TOLERANCE ANALYSIS WITH INTERVAL ARITHMETIC

by A. Neumaier
Institut für Angewandte Mathematik
Universität Freiburg
West - Germany

This paper describes an interval arithmetic approach to tolerance analysis, i.e. the computation of a domain for the variables of a function $F : D \subseteq \mathbb{R}^n \to \mathbb{R}^m$ such that the image of every point in the domain lies in a prescribed range.

A problem of considerable practical interest is the estimation of tolerances for certain parameters of a device which guarantee an intended behaviour of the device within prescribed bounds. In mathematical terms, suppose that we know a function F defined on $D \in \mathbb{R}^n$ and with values in $\mathbb{R}^m$ (where, usually, $m \geq n$), and an interval vector $c \in \mathbb{I}\mathbb{R}^n$. Wanted is a computable subset (as large as possible, but easy to describe) of the set

$$T_F(c) := \{\tilde{x} \in D \,|\, F(\tilde{x}) \in c\} \tag{1}$$

of preimages of elements of c. The interpretation is that $F(\tilde{x})$ describes the behaviour of the device with parameter vector $\tilde{x}$,

MOS Classification: 65G10

the midpoint $\check{c}$ of c is the intended behaviour, and the radius $\rho(c)$ of c is the maximal allowed deviation from the intended behaviour.

In Section 1 we consider the case where F is linear as a special case of the more general problem of finding inner approximations of linear interval equations. The general, nonlinear case is then treated in Section 2. The notation follows Neumaier [10],[11].

1. Inner approximations of linear interval equations

As in [11], we denote by

$$\Sigma(A,b) := \{\tilde{x} \in \mathbb{R}^n \mid \tilde{A}\tilde{x} = \tilde{b} \quad \text{for some } \tilde{A} \in A,\ \tilde{b} \in b\} \qquad (2a)$$

the solution set[1] of a system of linear interval equations with coefficient matrix $A \in \mathbb{IR}^{m\times n}$ and righthand side $b \in \mathbb{IR}^m$. Equivalently, the solution set can be expressed as

$$\Sigma(A,b) = \{\tilde{x} \in \mathbb{R}^n \mid A\tilde{x} \cap b \neq \emptyset\}, \text{ or} \qquad (2b)$$

$$\Sigma(A,b) = \{\tilde{x} \in \mathbb{R}^n \mid 0 \in b - A\tilde{x}\} . \qquad (2c)$$

We also define the restricted solution set[2]

$$\Sigma_0(A,b) := \{\tilde{x} \in \mathbb{R}^n \mid A\tilde{x} \subseteq b\}; \qquad (3)$$

this is always a subset of $\Sigma(A,b)$, and agrees with $\Sigma(A,b)$ if A is thin (i.e. contains only one real matrix).

1) "outer solution" in [13,17].

2) "inner solution" in [13,17], "set of inner solutions" in [15].

The relations to the tolerance problem (1) are as follows. If F is linear and $D = \mathbb{R}^n$ then $F(\tilde{x})$ can be written in the form $F(\tilde{x}) = a + A\tilde{x}$, where $a = F(0) \in \mathbb{R}^m$ and $A = F'(0) \in \mathbb{R}^{m\times n}$. Therefore

$$F(\tilde{x}) \in c \Longleftrightarrow A\tilde{x} \in b := c - a,$$

so that

$$T_F(c) = \Sigma_0(A,b) = \Sigma(A,b)$$

since A is thin. Thus the linear tolerance problem is equivalent to finding a computable subset of the solution set of a system of linear interval equations with thin coefficient matrix. (In contrast to this, most papers on linear interval equations consider only the problem of finding a computable set enclosing this solution set; cf. the survey [12].)

