Improved Iteration Schemes for Validation Algorithms for Dense and Sparse Nonlinear Systems *

S. M. Rump, Hamburg

Abstract

An iteration process for computing validated solutions of nonlinear systems is improved for the dense case and for the sparse case. The improvement may result in the reduction of the number of Jacobian or slope matrices to be computed. Possibly, without the improvement no inclusion is computed at all.

Zusammenfassung

Es werden Sätze angegeben, wie ein Iterationsprozeß zur Bestimmung von validierten Einschließungen der Lösung dichtbesetzter sowie dünnbesetzter nichtlinearer Gleichungssysteme verbessert werden kann. Die Anzahl der Berechnungen von Jacobi- oder Steigungsmatrizen kann reduziert werden. U.U. wird die Berechnung einer Einschließung überhaupt erst ermöglicht.

AMS Subject Classification 65G10, 65H10

0 Introduction

Let $f(x) = 0$, $f : D \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be a system of nonlinear equations with continuous f. For compact and convex $\emptyset \neq X \subseteq \mathbb{R}^n$ with fixed $\widetilde{x} \in D$ and $\widetilde{x} + X \subseteq D$, we assume a linearization of f with respect to some $\widetilde{x}$ to be given by means of a set-valued matrix $S(\widetilde{x}, X) \subseteq \mathbb{R}^{n \times n}$, i.e.

$$\forall\, x \in X \;\exists\, M \in S(\widetilde{x}, X): \quad f(\widetilde{x} + x) - f(\widetilde{x}) = M \cdot x. \tag{1}$$

*published in Computing, 57(1):77-84, 1996

For a wide class of functions, $S(\widetilde{x}, X)$ can be computed by means of slopes [7], [10] or, if f is differentiable, by means of automatic differentiation [11], [4] together with interval operations. This process is performed automatically, and essentially applies to functions given by means of a program. For details see [11], [4], [7], [10].

A widely used validation algorithm for nonlinear equations is based on the Krawcyk-operator [6], and the existence test by Moore [8]. There are several improvements of the original ideas, basically leading to the following algorithm (see [13] and the literature cited over there).

1) Let an approximate solution $\widetilde{x} \in D$ of $f(x) = 0$ be given.
2) Set $X := \widetilde{x} \cdot [-\varepsilon, \varepsilon]$, and calculate $S(\widetilde{x}, X)$ satisfying (1).
3) Let $R \approx \mathrm{mid}(S(\widetilde{x}, X))^{-1}$ be an approximate inverse, and $Z \supseteq -R \cdot f(\widetilde{x})$ be an inclusion of $-R \cdot f(\widetilde{x})$; $\quad \boldsymbol{C} := I - R \cdot S(\widetilde{x}, X)$; $\quad k = 0$.
4) Calculate an inclusion
$$Y \supseteq Z + \boldsymbol{C} \cdot X.$$
5) If $Y \subseteq \mathrm{int}(X)$ then
"there exists an $\widehat{x} \in Y$ with $f(\widetilde{x} + \widehat{x}) = 0$"; stop;
else
$k = k + 1$; $X := Y \cdot [1 - eps, 1 + eps] + [-\eta, \eta]$;
if $k \leq k_{\max}$ then calculate $S(\widetilde{x}, X)$ satisfying $(0, 1)$ and goto 4)

Algorithm 1. Validated inclusion of a solution of a nonlinear system

The constant ε is usually around machine precision, η is the smallest positive machine number and *eps* may be chosen around 0.1. For sparse nonlinear systems, the algorithm has to be modified to avoid the approximate inverse R which, in general, ist full. We come to this in Chapter 2. For more details and a convergence analysis see [13].

If $\widetilde{x}$ is sufficiently accurate, only few iterations of steps 4 and 5 are necessary. In many cases the algorithm finishes with $k = 0$. However, *if* an iteration is necessary, it is fairly expensive, because a new slope matrix $S(\widetilde{x}, X)$ or Jacobian has to be computed. In the following, an addition to algorithm 1 is presented which may calculate an inclusion of a solution of $f(x) = 0$ with the information available. In chapter 2, the iteration process for sparse systems is improved and accelerated.

