Journal of Computational and Applied Mathematics 359 (2019) 145–152

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

An improved verification algorithm for nonlinear systems of equations based on Krawczyk operator ☆

Guoliang Hou [a,b,c], Shugong Zhang [a,b,*]

[a] School of Mathematics, Jilin University, Changchun, Jilin, 130012, People's Republic of China
[b] Key Laboratory of Symbolic Computation and Knowledge Engineering (Ministry of Education), Jilin University, Changchun, Jilin, 130012, People's Republic of China
[c] School of Mathematics, Changchun Normal University, Changchun, Jilin, 130032, People's Republic of China

ARTICLE INFO

Article history:
Received 27 October 2018
Received in revised form 16 March 2019

MSC:
65G20
65H10
65-04

Keywords:
Verification algorithm
Krawczyk operator
Nonlinear system
INTLAB

ABSTRACT

In this paper an improved version of a verification algorithm for solving nonlinear systems of equations based on Krawczyk operator is presented. Compared with the original algorithm, the improved verification algorithm not only saves computing time, but also computes a narrower (or at least the same) inclusion of the solution to nonlinear systems of equations for certain classes of problems. Numerical results demonstrate the performance of the proposed algorithms.

1. Introduction

In this paper, we consider verification methods for general square nonlinear systems of equations

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

where $f : \mathbb{R}^n \to \mathbb{R}^n$ with $f(x) = (f_1(x), f_2(x), \ldots, f_n(x))^T$, $x = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$. The verification methods for nonlinear systems are widely applied in the various science (e.g., [1–5]) and engineering problems (e.g., [6,7]).

For nonlinear systems solving, one of the most fundamental goals of verification methods is computing verified bounds within which there exists a unique solution to (1). The working tools of verification methods are floating-point and interval arithmetic. Rump has given an overview on verification methods in [8].

With interval Newton operator first proposed by Sunaga [9] in the 1950s, a verification method was presented by Moore [10]. However, this method requires to solve an interval linear system at each verification step, which results in huge computational costs. Later, another verification method without solving interval linear systems was developed in terms of Krawczyk operator by Krawczyk [11] in 1969. The sufficient conditions for the success of the new verification method were given by Moore [12]. We call it Krawczyk–Moore method in this paper. Furthermore, Rump [13] improved the method so that it shows better performance in practice in 1983, and wrote the function `verifynlss` in INTLAB [14],

☆ The authors gratefully acknowledge the partial supports of the National Natural Science Foundation (Grant No. 11671169) of China and the Science and Technology Foundation for Education Department of Jilin Province, China (Grant No. JJKH20190502KJ).

* Corresponding author at: School of Mathematics, Jilin University, Changchun, Jilin, 130012, People's Republic of China.
E-mail address: sgzh@jlu.edu.cn (S. Zhang).

https://doi.org/10.1016/j.cam.2019.03.034
0377-0427/

the MATLAB Toolbox for Reliable Computing. If `verifynlss` function runs successfully for a given square nonlinear system of Eqs. (1), then the computed box (i.e., interval vector) contains a unique solution to the original system. Since then, based on Krawczyk–Moore method, various verification methods have been developed for general nonlinear problems such as over- and underdetermined nonlinear systems [15–17], multiple roots of nonlinear equations [18–22], nonlinear matrix equations [23,24] and so on. Up to now Krawczyk–Moore method still plays a very important role in the field of nonlinear verification problems.

In this paper we present an improved version of Krawczyk operator, and propose a new verification algorithm. Both theoretical results and numerical examples show that the improved verification algorithm not only saves more computing time, but also computes a narrower (or at least the same) inclusion of the solution to (1) than the original one for certain classes of problems.

The outline of this paper is as follows. Section 2 is devoted as a preparation for this paper. In this section we present several mathematical notations and technical results that are used in the paper. In Section 3, the main theoretical results of this paper and their proofs are given. In Section 4, we propose an improved verification algorithm based on improved Krawczyk operator (6) and Theorem 2.4. To illustrate the performance of our verification method, some numerical examples are given in Section 5. In Section 6, our further research in this direction is given and we express our gratitude to the Editors and the Reviewers for the warm work earnestly and the valuable suggestions and comments in the last section.