If, as happens often in practice, F (and hence the coefficients of $a = F(0)$ and $A = F'(0)$) are only known within certain bounds the appropriate interpretation of the tolerance problem is to restrict attention to the subset of those $\tilde{x} \in D$ for which the computable enclosure for $F(\tilde{x})$ is contained in c. In the linear case, this translates to consideration of $\Sigma_0(A,b)$, where now A is an interval matrix containing the unknown $F'(0)$, and $b = c \overset{\circ}{-} a$ where a is an interval vector containing the vector $F(0)$; here $\overset{\circ}{-}$ is the inner subtraction defined as

$$c \overset{\circ}{-} a = [\underline{c} - \underline{a},\ \overline{c} - \overline{a}],$$

where the result is interpreted as the empty set if $\underline{c} - \underline{a} \not\leq \overline{c} - \overline{a}$.

Since now A is no longer thin, we cannot equate $\Sigma_0(A,b)$ and $\Sigma(A,b)$; thus the problem is now to find a computable subset of the restricted solution set $\Sigma_0(A,b)$.

Therefore, we shall distinguish between an inner approximation [3] of the system in question, defined as an interval vector $x \in \mathbb{I}\mathbb{R}^n$ satisfying

$$x \subseteq \Sigma(A,b),$$

and a restricted inner approximation, defined as an interval vector $x \in \mathbb{I}\mathbb{R}^n$ satisfying

$$x \subseteq \Sigma_0(A,b).$$

Inner approximations for linear interval equations were first discussed by Chartres [2] in connection with Gauss elimination (cf. also Wisskirchen [16], and, for matrix inversion, Klein [6]). Then Kartheus [5] noticed that better results can be obtained by using the characterization of the solution set $\Sigma(A,b)$ by Oettli and Prager [14] to find an inner approximation symmetric around the solution of the midpoint equation $\check{A}\tilde{x} = \check{b}$. He noticed the non-uniqueness of maximal inner approximations (in terms of order by inclusion) and suggests to improve approximations by applying a gradient method to maximize the volume of the interval vector. (However, the volume has little meaning for most problems, and in practice one would have to consider objective functions more appropriate to the problem.)

Unfortunately, if the righthand side b contains some thin (or almost thin) component, the midpoint solution may give very poor results.

3) "Inneninterval" in [5].

For example, if

$$A = \begin{pmatrix} [2,4] & [-1,1] \\ [-1,1] & [2,4] \end{pmatrix}, \quad b = \begin{pmatrix} [-3,3] \\ 0 \end{pmatrix} \qquad (4)$$

then $\Sigma(A,b)$ is the butterfly-shaped region drawn in Fig. 1, and the midpoint solution (at the origin) is a boundary point of $\Sigma(A,b)$. Thus Kartheus' method only yields the point $\tilde{x} = 0$ as inner solution.