In the following we denote the set of real intervals by $\mathbb{IR} = \{\, [a, b] \mid a, b \in \mathbb{R}, a \leq b \,\}$. $\mathbb{IR}^n$, $\mathbb{IR}^{n \times n}$ are vectors, matrices over $\mathbb{IR}$, respectively. For $X \in \mathbb{IR}$ we define $|X| := \max\{\, |x| \mid x \in X \,\}$. $\mathrm{int}(X)$ denotes the topological interior of X. For vectors and matrices, absolute value, diameter and comparison is defined componentwise, while ρ denotes the spectral radius. For a set T, ∂T denotes the topological boundary of T.

For interval quantities A, B, operations between them are always *interval operations*. The result is the smallest interval quantity containing the corresponding result when using power set operations. For example,

$$A \in \mathbb{IIR}^{n\times n},\ B \in \mathbb{IIR}^n: \quad A \cdot B := \bigcap \{\, C \in \mathbb{IIR}^n \mid \forall\, a \in A\ \forall\, b \in B : a \cdot b \in C \,\}.$$

Interval operations are easy to implement on digital computers (see [9], [3], [10]; for a very fast and public domain implementation for PCs and workstations see [5]).

For simplicity of notation, we allow in the following for empty interval vectors. Any operation with those yields as the result the empty set. The basic property of interval operations is the isotonicity, cf. the cited literature.

1 Dense nonlinear systems

Algorithm 1 provides an existence test for some $\widehat{x} \in X$ with $f(\widetilde{x} + \widehat{x}) = 0$. In this case, $\widehat{x} \in Y \subseteq X$ can be shown. This is in the spirit of [8]. However, if the condition $Y \subseteq \mathrm{int}(X)$ is not satisfied, then an iteration has to be started. This iteration is expensive because each time a new expansion matrix (Jacobian or slope) has to be computed, and the matrix multiplication $R \cdot S(\widetilde{x}, X)$ has to be performed. This is because for $x \in Y \backslash X$ the matrix $S(\widetilde{x}, X)$ does no longer expand f w.r.t. $\widetilde{x}$.

In the following, we show how the information already available in step 4 of Algorithm 1 can be used to prove $\widehat{x} \in X$ with $f(\widetilde{x} + \widehat{x}) = 0$. A criterion for $f(\widetilde{x} + x) \neq 0$ for all $x \in X$ is given as well.

Theorem 1. Let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be continuous, $\widetilde{x} \in D$, and $R \in \mathbb{R}^{n\times n}$ fixed but arbitrary. Let $X \in \mathbb{IIR}^n$ with $\widetilde{x} + X \subseteq D$, and let $S(\widetilde{x}, X) \in \mathbb{IIR}^{n\times n}$ be given with

$$\forall\, x \in X \quad \exists\, M \in S(\widetilde{x}, X): \quad f(\widetilde{x} + x) - f(\widetilde{x}) = M \cdot x. \tag{2}$$

Let $\boldsymbol{C} \in \mathbb{IIR}^{n\times n}$, $Z \in \mathbb{IIR}^n$ be given with $I - R \cdot S(\widetilde{x}, X) \subseteq \boldsymbol{C}$ and $-R \cdot f(\widetilde{x}) \in Z$, and define

$$X^0 := X; \quad X^{i+1} := (Z + \boldsymbol{C} \cdot X^i) \cap X^i \quad \text{for } 0 \leq i \in \mathbb{N}. \tag{3}$$

Then for $0 \leq k \in \mathbb{N}$ holds

i) $X^{k+1} \neq \emptyset$ and $X^{k+1} \subseteq \mathrm{int}(X^0) \quad \Rightarrow \quad \forall\, 1 \leq \mu \in \mathbb{N} \quad \exists\, \widehat{x} \in X^{k+\mu} : f(\widetilde{x} + \widehat{x}) = 0.$

ii) $\forall\, x \in X^0 \backslash X^k : f(\widetilde{x} + x) \neq 0.$

Especially

iii) $X^{k+1} = \emptyset \quad \Rightarrow \quad \forall\, x \in X^0 : f(\widetilde{x} + x) \neq 0.$

With the assumptions of i) also follows regularity of R and every $M \in S(\widetilde{x}, X)$.

Remark. We want to stress that the matrix $S(\widetilde{x}, X)$ can be calculated by means of an automized process using forward or backward differentiation [11], [4] or slopes [7], [10]. I denotes the $n \times n$ identity matrix, all operations are interval operations. Especially, $Z + \boldsymbol{C} \cdot X^i \in \mathbb{IIR}^n$ and $X^i \in \mathbb{IIR}^n$ imply $(Z + \boldsymbol{C} \cdot X^i) \cap X^i \in \mathbb{IIR}^n$, that is the topological intersection of two interval vectors is a (possibly empty) interval vector. Therefore $X^{i+1} \in \mathbb{IIR}^n$ is well-defined.

