# Rigorous Sensitivity Analysis for Systems of Linear and Nonlinear Equations*

by Siegfried M. Rump

**Abstract**

Methods are presented for performing a rigorous sensitivity analysis of numerical problems with independent, noncorrelated data for general systems of linear and nonlinear equations. The methods may serve for the following two purposes. First, to bound the dependency of the solution on changes in the input data. In contrast to condition numbers a componentwise sensitivity analysis of the solution vector is performed. Second, to estimate the true solution set for problems the input data of which are afflicted with tolerances. The methods presented are very effective with the additional property that, due to an automatic error control mechanism, every computed result is guaranteed to be correct. Examples are given for linear systems demonstrating that the computed bounds are in general very sharp. Interesting comparisons to traditional condition numbers are given.

**Keywords and Phrases:** Sensitivity analysis, linear and nonlinear systems, guaranteed bounds, inner inclusions, condition numbers

**AMS Classification:** 65G10, 65G05, 65H10

# 0 Introduction

In the first part we concentrate on the theoretical results; the practical implementation is discussed in chapter 3.

Let $T$ denote one of the sets $\mathbb{R}$ (real numbers) or $\mathbb{C}$ (complex numbers). Vectors $v \in VT$ and matrices $A \in MT$ consists of $n$ resp. $n \times n$ components throughout this paper. Let $S$ denote one of the sets $T, VT$ or $MT$. The power set over one of these sets is denoted by $\mathbb{I}$PT, $\mathbb{I}$PVT, $\mathbb{I}$PMT, respectively.

*published in Math. of Comp., 54(10):721–736, 1990

If not stated otherwise operations $+, -, \cdot, /$ are power set operations throughout this paper, defined in the usual way. Sets occuring several times in an expression are treated indenpendently, e.g.

$$Z \in \mathbb{IPS}: \quad Z * Z := \{ z_1 * z_2 \mid z_1, z_2 \in Z \} \supseteq \{ z * z \mid z \in Z \}$$

for all suitable operations $* \in \{+, -, \cdot, /\}$.

Infimum $\inf(z)$ and supremum $\sup(z)$ of nonempty and bounded sets $Z \in \mathbb{IPS}$ are defined in the usual way, in case of vectors and matrices componentwise (so that $\inf(A) \in MT$ when $A \in \mathbb{IPMT}$). The diameter $d(Z)$ and the radius $r(Z)$ of some nonempty, bounded $Z \in \mathbb{IPS}$ are defined by

$$d(Z) := \sup(Z) - \inf(Z) \quad \text{and} \quad r(Z) := 0.5 \cdot d(Z).$$

The diameter of $A \in \mathbb{IPMT}$ is the matrix of diameters of its components. Throughout this paper we use partial ordering for complex numbers and for vectors and matrices over those, i.e. for $T \in \{\mathbb{C}, \mathrm{V}\mathbb{C}, \mathrm{M}\mathbb{C}\}$

$$z_1, z_2 \in T : z_1 \leq z_2 :\Leftrightarrow \ \mathrm{re}(z_1) \leq \mathrm{re}(z_2) \text{ and } \mathrm{im}(z_1) \leq \mathrm{im}(z_2).$$

# 1 Bounds on the sensitivity

In [15], [16] Neumaier gives estimations on the sensitivity of systems of linear equations. He computes those bounds together with an inclusion of the solution using methods described e.g. in [18], [20]. The bounds are sharp but require some additional computational effort. Especially the solution of the linear system with another right hand side is needed; for guaranteed estimations on the sensitivity this solution has to be guaranteed to be correct.

In the following we use ideas by Neumaier to design rigorous bounds on the sensitivity of a linear system together with an inclusion for the solution with very little additional computational effort. The sensitivity is bounded by estimating the interior of the solution set of a linear system the data of which are afflicted with tolerances.

We start with a lemma estimating the diameter of sets.

**Lemma 1.** Let $S \in \{\mathbb{IR}, \mathrm{V}\mathbb{IR}, \mathrm{M}\mathbb{IR}, \mathbb{C}, \mathrm{V}\mathbb{C}, \mathrm{M}\mathbb{C}\}$ and $Q, z, \Delta \in \mathbb{IPS}$ be nonempty and bounded subsets of $S$ with

$$Q \subseteq Z - \Delta \tag{1}$$

Then

$$\inf(Z) \leq \inf(Q) + \sup(\Delta) \quad \text{and} \tag{2}$$

$$\sup(Z) \geq \sup(Q) + \inf(\Delta). \tag{3}$$

**Proof.** W.l.o.g. we prove the real vector case only. The other cases derive similarly, especially the complex case by treating the real and imaginary part. Formula (1.1) states

$$\forall\ q \in Q\ \exists\ z \in Z\ \exists\ \delta \in \Delta : \quad q = z - \delta. \tag{4}$$

For every $1 \leq i \leq n$ there is a convergent sequence $\{q^k\}$, $k \in N$ with $q^k \in Q$ for all $k \in N$ and

$$\lim_{k\to\infty} q_i^k = (\inf(Q))_i \tag{5}$$

By (1.4) follows

$$\forall\ k \in N\ \exists\ z \in Z\ \exists\ \delta \in \Delta : \quad z = q^k + \delta.$$

Therefore, for fixed $i, 1 \leq i \leq n$ holds

$$\forall\ k \in N : (\inf(Z))_i \leq z_i = q_i^k + \delta_i \leq q_i^k + (\sup(\Delta))_i. \tag{6}$$

Since (1.6) holds true for every $k \in \mathbb{N}$ and $i \in \{1, \ldots, n\}$ assertion (1.2) follows by (1.5). Assertion (1.3) follows similarly. ■

In the following we use lemma 1 to derive an estimation on the infimum and supremum of the solution set of a set of linear systems, the latter being defined by

**Definition 2.** Let $T \in \{\mathbb{IR}, \mathbb{C}\}$, $[A] \in \mathbb{IPMT}$ and $[b] \in \mathbb{IPVT}$. Then

$$\sum([A], (b)) := \{\, x \in VT \mid \exists\ A \in [A]\ \exists\ b \in [b] : Ax)b \,\} \tag{7}$$

Definition 2 does not require all or even any $A \in [A]$ to be nonsingular; $\sum([A], [b])$ may be empty.

First we mention an outer estimation for $\sum([A],[b])$ given in [18]. All operations are power set operations.

