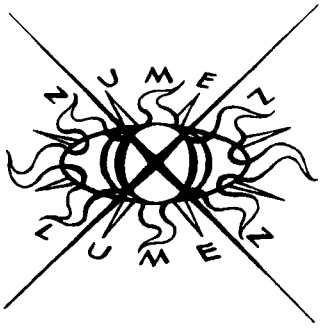
THE UNIVERSITY
OF WISCONSIN



Address:

Mathematics Research
Center
The University of Wisconsin
Madison, Wisconsin 53706
U.S.A.

23

THE UNIVERSITY OF WISCONSIN

MATHEMATICS RESEARCH CENTER

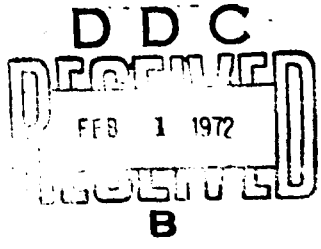
Contract No.: DA-31-124-ARO-D-462

ON THE NEWTON METHOD IN INTERVAL ANALYSIS

Karl Nickel

MRC Technical Summary Report # 1136
December 1971

Received March 25, 1971



Madison, Wisconsin 53706

ABSTRACT

A Newton method using interval arithmetic for an n-dimensional vector equation $f(x) = 0$ is treated. Some of the properties of the Newton operator are investigated. This leads to two criteria for the non-existence and for the existence of a zero of $f(x)$ in a given vector interval. Furthermore a posteriori error bounds for a given real vector sequence converging to a zero of $f(x)$ can be established. An example deals with the application to convex programming.

ON THE NEWTON METHOD IN INTERVAL ANALYSIS

Karl Nickel

1. Introduction.

Let $f(x)$ be a real valued n-dimensional vector function of the n-dimensional real vector x. The solutions $\hat{x}$ of the problem

$$f(x) = 0 \tag{1}$$

are treated. The Newton method for (1) defines a sequence $\{x^{(\nu)}\}$ by setting

$$\left.\begin{aligned} &x^{(0)} \quad \text{"appropriate"}, \\ &x^{(\nu+1)} := Nx^{(\nu)} \quad \text{for } \nu = 0, 1, \ldots \end{aligned}\right\} \tag{2}$$

where the Newton operator N is defined by

$$Nx := x - f'^{-1}(x)\, f(x) \tag{3}$$

(f' is the Fréchet derivative (Jacobian matrix)). Under certain assumptions ($\hat{x}$ exists, f'^{-1} exists in a neighbourhood of $\hat{x}$, $x^{(0)}$ is "close" to $\hat{x}$, ...) the sequence $\{x^{(\nu)}\}$ converges superlinearly or even quadratically to $\hat{x}$. Unfortunately this is not true for general starting vectors $x^{(0)}$, as the following one dimensional example shows:

Example: Let $0 < a < b$ and define

Sponsored by the United States Army under Contract No: DA-31-124-ARO-D-462.

$$f(x) := \begin{cases} \frac{3}{2}\frac{x}{\sqrt{a}} - \frac{1}{2}\frac{x^2}{\sqrt{a^3}} & \text{for } 0 \le x \le a\ ; \\ \sqrt{x} & \text{for } a \le x \le b\ ; \\ \sqrt[3]{3/2}\ b^{1/6}\ (x-b/3)^{1/3} & \text{for } b \le x\ ; \\ -f(-x) & \text{for } x \le 0\ . \end{cases}$$

There exists exactly one solution $\hat{x} = 0$ to (1) in $(-\infty, \infty)$. The function $f \in C_1(-\infty, \infty)$ with $f' > 0$ and $f'^{-1} > 0$, hence the operator N from (3) can always be applied with the result

$0 \le |x^{(0)}| < a$: $|x^{(\nu+1)}/x^{(\nu)}| \le 1-3(1-|x^{(0)}|/a)/(3-2|x^{(0)}|/a) < 1$, hence $\{x^{(\nu)}\}$ is always convergent with $x^{(\nu)} \to \hat{x} = 0$;

$a \le |x^{(0)}| \le b$: $x^{(2\nu)} = x^{(0)}, x^{(2\nu+1)} = -x^{(0)}$: $\{x^{(\nu)}\}$ is an alternating sequence with $|x^{(\nu)}| = |x^{(0)}|$ and is not convergent;

$b < |x^{(0)}|$: $|x^{(\nu+1)}/x^{(\nu)}| \ge 1+(1-b/|x^{(0)}|) > 1$: $\{x^{(\nu)}\}$ is always divergent with $|x^{(\nu)}| \to +\infty$ for $\nu \to \infty$.

R. E. Moore [4] first suggested replacing the real Newton method (2), (3) by the following interval Newton method:

$$\left.\begin{aligned} &x^{(0)} \in \overline{x}^{(0)} \text{ "appropriate"} , \\ &\overline{x}^{(\nu+1)} := \overline{x}^{(\nu)} \cap \overline{NT}\,(x^{(\nu)}, \overline{x}^{(\nu)}) , \\ &x^{(\nu+1)} \in \overline{x}^{(\nu+1)} \text{ arbitrarily} , \end{aligned}\right\} \tag{4}$$

where the interval Newton transformation operator is defined by

$$\overline{NT}(x, \overline{x}) := x - \overline{f}'^{-1}(\overline{x})\, f(x) . \tag{5}$$

Here $\{\overline{x}^{(\nu)}\}$ denotes a n-dimensional interval vector sequence, $\overline{f}'(\overline{x})$ is an interval extension of the Jacobian matrix.