2. Notation and preliminaries

In our text, the interval notation adheres to the recently adopted project of the international standard [25]. Specifically, we designate intervals and interval objects (vectors, matrices) by boldface letters. $\mathbb{IR}$ stands for the set of closed intervals of the real axis $\mathbb{R}$. $\mathbb{IR}^n$ means the set of n-dimensional interval vectors, whose geometric images are axes aligned boxes in $\mathbb{R}^n$, and $\mathbb{IR}^{n\times n}$ denotes the set of interval $n \times n$-matrices whose entries are all closed intervals of the real axis $\mathbb{R}$. We write wid $\boldsymbol{x}$, rad $\boldsymbol{x}$ and mid $\boldsymbol{x}$ for the width, radius and midpoint of an interval $\boldsymbol{x} \in \mathbb{IR}$, respectively. And the notations wid and mid for interval objects (vectors, matrices) are defined entrywise, i.e., the width and midpoint of $\boldsymbol{x}$ are the real vectors wid $\boldsymbol{x}$ and mid $\boldsymbol{x}$ whose entries are the widths and midpoints of the corresponding entries of $\boldsymbol{x}$: $(\text{wid } \boldsymbol{x})_i = \text{wid } \boldsymbol{x}_i$ and $(\text{mid } \boldsymbol{x})_i = \text{mid } \boldsymbol{x}_i$ when $\boldsymbol{x} \in \mathbb{IR}^n$, and the width and midpoint of $\boldsymbol{X}$ are the real matrices wid $\boldsymbol{X}$ and mid $\boldsymbol{X}$ whose entries are the widths and midpoints of the corresponding entries of $\boldsymbol{X}$: $(\text{wid } \boldsymbol{X})_{ij} = \text{wid } \boldsymbol{x}_{ij}$ and $(\text{mid } \boldsymbol{X})_{ij} = \text{mid } \boldsymbol{x}_{ij}$ when $\boldsymbol{X} \in \mathbb{IR}^{n\times n}$, where $\boldsymbol{x}_i,\ \boldsymbol{x}_{ij} \in \mathbb{IR}$, $i, j = 1, 2, \ldots, n$.

Let $x = (x_i) \in \mathbb{R}^n$, $y = (y_i) \in \mathbb{R}^n$. Then denote $x \geq y$ if $x_i \geq y_i$ for all $1 \leq i \leq n$, especially, if $x_i \geq 0$ for all i, we say that x is a nonnegative vector, denote it by $x \geq 0$, and $|y|$ denotes the nonnegative vector with entries $|y_i|$. Similarly, let $A = \left(a_{ij}\right) \in \mathbb{R}^{n\times n}$, $B = \left(b_{ij}\right) \in \mathbb{R}^{n\times n}$. Then denote $A \geq B$ if $a_{ij} \geq b_{ij}$ for all $1 \leq i,\ j \leq n$, especially, if $a_{ij} \geq 0$ for all i, j, we say that A is a nonnegative matrix, denote it by $A \geq 0$, and $|B|$ denotes the nonnegative matrix with entries $|b_{ij}|$.

Denote the Jacobian of f at x by $J_f(x)$. Let $\hat{x} \in \mathbb{R}^n$ be a solution to (1), if $J_f\left(\hat{x}\right)$ is nonsingular, then we call $\hat{x}$ a simple solution to (1) or a simple zero of f.

Krawczyk operator was first proposed by West German mathematician Rudolf Krawczyk [11] in 1969, and redefined for the solution existence tests by Moore [12], Neumaier [26], Kearfott [27], etc. Our exposition follows Shary [28].

Definition 2.1. Let some rules be defined that assign, to any box $\boldsymbol{x} \in \mathbb{IR}^n$, a point $\tilde{x} \in \boldsymbol{x}$ and a nonsingular real $n\times n$-matrix R, while $\boldsymbol{G} \in \mathbb{IR}^{n\times n}$ is an enclosure for the Jacobian $J_f(x)$ of the function f over the box $\boldsymbol{x}$. The mapping

$$K : \mathbb{IR}^n \times \mathbb{R}^n \to \mathbb{IR}^n,$$

defined by the rule

$$K\left(\boldsymbol{x}, \tilde{x}\right) := \tilde{x} - Rf\left(\tilde{x}\right) + (I - R\boldsymbol{G})\left(\boldsymbol{x} - \tilde{x}\right), \tag{2}$$

is called (interval) Krawczyk operator for the function f, where I denotes the identity matrix of order n.

