EXISTENCE VERIFICATION FOR SINGULAR ZEROS OF REAL NONLINEAR SYSTEMS

JIANWEI DIAN AND R. BAKER KEARFOTT

Abstract. Traditional computational fixed point theorems, such as the Kantorovich theorem (made rigorous with directed roundings), Krawczyk's method, or interval Newton methods use a computer's floating-point hardware computations to mathematically prove existence and uniqueness of a solution to a nonlinear system of equations within a given region of n-space. Such computations require the Jacobi matrix of the system to be nonsingular in a neighborhood of a solution. However, in previous work we showed how we could mathematically verify existence of singular solutions in a small region of complex n-space containing an approximate real solution. We verified existence of such singular solutions by verifying that the topological degree of a small region is non-zero; a non-zero topological degree implies existence of a solution in the interior of the region. Here, we show that, when the actual topological degree in complex space is odd and the rank defect of the Jacobi matrix is one, the topological degree of a small region containing the singular solution can be verified to be plus or minus one in real space. The algorithm for verification in real space is significantly simpler and more efficient. We demonstrate this efficiency with numerical experiments.

1. Introduction

Our fundamental problem is

$$\text{(1.1)}\qquad \boxed{\begin{array}{l}\text{Given } F:\boldsymbol{x}\to\mathbb{R}^n \text{ and } \boldsymbol{x}\in\mathbb{IR}^n,\ \textit{rigorously} \text{ verify:}\\ \quad\bullet\ \text{there exists a unique } x^*\in\boldsymbol{x} \text{ such that } F(x^*)=0, \text{ where}\\ \quad\bullet\ \boldsymbol{x}=\left\{(x_1,x_2,\ldots,x_n)^T\in\mathbb{R}^n\ \middle|\ \underline{x}_i\le x_i\le\overline{x}_i, 1\le i\le n\right\},\end{array}}$$

where the $\underline{x}_i$ and $\overline{x}_i$ represent constant bounds on the problem variables x_i. When the Jacobi matrix $F'(x^*)$ well-conditioned and not too quickly varying, interval computations have no trouble proving that there is a unique solution within small boxes with x^* reasonably near the center; see [5, 10, 14]. When $F'(x^*)$ is ill-conditioned or singular, in general, no computational techniques can verify the existence of a solution within a given region $\boldsymbol{x}$ of $\mathbb{R}^n$. However, in the singular case, computational but rigorous verification that a given number of true solutions exist within a region in complex space containing $\boldsymbol{x}$ is possible. In [13], we developed and experimentally validated algorithms for the case when the rank defect of the Jacobi matrix is one and the topological index is 2, while, in [11], we both generalized the theory and techniques from [13] to arbitrary topological index and presented a heuristic to efficiently determine the topological index at an approximate singular solution, prior to verification. (A non-zero topological index implies existence of a

Date: April 14, 2001.
1991 *Mathematics Subject Classification.* 65G10,65H10.
This work was supported in part by National Science Foundation grant DMS-9701540.

solution.) In [12], we outlined a possible generalization to higher rank defect, and observed that this generalization would lead to more complicated algorithms and a computational effort that grew exponentially with the size of the rank defect.

Here, we focus again on the rank-one defect case. Although we have presented algorithms for this case that succeed in general with an amount of effort that is approximately proportional to n^3, there is extra inefficiency in working in $\mathbb{C}^n$ rather than in $\mathbb{R}^n$. Furthermore, the answer obtained, namely, "there exist solutions within a small box in $\mathbb{C}^n$ containing the approximate solution $\check{x} \in \mathbb{R}^n$," says nothing about actual solutions in $\mathbb{R}^n$. However, in general, it is necessary to work in $\mathbb{C}^n$, since only in that case does existence of a solution to $F(x) = 0$ imply a non-zero topological index.

In fact, we show below that, if the rank defect of the Jacobi matrix is one and if the topological index of a point x^* in $\mathbb{C}^n$ is odd, the topological index of x^* in $\mathbb{R}^n$ must be ± 1. Furthermore, we present an algorithm for verifying the degree in real space. This algorithm is orders of magnitude faster than the verification algorithm in complex space, as we illustrate with experiments with a variable dimension problem.

Although we do not wish to repeat the development in [13] and [11], we introduce in §1.1 and §1.2 that notation and underlying theory necessary to comprehend the results in this paper. We precisely state and prove a theorem on the real topological index in §3, while the actual verification algorithm appears in §4. We present our experimental results in 5, and we summarize in §6.

1.1. **Notation.** We assume familiarity with the fundamentals of interval arithmetic; see [1, 5, 10, 14, 15] for introductory material.