The following surprising fact seems not to be widely known: In the one-dimensional case n = 1 the interval Newton method defined by (4) and (5) never fails! It delivers <u>always</u> two convergent sequences: $x^{(\nu)} \to \hat{x}$; $\overline{x}^{(\nu)} \to \hat{x}$ with $\hat{x}$, $x^{(\nu)} \in \overline{x}^{(\nu)}$ for $\nu = 0, 1, \ldots$, if only $f \in C_1$, if $\overline{f}'^{-1}$ exists and if $\hat{x} \in \overline{x}^{(0)}$ (see K. Nickel [5] and R. Krawczyk [3]). Hence:

a) While the real Newton method (2), (3) gives convergence only in the neighbourhood of $\hat{x}$, the interval Newton method (4), (5) provides convergence for <u>all</u> initial values $x^{(0)}$, if only $\overline{x}^{(0)}$ is chosen such that $\hat{x} \in \overline{x}^{(0)}$.

b) The interval Newton method (4), (5) provides not only an <u>approximating</u> real sequence $\{x^{(\nu)}\}$ with $x^{(\nu)} \to \hat{x}$, but computes at the same time a sequence of <u>error bounds</u> $\{\overline{x}^{(\nu)}\}$ with $\overline{x}^{(\nu)} \to \hat{x}$. In the case of the real Newton method (2), (3) no such error bounds are generally available.

Applied to the above example this shows that while the real Newton method (2), (3) failed if $|x^{(0)}| \geq a$, the interval Newton method (4), (5) gives a convergent approximating sequence $\{x^{(\nu)}\}$ and convergent error bounds $\{\overline{x}^{(\nu)}\}$, if only $0 = \hat{x} \in \overline{x}^{(0)}$, regardless of the initial value $x^{(0)}$.

Unfortunately a similar statement is not true in the case of more than one variable: if $n > 1$ then in general (4), (5) does not give two converging sequences $\{x^{(\nu)}\}$ and $\{\overline{x}^{(\nu)}\}$, as counterexamples show.

The purpose of the following paper is twofold: to investigate some of the properties of the operator $\overline{NT}$ and to find a way to guarantee convergence of the sequence $\overline{x}^{(\nu)} \to \hat{x}$. It will be shown that two very simple criteriae can be derived from the Newton operator which - when fulfilled - guarantee either the existence and uniqueness of a solution $\hat{x}$ of (1) in a given interval vector $\overline{x}$ or guarantee the nonexistence of such a solution $\hat{x}$ in $\overline{x}$. Both criteria can easily be programmed for a computer. In order to get unrestricted convergence of $\{\overline{x}^{(\nu)}\}$, a new approach is recommended: Since the definition of the real sequence $\{x^{(\nu)}\}$ by (4) does not always lead to a convergent series, it is suggested to compute this real sequence $\{x^{(\nu)}\}$ by any method providing convergence $x^{(\nu)} \to \hat{x}$. Then the sequence $\{\overline{x}^{(\nu)}\}$ constructed a posteriori by (4) is a sequence of error bounds, i.e. $x \in \overline{x}^{(\nu)}$ and it follows easily that it always converges, i.e. $\lim_{\nu \to \infty} \overline{x}^{(\nu)} = \hat{x}$. This leads to very satisfying a posteriori error bounds in linear algebra problems, for eigen-

value problems, for the finding of roots of polynomials etc. As a special example it will be shown, that this approach can successfully be used for the purpose of convex programming. A simple example will show how it works. Since the definition and use of the interval arithmetic and interval analysis is not yet known widely enough, special emphasis has been given to listing all assumptions and definitions.

In what follows a slightly different interval Newton operator $\overline{N}(x, \overline{x})$ will be used, which is a generalization of $\overline{NT}(x, \overline{x})$ from (5). This goes back to an idea of E. R. Hansen [2].

2. Notations.

Let $\mathbb{R}$ denote the field of real numbers. Let $\alpha, \beta \in \mathbb{R}$ with $\alpha < \beta$ and denote by $\overline{\gamma} := [\alpha, \beta]$ the closed interval with endpoints α and β. Let $\mathbb{I}(\overline{\gamma})$ be the set of all closed intervals in $\overline{\gamma}$ with endpoints in $\mathbb{R}$; denote analoguously by $\mathbb{I}(\mathbb{R})$ the set of all intervals in $\mathbb{R}$.

Let $\xi, \eta, \zeta \in \mathbb{R}$ with $\eta < \zeta$ and let $\overline{\xi} := [\eta, \zeta] \in \mathbb{I}(\mathbb{R})$. The special function intval $(\xi, \overline{\xi})$ with interval: $\mathbb{R} \times \mathbb{I}(\mathbb{R}) \to \mathbb{I}(\mathbb{R})$ is defined by

$$\text{intval } (\xi, \overline{\xi}) := \begin{cases} \overline{\xi} & \text{if } \xi \in \overline{\xi}\ ; \\ [\eta, \xi] & \text{if } \xi > \zeta\ ; \\ [\xi, \zeta] & \text{if } \xi < \eta\ , \end{cases}$$

i.e. intval $(\xi, \overline{\xi})$ is the smallest interval in $\mathbb{I}(\mathbb{R})$ containing

both ξ and $\overline{\xi}$.