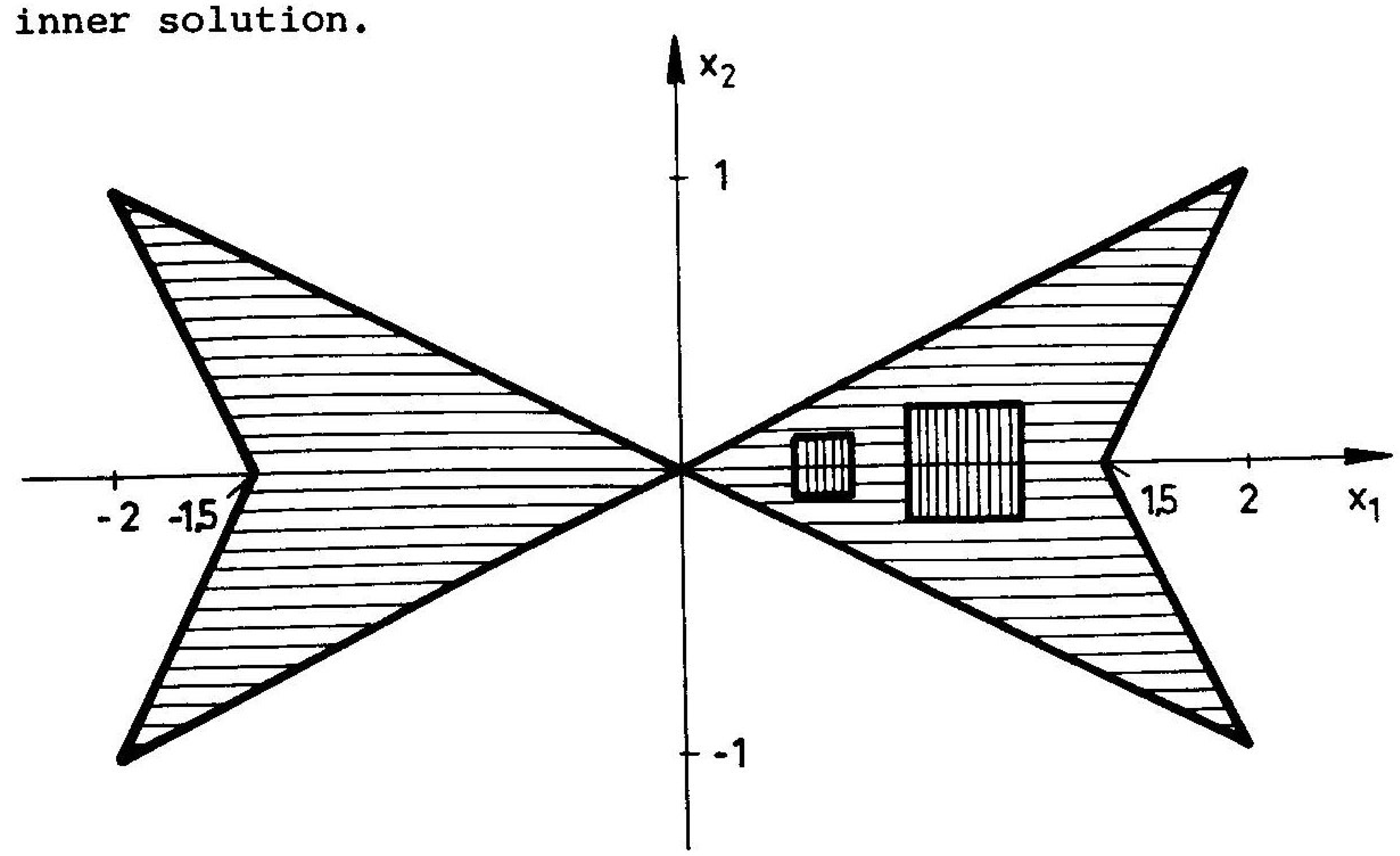

Fig. 1. A 'difficult' solution set and two inner approximations.

The natural modification of Kartheus' procedure is to use the Oettli-Prager bounds with an arbitrary interior solution point $\tilde{x}$. However, it is neater to use Beeck's [1] formulation of the Oettli-Prager bounds, given in equations (2b),(2c) above. Noting that (2c) implies

$$\tilde{x} \in \operatorname{int} \Sigma(A,b) \Longleftrightarrow 0 \in \operatorname{int}(A\tilde{x} - b)$$

(where int denotes the interior of a set), we have to start with a vector $\tilde{x} \in \mathbb{R}^n$ such that for its residual

$$r := A\tilde{x} - b,\ 0 \in \text{int } r. \qquad (5)$$

Unfortunately, it is not clear how to find such an $\tilde{x}$ if an approximate midpoint solution doesn't work; but trial and error in a suitable neighbourhood of the midpoint solution usually works quickly.

With $\tilde{x}$ satisfying (5), we choose an interval vector $d \in \mathbb{IR}^n$ with $0 \in d$ and determine the largest positive number α with the property

$$[-\alpha,\alpha](Ad) \subseteq r; \qquad (6)$$

in computational terms,

$$\alpha = \min\{|Ad|_i^{-1} \min(|\underline{r}_i|, |\overline{r}_i|) \ \ |(Ad)_i \neq 0\}. \qquad (6a)$$

Clearly, α is welldefined (and positive by (5)) unless $Ad = 0$ when any positive α satisfies (6).