Throughout, lower case denotes scalars and vectors, while upper case denotes matrices. Boldface denotes intervals, interval vectors (also called "boxes") and interval matrices. For instance, $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ denotes an interval vector, $A = (a_{i,j})$ denotes a point matrix, and $\boldsymbol{A} = (\boldsymbol{a}_{i,j})$ denotes an interval matrix. The midpoint of an interval or interval vector $\boldsymbol{x}$ will be denoted by $\mathrm{m}(\boldsymbol{x})$. Real n-space will be denoted by $\mathbb{R}^n$, while complex n-space will be denoted by $\mathbb{C}^n$.

Suppose $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is an n-dimensional real box, where $\boldsymbol{x}_k = [\underline{x}_k, \overline{x}_k]$. The non-oriented boundary of $\boldsymbol{x}$, denoted by $\boldsymbol{\partial x}$, consists of $2n$ $(n-1)$-dimensional real boxes

$$\boldsymbol{x}_{\underline{k}} \equiv (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{k-1}, \underline{x}_k, \boldsymbol{x}_{k+1}, \ldots, \boldsymbol{x}_n) \text{ and } \boldsymbol{x}_{\overline{k}} \equiv (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{k-1}, \overline{x}_k, \boldsymbol{x}_{k+1}, \ldots, \boldsymbol{x}_n),$$

where $k = 1, \ldots, n$. If $\boldsymbol{x}$ is positively oriented, then the derived orientation of $\boldsymbol{x}_{\underline{k}}$ is $(-1)^k$ and the derived orientation of $\boldsymbol{x}_{\overline{k}}$ is $(-1)^{k+1}$; see [13]. Also let

$$F_{\neg k}(\boldsymbol{x}) \quad \equiv \quad \big(f_1(\boldsymbol{x}), \ldots, f_{k-1}(\boldsymbol{x}), f_{k+1}(\boldsymbol{x}), \ldots, f_n(\boldsymbol{x})\big)$$

for $1 \leq k \leq n$.

Throughout, $\frac{\partial F}{\partial x_1 \ldots \partial x_n}(\check{x})$ denotes the Jacobi matrix of F, and $\left|\frac{\partial F}{\partial x_1 \ldots \partial x_n}(\check{x})\right|$ denotes its determinant.

1.2. **Formulas from Degree Theory.** In [13], we reviewed the topological degree in the context of this paper. Also see [2, 4, 7, 8, 16, 17]. We repeat here and in §1.3 only those properties necessary for a minimal intuition and for those properties used in the proof of our main theorem.

Theorem 1.1 ([16, p. 150], etc.)**.** *Suppose that the Jacobi matrix $F'(x)$ is nonsingular at each zero of F, and $F \neq 0$ on the boundary of $\mathbf{D}$. Then the topological degree $\mathrm{d}(F, \mathbf{D}, 0)$ is equal to the number of zeros of F in $\mathbf{D}$ at which the determinant of the Jacobian matrix $F'(x)$ is positive minus the number of zeros of F in $\mathbf{D}$ at which the determinant of the Jacobian matrix $F'(x)$ is negative.*

Theorem 1.2 (Kronecker existence theorem, [16, p. 161])**.** *Suppose $F : \boldsymbol{x} \to \mathbb{R}^n$ is continuous on the closure $\boldsymbol{x} \subset \mathbb{R}^n$ of an open domain, and suppose $\mathrm{d}(F, \boldsymbol{x}, 0) \neq 0$. Then $F(x) = 0$ has at least one solution in the interior of $\boldsymbol{x}$.*

1.3. **A Basic Degree Computation Formula.**

If we select $s \in \{-1, 1\}$, then it can be shown ([2], etc.) that $\mathrm{d}(F, \boldsymbol{x}, 0)$ is equal to the number of zeros of $F_{\neg k}$ on $\partial \boldsymbol{x}$ with positive orientation at which $\mathrm{sgn}(f_k) = s$, minus the number of zeros of $F_{\neg k}$ on $\partial \boldsymbol{x}$ with negative orientation at which $\mathrm{sgn}(f_k) = s$. The orientation of each zero can be computed by computing the sign of the determinant of the Jacobi matrix of $F_{\neg k}$ and by taking into account the orientation of the face of $\boldsymbol{x}$ on which the zero lies.

Next, we present a degree computation formula common to both this work and the work in [13] and [11]; see Theorem 2.5 of [13]. We can obtain the formula in this theorem by noticing formulas (4.12) and (4.14) in [17] and by taking the orientations of the faces of $\boldsymbol{x}$ into account. We will use this formula to derive the computational procedures in §4.