**Theorem 3.** Let $T \in \{\mathbb{IR}, \mathbb{C}\}$, $\emptyset \neq [A] \in \mathbb{IP}\mathrm{MT}$, $\emptyset \neq [b] \in \mathbb{IP}\mathrm{VT}$, $\widetilde{x} \in VT$, $R \in MT$ and $\emptyset \neq X \in \mathbb{IP}\mathrm{VT}$ being compact. Define

$$Y := \widetilde{x} + R \cdot ([b] - [A] \cdot \widetilde{x}) + \{I - R \cdot [A]\} \cdot (X - \widetilde{x}). \tag{8}$$

If

$$Y \subseteq \mathrm{int}(X) \tag{9}$$

then $R$ and every matrix $A \in (A)$ is nonsingular, $\sum([A],[b])$, is nonempty and

$$\sum([A],[b]) \subseteq Y. \tag{10}$$

**Remark.** I denotes the identity matrix, $\mathrm{int}(X)$ the interior of $X$.

Estimations of the set $Y$ described by theorem 3 are effectively computable as has been shown in [18], [20]. Note that there are no a priori assumptions on the non-singularity of $R$ or of matrices within $[A]$. Next we give an upper and lower estimation of the infimum and supremum of $\sum([A],[b])$ using lemma 1 coming virtually free of cost together with the outer estimation (1.10).

**Theorem 4.** Let $T \in \{\mathbb{IR}, \mathbb{C}\}$, $\emptyset \neq [A] \in \mathbb{IP}\mathrm{MT}$, $\emptyset \neq [b] \in \mathbb{IP}\mathrm{VT}$, $\widetilde{x} \in VT$, $R \in MT$ and every matrix within $[A]$ being nonsingular.

Define

$$\begin{array}{l} Q := \widetilde{x} + R \cdot ([b] - [A] \cdot \widetilde{x}) \quad \text{and} \\ \Delta := (I - R \cdot [A]) \cdot (\sum([A],[b]) - \widetilde{x}). \end{array} \tag{11}$$

Then

$$\inf(\sum([A],[b])) \leq \inf(Q) + \sup(\Delta) \quad \text{and} \tag{12}$$

$$\sup(\sum([A],[b])) \geq \sup(Q) + \inf(\Delta). \tag{13}$$

**Remark.** All operations in theorem 4 are power set operations.

**Proof.** For nonsingular $A \in MT$ and for $b \in VT$ holds

$$\widetilde{x} + R \cdot (b - A\widetilde{x}) = A^{-1}b - (I - R \cdot A) \cdot (A^{-1} \cdot b - \widetilde{x}). \tag{14}$$

$\sum([A],[b])$ is nonempty and

$$Q = \{\, \widetilde{x} + R \cdot (b - A \cdot \widetilde{x}) \mid A \in [A], b \in [b] \,\} =$$

$$\{\, A^{-1} \cdot b - (I - R \cdot A) \cdot (A^{-1} \cdot b - \widetilde{x}) \mid A \in [A], b \in [b] \,\} \subseteq$$

$$\sum([A],[b]) - \{\, (I - R \cdot A) \cdot (A^{-1} \cdot b - \widetilde{x}) \mid A \in [A], b \in [b] \,\} \subseteq$$

$$\sum([A],[b]) - \{\, (I - R \cdot A_1) \cdot A_2^{-1} \cdot b - \widetilde{x}) \mid A_1, A_2 \in [A], b \in [b] \,\} =$$

$$\sum([A],[b]) - (I - R \cdot [A]) \cdot (\sum([A],[b]) - \widetilde{x}) =$$

$$\sum([A],[b]) - \Delta$$

according to definition 2 and the non-singularita of every $A \in [A]$. Applying lemma 1 with $Z := \sum([A],[b])$ finishes the proof. ■

The infimum and supremum of $\sum([A],[b])$ are the (componentwise) variations of the set of solutions of $Ax = b$ for $A \in [A]$, $b \in [b]$.

It looks like a vicious circle to estimate $\sum([A],[b])$ by some expression depending on $\sum([A],[b])$, e.g. some outer estimation computed by using theorem 3 and defining $\Delta^* := (I - R \cdot [A]) \cdot (Y - \widetilde{x})$ yields

$$\Delta = (I - R \cdot [A]) \cdot (\sum([A],[b]) - \widetilde{x}) \subseteq \Delta^* \quad \text{implying}$$

$$Q \subseteq \sum([A],[b]) - \Delta^*. \tag{15}$$

Hence applying lemma 1 proves estimations (1.12) and (1.13) replacing $\Delta$ by $\Delta^*$.

To obtain very sharp estimations of the infimum and supremum of $\sum([A],[b])$ it is necessary that

$$d((I - R \cdot [A]) \cdot (Y - \widetilde{x}) \ll d(\widetilde{x} + R \cdot ([b] - [A] \cdot \widetilde{x})).$$

If $[A]$ and $[b]$ are of small diameter this, in general is true because with $R \approx A^{-1}$ for some $A \in [A]$ and $\widetilde{x} := R \cdot b$ for some $b \in [b]$ the quantities

$$I - R \cdot [A], Y - \widetilde{x} \quad \text{and} \quad [b] - [A] \cdot \widetilde{x}$$

are of the same small order of magnitude.

It should be mentioned that if, by some source, there is further knowledge on the structure and especially on the interior of $(I - R \cdot [A]) \cdot (Y - \widetilde{x})$ estimations (1.12) and (1.13) can be further sharpened.

# 2 Systems of nonlinear equations

In [18], [20] methods have been given for computing guaranteed bounds of the solution of nonlinear equations involving differentiable real or complex functions in one or more variables. The formulation of such a method requires an adapted formulation of the derivative.

**Definition 5.** Let $T \in \{\mathbb{IR}, \mathbb{C}\}$ and $f : D \to VT$, $D \subseteq VT$ continuously differentiable. Let $f' : VT \to MT$ denote the Jacobian of $f$. Then for $X \in \mathbb{I}\mathrm{PVT}$, $X \subseteq D$

$$f'(X) := \left\{ \Big( \frac{\partial f_1}{\partial x}(\zeta_1), \ldots, \frac{\partial f_n}{\partial x}(\zeta_n) \Big)^t \mid \zeta_1, \ldots, \zeta_1 \in X \right\}. \tag{16}$$

$f'$ coincides with the Jacobian if $X$ consists of one point. $f'$ is choosen such that for $x \cup y \subseteq D$

$$\forall\, x, y \in D\ \exists\, Q \in f'(x \cup y) : f(y) = f(x) + Q \cdot (y - x), \tag{17}$$

where $\cup$ denotes the convex union and $^t$ denotes transposition (see [20]).

