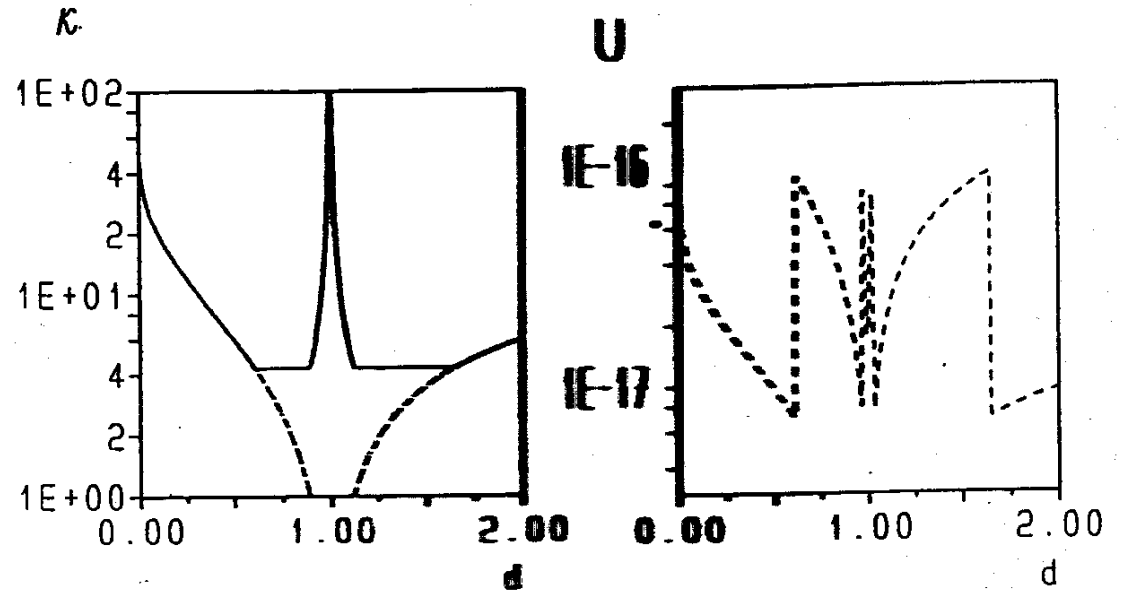

Figure I–5: See Example I–5. As functions of d are shown: left picture: κ_{ord} (———), κ_{eff} and κ_{loc} (coinciding) $(- \cdot -)$; right picture: U $(- - -)$.

$d = 1$, while $1/\|\underline{\check{x}}\|$ and also $\kappa_{\mathit{eff}} = \kappa_{loc}$ are proportional to $\ln(d)$. Therefore U and κ_{loc} are not unrelated, when the calculations are carried out using ACRITH–XSC and DLIN0 for point value matrices.

5.2 Experiments with matrix II

We here use an HP 9000 715/75 workstation to treat the system with the matrix (4–5) and the right-hand sides (4–1). For the point value computations we here use the (very convenient) system MATLAB [16, 17], where also the SVD is found; for the interval value computations we here use the language PROFIL [18] (which contains several convenient built-in functions). For the perturbation $\delta\underline{b}$ we here use the relative one (3–3b) which leads to the quantity K_b in the simple form (3–10b). For the determination of K_A, K_b, and U we solve the equations (3–4), (3–5b), and (3–11b), respectively, using ILSS [18, p. 25] for interval value matrices. From the various interval value solutions the three ratios (3–9), (3–10b), and (3–12) are easily computed using the built in functions [18]. Below we give some examples where the various computed quantities are shown graphically as function of d or n, using the plotting system of MATLAB [16, 17]. The choice of the parameters used is guided by the results of an investigation presented in Section 4.2.