Theorem 1.3. *Suppose $F \neq 0$ on $\boldsymbol{\partial x}$, and suppose there is a p, $1 \leq p \leq n$, such that:*

1. *$F_{\neg p} \equiv (f_1, \ldots, f_{p-1}, f_{p+1}, \ldots, f_n) \neq 0$ on $\boldsymbol{\partial x}_{\underline{k}}$ or $\boldsymbol{\partial x}_{\overline{k}}$, $k = 1, \ldots, n$; and*
2. *the Jacobi matrices of $F_{\neg p}$ are non-singular at all solutions of $F_{\neg p} = 0$ on $\boldsymbol{\partial x}$.*

Then

$$\begin{aligned}\mathrm{d}(F, \boldsymbol{x}, 0) \quad = \quad & (-1)^{p-1} s \left\{ \sum_{k=1}^{n} (-1)^k \sum_{\substack{x \in \boldsymbol{x}_{\underline{k}} \\ F_{\neg p}(x)=0 \\ \mathrm{sgn}(f_p(x))=s}} \mathrm{sgn} \left| \frac{\partial F_{\neg p}}{\partial x_1 x_2 \ldots x_{k-1} x_{k+1} \ldots x_n}(x) \right| \right. \\ & \left. + \sum_{k=1}^{n} (-1)^{k+1} \sum_{\substack{x \in \boldsymbol{x}_{\overline{k}} \\ F_{\neg p}(x)=0 \\ \mathrm{sgn}(f_p(x))=s}} \mathrm{sgn} \left| \frac{\partial F_{\neg p}}{\partial x_1 x_2 \ldots x_{k-1} x_{k+1} \ldots x_n}(x) \right| \right\},\end{aligned}$$

where $s = +1$ or -1.

2. Assumptions and Choice of Box

Here, we present the basic assumptions, and also show how to choose the coordinate bounds $\boldsymbol{x}_i = [\underline{x}_i, \overline{x}_i]$ to satisfy the assumptions and enable a more efficient algorithm. When the rank of $F'(x^*)$ is $n - p$ for some $p > 0$, an appropriate preconditioner can be used to reduce $\boldsymbol{F'}(\boldsymbol{x})$ to approximately the pattern shown in Figure 1; for details, see [10] and [13]. In this paper, we assume that the system has already been preconditioned, so that it is, to within second-order terms with respect to $\mathrm{w}(\boldsymbol{x})$, of the form in Figure 1. As in [13] and [11], we also assume in this paper that $p = 1$.

$$Y\boldsymbol{F'}(\boldsymbol{x}) = \begin{pmatrix} 1 & 0 & \dots & 0 & \overbrace{*\cdots*}^{p} \\ 0 & 1 & 0\dots & 0 & *\cdots* \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 1 & *\cdots* \\ 0 & \dots & 0 & 0 & 0\dots0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & 0\dots0 \end{pmatrix}.$$

FIGURE 1. A preconditioned singular system of rank $n-p$, where "$*$" represents a non-zero element.

2.1. **The Basic Assumptions.** We assume

(1) $\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) = ([\underline{x}_1, \overline{x}_1], \dots, [\underline{x}_n, \overline{x}_n])$ is a small box constructed to be centered at an approximate solution $\check{x}$, i.e. $\mathrm{m}(\boldsymbol{x}) = (\check{x}_1, \dots, \check{x}_n)$.

(2) $\check{x}$ is near a point x^* with $F(x^*) = 0$, such that $\|\check{x} - x^*\|$ is much smaller than the width of the box $\boldsymbol{x}$, and width of the box $\boldsymbol{x}$ is small enough so that mean value interval extensions lead, after preconditioning, to a system like Figure 1, with small intervals replacing the zeros.

(3) F has been preconditioned as in Figure 1, and $F'(x^*)$ has null space of dimension 1.

Denote

$$\begin{aligned} \alpha_k &\equiv \frac{\partial f_k}{\partial x_n}(\check{x}), \qquad 1 \le k \le n-1, \\ \alpha_n &\equiv -1, \\ \Delta_1 &\equiv \left| \frac{\partial F}{\partial x_1 \dots \partial x_n}(\check{x}) \right| \\ \Delta_d &\equiv \sum_{k_1=1}^{n} \cdots \sum_{k_d=1}^{n} \frac{\partial^d f_n}{\partial x_{k_1} \dots \partial x_{k_d}}(\check{x}) \alpha_{k_1} \dots \alpha_{k_d}, \qquad 2 \le d. \end{aligned}$$

The following representation of $f(x)$ near $\check{x}$ is appropriate under these assumptions.

$$\begin{aligned} \text{(2.1)} \quad f_k(x) &= (x_k - \check{x}_k) + \alpha_k (x_n - \check{x}_n) + \mathcal{O}\left(\|x - \check{x}\|^2 \right) \\ & \quad \text{for } 1 \le k \le n-1. \\ \text{(2.2)} \quad f_n(x) &= \frac{1}{2!} \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \frac{\partial^2 f_n}{\partial x_{k_1} \partial x_{k_2}}(\check{x})(x_{k_1} - \check{x}_{k_1})(x_{k_2} - \check{x}_{k_2}) + \dots \\ & \quad + \frac{1}{d!} \sum_{k_1=1}^{n} \cdots \sum_{k_d=1}^{n} \frac{\partial^d f_n}{\partial x_{k_1} \dots \partial x_{k_d}}(\check{x})(x_{k_1} - \check{x}_{k_1}) \dots (x_{k_d} - \check{x}_{k_d}) \\ & \quad + \mathcal{O}\left(\|x - \check{x}\|^{d+1} \right), \end{aligned}$$