An inclusion of a zero of each individual function out of a set of functions can be effectively calculated according to the following theorem (see [18]):

**Theorem 6.** Let $T \in \{\mathbb{IR}, \mathbb{C}\}$ and let $[f]$ be a nonempty set of continuously differentiable functions $f : D \to VT$, $D \subseteq VT$. For $R \in MT$, $\widetilde{x} \in VT$ and compact and convex $\emptyset \neq X \in \mathbb{I}\mathrm{PVT}$ define

$$Y := \cup\{\widetilde{x} - R \cdot f(\widetilde{x}) + (I - R \cdot [Q]) \cdot (X - \widetilde{x}) \mid f \in [f]\} \tag{18}$$

where $[Q] := \bigcup \{\, f'(\widetilde{x} \cup X) \mid f \in [f] \,\}$.

If

$$Y \subseteq \mathrm{int}(X) \tag{19}$$

then the matrix $R$ and every matrix $Q \in [Q]$ is nonsingular. Furthermore, every nonlinear system $f \in [f]$ has exactly one zero in $Y$:

$$\forall\ f \in [f]\ \exists^{1-1}\ \widehat{x} \in Y : f(\widehat{x}) = 0.$$

For the proof cf. [18]. The methods derived in the previous chapter allow to give inner estimations, i.e. estimations of the infimum and supremum of the set $\{\, \widehat{x} \in Y \mid\ \exists\ f \in [f] : f(\widehat{x}) = 0 \,\}$.

Let the assumptions of theorem 6 be satisfied, especially (2.4) with (2.3). Then for every $f \in [f]$ there is one and only one $\widehat{x}_f \in Y$ with $f(\widehat{x}_f) = 0$. Therefore, the set

$$Z := \{\, x \in Y \mid\ \exists\ f \in [f] : f(x) = 0 \,\} \tag{20}$$

is nonempty. $f'$ is defined in such a way (see (2.2)) that for every $f \in [f]$ there exists a $Q_f \in [Q]$ with

$$f(\widetilde{x}) = f(\widehat{x}_f) + Q_f \cdot (\widetilde{x} - \widehat{x}_f). \tag{21}$$

$Q - f$ is, in general, not uniquely determined. By (2.5) and (2.6) the following is true for every $f \in [f]$

$$\begin{aligned} \widetilde{x} - R \cdot f(\widetilde{x}) &= \widetilde{x} - R \cdot \{f(\widehat{x}_f) + Q_f \cdot (\widetilde{x} - \widehat{x}_f)\} \\ &= \widehat{x}_f - \{I - R \cdot Q_f\} \cdot (\widehat{x}_f - \widetilde{x}) \\ &\subseteq Z - \{I - R \cdot f'(\widetilde{x} \mathbin{\underline{\cup}} X)\} \cdot (X - \widetilde{x}). \end{aligned}$$

Defining

$$\begin{aligned} &\Delta := \bigcup \Big\{\, (I - R \cdot f'(\widetilde{x} \mathbin{\underline{\cup}} X)) \cdot (X - \widetilde{x}) \mid f \in [f] \,\Big\} \quad \text{and} \\ &Q := \{\, \widetilde{x} - R \cdot f(\widetilde{x}) \mid f \in [f] \,\} \end{aligned} \tag{22}$$

yields

$$Q \subseteq Z - \Delta. \tag{23}$$

Together with lemma 1 this proves the following theorem.

**Theorem 7.** Let $T \in \{\mathbb{IR}, \mathbb{IC}\}$ and let $[f]$ be a nonempty set of continuously differentiable functions $f : D \to VT$, $D \subseteq VT$. For $R \in MT$, $\widetilde{x} \in VT$ and compact and convex $\emptyset \neq X \in \mathbb{IP}\mathrm{VT}$ define $Y$ by (2.3). If

$$Y \subseteq \mathrm{int}(X)$$

then the following holds true. The set $Z$ defined by (2.5) is nonempty and

$$\begin{aligned} &\inf(Z) \leq \inf(Q) + \sup(\Delta) \\ &\sup(Z) \geq \sup(Q) + \inf(\Delta) \end{aligned} \tag{24}$$

using $Q$ and $\Delta$ defined by (2.7).

Theorem 7 is applicable on digital computers. In particular, outer estimations for $f'(\widetilde{x} \cup X)$ can be computed automatically without calculating the Jacobian explicitly, by computing the **value** of the derivative of an arbitrary function implicitly (see e.g. [4], [12], [14], [17], [21]). Sets of functions can be stored on computers using interval techniques which we are going to describe in the next chapter.

As in the cas of linear systems, (2.9) estimates inner bounds for edges of the smallest hyperrectangle containing $Z$. The bounds are very sharp as long as $d(\Delta) \ll d(Q)$.

Having bounds for general systems of nonlinear equations similar estimations of the overestimation of calculated inclusions for other problem areas in numerical analysis such as eigenvalue problems, polynomial zeros, singular values etc. can be derived.

# 3 Application on digital computers

Stepping towards a practical implementation on digital computers a number of problems have to be solved. First we need an appropriate representation for sets: simple enough to allow efficient arithmetic operations and general enough not to be too restrictive for practical applications. Second an appropriate arithmetic has to be defined allowing simple and fast execution with the property that inner and outer estimations of the corresponding power set operations are possible. Third the arithmetic must handle rounding errors in an appropriate way to maintain the guarantee of correctness of all results.

We want to stress that **any** arithmetic having the above properties is suitable for the following discussions. Here we will concentrate on a rectangular interval arithmetic. First we discuss those over the set of real or complex numbers, postponing the problems of rounding errors over floating-point numbers.

The set of intervals, i.e. hyperrectangles over real resp. complex numbers is denoted by $\mathbb{I}\mathbb{R}$ resp. $\mathbb{I}\mathbb{C}$. Let $T \in \{\mathbb{R}, \mathbb{C}\}$ then we define

$$[A] \in \mathbb{I}\mathrm{T} \Leftrightarrow [A] = \{\, A \in T \mid \underline{A} \leq A \leq \overline{A} \text{ for some } \underline{A}, \overline{A} \in T \,\}.$$

Obviously $\underline{A} = \inf([A])$ and $\overline{A} = \sup([A9)$. Intervals over vectors resp. matrices($\mathbb{I}$VT resp. $\mathbb{I}$MT) are defined as vectors resp. matrices of intervals.