Example II–1. For a symmetric matrix, i.e., $\varepsilon = 1$, we have $\kappa_{ord} \to \infty$ for n or $n-1$ values of d with $|d| > 1/2$, and we therefore choose a small value of n. With $\underline{b} = \underline{b}_\alpha$, (4–1a), and $n = 4$, the quantities κ_{ord}, κ_{loc} and K_A, K_b are computed as functions of d for $-2.0 \le d \le +2.0$ and shown in Figure II–1. Also here we notice that K_A resembles κ_{ord}, while K_b resembles κ_{loc}. A similar problem, but with a non-symmetric matrix, is treated in the next Example.

Example II–2. For a non-symmetric matrix, i.e., $\varepsilon \neq 1$, we have $\kappa_{ord} \to \infty$ for n or $n-1$ values of d with $|d| > 1/(2\sqrt{\varepsilon})$. With $\underline{b} = \underline{b}_\gamma$, (4–1c), $\varepsilon = 0.5$, and $n = 4$, the quantities κ_{ord}, κ_{loc} and K_A, K_b are computed as functions of d for $-2.0 \le d \le +2.0$ and shown in Figure II–2. Also here we notice that K_A resembles κ_{ord}, while K_b resembles κ_{loc}.

Example II–3. With $\varepsilon = 2.0$ and $n = 4$ we have that $\kappa_{ord} \to \infty$ for four values of d. For $\underline{b} = \underline{b}_\gamma$, (4–1c), the quantities κ_{ord} and U are computed as functions of d for $-2.0 \le d \le +2.0$ and shown in Figure II–3. We notice that U can become extremely small ($\approx 10^{-15}$). Contrary to Example I–5 we

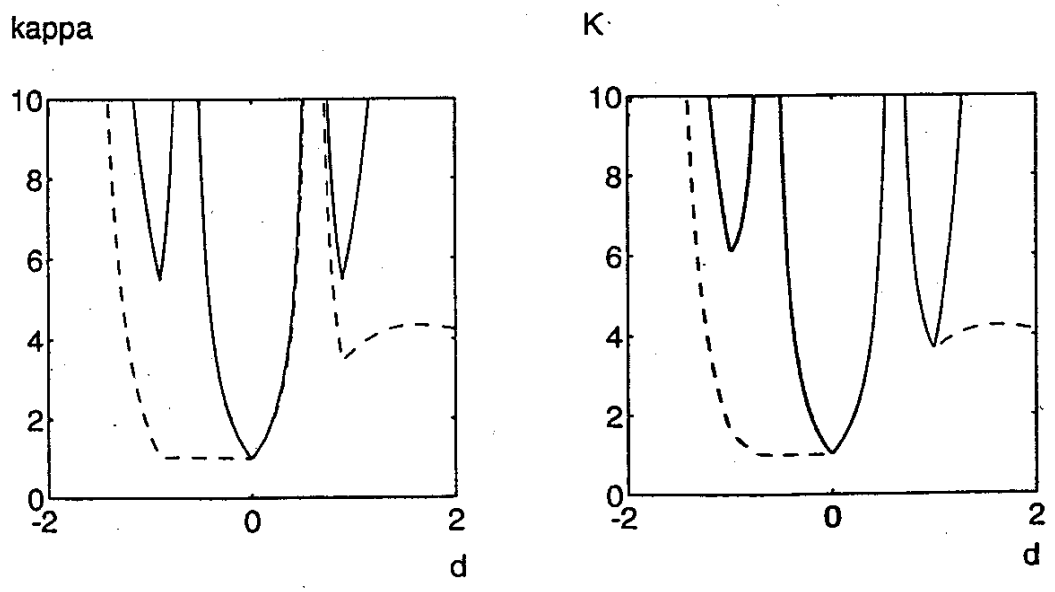

Figure II–1: See Example II–1. As functions of d are shown:
left picture: κ_{ord} (———), κ_{loc} (− − −);
right picture: K_A (———), K_b (− − −).