<u>Proposition</u>. If (5) and (6) hold then

$$\tilde{x} + \alpha d \subseteq \Sigma(A,b). \qquad (7)$$

<u>Proof</u>. If $\tilde{z} \in \tilde{x} + \alpha d$ then $\tilde{z} = \tilde{x} + \alpha\tilde{d}$ for some $\tilde{d} \in d$; hence the residual $r' := A\tilde{z} - b$ satisfies

$$|\check{r}'| = |\check{A}\tilde{z} - \check{b}| \leq |\check{A}\tilde{x} - \check{b}| + \alpha|\check{A}\tilde{d}| = |\check{r}| + \alpha|\check{A}\tilde{d}|,$$

$$\rho(r') = \rho(A)|\tilde{z}| + \rho(b) \geq \rho(A)|\tilde{x}| + \rho(b) - \alpha\rho(A)|\tilde{d}|$$

$$= \rho(r) - \alpha\rho(A\tilde{d}).$$

Hence

$$\rho(r') - |\check{r}'| \geq \rho|r| - |\check{r}| - \alpha(\rho(A\tilde{d}) + |\check{A}\tilde{d}|)$$

$$= \rho(r) - |\check{r}| - \alpha|A\tilde{d}| \geq \rho(r) - |\check{r}| - \alpha|Ad|.$$

But (6) is equivalent with $|\check{r}| + \alpha|Ad| \leq \rho(r)$; hence $\rho(r') - |\check{r}'| \geq 0$, which shows that $0 \in r'$. Therefore $\tilde{z} \in \Sigma(A,b)$, and since $\tilde{z} \in \tilde{x} + \alpha d$ was arbitrary, (7) follows. □

In absense of other information, the vector d with $d_i = [-1,1]$ for all i (for absolute tolerance errors) or $d_i = [-1,1]|\tilde{x}_i|$ for all i (for relative tolerance errors) is appropriate. In the presence of roundoff errors, the resulting inner approximation remains valid if r and Ad are rounded outwards, α is rounded downwards, and $\tilde{x} + \alpha d$ is rounded <u>inward</u> (this requires a nonstandard implementation of interval arithmetic, cf. Chartres [2], Krückeberg [9]).

<u>Example</u>. For the linear system determined by (4), the choice

$$\tilde{x} = \begin{pmatrix} 0.5 \\ 0 \end{pmatrix},\ d = \begin{pmatrix} [-1,1] \\ [-1,1] \end{pmatrix} \text{ yields } r = \begin{pmatrix} [\ -2,5\ \ ] \\ [-0.5,0.5] \end{pmatrix},\ Ad = \begin{pmatrix} [-5,5] \\ [-5,5] \end{pmatrix},$$

$\alpha = 0.1$, giving the inner approximation $\tilde{x} + \alpha d = \begin{pmatrix} [0.4,0.6\ ] \\ [-0.1,0.1] \end{pmatrix}$.

Replacing $\tilde{x}$ by $\begin{pmatrix}1\\0\end{pmatrix}$ gives instead $r = \begin{pmatrix}[-1,7]\\ [-1,1]\end{pmatrix}$, $\alpha = 0.2$ and as inner approximation the vector $\tilde{x} + \alpha d = \begin{pmatrix}[0.8,1.2]\\ [-0.2,0.2]\end{pmatrix}$. The comparison with Fig. 1 shows that the inner approximations computed cannot be regarded as optimal since they are not maximal with respect to inclusion. So there is room for further improvements of the method. One also sees that the approximation obtained depends much on the choice of $\tilde{x}$. (Less pronounced, this phenomenon is still observed in problems of practical nature where all radii are small.)