In contrast to usual definitions we do not require $\underline{A} \leq \overline{A}$ for intervals. That means for instance in the case of matrices that some components of an interval matrix may be empty. We do not need and do not define operations for those; such interval matrices are needed as results which contain useful information for the nondegenerated components.

The rules of interval arithmetic (see [2], [12]) define an arithmetic which is best possible in the sense that the result interval is the smallest interval containing the result of the power set operation. More precisely let $S_1, S_2, S_3 \in \{\mathbb{IR}, \mathrm{V}\mathbb{IR}, \mathrm{M}\mathbb{IR}, \mathbb{C}, \mathrm{V}\mathbb{C}, \mathrm{M}\mathbb{C}\}$ and $[A] \in \mathbb{I}\mathrm{S}_1$, $[B] \in \mathbb{I}\mathrm{S}_2$ be intervals such that for some fixed but arbitrary operation $* \in \{+, -, \cdot, /\}$

$$A * B \text{ for all } A \in [A],\ B \in [B]$$

is well-defined with result in $\mathbb{I}\mathrm{PS}_3$. Then the corresponding interval operation $\widehat{*}$ is defined by

$$[A]\widehat{*}[B] = \bigcup \{\, [C] \in \mathbb{I}\mathrm{S}_3 \mid A * B \in [C] \text{ for all } A \in [A], B \in [B] \,\} \tag{25}$$

It can be shown (see [2], [12]) that all operations $\widehat{*}$ according to (3.1) are well-defined and, most important, are effectively computable using the componentwise definition (except complex division, which we do not need here). For example the multiplication of two interval matrices $[A], [B] \in \mathbb{I}\mathrm{T}$ with $T \in \{\mathbb{IR}, \mathbb{C}\}$ can be performed by using

$$([A] \mathbin{\widehat{\cdot}} [B])_{ij} = [A]_{i1} \mathbin{\widehat{\cdot}} [B]_{1j} \widehat{+}, \ldots, \widehat{+} [A]_{in} \mathbin{\widehat{\cdot}} [B]_{nj} \tag{26}$$

where $n$ is the number of columns of every $A \in [A]$ and rows of every $B \in [B]$. This componentwise definition (3.2) is indeed identical with definition (3.1).

However, in an interval matrix multiplication the components of every matrix is in general an overestimation of the power set operation:

$$[A] \cdot [B] = \{\, a \cdot b \mid a \in [A], b \in [B] \,\} \subseteq [A] \mathbin{\widehat{\cdot}} [B].$$

This is not the case for multiplying intervals over $T$ or for performing a dot product of two interval vectors.

For our subsequent considerations it is especially important to notice that the multiplication of an interval matrix by a point vector does not imply an overestimation:

$$[A] \in \mathbb{I}\mathrm{MT}, b \in VT \Rightarrow [A] \cdot b = [A] \mathbin{\widehat{\cdot}} b. \tag{27}$$

This is true because every component of the interval matrix $[A]$ occurs only once in the process of the multiplication. Addition and subtraction of intervals over scalars, vectors or matrices are always identical with the power set operations without any overestimation.

For the multiplication of a point matrix $R \in MT$ an interval vector $[b] \in \mathbb{I}\mathrm{VT}$ we have at least

$$\inf(R \mathbin{\widehat{\cdot}} [b]) = \inf(R \cdot [b]) \quad \text{and} \quad \sup(R \mathbin{\widehat{\cdot}} [b]) = \sup(R \cdot [b]). \tag{28}$$

This can be shown for instance by estimating the multiplication componentwise.

According to the proof of theorem 4 we need to compute an inner estimation of $Q$ and an outer estimation of $\Delta$ and of $\sum([A],[b]) - \Delta$. The latter problem can be solved by replacing every operation in the computation of $\Delta$ by its corresponding interval operation:

$$[\Delta] \subseteq (I \mathbin{\widehat{-}} R \mathbin{\widehat{\cdot}} [A]) \mathbin{\widehat{\cdot}} (X \mathbin{\widehat{-}} \widetilde{x}). \tag{29}$$

The first problem is more difficult to solve. The proof of lemma 1 shows that for our purposes it suffices to have for fixed but arbitrary $i \in \{1, \ldots, n\}$ a sequence of $q^k \in Q$ with $\lim\limits_{k\to\infty} q_i^k = (\inf(Q))_i$ and a similar sequence for the supremum of $Q$.

Intervals are closed, therefore some $\underline{q}, \overline{q} \in Q$ with

$$(\underline{q})_i = (\inf(Q))_i \quad \text{and} \quad (\overline{q})_i = (\sup(Q))_i$$

can be found. However, such $\underline{q}, \overline{q}$ are effectively computable using interval operations and (3.3). By (3.3) we already know that

$$[b] - [A] \cdot \widetilde{x} = [b] \mathbin{\widehat{-}} [A] \mathbin{\widehat{\cdot}} \widetilde{x}.$$

Regarding this and (3.4) yields

$$\inf(Q) = \inf(\widetilde{x} + R \cdot ([b] - [A]\widetilde{x})) = \inf(\widetilde{x} \mathbin{\widehat{+}} R \mathbin{\widehat{\cdot}} ([b] \mathbin{\widehat{-}} [A] \mathbin{\widehat{\cdot}} \widetilde{x})) \text{ and}$$

$$\sup(Q) = \sup(\widetilde{x} + R \cdot ([b] - [A]\widetilde{x})) = \sup(\widetilde{x} \mathbin{\widehat{+}} R \mathbin{\widehat{\cdot}} ([b] \mathbin{\widehat{-}} [A] \mathbin{\widehat{\cdot}} \widetilde{x})).$$

Using definition (3.4) for $[\Delta]$ and $[Q] := \widetilde{x} \mathbin{\widehat{+}} R \mathbin{\widehat{\cdot}} ([b] \mathbin{\widehat{-}} [A] \mathbin{\widehat{\cdot}} \widetilde{x})$ and carefully following the proof of lemma 1 demonstrates

$$\begin{array}{l} [Q] \subseteq \sum([A],[b]) \mathbin{\widehat{-}} [\Delta] \text{ and} \\ \inf(\sum([A],[b])) \leq \inf([Q]) + \sup([\Delta]) \text{ and} \\ \sup(\sum([A],[b])) \geq \sup([Q]) + \inf([\Delta]). \end{array} \tag{30}$$

Bounds on the infimum and supremum of $\sum([A],[b])$ are therefore computable using hyper-rectangles for the representation of sets, by computing $[Q]$ and $[\Delta]$ using traditional interval operations and by applying (3.6). If the diameter of $[\Delta]$ gets too large, some or in extreme cases all inner estimations on the components of $\sum([A],[b])$ may become empty.