or

$$\begin{aligned}
(2.3)\quad f_k(x) &\approx (x_k - \check{x}_k) + \alpha_k(x_n - \check{x}_n) \quad \text{for } 1 \le k \le n-1. \\
(2.4)\quad f_n(x) &\approx \frac{1}{2!}\sum_{k_1=1}^{n}\sum_{k_2=1}^{n}\frac{\partial^2 f_n}{\partial x_{k_1}\partial x_{k_2}}(\check{x})(x_{k_1}-\check{x}_{k_1})(x_{k_2}-\check{x}_{k_2}) + \dots \\
&\quad + \frac{1}{d!}\sum_{k_1=1}^{n}\cdots\sum_{k_d=1}^{n}\frac{\partial^d f_n}{\partial x_{k_1}\dots\partial x_{k_d}}(\check{x})(x_{k_1}-\check{x}_{k_1})\dots(x_{k_d}-\check{x}_{k_d}).
\end{aligned}$$

2.2. **Choosing the Coordinate Bounds.** We use a similar scheme to that of §5 of [13] and [11], except we work only with real coordinates: to compute the degree $\mathrm{d}(F, \boldsymbol{x}, 0)$, we consider $F_{\neg n}$ on the boundary of $\boldsymbol{x}$. This boundary consists of the $2n$ faces $\boldsymbol{x}_{\underline{1}}$, $\boldsymbol{x}_{\overline{1}}$, $\boldsymbol{x}_{\underline{2}}$, $\boldsymbol{x}_{\overline{2}}$, ..., $\boldsymbol{x}_{\underline{n}}$, $\boldsymbol{x}_{\overline{n}}$. We set $\boldsymbol{x}_n$ in such a way that

$$\mathrm{w}(\boldsymbol{x}_n) \le \frac{1}{2}\min_{1\le k\le n-1}\left\{\frac{\mathrm{w}(\boldsymbol{x}_k)}{|\alpha_k|}\right\} \tag{2.5}$$

Constructing the box widths this way makes it is unlikely that $f_k(x) = 0$ on either $\boldsymbol{x}_{\underline{k}}$ or $\boldsymbol{x}_{\overline{k}}$ for any k with $k = 1, \dots, n-1$. This, in turn, allows us to replace searches on these $2n-2$ of the $2n$ faces of $\boldsymbol{\partial x}$ by simple interval evaluations, reducing the total computational cost dramatically. See [13] and §4 below for details.

3. Our Main Theorem

The following underlies our algorithm for computation of the topological index of $F(x)$ in real space.

Theorem 3.1. *Suppose*

1. *all the assumptions in §2.1 are true;*
2. (2.3) *and* (2.4) *are exact; and*
3. $\Delta_1 = \cdots = \Delta_{d-1} = 0$, $\Delta_d \neq 0$, *where* $2 \le d$.

Then $\mathrm{d}(F, \boldsymbol{x}, 0) = -\mathrm{sgn}(\Delta_d)$ *when d is an odd number and* $\mathrm{d}(F, \boldsymbol{x}, 0) = 0$ *when d is an even number.*

Proof. We will use Theorem 1.3 to compute the degree. Consider $F_{\neg n} \equiv (f_1, \dots, f_{n-1})$ on the boundary of $\boldsymbol{x}$.

When $\boldsymbol{x}_n$ is constructed to satisfy (2.5), it is easy to see that

$$f_k(x) = (x_k - \check{x}_k) + \alpha_k(x_n - \check{x}_n) \neq 0$$

on $\boldsymbol{x}_{\underline{k}}$ and $\boldsymbol{x}_{\overline{k}}$, $k = 1, \dots, n-1$. Thus, $F_{\neg n}$ has no zeros on $\boldsymbol{x}_{\underline{k}}$ and $\boldsymbol{x}_{\overline{k}}$, $k = 1, \dots, n-1$. Thus, to use Theorem 1.3, we need only consider the two faces $\boldsymbol{x}_{\underline{n}}$ and $\boldsymbol{x}_{\overline{n}}$.

Obviously, $F_{\neg n}$ has a unique zero point $\tilde{x} = (\tilde{x}_1, \dots, \tilde{x}_n)$ on $\boldsymbol{x}_{\underline{n}}$ with $\tilde{x}_k - \check{x}_k = -\alpha_k(\underline{x}_n - \check{x}_n)$, $k = 1, \dots, n$. Note $\alpha_n = -1$ and $\tilde{x}_n = \underline{x}$. The value of f_n at this zero point is