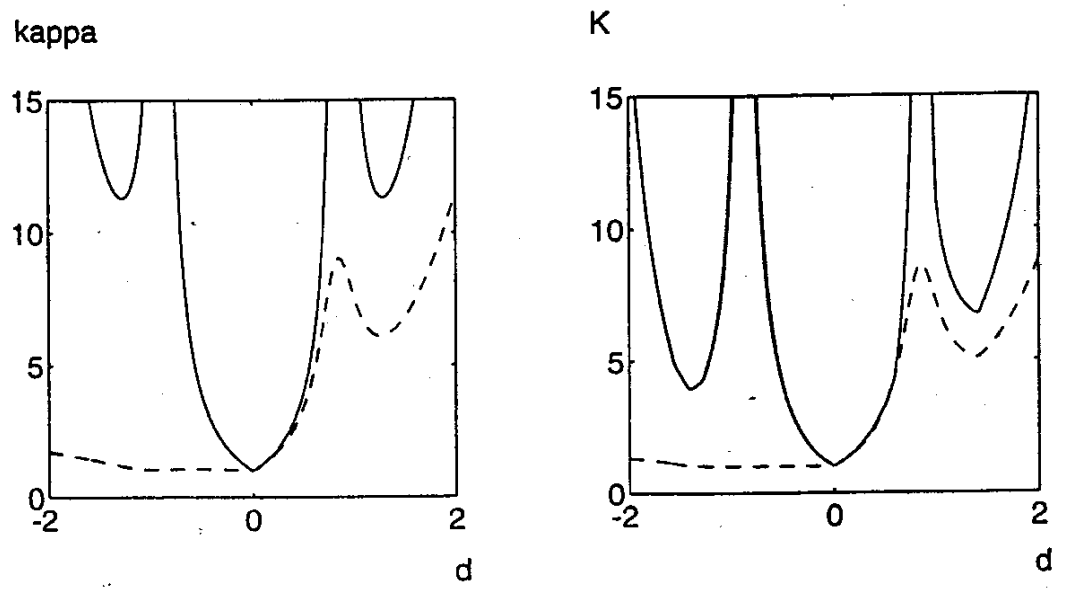

Figure II–2: See Example II–2. As functions of d are shown:
left picture: κ_{ord} (———), κ_{loc} (− − −);
right picture: K_A (———), K_b (− − −).

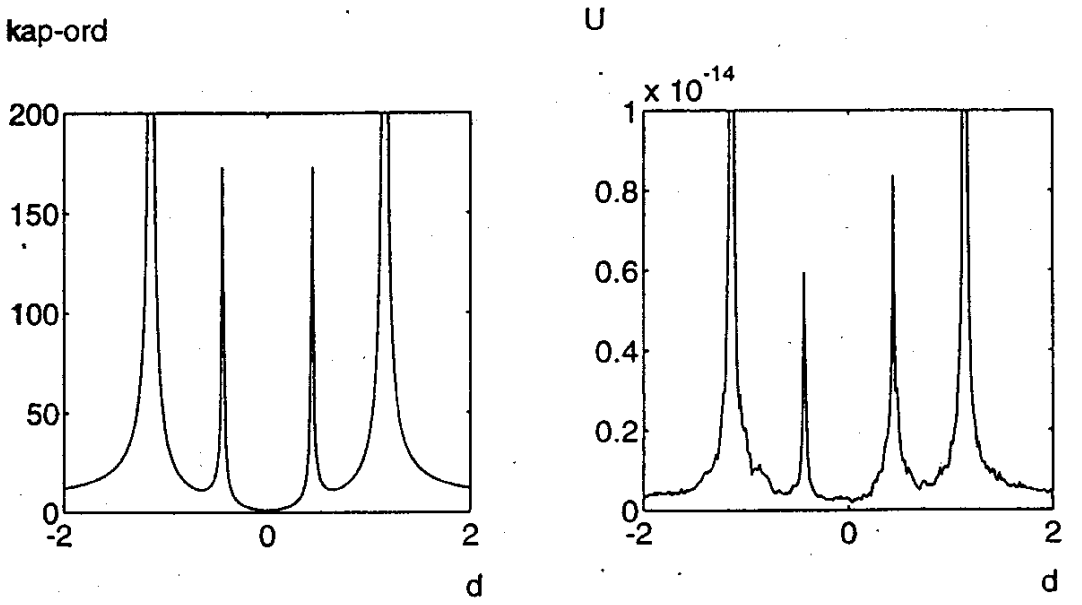

Figure II–3: See Example II–3. As functions of d are shown: left picture: κ_{ord} (———); right picture: U (———).

observe that U resembles κ_{ord} when the calculations are carried out using PROFIL and ILSS for interval value matrices.