In the following we use the definition of an interval in terms of its bounds for some $A_1, A_2 \in T \in \{\mathbb{IR}, \mathrm{V}\mathbb{IR}, \mathrm{M}\mathbb{IR}, C, \mathrm{V}\mathbb{C}, \mathrm{M}\mathbb{C}\}$:

$$[A_1, A_2] := \{\, A \in T \mid A_1 \leq A \leq A_2 \,\} \in \mathbb{I}\mathrm{T}.$$

In this notation (3.6) writes

$$[\inf([Q]) + \sup([\Delta]), \sup([Q]) + \inf([\Delta])] \subseteq \Big[\inf([A],[b])), \sup(\sum([A],[b]))\Big]. \tag{31}$$

The final global is to calculate bounds similar to (3.7) on a computer. This is hindered by the fact that digital computers only allow the exact representation of a finite set of floating-point numbers to approximate the infinite set of real or complex numbers.

Towards this goal we need appropriate roundings from the real numbers $\mathbb{R}$ into a set $\mathbb{F} \subseteq \mathbb{R}$ of floating-point numbers.

In order not to exclude certain arithmetics we state the mathematically necessary properties for those roundings and for an appropriate arithmetic. A rounding occurs always together with an arithmetic operator; therefore we add the rounding symbol to the operator.

Let $\mathbb{F} \subseteq \mathbb{R}$ denote some finite subset of $\mathbb{R}$ (which may be regarded as the set of floating-point numbers on a computer) and $\mathbb{C}\mathbb{F} := F + i \cdot F$ be a complex extension of $\mathbb{F}$. Vectors and matrices over $\mathbb{F}$ and $\mathbb{C}\mathbb{F}$ are defined as $n$-tupels resp. $n^2$-tupels forming the sets $\mathbf{V}\mathbb{F}$, $\mathbf{M}\mathbb{F}$ and $\mathbf{V}\mathbb{C}\mathbb{F}$, $\mathbf{M}\mathbb{C}\mathbb{F}$.

Let $T_1, T_2, T_3 \in \{\mathbb{F}, \mathbf{V}\mathbb{F}, \mathbf{M}\mathbb{F}, \mathbb{C}\mathbb{F}, \mathbf{V}\mathbb{C}\mathbb{F}, \mathbf{M}\mathbb{C}\mathbb{F}\}$ with corresponding sets $S_1, S_2, S_3 \in \{\mathbb{R}, \mathbf{V}\mathbb{R}, \mathbf{M}\mathbb{R}, \mathbb{C}, \mathbf{V}\mathbb{C}, \mathbf{M}\mathbb{C}\}$, resp. Using the canonical embedding $T_i \subseteq S_i$, $i)1, 2, 3$ let $* \in \{+, -, /, \cdot\}$ be an operator such that for $A_1 \in T_1$, $A - 2 \in T_2$, the image $A_1 * A_2$ is well-defined and $A_1 * A_2 \in S_3$. Let $[A]_1 \in \mathbb{I}\mathrm{T}_1$ and $[A]_2 \in \mathbb{I}\mathrm{T}_2$ be given. Then $\overset{\vee}{*}$ is an operator with $\overset{\vee}{*}\colon \mathbb{I}\mathrm{T}_1 \times \mathbb{I}\mathrm{T}_2 \to \mathbb{I}\mathrm{T}_3$ satisfying

$$[A]_1 \overset{\vee}{*} [A]_2 \subseteq [\inf([A]_1 * [A]_2), \sup([A]_1 * [A]_2)] \tag{32}$$

where the last two operations in (3.8) are the power set operations over $A_i$ in the cannonical embedding $T_i \subseteq S_i$, $i = 1, 2, 3$. The result may be the empty set for some components. By purpose we do not restrict the operators $\overset{\vee}{*}$ in any way except requiring property (3.8). Operators $\overset{\vee}{*}$ give inner estimations on the infimum and supremum of a power set operation.

In a practical implementation in principle it suffices to have operators $\overset{\vee}{*}\colon \mathbb{I}\mathbb{F} \times \mathbb{I}\mathbb{F} \to \mathbb{I}\mathbb{F}$ which can be implemented taking advantage of the different rounding modes. Such operations are, for instance, defined in the IEEE 754 standard for binary floating arithmetic [5] or in [10] [11]. Operations over $\mathbb{C}\mathbb{F}$ and vector and matrix operations over $\mathbb{F}$ and $\mathbb{C}\mathbb{F}$ can be defined componentwise using appropriate roundings.

For vector and matrix operations better results are achieved when using the inner product proposed by Kulisch (see [9], [10], [11]). Those inner product algorithms are especially advantageous for point vectors.

For interval operations, i.e. operations with outer roundings, over floating-point numbers we use (without fearing confusion) the same symbol $\widehat{*} : \mathbb{I}\mathrm{T}_1 \times \mathbb{I}\mathrm{T}_2 \to \mathbb{I}\mathrm{T}_3$ as for those over real numbers. Compared to interval operations over $\{\mathbb{R}, \mathbf{V}\mathbb{R}, \mathbf{M}\mathbb{R}, \mathbb{C}, \mathbf{V}\mathbb{C}, \mathbf{M}\mathbb{C}\}$ interval

operations over $\{\mathbb{IF}, \mathrm{V}\mathbb{IF}, \mathrm{M}\mathbb{IF}, \mathbb{CIF}, \mathrm{V}\mathbb{CIF}, \mathrm{M}\mathbb{CIF}\}$ deliver slightly wider results. They have the property

$$[A]_1 \mathbin{\widehat{*}} [A]_2 \supseteq [\inf([A]_1 * [A]_2), \sup([A]_1 * [A]_2)] \tag{33}$$

for every $[A]_1 \in \mathbb{IT}_1,\ [A]_2 \in \mathbb{IT}_2$.