Proof. Define $g(x) := x - R \cdot f(\widetilde{x} + x)$. Then g is continuous and

$$\forall\, x \in X \quad \exists\, z \in Z \quad \exists\, M \in S(\widetilde{x}, X) : \quad g(x) = x - R \cdot \{f(\widetilde{x}) + M \cdot x\} = z + \{I - RM\}x.$$

Hence

$$\forall\, x \in X^0 \quad g(x) \in Z + \boldsymbol{C} \cdot x. \tag{4}$$

For $\widehat{x} \in X^0$ with $f(\widetilde{x} + \widehat{x}) = 0$ follows $g(\widehat{x}) = \widehat{x}$, and by (3) and (4)

$$\widehat{x} \in X^0 \quad \text{and} \quad f(\widetilde{x} + \widehat{x}) = 0 \quad \Rightarrow \quad \widehat{x} \in X^i \quad \text{for} \quad 0 \leq i \in \mathbb{IN}. \tag{5}$$

This proves proposition ii) and also iii). Assume $X^{k+1} \neq \emptyset$ and $X^{k+1} \subseteq \mathrm{int}(X^0)$ for some $0 \leq k \in \mathbb{IN}$. Then

$$\emptyset \neq X^{k+1} \subseteq X^k \subseteq \cdots \subseteq X^1 \subseteq X^0. \tag{6}$$

First we prove $Z + \boldsymbol{C} \cdot X^k \subseteq \mathrm{int}(X^0)$. Assume $x \in (Z + \boldsymbol{C} \cdot X^k) \cap \partial X^0$. Then we show by induction that $x \in X^i$ for $0 \leq i \leq k+1$. This is true for $i = 0$, and assuming $x \in X^i$ and using $X^k \subseteq X^i$ yields

$$x \in (Z + \boldsymbol{C} \cdot X^k) \cap X^i \subseteq (Z + \boldsymbol{C} \cdot X^i) \cap X^i = X^{i+1}.$$

Therefore, $x \in X^{k+1}$ which condradicts our assumption $X^{k+1} \subseteq \mathrm{int}(X^0)$. This implies $(Z + \boldsymbol{C} \cdot X^k) \cap \partial X^0 = \emptyset$. Now, $Z + \boldsymbol{C} \cdot X^k$ is nonempty and connected, and $\emptyset \neq (Z + \boldsymbol{C} \cdot$

$X^k) \cap X^k = X^{k+1} \subseteq \mathrm{int}(X^0)$ shows that there are points of $Z + \boldsymbol{C} \cdot X^k$ in the interior of X^0. This proves $Z + \boldsymbol{C} \cdot X^k \subseteq \mathrm{int}(X^0)$. Extensively using (6), this implies

$$\begin{aligned} Z + \boldsymbol{C} \cdot X^{k+1} &\subseteq Z + \boldsymbol{C} \cdot (\bigcap_{i=0}^{k} X^i) \subseteq \bigcap_{i=0}^{k} (Z + \boldsymbol{C} \cdot X^i) \subseteq X^0 \cap \bigcap_{i=0}^{k} (Z + \boldsymbol{C} \cdot X^i) \\ &= X^1 \cap \bigcap_{i=1}^{k} (Z + \boldsymbol{C} \cdot X^i) = \cdots = X^k \cap (Z + \boldsymbol{C} \cdot X^k) = X^{k+1}. \end{aligned} \tag{7}$$

Now basic properties of interval analysis (Proposition (17), p. 153 in [3], see also [10]) show

$$\mathrm{diam}(X^{k+1}) \geq \mathrm{diam}(Z + \boldsymbol{C} \cdot X^{k+1}) \geq \mathrm{diam}(\boldsymbol{C} \cdot X^{k+1}) \geq |\boldsymbol{C}| \cdot \mathrm{diam}(X^{k+1}).$$