$$\begin{aligned}
f_n(\tilde{x}) &= \frac{1}{2!}\sum_{k_1=1}^{n}\sum_{k_2=1}^{n}\frac{\partial^2 f_n}{\partial x_{k_1}\partial x_{k_2}}(\check{x})(\tilde{x}_{k_1}-\check{x}_{k_1})(\tilde{x}_{k_2}-\check{x}_{k_2})+\cdots+ \\
&\quad \frac{1}{d!}\sum_{k_1=1}^{n}\cdots\sum_{k_d=1}^{n}\frac{\partial^d f_n}{\partial x_{k_1}\dots\partial x_{k_d}}(\check{x})(\tilde{x}_{k_1}-\check{x}_{k_1})\dots(\tilde{x}_{k_d}-\check{x}_{k_d}) \\
&= \frac{1}{2!}\sum_{k_1=1}^{n}\sum_{k_2=1}^{n}\frac{\partial^2 f_n}{\partial x_{k_1}\partial x_{k_2}}(\check{x})(-1)^2\alpha_{k_1}\alpha_{k_2}(\underline{x}_n-\check{x}_n)^2+\cdots+ \\
&\quad \frac{1}{d!}\sum_{k_1=1}^{n}\cdots\sum_{k_d=1}^{n}\frac{\partial^d f_n}{\partial x_{k_1}\dots\partial x_{k_d}}(\check{x})(-1)^d\alpha_{k_1}\dots\alpha_{k_d}(\underline{x}_n-\check{x}_n)^d \\
&= \frac{(-1)^2}{2!}(\underline{x}_n-\check{x}_n)^2\sum_{k_1=1}^{n}\sum_{k_2=1}^{n}\frac{\partial^2 f_n}{\partial x_{k_1}\partial x_{k_2}}(\check{x})\alpha_{k_1}\alpha_{k_2}+\dots \\
&\quad +\frac{(-1)^d}{d!}(\underline{x}_n-\check{x}_n)^d\sum_{k_1=1}^{n}\cdots\sum_{k_d=1}^{n}\frac{\partial^d f_n}{\partial x_{k_1}\dots\partial x_{k_d}}(\check{x})\alpha_{k_1}\dots\alpha_{k_d} \\
&= \frac{(-1)^2\Delta_2}{2!}(\underline{x}_n-\check{x}_n)^2+\cdots+\frac{(-1)^{d-1}\Delta_{d-1}}{(d-1)!}(\underline{x}_n-\check{x}_n)^{d-1} \\
&\quad +\frac{(-1)^d\Delta_d}{d!}(\underline{x}_n-\check{x}_n)^d \\
&= \frac{(-1)^d\Delta_d}{d!}(\underline{x}_n-\check{x}_n)^d=\frac{(-1)^d\Delta_d}{d!}(-1)^d(\check{x}_n-\underline{x}_n)^d. \\
&= \frac{\Delta_d}{d!}(\check{x}_n-\underline{x}_n)^d.
\end{aligned}$$

The determinant of the Jacobi matrix at this zero point is

$$\left|\frac{\partial F_{\neg n}}{\partial x_1\dots\partial x_{n-1}}(\tilde{x})\right| = \begin{vmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{vmatrix} = 1.$$

Similarly, $F_{\neg n}$ has a unique zero point $\tilde{x}=(\tilde{x}_1,\dots,\tilde{x}_n)$ on $\boldsymbol{x}_{\overline{n}}$ with $\tilde{x}_k-\check{x}_k=-\alpha_k(\overline{x}_n-\check{x}_n), k=1,\dots,n$. With a similar computation to above, the value of f_n at this zero point is

$$f_n(\tilde{x})=\frac{\Delta_d}{d!}(\check{x}_n-\overline{x}_n)^d.$$

The determinant of the Jacobi matrix at this zero point is also

$$\left|\tfrac{\partial F_{\neg n}}{\partial x_1\dots\partial x_{n-1}}(\tilde{x})\right| = \begin{vmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{vmatrix} = 1.$$

Next, we apply Theorem 1.3 with $s=+1$: When d is odd, if $\Delta_d>0$, then $f_n(\tilde{x})>0$ at the zero point $\tilde{x}$ of $F_{\neg n}$ on $\boldsymbol{x}_{\underline{n}}$ and thus $\mathrm{d}(F,\boldsymbol{x},0)=(-1)^{n-1}\times(-1)^n\times 1=-1$. If $\Delta_d<0$, then $f_n(\tilde{x})>0$ at the zero point $\tilde{x}$ of $F_{\neg n}$ on $\boldsymbol{x}_{\overline{n}}$ and thus $\mathrm{d}(F,\boldsymbol{x},0)=(-1)^{n-1}\times(-1)^{n+1}\times 1=+1$. Thus, in any case, when d is odd, we have $\mathrm{d}(F, \boldsymbol{x}, 0) = -\mathrm{sgn}(\Delta_d)$. When d is even, $f_n(\tilde{x})$ is either positive at the two zero points $\tilde{x}$'s of $F_{\neg n}$ on $\boldsymbol{x}_{\underline{n}}$ and $\boldsymbol{x}_{\overline{n}}$ or negative at the two zero points $\tilde{x}$'s of $F_{\neg n}$ on $\boldsymbol{x}_{\underline{n}}$ and $\boldsymbol{x}_{\overline{n}}$. In the first case, it's obvious that $\mathrm{d}(F, \boldsymbol{x}, 0) = 0$. In the second case, $\mathrm{d}(F, \boldsymbol{x}, 0) = (-1)^{n-1}[(-1)^n \times 1 + (-1)^{n+1} \times 1] = 0$. Thus, in any case, when d is even, we have $\mathrm{d}(F, \boldsymbol{x}, 0) = 0$.