Summarizing the discussion above and using interval floating-point operations $\overset{\vee}{ast}$ and $\widehat{*}$ yields

**Theorem 8.** Let $\mathbb{IF} \subseteq \mathbb{IR}$, $\mathbb{CIF} \subseteq \mathbb{C}$ and $T \in \{\mathbb{IF}, \mathbb{CIF}\}$ with operations $\overset{\vee}{*}$ and $\widehat{*}$ having the properties (3.8) and (3.9), respectively. Let $[A] \in \mathbb{IMT}$, $[b] \in \mathbb{IVT}$, $\widetilde{x} \in VT$, $R \in MT$ and $X \in \mathbb{IVT}$ be given where every component of $[A]$, $[b]$ and $X$ is nonempty. Let

$$Y := \widetilde{x} \mathbin{\widehat{+}} R \mathbin{\widehat{\cdot}} ([b] \mathbin{\widehat{-}} [A] \mathbin{\widehat{\cdot}} \widetilde{x}) \mathbin{\widehat{+}} (I \mathbin{\widehat{-}} R \mathbin{\widehat{\cdot}} [A]) \mathbin{\widehat{\cdot}} (X \mathbin{\widehat{-}} \widetilde{x}). \tag{34}$$

If

$$Y \subseteq \mathrm{int}(X)$$

then $R$ and every matrix $A \in \mathbf{M}\mathbb{C}$ with $A \in [A]$ is nonsingular, the solution set $\sum([A], [b])$ according to definition 2 satisfies

$$\sum([A], [b]) \subseteq Y$$

and for $[Q] := \widetilde{x} \overset{\vee}{+} R \overset{\vee}{\cdot} ([b] \overset{\vee}{-} [A] \overset{\vee}{\cdot} \widetilde{x})$ and $[\Delta] := (i \mathbin{\widehat{-}} R \mathbin{\widehat{\cdot}} [A]) \mathbin{\widehat{\cdot}} (X \mathbin{\widehat{-}} \widetilde{x})$ holds

$$[\inf([Q]) \mathbin{\triangle} \sup([\Delta]), \sup([Q]) \mathbin{\triangledown} \inf([\Delta])] \subseteq \left[\inf(\sum([A], [b])), \sup(\sum([A], [b]))\right] \tag{35}$$

Here $\triangle, \triangledown$ denotes the floating-point addition rounded upwards resp. downwards. Components of $[Q]$ may be empty in which case no lower bound on the sensitivity for this component is given. Computing $[Q]$ is practically free of cost, at least compared to the costs for solving the linear system. $[\Delta]$ has already been computed in (3.10).

In this case of systems of nonlinear equations there is the problem of computing a sharp inner estimation of $\widetilde{x} - R \cdot f(\widetilde{x})$. This can be done using operations $\overset{\vee}{*}$ for $* \in \{+, -, \cdot, /\}$ or other methods. e.g. [8]. For special nonlinear methods such as eigenproblems methods similar to the ones described above can be used.

# 4 Practical results

In the following we display results for linear systems with well-conditioned and ill-conditioned matrices and varying diameter of matrix and right hand side. Matrices in use are Hilbert matrices, which are, for the sake of being exactly representable on digital computers, multiplied by the least common multiple of all denominators:

$$\text{Hilbert } (H_n)_{ij} := (1cm(1,\ldots,2n-1))/(i+j-1),$$

Pascal and Zielke matrices defined by

$$\text{Pascal } (P_n)_{ij} := \begin{pmatrix} i+j \\ i \end{pmatrix}$$

$$\text{Zielke } (Z_n)_{ij} := \frac{\begin{pmatrix} n+i-1 \\ i-1 \end{pmatrix} \cdot n \cdot \begin{pmatrix} n-1 \\ j-1 \end{pmatrix}}{i+j-1}$$

Especially Zielke matrices are extremely ill-conditioned with the interesting property that a checkerboard-like distribution of $+/-$ signs over $Z_n$ generates the inverse matrix of $Z_n$. In the following randomly generated matrices have components uniformly distributed in $[0,1]$.

In the following tables we list the "overestimation" $\delta$ calculated by (3.11) in per cent. With the notation of theorem 4 and the abbreviation $Z := [\inf([Q]) + \sup([\Delta]), \sup([Q]) + \inf([\Delta])]$ we define

$$\delta := \begin{cases} 100 & \text{if } d(Z_i) = 0 \text{ for some } 1 \le i \le n \\ \max\limits_{1\le i\le n}(1 - \dfrac{d(Z-i)}{d(Y_i)}) \cdot 100 & \text{otherwise.} \end{cases} \tag{36}$$

$\delta$ gives the percentage of the inner estimation $Z$ with respect to the outer incusion $Y$. Inclusion $Y$ is calculated using theorem 7 and the inclusion methods described in [18], [20].

It should be mentioned that $\delta$ is an upper bound on the true "overestimation" of a computed inclusion. If $\delta$ is poor, i.e. near or equal 100 %, the true overestimation might still be reasonable.

The first example are Zielke matrices with tolerances

$$M_i := Z_i \cdot (1 \pm 10^{-1.4i}) \quad \text{for } i = 5 \ldots 10. \tag{37}$$

The right hand side $b$ is randomly chosen (denoted by rand) with proper dimension:

$$b_i := \text{rand} \cdot (1 \pm 10^{-8+2i}) \quad \text{for } i = 1 \ldots 5.$$

The computer in use is an IBM 3090 using double precision equivalent to 14 hex or 16 to 17 decimal digits in the mantissa.

| $\delta$[%] | b1 | b2 | b3 | b4 | b5 |
|---|---|---|---|---|---|
| $M_5$ | 3.0 | 3.0 | 3.1 | 3.2 | 3.2 |
| $M_6$ | 3.3 | 3.3 | 3.4 | 3.5 | 3.4 |
| $M_7$ | 3.8 | 3.8 | 3.9 | 3.9 | 3.8 |
| $M_8$ | 4.4 | 4.4 | 4.4 | 4.5 | 4.4 |
| $M_9$ | 5.2 | 5.2 | 5.3 | 5.3 | 5.4 |
| $M_{10}$ | 6.4 | 6.4 | 6.5 | 6.6 | 6.5 |

**Table 4.1** Overestimation for Zielke matrices (4.2)

Obviously there is only a small dependency on the diameters of the right hand side. The difference between inner and outer inclusion is less than 7 %. This is an excellent value since even a value of 90 % suffices for the purpose of estimating the magnitude of the condition number (which is usually sufficient).