Example II–4. With $d = 0.45$ and $\varepsilon = 1.0$ (a symmetric matrix) we have, according to (4–7), that $\kappa_{ord} \to 18.0$ for $n \to \infty$. For $\underline{b} = \underline{b}_\beta$, (4–1b), the quantities κ_{ord}, κ_{loc} and K_A, K_b are computed as functions of n for $2 \leq n \leq 60$ and shown in Figure II–4. Also here we notice that K_A resembles κ_{ord}, while K_b resembles κ_{loc}.

Example II–5. With $d = 0.5$, $\varepsilon = 1.1$ (a non-symmetric matrix), and $\underline{b} = \underline{b}_\beta$, (4–1b), are the quantities κ_{loc} and K_b computed as functions of n for $2 \leq n \leq 40$ and shown in Figure II–5. Also here we notice that K_b resembles κ_{loc}.

Example II–6. With $d = 0.5$, $\varepsilon = 0.9$ (a non-symmetric matrix), and $\underline{b} = \underline{b}_\beta$, (4–1b), are the quantities κ_{ord} and U computed as functions of n for $2 \leq n \leq 40$ and shown in Figure II–6. Similar to Example II–3 we observe that U is extremely small ($\approx 10^{-15}$), and that U resembles κ_{ord}.

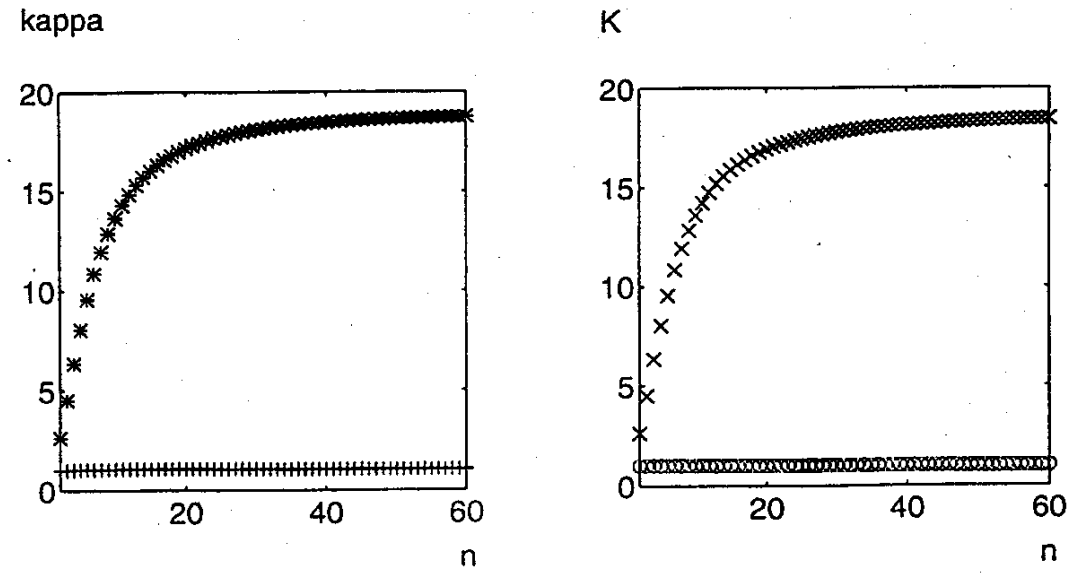

Figure II–4: See Example II–4. As functions of n are shown:
left picture: κ_{ord} $(* * *)$, κ_{loc} $(+ + +)$;
right picture: K_A $(\times \times \times)$, K_b $(\circ \circ \circ)$.