□

4. ALGORITHM

In actual problems, the equalities (2.3) and (2.4) in §2.1 are only approximately true. However, the analysis in Theorem 3.1 is still valid if the approximations are accurate. In that case, proof of Theorem 3.1 leads to a practical computational technique. First, it is easy to apply simple interval evaluations of

$$f_k(x) = (x_k - \check{x}_k) + \alpha_k(x_n - \check{x}_n)$$

to verify $f_k(x) \neq 0$ on $\boldsymbol{x}_{\underline{k}}$ and $\boldsymbol{x}_{\overline{k}}$, $k = 1, \ldots, n-1$, since we have arranged $\boldsymbol{x}_n$ according to (2.5) to achieve this. Second, it is not difficult to apply an interval Newton method to verify the unique zero point of $F_{\neg n}$ on $\boldsymbol{x}_{\underline{n}}$ and on $\boldsymbol{x}_{\overline{n}}$, since $F_{\neg n}$ is approximately linear: On $\boldsymbol{x}_{\underline{n}}$, $x_n = \underline{x}$ is known precisely, and formally solving $\boldsymbol{f}_k(\boldsymbol{x}) = 0$ for x_k gives sharper bounds $\tilde{\boldsymbol{x}}_k$ with $\mathrm{w}(\tilde{\boldsymbol{x}}_k) = \mathcal{O}\left(\|\boldsymbol{x} - \check{x}\|^2\right)$, $1 \leq k \leq n-1$, and thus gives a small subspace $\boldsymbol{x}_{\underline{n}}^0$ over which we can set up an interval Newton method for $F_{\neg n}$ to verify existence and uniqueness of the zero. Once we have verified existence of the solution, an interval evaluation gives us $\mathrm{sgn}(f_n)$ at the zero point. We then calculate and verify the sign of the determinant of the Jacobi matrix of $F_{\neg n}$ at the zero point[1]. We process $\boldsymbol{x}_{\overline{n}}$ similarly, then we use the formula in Theorem 1.3 to compute the degree.

```
ALGORITHM 1
  Initialization
      Input the approximate solution $\check{x}$ and a tolerance $\epsilon_{x_n}$ that determines the
      size of the box constructed about $\check{x}$ in which existence of a solution is to
      be verified.
  Box-setting Phase
    (1) Compute the preconditioner of the original system, using Gaussian elimi-
        nation with full pivoting. (In the remainder of this algorithm, the notation
        $f_k$ will refer to the $k$-th component of the preconditioned system.)
    (2) $\boldsymbol{x}_n \leftarrow [\check{x}_n - \epsilon_{x_n}, \check{x}_n + \epsilon_{x_n}]$.
    (3) For $1 \leq k \leq n-1$:
         (a) $r_k \leftarrow \max\{\epsilon_{x_n}, |\alpha_k|\mathrm{w}(\boldsymbol{x}_n)\}$.
         (b) $\boldsymbol{x}_k \leftarrow [\check{x} - r_k, \check{x} + r_k]$.
  Elimination Phase
      Do for $1 \leq k \leq n-1$:
          Do for $\boldsymbol{x}_{\underline{k}}$ and $\boldsymbol{x}_{\overline{k}}$
             (i) Compute the mean-value extension of $\boldsymbol{f}_k$ over that face.
            (ii) If $0 \in \boldsymbol{f}_k$, then stop and signal failure.
```

[1] At present, we only verify the sign of the determinant of the preconditioned Jacobi matrix, and do not take account of the sign of the determinant of the preconditioner matrix. Thus, we only obtain a verified value of the absolute value of the degree. This, however, is sufficient to verify existence of a solution.