Let $\mathbb{R}^n$ be the n-dimensional Euclidean vector space with elements a, b, ..., z and the corresponding components a_ν, b_ν, ..., $z_\nu \in \mathbb{R}$ for $\nu = 1(1)n$. Interval vectors and interval matrices are vectors and matrices with intervals from $\mathbb{I}(\mathbb{R})$ as components. In what follows they will be distinguished from real vectors and real matrices by a bar; hence x denotes a real vector, $\overline{x}$ an interval vector; X resp. $\overline{X}$ is a real resp. interval matrix. An $n \times n$ interval matrix $\overline{A}$ is said to be invertible and to have an inverse $\overline{B}$, iff any real $n \times n$ matrix $A \in \overline{A}$ has an inverse A^{-1} and if $A^{-1} \in \overline{B}$. Kindly note that an interval inverse in general is not unique. Let $\mathbb{I}^n(\mathbb{R}^n)$ be the set of all interval vectors $\overline{a}$, $\overline{b}$, ..., $\overline{z}$ with the corresponding components $\overline{a}_\nu$, $\overline{b}_\nu$, ..., $\overline{z}_\nu \in \mathbb{I}(\mathbb{R})$ for $\nu = 1(1)n$. Since real vectors can be treated as degenerate interval vectors it follows that $\mathbb{R}^n \subset \mathbb{I}^n(\mathbb{R}^n)$. Let $x \in \mathbb{R}^n$, $\overline{x} \in \mathbb{I}^n(\mathbb{R}^n)$ with components x_ν and $\overline{x}_\nu$ for $\nu = 1(1)n$. The special function Intval $(x, \overline{x})$ with Intval : $\mathbb{R}^n \times \mathbb{I}^n(\mathbb{R}^n) \to \mathbb{I}^n(\mathbb{R}^n)$ is the n-dimensional analogue of intval $(\xi, \overline{\xi})$ and is defined by $\overline{y}$:= Intval $(x, \overline{x}) \in \mathbb{I}^n(\mathbb{R}^n)$ has the components $\overline{y}_\nu$:= Intval $(x_\nu, \overline{x}_\nu)$ for $\nu = 1(1)n$. The interval arithmetic operations in $\mathbb{I}^n(\mathbb{R}^n)$ are defined as usual (see R. E. Moore [4]); they will be not distinguished from the real arithmetic operations since no confusion should occur.

Let $\overline{u} \in \mathbb{I}^n(\mathbb{R}^n)$ be the special interval vector with components $\overline{u}_\nu := \overline{\gamma} := [\alpha, \beta]$ for $\nu = 1(1)n$. $\mathbb{I}^n(\overline{u})$ is the set of all interval vectors in $\overline{u}$ (i.e. whose components are contained in $\overline{\gamma}$).

Let $f(x) = f(x_1, x_2, \ldots, x_n)$ with $f: \bar{u} \to \mathbb{R}^n$ and $\bar{f}(\bar{x}) = \bar{f}(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)$ with $\bar{f}: \mathbb{I}^n(\bar{u}) \to \mathbb{I}^n(\mathbb{R}^n)$ denote a real vector function and an interval vector function respectively. The corresponding components are denoted by $f_\mu(x) = f_\mu(x_1, \ldots, x_n)$ and $\bar{f}_\mu(\bar{x}) = \bar{f}_\mu(\bar{x}_1, \ldots, \bar{x}_n)$ respectively with $f_\mu: \bar{u} \to \mathbb{R}$ and $\bar{f}_\mu: \mathbb{I}^n(\bar{u}) \to \mathbb{I}(\mathbb{R})$. The function $\bar{f}(\bar{x})$ is called an <u>interval extension</u> of $f(x)$ on $\bar{u}$, if

$$\bar{f}(x) := f(x) \qquad \text{for all } x \in \bar{u};$$

and if

$$\bar{f}(\bar{x}) \subset \bar{f}(\bar{y}) \quad \text{for } \bar{x} \subset \bar{y} \quad \text{and for all } \bar{x}, \bar{y} \in \bar{u}\ . \tag{6}$$

If $f(x)$ is a rational function then an interval extension $\bar{f}(\bar{x})$ of $f(x)$ (the "natural" extension) is obtained by replacing the real variable x by an interval variable $\bar{x}$ and the real arithmetic operators in $f(x)$ by the corresponding interval arithmetic operators.

Let $f: \bar{u} \to \mathbb{R}^n$ and let $f \in C^1(\bar{u})$. Let the (real) $n \times n$ matrix $\phi(x, y)$ have the components $\varphi_{\mu\nu} = \varphi_{\mu\nu}(x_1, x_2, \ldots, x_\nu, y_\nu, y_{\nu+1}, \ldots, y_n)$ for $\mu, \nu = 1(1)n$, which are defined as divided differences by ($f_{\mu\nu} := \partial f_\mu / \partial x_\nu$)

$$\varphi_{\mu\nu} := \begin{cases} \dfrac{f_\mu(x_1,\ldots,x_{\nu-1},y_\nu,y_{\nu+1},\ldots,y_n) - f_\mu(x_1,\ldots,x_{\nu-1},x_\nu,y_{\nu+1},\ldots,y_n)}{y_\nu - x_\nu} & \\ \qquad\qquad \text{for } x_\nu \neq y_\nu & \\ f_{\mu\nu}(x_1,\ldots,x_{\nu-1},x_\nu,y_{\nu+1},\ldots,y_n) \qquad \text{for } x_\nu = y_\nu\ . & \end{cases}$$