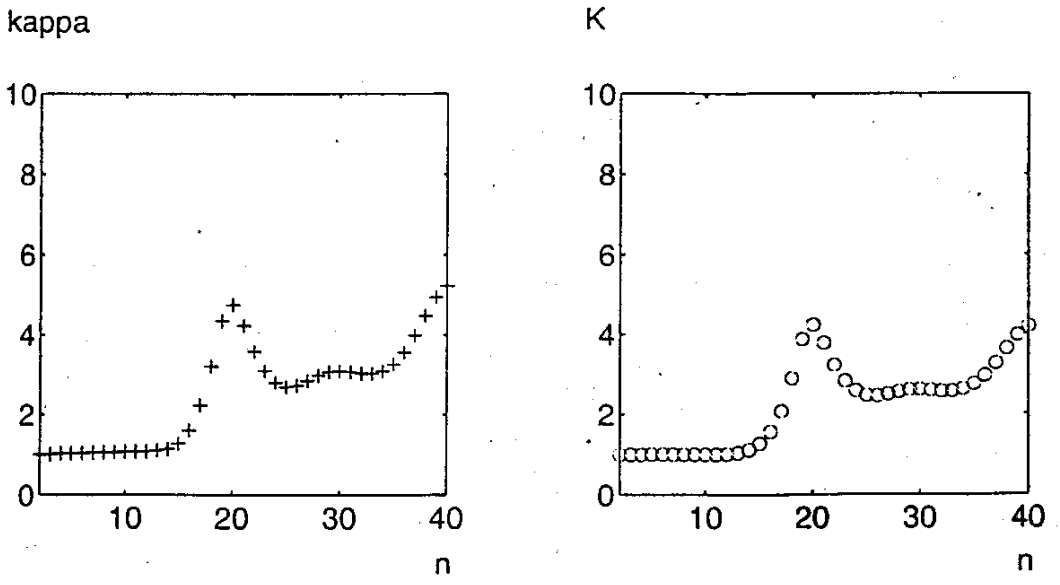

Figure II–5: See Example II–5. As functions of n are shown:
left picture: κ_{loc} $(+ + +)$; right picture: K_b $(\circ \circ \circ)$.

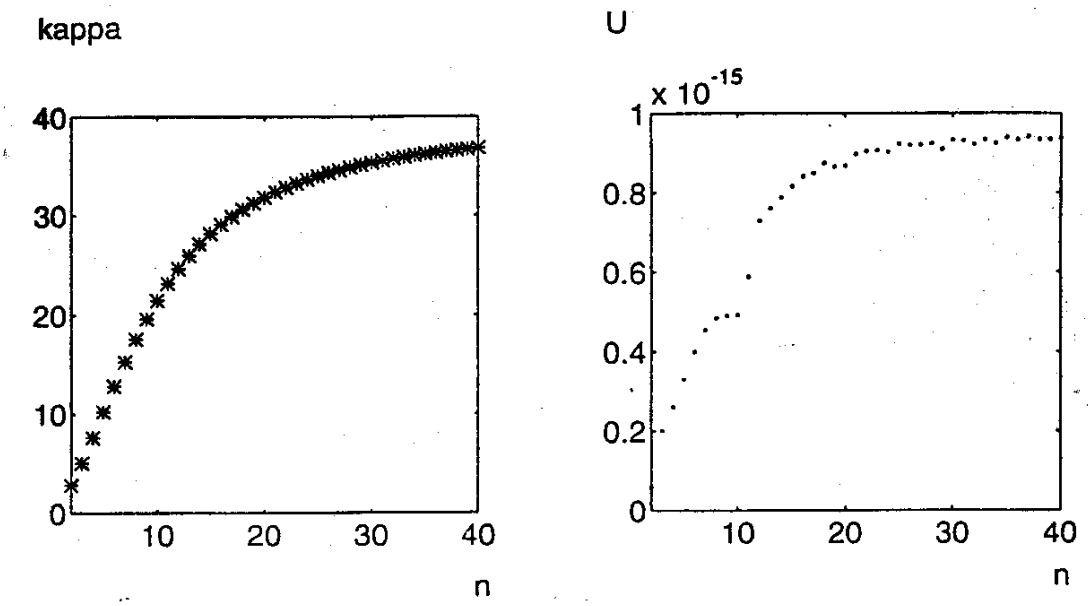

Figure II–6: See Example II–6. As functions of n are shown: left picture: κ_{ord} $(* * *)$; right picture: U $(\cdot \cdot \cdot)$.