For different diameters of the matrix things change. In the following table in the first row the matrices

$$Z_5 \cdot (1 \pm 10^{-i}) \quad \text{for } i = -10(1) - 6 \tag{38}$$

and in the second row the matrices

$$Z_{10} \cdot (1 \pm 10^{-i}) \quad \text{for } i = -15(1) - 11 \tag{39}$$

are treated using theorem 7 and randomly chosen right hand side of relative diameter $10^{-2}$. The results for $\delta$ are

| $\delta$[%] | i=0 | i=1 | i=2 | i=3 | i=4 |
|---|---|---|---|---|---|
| $Z_5 \cdot (1 \pm 10^{-10+i})$ | 0.0 | 0.0 | 0.3 | 3.0 | 30.4 |
| $Z_{10} \cdot (1 \pm 10^{-15+i})$ | 0.7 | 6.5 | 62.8 | -1.0 | -1.0 |

**Table 4.2** Overestimation for Zielke matrices (4.3) and (4.4)

where $-1.0$ indicates that no inclusion $Y$ using theorem 7 and the methods described in [18], [20] could be computed. Finer methods proved that in those cases there was indeed a singular matrix within the tolerance matrix. An entry 0.0 indicates that $\delta$ was less than 0.05 %.

Table 4.2 indicates an almost linear dependency of the overestimation $\delta$ on the diameter of the matrix of the linear system. $\delta$ gets big when the diameter of the matrix gets so big that nearly singular matrices are enclosed.

We omit corresponding tables for Hilbert and Pascal matrices because they look very similar, in fact almost identical.

For random right hand side of proper dimension and of constant relative diameter $10^{-2}$ and for a linear system with matrices

$$R_n \cdot (1 \pm 10^{-i}) \quad \text{for } i = 5(-1)2 \tag{40}$$

with random matrices $R_n$ of dimension 10(10)50 gives the following the values for $\delta$ in %:

| $\delta$[%] | i=5 | i=4 | i=3 | i=2 |
|---|---|---|---|---|
| $R_{10} \cdot (1 \pm 10^{-i})$ | 0.1 | 2.8 | 12.2 | -1.0 |
| $R_{20} \cdot (1 \pm 10^{-i})$ | 0.2 | 1.0 | 62.8 | -1.0 |
| $R_{30} \cdot (1 \pm 10^{-i})$ | 2.0 | 6.6 | 100.0 | -1.0 |
| $R_{40} \cdot (1 \pm 10^{-i})$ | 3.7 | 13.6 | 100.0 | -1.0 |
| $R_{50} \cdot (1 \pm 10^{-i})$ | 2.4 | 28.7 | 100.0 | -1.0 |

**Table 4.3** Overestimation for random matrices (4.5)

The entry 100.0 indicates that an inclusion $Y$ was computed according to theorem 7 but at least one component of the inner estimation $Z$ was empty.

The tables above indicate that there is a small area where the matrix of the linear system does not contain singular matrices but nearly singular matrices where the overestimation $\delta$ is poor.

Following we display estimations of the sensitivity of a linear system with respect to perturbations of the input data. Let a linear systeme $A \cdot x = b$ with $A \in MT$, $b \in VT$ for $T \in \{\mathbb{IR}, \mathbb{C}\}$ be given, $A$ invertible. The diameter of $\sum([A], [b])$ for $[A] = A \cdot (1 \pm \varepsilon)$ and $[b] = b \cdot (1 \pm \varepsilon)$ for small $\varepsilon > 0$ gives a componentwise measure of the sensitivity of $A^{-1} \cdot b$ w.r.t. small changes in $A$ and $b$. In the following tables we display

$$f := r(Y)/\varepsilon \geq r(\sum([A], [b])/\varepsilon \tag{41}$$

using the upper bound $Y$ of $\sum([A], [b])$, cf. theorem 8. The quality of $Y$ is estimated by $\delta$ from (6.1). $f$ bounds the maximum factor by which an $\varepsilon$-perturbation of $A$ and $b$ is amplified in terms of variations in the solution. An algorithm based on theorem 7 gives $f$ for every component independently. In the following we solve the linear system $Ax = I$, i.e. compute a full inverse of $A$. The algorithm yields $n^2$ amplification factors the maximum of which, namely $f$ is displayed. Beside $f$ we display the condition number $c$ defined by

$$c := s_1/s_n \quad \text{where} \quad s_1 \geq \ldots \geq s_n > 0$$

where $s_i$ are the singular values of $A$.

In the following table we choose $\varepsilon := 10^{-15}$, one order of magnitude above the relative rounding error unit. Matrices in use are Hilbert matrices $H - n$, Pascal matrices $P_n$ and Zielke matrices $Z - n$ for different dimensions.

| | $f(H_n)$ | $c(H_n)$ | $f(P_n)$ | $c(P_n)$ | $f(Z_n)$ | $c(Z_n)$ |
|---|---|---|---|---|---|---|
| n=5 | 2.0e5 | 4.8e5 | 1.5e4 | 6.3e4 | 1.9e5 | 7.9e5 |
| 6 | 5.3e6 | 1.5e7 | 1.4e5 | 9.3e5 | 5.3e6 | 3.7e7 |
| 7 | 1.5e8 | 4.8e8 | 1.2e6 | 1.4e7 | 1.5e8 | 1.8e9 |
| 8 | 4.4e9 | 1.5e10 | 1.1e7 | 2.0e8 | 4.4e9 | 9.4e10 |
| 9 | 1.3e11 | 4.9e11 | 9.5e7 | 3.0e9 | 2.6e11 | 5.0e12 |
| 10 | 4.0e12 | 1.6e13 | 8.4e8 | 4.5e10 | 8.0e12 | 2.7e14 |

**Table 4.4** Sensitivity of linear systems vs. traditional condition number

The uncertainty $\delta$ of $f$ in all cases of table 4.4 is less than 1 %, i.e. all estimations on $f$ are correct to two figures and, by the general principle underlying the methods, are guaranteed to be correct.

In the examples above the condition number $c(.)$ is an overestimation of the true sensitivity $f(.)$. Of course, usually the condition number would not be computed in this way because a singular value decomposition is too expensive solely for the purpose of giving an estimation on the condition of the matrix.