Search Phase

(1) For $\boldsymbol{x}_{\underline{n}}$ (or $\boldsymbol{x}_{\overline{n}}$.)
   (a) Use mean-value extensions for $\boldsymbol{f}_k(\boldsymbol{x}) = 0$ to solve for x_k to get sharper bounds $\tilde{\boldsymbol{x}}_k$ with width $\mathcal{O}\left(\|\boldsymbol{x} - \check{x}\|^2\right)$, $1 \le k \le n-1$, and thus to get a small subface $\boldsymbol{x}^0_{\underline{n}}$ (or $\boldsymbol{x}^0_{\overline{n}}$) of $\boldsymbol{x}_{\underline{n}}$ (or $\boldsymbol{x}_{\overline{n}}$).
   (b) Set up an interval Newton method for $F_{\neg n}$ to verify existence and uniqueness of a zero over $\boldsymbol{x}^0_{\underline{n}}$ (or $\boldsymbol{x}^0_{\overline{n}}$.)
   (c) If the zero can not be verified, then stop and signal failure.
   (d) Compute the mean-value extension of $\boldsymbol{f}_n$ over $\boldsymbol{x}^0_{\underline{n}}$ (or $\boldsymbol{x}^0_{\overline{n}}$).
   (e) If $0 \in \boldsymbol{f}_n$, then stop and signal failure.
   (f) Compute $\left|\frac{\partial F_{\neg n}}{\partial x_1 \ldots x_{n-1}}(\boldsymbol{x}^0_{\underline{n}})\right|$ (or $\left|\frac{\partial F_{\neg n}}{\partial x_1 \ldots x_{n-1}}(\boldsymbol{x}^0_{\overline{n}})\right|$).
   (g) If $0 \in \left|\frac{\partial F_{\neg n}}{\partial x_1 \ldots x_{n-1}}(\boldsymbol{x}^0_{\underline{n}})\right|$ (or $0 \in \left|\frac{\partial F_{\neg n}}{\partial x_1 \ldots x_{n-1}}(\boldsymbol{x}^0_{\overline{n}})\right|$), then stop and signal failure.
   (h) Use the formula in Theorem 3.1 with $s = +1$ to compute the degree contribution of $\boldsymbol{x}_{\underline{n}}$ (or $\boldsymbol{x}_{\overline{n}}$.)
(2) Add the degree contributions of $\boldsymbol{x}_{\underline{n}}$ and $\boldsymbol{x}_{\overline{n}}$ to get the degree.

END OF ALGORITHM 1

The computational complexity of this algorithm is $\mathcal{O}\left(n^3\right)$. (See the computational complexity analysis in [13] for details.)

From Theorem 3.1, Algorithm 1 can be used to verify a non-zero topological degree (and hence, by the Kronecker existence theorem, existence of a solution), provided d is odd. However, we have demonstrated that no general computational method can verify existence in real space when d is even, although, in that case, we can verify existence when we embed the approximate solution in a small surrounding region of complex space; see [13] and [11]. In particular, we can use the general algorithm, [11, Algorithm 1], for any d. However, for odd d, Algorithm 1 above is much more efficient, and also verifies a stronger assertion (existence of a real solution, versus existence of a complex solution). If d is initially unknown, we can use the heuristic in [11, §5] to guess d. The value of d, and hence, existence of a solution along with rigorous bounds on its coordinates, can then be verified with Algorithm 1 above when d is odd and with [11, Algorithm 1] when d is even.

5. Experimental Results

We programmed Algorithm 1 in Fortran 90 in the same interval arithmetic environment and on the same machine as the experiments in [13] and [11]. Namely, we used the Fortran 90 system described in [9] and [10] with subsequent improvements within the GlobSol project [3], and we used the Sun Fortran 95 compiler version 6.0 with optimization level 0 on a Sparc Ultra-1 model 140. We tested Algorithm 1 above with Example 2 from [11], that is, with

Example 5.1 (Example 2 from [11], motivated from considerations in [6])**.** Set $f(x) = h(x,t) = (1-t)(Ax - x^3) - tx$, where $A \in \mathbb{R}^{n \times n}$ is the matrix corresponding to central difference discretization of the boundary value problem $-u'' = 0$, $u(0) = u(1) = 0$ and $x^3 = (x_1^3, \ldots, x_n^3)^T$, and t is chosen equal to $t_1 = \lambda_1/(1+\lambda_1)$, where λ_1 is the largest eigenvalue of A.

As in our tests of the algorithms in [11], we tried Algorithm 1 above with the example, with $n = 5$, 10, 20, 40, 80, and 160. We also tried Algorithm 1 above with $n = 320$ and $n = 640$, dimensions that were impractical within our experimental settingfor the algorithms in complex space from [11]. In all cases, we used $\epsilon_{x_n} = 10^{-2}$. (Considerably larger or smaller values of ϵ_{x_n} seemed to lead to less desirable results.) Algorithm 1 above succeeded in verifying existence of a solution in all cases. We compare execution times of Algorithm 1 with that of the corresponding algorithm in complex space (from experiments in [11]) in Table 1.

Table 1. Numerical Results

n	CPU Time (Complex)	Time Ratio (Complex)	CPU Time (Real)	Time Ratio (Real)
5	39.27		0.06	
10	10.31	0.26	0.15	2.50
20	74.32	7.21	0.60	4.00
40	481.23	6.48	3.61	6.02
80	3805.06	7.91	25.35	7.02
160	33944.20	8.92	192.20	7.58
320	—		1569.95	8.17
640	—		12956.11	8.25

The following are seen from Table 1.