5.3 Experiments with matrix III

We use the same computer and the same program packages to treat the system with the matrix (4–8) and the right-hand sides (4–1) in the same way as was done in Section 5.2. The choice of the parameters used is guided by the results of an investigation presented in Section 4.3.

Example III–1. With $n = 5$, $\varepsilon = 0.5$ (a non-symmetric matrix), and $\underline{b} = \underline{b}_\alpha$, (4–1a), are the quantities κ_{ord}, κ_{loc} and K_A, K_b computed as functions of d for $-3.0 \leq d \leq +3.0$ and shown in Figure III–1. Also here we notice that K_A resembles κ_{ord}, while K_b resembles κ_{loc}.

6 Conclusions

When interval methods are applied to systems of linear algebraic equations $\underline{\underline{A}}\underline{x} = \underline{b}$ we have found, by experimenting with three particularly suitable systems:

- that the uncertainty of the interval solution, per se, is extremely small for both the interval computation packages used, and it seems to

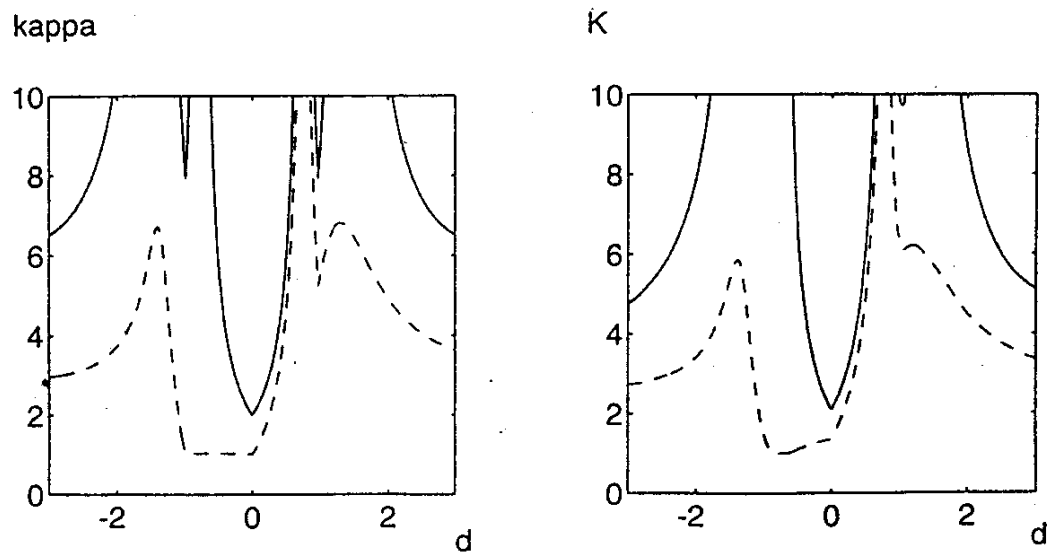

Figure III–1: See Example III–1. As functions of d are shown:
left picture: κ_{ord} (———), κ_{loc} (− − −);
right picture: K_A (———), K_b (− − −).

have some relation to the local condition number of the system when ACRITH–XSC is used and to have relation to the ordinary condition number when PROFIL is used,

- that the sensitivity of the linear system, per se, can be determined experimentally by interval methods, without using Singular Value Decomposition (SVD), and without performing "many" experiments,

- that the sensitivity with respect to $\underline{\underline{A}}$ (for a certain perturbation) depends on the ordinary condition number, while the sensitivity with respect to $\underline{b}$ depends on the local condition number, and not on the ordinary nor on the effective condition number, and

- that the singularity of the matrix $\underline{\underline{A}}$ can be detected using interval methods, without using SVD, by considering the sensitivity with respect to $\underline{\underline{A}}$.