The matrix $\phi(x, y)$ is continuous for $x, y \in \bar{u}$. Define the class of functions

$$\mathfrak{F} := \{f(x) \in C^1(\bar{u}) \mid \phi^{-1}(x, y) \text{ exists with}$$

$$\phi^{-1}(x, y) \in \bar{B}(x, \bar{x}) \subset \bar{B}^* \qquad (7)$$

for all $x, y \in \bar{x} \subset \bar{u}$, where $\bar{B}(x, \bar{x})$ has the property (6) with respect to the variable $\bar{x}$ and
$\bar{B}: \bar{u} \times \mathbb{I}^n(\bar{u}) \to \mathbb{I}^n(\mathbb{R}^n) \times \mathbb{I}^n(\mathbb{R}^n)$ and
$\bar{B}^* \in \mathbb{I}^n(\mathbb{R}^n) \times \mathbb{I}^n(\mathbb{R}^n)\}$.

3. Results.

It is in general difficult to decide if $\phi^{-1}(x, y)$ exists for a given function $f(x)$ and if there are bounds $\bar{B}(x, \bar{x})$ and $\bar{B}^*$ such that (7) holds. The following criterion provides a sufficient condition which can easily be checked on a computer:

Criterion: Let $\bar{f}_{\mu\nu}$ for $\mu, \nu = 1(1)n$ be interval extensions of $f_{\mu x_\nu} := \partial f_\mu / \partial x_\nu$ on $\bar{u}$. Let the interval matrix $\bar{F}'(x, \bar{x})$ have the components $\bar{f}_{\mu\nu}(x_1, \ldots, x_{\nu-1}, \bar{x}_\nu, \ldots, \bar{x}_n)$. Let $\bar{F}'(\bar{u}, \bar{u})$ be nonsingular. Then $f \in \mathfrak{F}$ with $\bar{B}(x, \bar{x}) := \bar{F}'^{-1}(x, \bar{x})$ and $\bar{B}^* := \bar{F}'^{-1}(\bar{u}, \bar{u})$.

Proof: By the (one dimensional) mean value theorem there exist real numbers $z_{\mu\nu} \in [\min(x_\nu, y_\nu), \max(x_\nu, y_\nu)] \subset \bar{u}$ such that for $\mu, \nu = 1(1)n$

$$\varphi_{\mu\nu} = \varphi_{\mu\nu}(x_1, \ldots, x_\nu, y_\nu, \ldots, y_n)$$

$$= f_{\mu\nu}\ (x_1, \ldots, x_{\nu-1}, z_{\mu\nu}, y_{\nu+1}, \ldots, y_n)$$

$$= \bar{f}_{\mu\nu}\ (x_1, \ldots, x_{\nu-1}, z_{\mu\nu}, y_{\nu+1}, \ldots, y_n)$$

$$\in \bar{f}_{\mu\nu}\ (x_1, \ldots, x_{\nu-1}, \bar{x}_\nu, \bar{x}_{\nu+1}, \ldots, \bar{x}_n)$$

$$\subset \bar{f}_{\mu\nu}(\bar{x}) \subset \bar{f}_{\mu\nu}(\bar{u})\ .$$

Hence $\phi(x, y) \in \bar{F}'(x, \bar{x}) \subset \bar{F}'(\bar{x}, \bar{x}) \subset \bar{F}'(\bar{u}, \bar{u})$. Since $\bar{F}'^{-1}(\bar{u}, \bar{u})$ exists, there exist also the other inverses with

$$\phi^{-1}(x, y) \in \bar{B}(x, \bar{x}) := \bar{F}'^{-1}(x, \bar{x}) \subset \bar{F}'^{-1}(\bar{x}, \bar{x}) \subset \bar{B}^*$$

$$:= \bar{F}'^{-1}(\bar{u}, \bar{u})\ .$$

The basic idea of this criterion comes from a paper from E. R. Hansen [2].

Lemma: Let $f \in \mathfrak{F}$ and let $x, y \in \bar{u}$. Then the following identity holds:

$$f(y) = f(x) + \phi(x, y)\ (y - x)\ . \tag{8}$$

The proof is very simple: By using the definition of the $\varphi_{\mu\nu}$ one gets for $\mu, \nu = 1(1)n$:

$$f_\mu(y) = f_\mu(y_1, y_2, y_3, \ldots, y_n)$$

$$= f_\mu(x_1, y_2, y_3, \ldots, y_n) + (y_1 - x_1)\varphi_{\mu 1}(x_1, y_1, y_2, \ldots, y_n)$$

$$= f_\mu(x_1, x_2, y_3, \ldots, y_n) + (y_1 - x_1)\varphi_{\mu 1}(x_1, y_1, y_2, \ldots, y_n)$$

$$+ (y_2 - x_2)\varphi_{\mu 2}(x_1, x_2, y_2, \ldots, y_n)$$

$$= \ldots$$

$$= f_\mu(x) + \sum_{\nu=1}^{n} (y_\nu - x_\nu)\, \varphi_{\mu\nu}(x_1, \ldots, x_\nu, y_\nu, \ldots, y_n) \ .$$

These are the components of (8).