The following important statement concerning Krawczyk operator is valid [11,12,26,27].

Theorem 2.2. *Under the assumption of Definition 2.1, for a box $\boldsymbol{x} \in \mathbb{IR}^n$, if*

$$K\left(\boldsymbol{x}, \tilde{x}\right) \subseteq \boldsymbol{x}, \tag{3}$$

then there exists at least one $\hat{x} \in \boldsymbol{x}$ with $f(\hat{x}) = 0$.

Remark 2.3. According to the working principle of verification methods, for a box $\boldsymbol{x}$, if the inclusion (3) is rigorously verified on the computers, then the box $\boldsymbol{x}$ contains a solution to (1). In general, there are no other restrictions on $\tilde{x}$ and R in Krawczyk operator (2). However, in order to make the inclusion (3) easier to satisfy, people usually take $\tilde{x}$ as an approximate solution to (1) and $R = J_f\left(\tilde{x}\right)^{-1}$, where $J_f\left(\tilde{x}\right)^{-1}$ denotes the inverse matrix of $J_f\left(\tilde{x}\right)$. Besides, since the overestimation caused by interval operations can also make the inclusion (3) hard to verify successfully, the narrower the width of $\boldsymbol{x}$ the better.

From a practical point of view, it is preferable for $\boldsymbol{x}$ not to include a zero $\hat{x}$ of f itself but the difference between an approximate zero $\tilde{x}$ and $\hat{x}$, because calculating an inclusion of $\hat{x}$ or $\hat{x}-\tilde{x}$ requires the same computing time and $\hat{x}-\tilde{x}$ is more suitable for programming than $\hat{x}$. To this purpose, in 1983 Rump gave a more practical verification theorem as follows.

Theorem 2.4 ([8]). *Let $f: D \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable functions, and $\tilde{x} \in \mathbb{R}^n$, $R \in \mathbb{R}^{n\times n}$, $\boldsymbol{x} \in \mathbb{IR}^n$ with $0 \in \boldsymbol{x}$ and $\tilde{x}+\boldsymbol{x} \subseteq D$, and define $J_f\left(\tilde{x}+\boldsymbol{x}\right) := \bigcap\{\boldsymbol{M} \in \mathbb{IR}^{n\times n} : \forall x \in \tilde{x}+\boldsymbol{x}, J_f(x) \in \boldsymbol{M}\}$. Suppose*

$$S\left(\boldsymbol{x},\tilde{x}\right) := -Rf\left(\tilde{x}\right) + \left(I - RJ_f\left(\tilde{x}+\boldsymbol{x}\right)\right)\boldsymbol{x} \subseteq int\ \boldsymbol{x}, \tag{4}$$

where int $\boldsymbol{x}$ denotes the interior of $\boldsymbol{x}$.

Then R and all matrices $M \in J_f\left(\tilde{x}+\boldsymbol{x}\right)$ are nonsingular, and there is a unique solution $\hat{x}$ to $f(x)=0$ in $\tilde{x}+S\left(\boldsymbol{x},\tilde{x}\right)$.

For similar considerations as mentioned in Remark 2.3 and based on Theorem 2.4, Rump gave the following verification algorithm written in executable MATLAB/INTLAB codes. This algorithm computes the inclusion of the solution to (1) near some numerical approximation $\tilde{x}$. In the algorithm Rump takes R as an approximate inverse of $J_f\left(\tilde{x}\right)$ calculated in floating-point arithmetic.

Algorithm 2.5 [8] Verified bounds for the solution of a nonlinear system:

```
function XX = VerifyNonLinSys(f,xs)
XX = NaN;    % initialization
y = f(gradientinit(xs));
R = inv(y.dx);    % approximate inverse of J_f(xs)
Y = f(gradientinit(intval(xs)));
Z = -R*Y.x;    % inclusion of -R*f(xs)
X = Z; iter = 0;
while iter< 15
iter = iter+1;
Y = hull(X*infsup(0.9,1.1)+1e-20*infsup(-1,1),0);
YY = f(gradientinit(xs+Y));    % YY.dx inclusion of J_f(xs+Y)
X = Z + (eye(n)-R*YY.dx)*Y;    % interval iteration
if all(in0(X,Y)), XX = xs+X; return; end
end
```

Algorithm 2.5 is the core of `verifynlss` function.