- Computations in the real space are orders-of-magnitude more efficient than computations in complex space.
- The real and complex algorithms exhibit approximately the same dependence on the number of variables n: the amount of work increases slightly faster than n^3.

Analysis in [13] and [11] exhibits that the total work, for equivalent systems, increases as $\mathcal{O}\left(n^3\right)$, everything else about the system being the same. However, both the real and complex algorithms use (possibly) multiple sweeps of the interval Gauss–Seidel method; more sweeps may be necessary to assure reduction of each coordinate width for higher condition numbers. Since the condition numbers of the Jacobi matrices increases for Example 5.1 as n increases, this could account for the growth in effort that is slightly greater than $\mathcal{O}\left(n^3\right)$.

Besides being carried out in a space with approximately double the number of variables, verification of the degree in complex space (with the algorithms in [13] and [11]) requires an expensive one-dimensional search. In contrast, the verification in $\mathbb{R}^n$ requires no such search.

6. Summary and Future Work

We have presented an algorithm for verifying that the topological index at a singular solution $x^* \in \boldsymbol{x}$, $F(x^*) = 0$ of a map $F : \boldsymbol{x} \subset \mathbb{R}^n \to \mathbb{R}^n$ is non-zero, in those cases in which the degree of the first non-zero tensor in the Taylor expansion of the singular part of F at x^* is odd, and the dimension of the null space of the Jacobi matrix of F at x^* is one. This algorithm is much more efficient than previously proposed algorithms for verifying the index in $\mathbb{C}^n$, and demonstrably has a running time approximately proportional to n^3.

The algorithm can be used to provide rigorous bounds on an actual singular solution. It can be incorporated as a post-processing step in traditional floating-point algorithms for solving nonlinear systems, or can be incorporated in global branch-and-bound algorithms.

The main computations in the algorithm are linear-algebra based. These computations may be made more efficient by taking account of structure, such as bandedness or sparsity, in the Jacobi matrix. This would reduce the computational effort below $\mathcal{O}\left(n^3\right)$.

References

1. G. Alefeld and J. Herzberger, *Introduction to interval computations*, Academic Press, New York, 1983.
2. P. Alexandroff and H. Hopf, *Topologie*, Chelsea, 1935.
3. G. F. Corliss, *Globsol entry page*, 1998, `http://www.mscs.mu.edu/∼globsol/`.
4. J. Cronin, *Fixed points and topological degree in nonlinear analysis*, American Mathematical Society, Providence, RI, 1964.
5. E. R. Hansen, *Global optimization using interval analysis*, Marcel Dekker, Inc., New York, 1992.
6. H. Jürgens, H.-O. Peitgen, and D. Saupe, *Topological perturbations in the numerical nonlinear eigenvalue and bifurcation problems*, Analysis and Computation of Fixed Points (New York) (S. M. Robinson, ed.), Academic Press, 1980, pp. 139–181.
7. R. B. Kearfott, *Computing the degree of maps and a generalized method of bisection*, Ph.D. thesis, University of Utah, Department of Mathematics, 1977.
8. ———, *An efficient degree-computation method for a generalized method of bisection*, Numer. Math. **32** (1979), 109–127.
9. ———, *A Fortran 90 environment for research and prototyping of enclosure algorithms for nonlinear equations and global optimization*, ACM Trans. Math. Software **21** (1995), no. 1, 63–78.
10. ———, *Rigorous global search: Continuous problems*, Kluwer, Dordrecht, Netherlands, 1996.
11. R. B. Kearfott and J. Dian, *Existence verification for higher-degree singular zeros of complex nonlinear systems*, 2000, Preprint, Department of Mathematics, Univ. of Louisiana at Lafayette, U.L. Box 4-1010, Lafayette, La 70504.
12. R. B. Kearfott and J. Dian., *Verifying topological indices for higher-order rank deficiencies*, 2000, Preprint, Department of Mathematics, Univ. of Louisiana at Lafayette, U.L. Box 4-1010, Lafayette, La 70504.
13. R. B. Kearfott, J. Dian, and A. Neumaier, *Existence verification for singular zeros of complex nonlinear systems*, SIAM J. Numer. Anal. **38** (2000), no. 2, 360–379.
14. A. Neumaier, *Interval methods for systems of equations*, Cambridge University Press, Cambridge, England, 1990.
15. H. Ratschek and J. Rokne, *New computer methods for global optimization*, Wiley, New York, 1988.
16. W. C. Rheinboldt and J. M. Ortega, *Iterative solution of nonlinear equations in several variables*, Academic Press, New York, 1970.
17. F. Stenger, *An algorithm for the topological degree of a mapping in $\mathbb{R}^n$*, Numer. Math. **25** (1976), 23–38.

Hewlett–Packard Company, 3000 Waterview Parkway, Richardson, TX 75080
E-mail address: `jianwei_dian@hp.com`

Department of Mathematics, University of Louisiana at Lafayette, Box 4-1010, Lafayette, LA 70504-1010
E-mail address: `rbk@louisiana.edu`