and Perron-Frobenius Theory yields

$$\rho(C) \leq 1 \quad \text{for all} \quad C \in \boldsymbol{C}.$$

Suppose R or some $M \in S(\widetilde{x}, X)$ are singular. Then by definition, there is a $C \in \boldsymbol{C}$ with eigenvalue 1, i.e. $Cy = y$ for $0 \neq y \in \mathbb{R}^n$. Let $z \in Z$ fixed but arbitrary and define $h(x) := z + C \cdot x$. Then (7) yields $h(X^{k+1}) \subseteq X^{k+1}$, and Brouwer's Fixed Point Theorem implies existence of some $\widehat{x} \in X^{k+1} \subseteq X^0$ with $h(\widehat{x}) = \widehat{x} = z + C \cdot \widehat{x}$. There exists $0 \neq \lambda \in \mathbb{R}$ with $\widehat{x} + \lambda y \in \partial X^0$. Then $h(\widehat{x} + \lambda y) = \widehat{x} + \lambda y$, and by definition (3) we have $\widehat{x} + \lambda y \in X^i$ for $0 \leq i \in \mathbb{N}$. This contradicts $X^{k+1} \subseteq \mathrm{int}(X^0)$, and henceforth shows regularity of R and every matrix $M \in S(\widetilde{x}, X)$.

Finally, (4) and (7) show

$$\forall\, x \in X^{k+1}: \quad g(x) \in Z + \boldsymbol{C} \cdot x \subseteq X^{k+1}. \tag{8}$$

Therefore, g is a self-mapping of the nonempty, compact and convex set X^{k+1}. Brouwer's Fixed Point Theorem implies existence of some $\widehat{x} \in X^{k+1}$ with $g(\widehat{x}) = \widehat{x} = \widehat{x} - R \cdot f(\widetilde{x} + \widehat{x})$. The regularity of R implies $f(\widetilde{x} + \widehat{x}) = 0$. (8) also shows $\widehat{x} = g(\widehat{x}) \in X^{k+2}$ and $\widehat{x} \in X^{k+\mu}$ for all $1 \leq \mu \in \mathbb{N}$. This proves i) and the theorem. ■

Computing the X^i of Theorem 1 is usually inexpensive compared to calculating a new expansion matrix. Note that for $x^0 \in X^0 \backslash X^{k+1}$ the iteration $x^{i+1} := g(x^i)$ may leave X^0. Nevertheless, Theorem 1 identifies in X^{k+1} a subset of X^0 containing a fixed point $\widehat{x}$ of g, and therefore implying $f(\widetilde{x} + \widehat{x}) = 0$.

Summarizing this means that steps 4) and 5) of Algorithm 1 can be replaced by the following. We also improved the overall iteration by using $Z + \boldsymbol{C} \cdot X^k, \emptyset \neq X^k$ as a new starting value. This is backed by proposition ii) in Theorem 1. Furthermore, the iteration in step 4) is stopped immediately if $X^{k+1} = X^k$.

4) $\boldsymbol{C} := I - R \cdot S(\widetilde{x}, X)$; $\quad Y := X$; $\quad l = 0$;
repeat
$l := l + 1$;
$Y^* := (Z + \boldsymbol{C} \cdot Y) \cap Y$;
if $Y^* \subseteq \mathrm{int}(X)$ then
"there exists an $\widehat{x} \in Y^*$ with $f(\widetilde{x} + \widehat{x}) = 0$"; stop;
if $Y^* \cap X = \emptyset$ then
"for every $\widehat{x} \in X$ holds $f(\widetilde{x} + \widehat{x}) \neq 0$"; $l = l_{\max}$;
if $Y^* = Y$ then $l = l_{\max}$
else $Y := Y^*$;
until $l = l_{\max}$;
5) $k := k + 1$; $\quad X := Y \cdot [1 - eps, 1 + eps] + [-\eta, \eta]$;
if $k \leq k_{\max}$ then calculate $S(\widetilde{x}, X)$ satisfying $(0, 1)$ and goto 4)

Algorithm 2. Improved steps 4) and 5) for Algorithm 1.

It can be shown by means of simple examples that the so-called epsilon-inflation (introduced in [12]) in the *else*-branch of step 5 in Algorithm 1 is necessary to achieve inclusion. In the linear case *eps* can be arbitrarily large, because we have either no convergence or global convergence. In this case the algorithm stops if and only if $\rho(|C|) < 1$ for every $C \in \boldsymbol{C}$ (for details see [13]). In the nonlinear case, *eps* should be large enough to avoid excessive iterations but small enough not to leave the range of attraction.