3. Main theoretical results

In accordance with the second part, in the current dominant verification methods for nonlinear systems of equations the interval operator $S\left(\boldsymbol{x},\tilde{x}\right)$ can be written as

$$-J_f\left(\tilde{x}\right)^{-1} f\left(\tilde{x}\right) + \left(I - J_f\left(\tilde{x}\right)^{-1} J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)\boldsymbol{x} \triangleq S_{\mathrm{R}}\left(\boldsymbol{x},\tilde{x}\right), \tag{5}$$

where $\boldsymbol{x} \in \mathbb{IR}^n$ with $0 \in \boldsymbol{x}$.

For continuously differentiable functions $f: D \subseteq \mathbb{R}^n \to \mathbb{R}^n$, if $J_f\left(\tilde{x}\right)$, $\tilde{x} \in D$ is nonsingular, then $J_f(x)$ is also nonsingular for x close to $\tilde{x}$ sufficiently. Therefore, we may assume that J_f is nonsingular thereafter. In this paper we take $R = \left(\mathrm{mid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)^{-1}$, and the interval operator $S\left(\boldsymbol{x},\tilde{x}\right)$ can be rewritten as

$$-\left(\mathrm{mid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)^{-1} f\left(\tilde{x}\right) + \left(I - \left(\mathrm{mid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)^{-1} J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)\boldsymbol{x} \triangleq S_{\mathrm{H}}\left(\boldsymbol{x},\tilde{x}\right).$$

Noted that $J_f\left(\tilde{x}+\boldsymbol{x}\right) = \mathrm{mid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right) + \frac{1}{2}\mathrm{wid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right)[-1,1]$ and $\boldsymbol{x} = \mathrm{mid}\, \boldsymbol{x} + \frac{1}{2}\mathrm{wid}\, \boldsymbol{x}\ [-1,1]$, $S_{\mathrm{H}}\left(\boldsymbol{x},\tilde{x}\right)$ can be formulated as

$$\begin{aligned} S_{\mathrm{H}}\left(\boldsymbol{x},\tilde{x}\right) &= -\left(\mathrm{mid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)^{-1} f\left(\tilde{x}\right) + \left(I - \left(\mathrm{mid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)^{-1} J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)\boldsymbol{x} \\ &= -\left(\mathrm{mid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)^{-1} f\left(\tilde{x}\right) + \frac{1}{4}\left|\left(\mathrm{mid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)^{-1}\right| \mathrm{wid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right) \mathrm{wid}\, \boldsymbol{x}\ [-1,1] \\ &\quad + \frac{1}{2}\left|\left(\mathrm{mid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)^{-1}\right| \mathrm{wid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right) |\mathrm{mid}\, \boldsymbol{x}|\ [-1,1]. \end{aligned} \tag{6}$$

Comparing the expressions

$$\begin{aligned} S_{\mathrm{H}}\left(\boldsymbol{x},\tilde{x}\right) = -\left(\mathrm{mid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)^{-1} f\left(\tilde{x}\right) + \tfrac{1}{4}\left|\left(\mathrm{mid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)^{-1}\right| \mathrm{wid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right) \mathrm{wid}\, \boldsymbol{x}\ [-1,1] \\ + \frac{1}{2}\left|\left(\mathrm{mid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)^{-1}\right| \mathrm{wid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right) |\mathrm{mid}\, \boldsymbol{x}|\ [-1,1] \end{aligned}$$

with

$$S_{\mathrm{R}}\left(\boldsymbol{x}, \tilde{x}\right) = -J_f\left(\tilde{x}\right)^{-1} f\left(\tilde{x}\right) + \left(I - J_f\left(\tilde{x}\right)^{-1} J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)\boldsymbol{x},$$