The restricted solution set $\Sigma_0(A,b)$ was first discussed by Beeck [1] who showed that the intersection of $\Sigma_0(A,b)$ with each orthant is convex and thus its hull can be found by linear programming. Rohn [15 (cf. also Nuding [17]) simplifies Beeck's approach by showing that $\Sigma_0(A,b)$ itself is convex and gives a corresponding linear program. Deif [3] also suggests a hull approach in view of the non-uniqueness of restricted inner approximations. However, the hull of $\Sigma_0(A,b)$ may considerably overestimate the admissible tolerances, and can thus not be regarded as a satisfactory solution of the linear tolerance problem.

Deif also noticed that, even for regular, square A, the restricted solution set can be empty, a trivial example being $A = [1,2]$, $b = [2,3]$ $(m = n = 1)$. Moreover, he notes that

$$\Sigma_0(A,b) \subseteq \Sigma(\underline{A},b) \cap \Sigma(\overline{A},b),$$

with equality if the righthand side contains only nonnegative (or only nonpositive) vectors. That this sign condition is required

can be seen from the example

$$A = \begin{pmatrix}[2,4] & [1,2]\\ [3,6] & [2,4]\end{pmatrix}, \quad b = \begin{pmatrix}[1,2]\\ [1,2]\end{pmatrix}$$

where $\Sigma_0(A,b)$ is empty but the righthand side contains the vector $\begin{pmatrix}1\\-1\end{pmatrix}$.

Restricted inner approximations can be computed by modifying the approach to inner approximations slightly. Clearly we also have to assume knowledge of an interior point $\tilde{x} \in \Sigma_0(A,b)$, a condition equivalent with

$$A\tilde{x} \subseteq \text{int } b. \tag{8}$$

Note that this forces b to be thick (i.e. $\rho(b) > 0$), a natural assumption in tolerance problems. To find such a point $\tilde{x}$, a natural trial point is still an approximate midpoint solution (or a point close to it if this fails). Again we choose an interval vector $d \in \mathbb{IR}^n$ with $0 \in d$, determine now the largest positive number α such that

$$A\tilde{x} + \alpha(Ad) \subseteq b, \tag{9}$$

and obtain a restricted inner approximation

$$\tilde{x} + \alpha d \subseteq \Sigma_0(A,b); \tag{10}$$

indeed, for all $\tilde{d} \in d$,

$$A(\tilde{x} + \alpha\tilde{d}) \subseteq A(\tilde{x} + \alpha d) \subseteq A\tilde{x} + \alpha A d \subseteq b.$$

Example. Suppose that $\check{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ is accurate to within 1% relative error, and $\check{b} = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$ has to be obtained to within 2% relative error. What are admissible tolerances for the variables x? We have

$$A = \begin{pmatrix} [1.98,2.02] & [0.99,1.01] \\ [0.99,1.01] & [1.98,2.02] \end{pmatrix}, \quad b = \begin{pmatrix} [6.86,7.14] \\ [7.84,8.16] \end{pmatrix}.$$

The midpoint solution is $\tilde{x} = \check{A}^{-1}\check{b} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$, and for $d = \begin{pmatrix} [-1,1] \\ [-1,1] \end{pmatrix}$ we get

$$A\tilde{x} = \begin{pmatrix} [6.93,7.07] \\ [7.92,8.08] \end{pmatrix}, \quad Ad = \begin{pmatrix} [-3.03,3.03] \\ [-3.03,3.03] \end{pmatrix}.$$

Hence (9) requires $\alpha \leq 0.07/3.03 = 0.0231 \ldots$, and with the downward rounded value $\alpha = 0.023$ we get the tolerance interval

$$\tilde{x} + \alpha d = \begin{pmatrix} [1.977,2.023] \\ [2.977,3.023] \end{pmatrix}$$

We end this section by pointing out that there is no essential difference between the square case $(m = n)$ and the overdetermined case $(m > n)$; the only change is that for $m > n$ one has to determine $\tilde{x}$ by approximately "solving" the overdetermined midpoint equation $\check{A}\tilde{x} = \check{b}$ (e.g. by least squares). Note also that the tolerance problem is much simpler than the problem of enclosing the hull in that it can be solved in $O(mn)$ operations once a point $\tilde{x}$ with (8) is known.