Acknowledgements

For valuable discussions I want to thank: From the Technical University of Denmark: Niels Christian Albertsen and Kaj Madsen, Institute of Mathematical Modelling; Per Christian Hansen, UNI C; Poul Wulff Pedersen, Mathematical Institute; and from the University of Oulu, Finland: Jukka Saranen, Department of Mathematics. The instigation to try to connect condition numbers and interval methods is due to Per Christian Hansen.

The Editor-in-Chief and two unknown referees of "Interval Computations" are thanked for suggestions, comments and advice which led to a revised and expanded version, giving a clearer and more thorough treatment of the subject.

References

[1] Golub, G. H. and van Loan, Ch. F. *Matrix computations*. The John Hopkins University Press, Baltimore, London, 1989.

[2] Chan, T. F. and Foulser, D. E. *Effectively well-conditioned linear systems*. SIAM J. Scientific Statistical Computing **9** (1988), pp. 963–969.

[3] de Boor, C. and Kreiss, H.–O. *On the condition of the linear systems associated with discretized BVPs of ODEs*. SIAM J. Numer. Analysis **23** (1986), pp. 936–939.

[4] Rice, J. R. *Numerical methods, software, and analysis*. McGraw-Hill Book Company, New York et al., 1983.

[5] Noble, B. and Daniel, J. W. *Applied linear algebra*. Prentice-Hall International, Inc., Englewood Cliffs, N.J., 1988.

[6] Christiansen, S. *Numerical solution of an integral equation with a logarithmic kernel*. BIT, Nordisk Tidskr. Inform. **11** (1971), pp. 276–287.

[7] Christiansen, S. and Saranen, J. *The conditioning of some numerical methods for first kind boundary integral equations*. J. Computational Appl. Math (submitted).

[8] Saranen, J. *The modified quadrature method for logarithmic-kernel integral equations on closed curves.* J. Integral Eq. Applications **3** (1991), pp. 575–600.

[9] Gregory, R. T. and Karney, D. L. *A collection of matrices for testing computational algorithms.* Wiley-Interscience, John Wiley & Sons, New York et al., 1969.

[10] *IBM VS FORTRAN version 2, library and language. Reference.* SC 26–4221–3, Release 3, Fourth edition, March 1988.

[11] *Library manual. Mark 15. F02WEF, Singular value decomposition of a general matrix.* NAG, The Numerical Algorithms Group Limited, 1991.

[12] *IBM high accuracy arithmetic—extended scientific computation. General information.* GC 33–6461–01, Version 1, Release 1, August 1990.

[13] *IBM high accuracy arithmetic—extended scientific computation. Reference.* SC 33–6462–00, Version 1, Release 1, First edition, September 1990.

[14] *Library manual. Mark 15. F06EJF, Computes the Euclidian length of a real vector.* NAG, The Numerical Algorithms Group Limited, 1991.

[15] If, F. *AMFPLOT, an easy and fast plotting routine.* In: "Lecture Notes for the Course 0310+0320", LAMF, The Technical University of Denmark, June 1993, pp. 223–229.

[16] *MATLAB user's guide.* The MathWorks, Inc., Natick, MA, August 1992.

[17] *MATLAB reference guide.* The MathWorks, Inc., Natick, MA, August 1992.

[18] Knüppel, O. *PROFIL—programmer's runtime optimized fast interval library.* Technische Universität Hamburg-Harburg, Berichte des Forschungsschwerpunktes Informations und Kommunikationstechnik. Bericht 93.4, July 1993.

Received: November 8, 1993
Revised version: August 22, 1994

Institute of Mathematical Modelling
The Technical University of Denmark
Building 321
DK–2800 Lyngby
Denmark
E-mail: lamfsc@unidhp.uni-c.dk