Theorem 1 allows to choose a smaller *eps*, because even if $Y \not\subseteq \mathrm{int}(X)$ an inclusion can still be achieved. There are simple examples where the use of Theorem 1 achieves a validated inclusion, whereas the original version of Algorithm 1 does not at all. Consider the well-known Rosenbrock function [1]

$$\begin{aligned} 400x_1(x_1^2 - x_2) + 2(x_1 - 1) &= 0 \\ 200x_1(x_1^2 - x_2) &= 0 \end{aligned}$$

with solution $\widehat{x} = (1, 1)^T$ and

$$\widetilde{x} = \begin{pmatrix} 0.999\,99 \\ 1.000\,40 \end{pmatrix} \quad \text{and} \quad R = \begin{pmatrix} 0.4998 & -0.9990 \\ 1 & -2.0030 \end{pmatrix}.$$

We set $X^0 := Z$, where $-R \cdot f(\widetilde{x}) \subseteq Z$ is computed by replacing real by interval operations. Then in the first iteration of steps 4, 5 ($k = 0$) in Algorithm 1 we have $\rho(|\boldsymbol{C}|) = 0.764$ but $Y \not\subseteq X$. In the following steps the spectral radius increases to 0.701, 0.828, 0.866, and

comes to 1.02 in the 9^{th} iteration, and increases rapidly from there on. Using Algorithm 1 no inclusion can be calculated. With the aid of Algorithm 2 we obtain

$$\widetilde{x} + \widehat{x} \in \begin{pmatrix} [0.999993, & 1.000006] \\ [0.999982, & 1.000016] \end{pmatrix} \tag{9}$$

in the 3^{rd} iteration. The 1^{st} iteration verifie that there is no zero in

$$\begin{pmatrix} [0.999990, & 1.000051] \\ [1.000165, & 1.000400] \end{pmatrix}.$$

Subsequent intersection narrow the inclusion (9) slightly. In other examples, the inclusion obtained by the new Algotihm was sharper due to intersection. The approach becomes valuable when the expansion matrix $S(\widetilde{x}, X)$ is expensive to compute.

2 Sparse nonlinear systems

For sparse systems an approximate inverse of the Jacobian or slope matrix cannot be used as a preconditioner because, in general, it will be full. Therefore, we rewrite the inclusion formula and use norms. *Throughout this chapter, we only use Euclidean norms.* For $M \in S(\widetilde{x}, X)$ and $x \in X$ holds

$$-R \cdot f(\widetilde{x}) + \{I - R \cdot M\} \cdot x = R \cdot \{-f(\widetilde{x}) + (R^{-1} - M)x\},$$

provided R is regular (which is demonstrated a posteriori). If X is symmetric to the origin ($X = -X$), then the following formula with norms can be used to verify existence of a zero of f:

$$\begin{aligned} r := \|R\| \cdot \{\, \|f(\widetilde{x})\| + \|\,(R^{-1} - S(\widetilde{x}, X)) \cdot X\|\,\} < \|X\| \\ \Rightarrow\ \exists\, \widehat{x} : \|\widetilde{x} - \widehat{x}\| \le r \quad \text{and } f(\widetilde{x} + \widehat{x}) = 0. \end{aligned}$$

Now R is taken to be the *optimal preconditioner*, namely $R := \text{mid}(S(\widetilde{x}, X))^{-1}$. Then $\|R\|$ is equal to the inverse of the smallest singular value of $\text{mid}(S(\widetilde{x}, X))^{-1}$:

$$\begin{aligned} r^* := \sigma_n\Big(\text{mid}(S(\widetilde{x}, X))\Big)^{-1} \cdot \{\, \|f(\widetilde{x})\| + \|\text{rad}(S(\widetilde{x}, X)) \cdot X\|\,\} < \|X\| \\ \Rightarrow\ \exists\, \widehat{x} : \|\widehat{x} - \widetilde{x}\| \le r^* \quad \text{and } f(\widetilde{x} + \widehat{x}) = 0. \end{aligned} \tag{10}$$

For more details see [13], where efficient techniques for estimating the smallest singular value of a large sparse or banded matrix are described.

If the condition in (10) is not satisfied for some X, we run into the same problem as in the dense case: An iteration is started. However, every step is expensive because it needs to calculate a new slope matrix. Using techniques of automatic differentiation in the backward mode [4] this can be accomplished at the cost of 3 to 5 function evaluations. However, the backward mode is not easy to implement. Therefore many people like to use the forward mode because of its simplicity. Truly, implementation of the forward mode reduces to replacing every operator by its corresponding gradient operator. However, the costs increase to $3n$ to $5n$ times the time for one function evaluation.