2. The nonlinear tolerance problem

In the general case (and with interval valued F), the problem of finding a tolerance interval contained in (1) was first posed by Nuding [13]; however, no constructive approach to solving the problem was given. Later, Garloff [4] observed that

$$F(\tilde{x}) \in c \iff 0 \in G(\tilde{x}) := c - F(\tilde{x})$$

(where, if F is interval valued, the inner subtraction has to be used instead). Thus $T_F(c)$ is the zero set of a function strip G of special kind, and the tolerance problem becomes a special case of the problem of finding a computable subset of the zero set of a function strip (cf. [7],[8] for a further discussion of zero sets and for methods for their enclosure).

Garloff solves this problem for the square case $(m = n)$ by an algorithm which (for the original question to find a computable subset of (1) for a real vector valued F), amounts to the following. One starts with a vector $\tilde{x}$ satisfying.

$$F(\tilde{x}) \in \operatorname{int} c, \tag{11}$$

chooses a number $\gamma > 0$ such that the interval vector $z := \tilde{x} + \gamma[-e,e]$ (where $e = (1,\ldots,1)^T$) bounds the expected tolerance intervals, and computes a bound σ for the norms $\|F'(\tilde{z})\|_\infty$ $(\tilde{z} \in z)$. Then one iterates according to

$$x^0 = \tilde{x}, \quad x^{\ell+1} = z \cap (x^\ell + \sigma^{-1}\alpha_\ell[-e,e]),$$

where $\alpha_\ell \geq 0$ is a lower bound for the set of all

$$|F_i(\tilde{z}) - \underline{c}_i|, \quad |F_i(\tilde{z}) - \overline{c}_i| \quad (i = 1,\ldots,n)$$

over all $\tilde{z}$ on the boundary ∂x^ℓ of x^ℓ, and obtains a sequence $x^0 \subseteq x^1 \subseteq x^2 \subseteq \ldots$ of interval vectors contained in (1). The calculation of σ and the α_ℓ involves interval extensions of F and its derivative F', and the splitting of the boundary ∂x^ℓ into the 2n faces which are the interval vectors obtained from x^ℓ by replacing some entry x_i^ℓ by one of its endpoints $\underline{x}_i^\ell$ or $\overline{x}_i^\ell$.

The problem was taken up again by Deif [3] who reduced it to the linear case by linearization. However, Deif discards higher order terms and only gets interval vectors which solve the tolerance problem only approximately.

We show here that the nonlinear problem can be solved quite simply by imitating the methods for the linear case. To this end we assume the knowledge of a vector $\tilde{x}$ satisfying (11); as in the linear case, such a vector can usually be found by solving $F(\tilde{x}) = \check{c}$ approximately. Then we choose a number $\gamma > 0$ and an interval vector $d \in \mathbb{IR}^n$ with $0 \in d$, and determine an interval matrix $A \in \mathbb{IR}^{m\times n}$ such that

$$F'(\xi) \in A \text{ for all } \xi \in z := \tilde{x} + \gamma d. \qquad (12)$$

Then we compute the largest number α satisfying

$$F(\tilde{x}) + \alpha\, Ad \subseteq c,\ \alpha \leq \gamma. \qquad (13)$$

The mean value theorem (applied for each component of F separately) then implies that for every $\tilde{z} \in z$ there are vectors $\xi^i \in z$ such that

$$F_i(\tilde{z}) = F_i(\tilde{x}) + [F'(\xi^i)(\tilde{z}-\tilde{x})]_i$$

$$\in F_i(\tilde{x}) + [A(\alpha d)]_i = (F(\tilde{x}) + \alpha\, Ad)_i \subseteq c_i,$$

so that $F(\tilde{z}) \in c$. Therefore, (11), (12) and (13) imply that

$$\tilde{x} + \alpha d \in T_F(c),$$

i.e. $x = \tilde{x} + \alpha d$ is an "inner approximation" for (1). In view of $F'(\tilde{x}) \in A$, condition (13) also suggests that a natural choice for γ is the largest number satisfying