Definition (Interval Newton Operator):

Let $f \in \mathfrak{F}$ and $x \in \overline{x} \subset \overline{u}$. Let the interval Newton operator $\overline{N}(x, \overline{x})$ with $\overline{N}$: $\overline{u} \times \mathbb{I}^n(\mathbb{R}^n) \to \mathbb{I}^n(\mathbb{R}^n)$ be defined by

$$\overline{N}(x, \overline{x}) := x - \overline{B}(x, \overline{x})\, f(x) \ . \tag{9}$$

Theorem: Let $x \in \overline{x} \subset \overline{u}$, let $f \in \mathfrak{F}$, let $\overline{N}(x, \overline{x})$ be defined by (9) and let $\hat{x} \in \overline{u}$ with $f(\hat{x}) = 0$. Then the following 5 statements hold:

a) Any solution $\hat{x}$ of (1) in $\overline{u}$ is unique.

b) If $\hat{x} \in \overline{x}$ then $\hat{x} \in \overline{N}(x, \overline{x})$.

c) If $\overline{x} \cap \overline{N}(x, \overline{x}) = \phi$, then $\hat{x} \notin \overline{x}$.

d) If $\overline{N}(x, \overline{x}) \subset \overline{x}$ then $\hat{x} \in \overline{x}$.

e) Let the real vector sequence $\{x^{(\nu)}\}$ with $x^{(\nu)} \in \overline{u}$ for $\nu = 0, 1, \ldots$ have the property

$$\lim_{\nu \to \infty} x^{(\nu)} = \hat{x} \ .$$

Then the interval vector sequence $\{\overline{x}^{(\nu)}\}$, defined for $\nu = 0, 1, \ldots$ by

$$\left.\begin{aligned} \overline{x}^{(-1)} &:= \overline{u} \ , \\ \overline{y}^{(\nu)} &:= \text{Intval}\ (x^{(\nu)}, \overline{x}^{(\nu-1)}) \ , \\ \overline{x}^{(\nu)} &:= \overline{x}^{(\nu-1)} \cap \overline{N}(x^{(\nu)}, \overline{y}^{(\nu)}) \end{aligned}\right\} \tag{10}$$

has the two properties

$$\hat{x} \in \bar{x}^{(\nu)} \text{ for } \nu = 0, 1, \ldots, \tag{11}$$

and

$$\lim_{\nu \to \infty} \bar{x}^{(\nu)} = \hat{x} . \tag{12}$$

Proof: Ad a) Let $\hat{x}, \hat{y} \in \bar{u}$ be two zeros of $f(x)$. Then from (8) it follows that $\phi(\hat{x}, \hat{y})(\hat{y} - \hat{x}) = 0$, but since ϕ^{-1} exists for all $\hat{x}, \hat{y} \in \bar{u}$ this implies $\hat{x} = \hat{y}$.

Ad b) Put $y := \hat{x}$ in (8). This gives $0 = f(x) + \phi(x, \hat{x})(\hat{x} - x)$. Since $\phi^{-1}(x, \hat{x})$ exists and with (7) one gets immediately $\hat{x} = x - \phi^{-1}(x, \hat{x})\, f(x) \in x - \bar{B}(x, \bar{x})\, f(x) = \bar{N}(x, \bar{x})$.

Ad c) Suppose $\hat{x} \in \bar{x}$. Then from b) it follows that also $\hat{x} \in \bar{N}(x, \bar{x})$ which contradicts the assumption $\bar{x} \cap \bar{N}(x, \bar{x}) = \phi$. Hence $\hat{x} \notin \bar{x}$.

Ad d) Define for fixed $x \in \bar{x}$ the real operator $Ty := x - \phi^{-1}(x, y)\, f(x) = y - \phi^{-1}(x,y)$. Clearly T is continuous for all $y \in \bar{x}$. Furthermore because of $Ty \in x - \bar{B}(x, \bar{x})\, f(x) = \bar{N}(x, \bar{x}) \subset \bar{x}$, the operator T maps $\bar{x}$ into $\bar{x}$. Since $\bar{x}$ is a convex set, the Schauder fixpoint theorem can be applied with the result: there exists a real vector $\hat{x} \in \bar{x}$ such that $\hat{x} = T\hat{x}$. This implies $\phi^{-1}(x, \hat{x})\, f(\hat{x}) = 0$ and therefore $f(\hat{x}) = 0$, i.e. $\hat{x}$ is also a zero of $f(x)$.

Ad e) The property (11) follows immediately from b). Furthermore since $x^{(\nu)} \in \bar{y}^{(\nu)} \subset \bar{u}$ and $\bar{B}(x, \bar{x})$ has the property (6) in $\bar{x}$ it follows that

$$\overline{N}(x^{(\nu)}, \overline{y}^{(\nu)}) \in \overline{N}(x^{(\nu)}, \overline{u})$$

$$= x^{(\nu)} - \overline{B}(x^{(\nu)}, \overline{u})\ f(x^{(\nu)})$$

$$\subset x^{(\nu)} - \overline{B}^*\ f(x^{(\nu)}) ,$$

hence $\lim_{\nu\to\infty} \overline{N}(x^{(\nu)}, \overline{y}^{(\nu)}) = \hat{x}$. Therefore with $\overline{x}^{(\nu-1)} \cap \overline{N}(x^{(\nu)}, \overline{y}^{(\nu)}) \subset \overline{N}(x^{(\nu)}, \overline{y}^{(\nu)})$ it follows also that $\lim_{\nu\to\infty} \overline{x}^{(\nu-1)} \cap \overline{N}(x^{(\nu)}, \overline{y}^{(\nu)}) = \hat{x}$, which completes the proof of the theorem.