The next example shows a significant underestimation of the sensitivity $r(\sum([A],[b]))/\varepsilon$ by the condition number $c$ due to equilibration effect of norms. Let $R_n$ be a random matrix with $n$ rows and columns, $\varepsilon$ is again $10^{-15}$.

| $n$ | $f(R_n)$ | $c(R_n)$ |
|---|---|---|
| 10 | 5.4e3 | 5.3e1 |
| 30 | 1.7e5 | 2.0e2 |

**Table 4.5** Sensitivity analysis of linear systems with random matrix vs. traditional condition number

Using the traditional definition of the quotient of largest and smallest singular value the true sensitivity of the linear system is underestimated by 2 resp. 3 orders of magnitude. Moreover, the traditional methods does only consider the matrix of the linear system, not the right hand side. The true condition number depends significantly on the right hand side.

The reason why the componentwise estimations $f$ defined in (4.6) are much larger than $c$ is that some individual components of the inverse are much more responsive to small perturbations in the input data than others. In the second example with the matrix $R_{30}$ the componentwise estimation on $f$ (which is delivered by theorem 7) is

$\leq$ 1e3 for 95 % of all components

$\leq$ 1e4 for 98 % of all components

and only in 2 cases greater than 1e5.

In the example of random matrices the greater sensitivity of an individual component of the inverse occured exactly for those components being of significantly smaller absolute value compared to all others. This need not to be the case as has been demonstrated by the examples for Hilbert, Pascal and Zielke matrices. The new methods allow a componentwise sensitivity analysis.

# 5 Conclusion

Methods have been described for the computation of inner and outer bounds of the solution set of linear and nonlinear systems the data of which are afflicted with tolerances. The bounds computed are sharp and guaranteed to be correct. It turns out that in most cases the differences of inner and outer bounds is negligible. This difference becomes larger only in extreme cases where (in the examples) a singular matrix is very close to the set of matrices of a linear system.

A criticism of inclusion algorithms for data afflicted with tolerances was that correct bounds for the solution set are computed and all experiences showed that those bounds are sharp, but the degree of sharpness could not be estimated (see [6]). The presented theorems and practical results fill this gap.

The inner estimations come virtually free of cost together with outer estimations. They allow a sensitivity analysis of problems with the additional advantage that rather than a single number estimating the condition of the problem in use a whole sensitivity matrix can be computed estimating variations of individual components of the solution for perturbations in the input data.

The estimation on the sensitivity of the linear system is guaranteed to be correct and reflects the true sensitivity of the linear system, i.e. of the matrix in combination with the particular right hand side. It has been shown by means of examples that traditional condition numbers do not necessarily reflect the true sensitivity of individual components of a solution.

The methods described can be implemented very effectively on digital computers. No special computer arithmetic is necessary; a state of the art arithmetic e.g. described in the IEEE 754 binary floating-point standard suffices. Especially all kinds of computer arithmetic allowing the representation of sets on computers are suitable; in our implementation we used a rectangular real or complex arithmetic. A computer implementation for non-linear systems is somewhat more involved and will be described later.

# References


[1] ACRITH, High-Accuracy Arithmetic Soubroutine Library, Program Description and User's Guide, IBM Publications, Document Number SC 33-6164-3 (1986)

[2] Alefeld, G. and Herzberger, J.: Introduction to Interval Computations, Academic Press (1983)

[3] Bauch, Jahn, Oelschlägel, Süsse, Wiebigke: Intervallmathematik, Theorie und Anwendungen, Mathematisch-Naturwissenschaftliche Bibliothek, Band 72, B.G. Teubner, Leipzig (1987)

[4] Baur, W. and Strassen, V.: The Complexity of Partial Derivatives. Theoretical Computer Science 22, 317–330 (1983)

[5] IEEE 754 Standard for Floating-Point Arithmetic (1986)

[6] Kahan, W. and LeBlanc, E.: Anomalies in the IBM ACRITH Package, Proceedings of the $7^{\text{th}}$ Symposium on Computer Arithmetic, edited by Kai Hwang, Urbana, Illinois (1985)

[7] Krawczyk, R.: Newton-Algorithmen zur Bestimmung von Nullstellen mit Fehlerschranken, Computing 4, 187–201 (1969)

[8] Krawczyk, R. and Neumaier, A.: Interval Slopes for Rational Functions and Associated Centered Forms, SIAM J. Numer. Anal. 22, No. 3, 604–616 (1985)

[9] Kulisch, U.: Grundlagen des numerischen Rechnens (Reihe Informatik 19), Mannheim Wien Zürich, Bibliographisches Institut (1976)

[10] Kulisch, U. and Miranker, W.L.: Computer Arithmetic in Theory and Practice. Academic Press, New York (1981)

[11] Kulisch, U. and Miranker, W.L. (Eds.): A New Approach to Scientific Computation, Academic Press (1983)

[12] Moore, R.E.: Interval Analysis, Englewood Cliffs: Prentice Hall (1966)

[13] Moore, R.E.: A Test for Existence of Solutions for Non-Linear Systems, SIAM J. Numer. Anal. 4 (1977)

[14] Moore, R.E.: Methods and Applications of Interval Analysis, Philadelphia: SIAM (1979)

[15] Neumaier, A.: Overestimation in Linear Interval Equations, SIAM J. Numer. Anal. 24, No. 1, 207–214 (1987)

[16] Neumaier, A.: Rigorous Sensitivity Analysis for Parameter-Dependent Systems of Equations, to appear

[17] Rall, L.B.: Automatic Differentiation: Techniques and Applications, Lecture Notes in Computer Science, No. 120, Springer-Verlag, Berlin Heidelberg New York (1981)

[18] Rump, S.M.: Solving Algebraic Problems with High Accuracy, in: A New Approach to Scientific Computation, eds. U. Kulisch and W.L. Miranker, Academic Press, 51–120 (1983)

[19] Rump, S.M.: Solving Non-Linear Systems with Least Significant Bit Accuracy, Computing 29, 183–200 (1982)

[20] Rump, S.M.: New Results on Verified Inclusions, in: Accurate Scientific Computations, eds. W.L. Miranker and R. Toupin, Springer Lecture Notes in Computer Science 235 (1986)

[21] Speelpennig, B.: Compiling Fast Partial Derivatives of Functions given by Algorithms, Ph. D., Urbana, Illinois (1980)

[22] Wongwises, P.: Experimentelle Untersuchungen zur numerischen Auflösung von linearen Gleichungssystemen mit Fehlererfassung, Interner Bericht 75/1, Institut für Praktische Mathematik, Universität Karlsruhe (1975)