$$F(\tilde{x}) + \gamma F'(\tilde{x})\, d \subseteq c. \qquad (13a)$$

The method described here works for all dimensions m,n; moreover, an inaccurate knowledge of F (i.e. in mathematical terms, an interval valued F) does not prevent the method from giving valid results. Moreover, outward rounding of intervals, downward rounding of α in (13), and inward rounding of the result $\tilde{x} + \alpha d$ are sufficient to guarantee correctness even in the presence of roundoff error.

References

1. Beeck, H.: Über intervallanalytische Methoden bei linearen Gleichungssystemen mit Intervallkoeffizienten und Zusammenhänge mit der Fehleranalysis.Dissertation, Techn. Hochschule München (1971).

2. Chartres, B.A.: Automatic controlled precision calculations. J. Assoc. Comput. Mach. 13, 386-403 (1966).

3. Deif, A.S.: Methods of interval analysis. Lectures, delivered at the Univ. Cairo (1984).

4. Garloff, J.: Ein Verfahren zur Bestimmung der inneren Lösungen von Intervall-Gleichungssystemen. Diplomarbeit, Inst. f. Angew. Math., Univ. Heidelberg (1976).

5. Kartheus, V.: Zur intervallanalytischen Behandlung linearer Gleichungssysteme. Mitt. der GMD Bonn 16 (1972).

6. Klein, H.-O.: Inversion von parameterabhängigen Matrizen. Diplomarbeit, Inst. f. Angew. Math. u. Informatik, Univ. Bonn (1968).

7. Krawczyk, R.: Intervalliterationen zur Einschliessung einer Lösungsmenge. 'Collection scientific papers honouring Prof. Dr. K. Nickel on occasion of his 60th birthday, Part I' ed. by J. Garloff et al., Inst. f. Angew. Math., Univ. Freiburg 1984, pp. 183-211.

8. Krawczyk, R. and Neumaier, A.: Interval Newton operators for function strips. Freiburger Intervall-Berichte 85(7), 1-34 (1985).

9. Krückeberg, F.: Numerische Intervallrechnung und deren Anwendung. Int. Bericht, Inst. f. Instrum. Math., Univ. Bonn (1966).

10. Neumaier, A.: Interval iteration for zeros of systems of equations. BIT 25, 256-273 (1985).

11. Neumaier, A.: Overestimation in linear interval equations. Freiburger Intervall-Berichte 85/4, 75-91 (1985).

12. Neumaier, A.: Linear interval equations. In: 'Interval Mathematics 1985', ed. by K. Nickel. Lecture Notes Comp. Sci. 212, Springer Verlag, Berlin - Heidelberg 1986, pp.109-120.

13. Nuding, E.: Innere Lösungen von Intervallgleichungen. Paper presented at the interval meeting in Oberwolfach (1972).

14. Oettli, W. and Prager, W.: Compatibilty of approximate solution of linear equations with given error bounds for coefficients and right-hand sides. Numer. Math. 6, 405-409 (1964).

15. Rohn, J.: Inner solutions of linear interval systems. In: 'Interval Mathematics 1985', ed. by K. Nickel. Lecture Notes Comp. Sci. 212, Springer Verlag, Berlin - Heidelberg 1986, pp. 157-158.

16. Wisskirchen, P.: Ein Steuerungsprinzip der Intervallrechnung und dessen Anwendung auf den Gauss'schen Algorithmus. Berichte der GMD Bonn 20 (1969).

17. Nuding, E.: Ein einfacher Beweis der Sätze von Oettli-Prager und J. Rohn, Freiburger Intervall-Berichte 86/9, 1-3 (1986).

Address of the Author:

Dr. A. Neumaier
Institut für Angewandte Mathematik
Universität Freiburg i.Br.
Hermann-Herder-Str. 10
D-7800 Freiburg i.Br.