Remarks: Ad a) The class $\mathfrak{F}$ is so small that it contains only functions $f(x)$ with at most one zero $\hat{x}$ in $\overline{u}$.

Ad b) Any solution $\hat{x}$ of (1) contained in the interval vector $\overline{x}$ is also contained in the image of the Newton transformation: i.e., the Newton transformation "does not lose" any zero of $f(x)$.

Ad c) This is a criterion for the non-existence of a zero $\hat{x}$ of $f(x)$ in the interval vector $\overline{x}$. This criterion can be programmed to be checked by a computer!

Ad d) This is a criterion for the existence of a zero $\hat{x}$ of $f(x)$ in the interval vector $\overline{x}$. This criterion can also be checked for by a computer.

Ad e) If somehow a convergent real vector sequence $\{x^{(\nu)}\}$ with $x^{(\nu)} \to \hat{x}$ for $\nu \to \infty$ has been determined, the application of the Newton operator by (10) produces immediately a sequence $\{\overline{x}^{(\nu)}\}$ of bounds to $\{x^{(\nu)}\}$ and $\hat{x}$ (a posteriori error bounds), such that $x^{(\nu)} \in \overline{x}^{(\nu)}$ for $\nu = 0, 1, \ldots$. This sequence converges too toward $\hat{x}$. The speed of convergence

(linear, superlinear, quadratic, ...) can easily be related to the speed of convergence of $\{x^{(\nu)}\}$.

If $x \in \overline{x}$ then the Newton Operator $\overline{N}$ can be applied to the pair x, $\overline{x}$ and the result b) of the theorem holds. Assume that the interval vector $\overline{x}^{(\nu-1)}$ has already been computed. Since the sequence $\{x^{(\nu)}\}$ is defined independently of the sequence $\{\overline{x}^{(\nu)}\}$, no guarantee is given that the next term $x^{(\nu)} \in \overline{x}^{(\nu-1)}$ and that therefore $\hat{x} \in \overline{N}(x^{(\nu)}, \overline{x}^{(\nu-1)})$. This is the reason why the intermediate interval vector $\overline{y}^{(\nu)}$ is introduced in (10).

4. Example. Convex programming.

Let $\psi = \psi(x)$ with ψ: $\mathbb{R}^n \to \mathbb{R}$ and $\psi \in C_2(\mathbb{R}^n)$ and let ψ be bounded from below on $\mathbb{R}^n$. The unconstrained minimization problem

$$\psi(x) = \text{minimum} \tag{13}$$

is considered. Define the function

$$f(x) := \partial\psi/\partial x$$

with f: $\mathbb{R}^n \to \mathbb{R}^n$ and $f \in C_1(\mathbb{R}^n)$ and define the set

$$S := \{x \in \mathbb{R}^n \mid \psi(x) \leq \psi(x^{(0)})\}$$

for a given real vector $x^{(0)} \in \mathbb{R}^n$. Let $\hat{S}$ be a convex set containing S. Let the Hessian matrix

$$H(x) := \partial^2\psi/\partial x^2 = \partial f/\partial x$$

satisfy the condition

$$\mu(x, x) \leq (x, H(y)x) \leq \lambda(x, x)$$

with $\mu > 0$ (where $(.,.)$ is the inner product in $\mathbb{R}^n$) for any vector $x \in \mathbb{R}^n$ and $y \in \hat{S}$. Then S is bounded and the minimum problem (13) is in S a convex problem and has exactly one solution which is the only zero $\hat{x}$ of $f(x)$ in S. Furthermore the real Newton method (2), (3) applied to the function $f(x)$ and starting with $x^{(0)}$ produces a sequence $\{x^{(\nu)}\}$ with $x^{(\nu)} \in S$ which always converges to $\hat{x}$ (where perhaps a finite number of steps have to be changed in order to get always $\psi(x^{(\nu+1)}) < \psi(x^{(\nu)})$), see Goldstein [1].

Hence the above suggestion can be used, provided $f \in \mathfrak{F}$ for a certain interval vector $\overline{u}$ containing S. This being true the interval vector sequence $\{\overline{x}^{(\nu)}\}$ defined by (10) has the properties (11) and (12), i.e., provides error bounds and converges toward $\hat{x}$. To the author's knowledge to the problem (13) no other procedure is yet known which gives an interval vector sequence $\overline{x}^{(\nu)}$ with the properties (11) and (12).

Numerical example.

The following example is chosen to be as simple as possible in order to show all the steps of the above method. Let $n = 2$, $\varepsilon > 0$ and $2\psi(x_1, x_2) := x_1^2 + x_2^2 + \varepsilon\, x_1^2\, x_2^2$. Then

$$f(x_1, x_2) := \partial\psi/\partial x := \begin{pmatrix} x_1(1 + \varepsilon\, x_2^2) \\ x_2(1 + \varepsilon\, x_1^2) \end{pmatrix}$$

and

$$\partial f/\partial x := \partial^2\psi/\partial x^2 := \begin{pmatrix} 1 + \varepsilon\, x_2^2 & 2\,\varepsilon\, x_1\, x_2 \\ 2\,\varepsilon\, x_1\, x_2 & 1 + \varepsilon\, x_1^2 \end{pmatrix}.$$