This can be reduced by observing that in (10) we only need the product of the radius of the slope matrix or Jacobian, times a vector. In [**?**] an efficient way for doing this is given. For gradients, Fischer uses $\varphi(t) := f(u + t \cdot v)$ with $\varphi'(0) = \bigtriangledown f(u)^T \cdot v$. That means, the automatic differentiation process is applied to φ, a function in one variable, and the product $\bigtriangledown f(u)^T \cdot v$ is computed directly rather than going the expensive way of computing $\bigtriangledown f(u)$ first and then multiply by v. The computational cost for differentiation of φ in the forward mode is again 3 to 5 times the time for one function evaluation of f. The same method is applicable for computing slopes.

To bound $\mathrm{rad}(S(\widetilde{x}, X)) \cdot X$, we look at one component, which writes $\mathrm{rad}(V) \cdot X$ for $V \in \mathbb{IIR}^n$. Now X yields a normwise inclusion of the error, and can therefore assumed to be symmetric to the origin: $X = [-x, x]$ with $0 < x \in \mathbb{IR}^n$. Then

$$\begin{aligned} |VX| &= |\sum_{i=1}^{n} V_i X_i| = \sum_{i=1}^{n} |V_i||X_i| = \sum_{i=1}^{n} (\mathrm{mid}(V_i) + \mathrm{rad}(V_i)) \cdot x_i \\ &= \sum_{i=1}^{n} \mathrm{mid}(V_i) x_i + \sum \mathrm{rad}(V_i) x_i = |\mathrm{mid}(V) \cdot X| + |\mathrm{rad}(V) \cdot X| \end{aligned}$$

and abbreviating $\mathrm{rad}(V) \cdot X = [-y, y]$ we obtain

$$y = |V \cdot X| - |\mathrm{mid}(V) \cdot X|.$$

Now $|V \cdot X|$ and $|\mathrm{mid}(V) \cdot X|$ can be computed together using automatic differentiation in the forward mode, and Fischer's observation at the effort of the equivalent of 6 to 10 function evaluations of f. This compares to $3n$ to $5n$ function evaluation without this observation.

Acknowledgement. The author likes to thank Prof. Dr. J. Albrecht, Clausthal-Zellerfeld, who inspired these notes with his very interesting talk in Hamburg in late 1994 (see [**?**] and [2]). The author is also indepted to Dr. C. Jansson and Prof. Dr. M. Plum for fruitful discussions.

References

[1] J.P. Abbott and R.P. Brent. Fast Local Convergence with Single and Multistep Methods for Nonlinear Equations. *Austr. Math. Soc. 19 (Series B)*, pages 173–199, 1975.

[2] J. Albrecht. Fehlerschranken und Konvergenzbeschleunigung bei einer monotonen oder alternierenden Iterationsfolge. *Numerische Mathematik 4*, pages 196–208, 1962.

[3] G. Alefeld and J. Herzberger. *Einführung in die Intervallrechnung.* B.I. Wissenschaftsverlag, 1974.

[4] A. Griewank. *On Automatic Differentiation*, volume 88 of *Mathematical Programming.* Kluwer Academic Publishers, Boston, 1989.

[5] O. Knüppel. PROFIL / BIAS — A Fast Interval Library. *Computing*, 53:277–287, 1994.

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

[7] R. Krawczyk and A. Neumaier. Interval Slopes for Rational Functions and Associated Centered Forms. *SIAM J. Numer. Anal.*, 22(3):604–616, 1985.

[8] R.E. Moore. A Test for Existence of Solutions for Non-Linear Systems. *SIAM J. Numer. Anal. 4*, pages 611–615, 1977.

[9] R.E. Moore. *Methods and Applications of Interval Analysis.* SIAM, Philadelphia, 1979.

[10] A. Neumaier. *Interval Methods for Systems of Equations.* Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1990.

[11] L.B. Rall. Automatic Differentiation: Techniques and Applications. In *Lecture Notes in Computer Science 120.* Springer Verlag, Berlin-Heidelberg-New York, 1981.

[12] S.M. Rump. *Kleine Fehlerschranken bei Matrixproblemen.* PhD thesis, Universität Karlsruhe, 1980.

[13] S.M. Rump. Verification Methods for Dense and Sparse Systems of Equations. In J. Herzberger, editor, *Topics in Validated Computations — Studies in Computational Mathematics*, pages 63–136, Elsevier, Amsterdam, 1994.