It is trivial that $\hat{x}_1 = \hat{x}_2 = 0$ is the unique solution of (13) in $\mathbb{R}^2$. Let $\overline{\gamma} := [-1, 1]$ and let $\overline{u}$ have the components $\overline{u}_1 := \overline{u}_2 := \overline{\gamma}$. Define $\hat{S} := \{x \in \mathbb{R}^2 \,|\, |x_1|, |x_2| \leq \sqrt{2+\varepsilon}\}$. Then for the initial values $x_1^{(0)} := x_2^{(0)} := 1$ it follows that $\overline{u} \subset S := \{x \in \mathbb{R}^2 \,|\, \psi(x) \leq \psi(x^{(0)})\} \subset \hat{S}$. One sees easily that $\mu := 1 > 0$ and $\lambda := 1 + 6\varepsilon + 3\varepsilon^2$ can be chosen. Furthermore

$$(\partial f/\partial x)^{-1} := \begin{pmatrix} 1 + \varepsilon\, x_1^2 & -2\,\varepsilon\, x_1\, x_2 \\ -2\,\varepsilon\, x_1\, x_2 & 1 + \varepsilon\, x_2^2 \end{pmatrix} /\det\,,$$

where $\det := 1 + \varepsilon(x_1^2 + x_2^2) - 3\,\varepsilon^2\, x_1^2\, x_2^2 > 0$ for $\max(x_1^2, x_2^2) < 1/\varepsilon$, thus $\varepsilon < 1$ in $\overline{u}$. From this result it is simply shown that the real Newton method (2), (3), starting with $x^{(0)}$ gives a convergent sequence $\{x^{(\nu)}\}$ with $x^{(\nu)} \in \overline{u}$ and $x^{(\nu)} \to \hat{x}$ for $\nu \to \infty$, where $x_1^{(\nu+1)} := x_2^{(\nu+1)} := 2\,\varepsilon\, x_1^{(\nu)3}/(1 + 3\,\varepsilon\, x_1^{(\nu)2})$ for $\nu = 0, 1, \ldots$.

To apply the interval Newton method (4), (5), compute

$$\phi(x, y) := \begin{pmatrix} 1 + \varepsilon\, y_2^2 & \varepsilon\, x_1(x_2 + y_2) \\ \varepsilon\, y_2(x_1 + y_1) & 1 + \varepsilon\, x_1^2 \end{pmatrix}$$

and

$$\phi^{-1}(x,\ y)\ :=\ \begin{pmatrix} 1 + \varepsilon\, x_1^2 & -\varepsilon\, x_1(x_2 + y_2) \\ -\varepsilon\, y_2(x_1 + y_1) & 1 + \varepsilon\, y_2^2 \end{pmatrix} \Big/ \det\, \phi(x,y),$$

with $\det\, \phi(x,\ y)\ :=\ 1 + \varepsilon(x_1^2 + y_2^2) - \varepsilon^2\, x_1\, y_2(x_1\, x_2 + y_2\, x_2 + y_1\, y_2)$. Hence $\phi^{-1}(x,\ y)\ \in\ \overline{B}(x,\ \overline{x})\ :=\ \overline{\phi}^{-1}(x,\ \overline{x})$, where $\overline{\phi}^{-1}$ is the natural interval extension of ϕ^{-1}, i.e.,

$$\overline{B}(x,\ \overline{x})\ :=\ \begin{pmatrix} 1 + \varepsilon\, x_1^2 & -\varepsilon\, x_1(x_2 + \overline{x}_2) \\ -\varepsilon\, \overline{x}_2(x_1 + \overline{x}_1) & 1 + \varepsilon\, \overline{x}_2^2 \end{pmatrix} \Big/ \overline{\det\, \phi}(x,\overline{x}),$$

with $\overline{\det\, \phi}(x,\ \overline{x})\ :=\ 1 + \varepsilon(x_1^2 + \overline{x}_2^2) - \varepsilon^2\, x_1\, \overline{x}_2(x_1\, x_2 + \overline{x}_1\, x_2 + \overline{x}_1\, \overline{x}_2)$.

Finally $\overline{B}(x,\ \overline{x})\ \in\ \overline{B}^*\ :=\ \overline{B}(\overline{u},\ \overline{u})$ with

$$\overline{B}^*\ :=\ \begin{pmatrix} 1 + \varepsilon\, \overline{\gamma}^2 & 2\, \varepsilon\, \overline{\gamma}^2 \\ 2\, \varepsilon\, \overline{\gamma}^2 & 1 + \varepsilon\, \overline{\gamma}^2 \end{pmatrix} \Big/ \overline{\det\, B^*}\ ,\ \text{and}$$

$\overline{\det\, B^*}\ :=\ 1 + 2\, \varepsilon\, \overline{\gamma}^2 + 3\, \varepsilon^2\, \overline{\gamma}^4 = 1 + \varepsilon(2 + 3\varepsilon)\ [-1,\ 1] > 0$ for $0 \leq \varepsilon < 1/3$. Kindly note that the interval matrix $\overline{B}^*$ is only defined for $0 \leq \varepsilon < 1/3$, while ϕ^{-1} was defined in the much larger region $0 \leq \varepsilon < 1$. This loss is quite typical for the switch from real to interval arithmetic in the case of matrices.

To apply the above criterion one computes

$$\overline{F}'(x,\ \overline{x})\ :=\ \begin{pmatrix} 1 + \varepsilon\, \overline{x}_2^2 & 2\, \varepsilon\, x_1\, \overline{x}_2 \\ 2\, \varepsilon\, \overline{x}_1\, \overline{x}_2 & 1 + \varepsilon\, x_1^2 \end{pmatrix}\ .$$

Hence

$$\overline{F}'^{-1}(x, \overline{x}) := \begin{pmatrix} 1 + \varepsilon\, x_1^2 & -2\,\varepsilon\, x_1\, \overline{x}_2 \\ -2\,\varepsilon\, \overline{x}_1\, \overline{x}_2 & 1 + \varepsilon\, \overline{x}_2^2 \end{pmatrix} \Big/ \overline{\det F}(x, \overline{x}),$$

with $\overline{\det F}(x, \overline{x}) := (1 + \varepsilon\, x_1^2)(1 + \varepsilon\, \overline{x}_2^2) - 4\,\varepsilon^2\, x_1\, \overline{x}_1\, \overline{x}_2^2.$

This gives

$$\overline{F}'^{-1}(\overline{u}, \overline{u}) := \begin{pmatrix} 1 + \varepsilon\, \overline{\gamma}^2 & 2\,\varepsilon\, \overline{\gamma}^2 \\ 2\,\varepsilon\, \overline{\gamma}^2 & 1 + \varepsilon\, \overline{\gamma}^2 \end{pmatrix} \Big/ \overline{\det F}(\overline{u}, \overline{u}),$$

with $\overline{\det F}(\overline{u}, \overline{u}) = 1 + \varepsilon(2 + 5\varepsilon)\ [-1\ 1] > 0$ for $0 \leq \varepsilon < (\sqrt{6}-1)/5$.
By comparing these results one sees clearly, as stated in the above criterion, that $\overline{B}(x, \overline{x}) \subset \overline{F}'^{-1}(x, \overline{x})$ and $\overline{B}^* \subset \overline{F}'^{-1}(\overline{u}, \overline{u})$. Kindly note that $\overline{B}^*$ is defined for $0 \leq \varepsilon < 1/3$ while $\overline{F}'^{-1}(\overline{u}, \overline{u})$ is defined only in the still smaller interval $0 \leq \varepsilon < (\sqrt{6}-1)/5 \approx 0.29(!)$.

Table 1 shows some numerical results for $\varepsilon = 0.5$ and $\nu = 0(1)5$. We remark that the symmetry of the problem which is illustrated in the real sequence $\{x^{(\nu)}\}$ by $x_1^{(\nu)} = x_2^{(\nu)}$ is not carried over to the sequence $\{\overline{x}^{(\nu)}\}$, due to the asymmetry of $\overline{B}(x, \overline{x})$ and therefore of $\overline{N}(x, \overline{x})$. Please note also, that for the higher values of ν it is <u>not</u> true, that $x^{(\nu)} \in \overline{x}^{(\nu)}$! This means: the error interval $\overline{x}^{(\nu)}$ containing the solution $\hat{x}$ is not only an error bound for $\hat{x}$ but also an <u>improvement</u> over the approximation $x^{(\nu)}$. This very favorable property is quite typical for the use of the Newton operator, at least in the neighbourhood of the solution.

Table 1.

The first 5 iteration steps to the Newton method for $\varepsilon = 0.5$.

ν	$x_1^{(\nu)}=x_2^{(\nu)}$ *)	$\bar{x}_1^{(\nu)}$ **)	$\bar{x}_2^{(\nu)}$ **)
-1	×	[-1.000 , 1.000]	[-1.000 , 1.000]
0	1.000	[-1.000 , 0.728]	[-1.000 , 1.000]
1	0.400	[-0.155 , 0.200]	[-0.554 , 0.291]
2	$0.516_{10}-1$	$[-0.781_{10}-3$, $0.722_{10}-2$]	$[-0.116_{10}-1$, $0.852_{10}-2$]
3	$0.136_{10}-3$	$[-0.117_{10}-9$, $0.923_{10}-8$]	$[-0.150_{10}-7$, $0.135_{10}-7$]
4	$0.256_{10}-11$	$[-0.136_{10}-18$, $0.109_{10}-18$]	$[-0.163_{10}-18$, $0.109_{10}-18$]

*) only the first three significant digits given.

**) rounded to three significant digits.

REFERENCES

[1] A. A. Goldstein: Constructive Real Analysis. Harper & Row, New York 1967, page 37.

[2] E. R. Hansen: On solving systems of equations using interval arithmetic. Math. of Comp. 22, 374-384 (1968).

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

[4] R. E. Moore: Interval Analysis. Prentice-Hall, Englewood Cliffs, N. J., 1966.

[5] K. Nickel: Triplex-ALGOL and applications. In: E. R. Hansen: Topics in Interval Analysis. Oxford at the Clarendon Press, 1969, pages 10-24.

AR 70-31 Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Mathematics Research Center, University of Wisconsin, Madison, Wis. 53706	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP None

3. REPORT TITLE

ON THE NEWTON METHOD IN INTERVAL ANALYSIS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Summary Report: no specific reporting period.

5. AUTHOR(S) (First name, middle initial, last name)

Karl Nickel

6. REPORT DATE December 1971	7a. TOTAL NO. OF PAGES 19	7b. NO. OF REFS 5
8a. CONTRACT OR GRANT NO. Contract No. DA-31-124-ARO-D-462 b. PROJECT NO. None c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) #1136 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) None	

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES None	12. SPONSORING MILITARY ACTIVITY Army Research Office-Durham, N.C.

13. ABSTRACT

An interval type Newton method for vector equations in the n-dimensional Euclidean space is investigated.

DD FORM 1 NOV 68 1473

Unclassified

Security Classification

AGO 5998A