we can see that, in $S_{\mathrm{H}}\left(\boldsymbol{x}, \tilde{x}\right)$, $J_f\left(\tilde{x}\right)$ does not need to be computed and $\operatorname{mid} J_f\left(\tilde{x}+\boldsymbol{x}\right)$ can be obtained directly from $J_f\left(\tilde{x}+\boldsymbol{x}\right)$, and interval multiplication is not involved because of $\left(\operatorname{mid} J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)^{-1}$, $\operatorname{wid} J_f\left(\tilde{x}+\boldsymbol{x}\right) \in \mathbb{R}^{n\times n}$ and $\operatorname{mid} \boldsymbol{x}$, $\operatorname{wid} \boldsymbol{x} \in \mathbb{R}^n$, i.e., they are point matrices and point vectors, whereas $I - J_f\left(\tilde{x}\right)^{-1} J_f\left(\tilde{x}+\boldsymbol{x}\right)$ and $\boldsymbol{x}$ presented in $S_{\mathrm{R}}\left(\boldsymbol{x}, \tilde{x}\right)$ are interval matrix and interval vector, respectively. Hence, the verification algorithm with $S_{\mathrm{H}}\left(\boldsymbol{x}, \tilde{x}\right)$ instead of $S_{\mathrm{R}}\left(\boldsymbol{x}, \tilde{x}\right)$ will spend less calculations.

In addition to the advantages mentioned above, the inclusion $S_{\mathrm{H}}\left(\boldsymbol{x}, \hat{x}\right) \subseteq S_{\mathrm{R}}\left(\boldsymbol{x}, \hat{x}\right)$ can be satisfied under some assumptions, where $\hat{x}$ denotes a true solution to (1). For an easier theoretical analysis, in the following theorems (i.e., Theorem 3.1 and its Corollary 3.3) we took $\boldsymbol{x}$ as symmetric (i.e., $\operatorname{mid} \boldsymbol{x} = 0$) because of $0 \in \boldsymbol{x}$.

Theorem 3.1. *Let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable functions and $\hat{x} \in \mathbb{R}^n$ be a simple zero of f, $\boldsymbol{x} \in \mathbb{IR}^n$ be symmetric with $\hat{x} + \boldsymbol{x} \subseteq D$ such that any matrix in $J_f\left(\hat{x}+\boldsymbol{x}\right)$ is nonsingular. If $\left|J_f\left(\hat{x}\right)^{-1}\right| \geq \left|\left(\operatorname{mid} J_f\left(\hat{x}+\boldsymbol{x}\right)\right)^{-1}\right|$, then*

$$S_H\left(\boldsymbol{x}, \hat{x}\right) \subseteq S_R\left(\boldsymbol{x}, \hat{x}\right). \tag{7}$$

Proof. Since $\left|J_f\left(\hat{x}\right)^{-1}\right| \geq \left|\left(\operatorname{mid} J_f\left(\hat{x}+\boldsymbol{x}\right)\right)^{-1}\right|$, we have

$$\frac{1}{4}\left|J_f\left(\hat{x}\right)^{-1}\right| \operatorname{wid} J_f\left(\hat{x}+\boldsymbol{x}\right) \operatorname{wid} \boldsymbol{x} \geq \frac{1}{4}\left|\left(\operatorname{mid} J_f\left(\hat{x}+\boldsymbol{x}\right)\right)^{-1}\right| \operatorname{wid} J_f\left(\hat{x}+\boldsymbol{x}\right) \operatorname{wid} \boldsymbol{x},$$

subsequently, we get

$$\frac{1}{4}\left|\left(\operatorname{mid} J_f\left(\hat{x}+\boldsymbol{x}\right)\right)^{-1}\right| \operatorname{wid} J_f\left(\hat{x}+\boldsymbol{x}\right) \operatorname{wid} \boldsymbol{x}\,[-1,1] \subseteq \frac{1}{4}\left|J_f\left(\hat{x}\right)^{-1}\right| \operatorname{wid} J_f\left(\hat{x}+\boldsymbol{x}\right) \operatorname{wid} \boldsymbol{x}\,[-1,1].$$

Moreover, using $f\left(\hat{x}\right) = 0$ we have

$$\begin{aligned} S_{\mathrm{H}}\left(\boldsymbol{x}, \hat{x}\right) &= \frac{1}{4}\left|\left(\operatorname{mid} J_f\left(\hat{x}+\boldsymbol{x}\right)\right)^{-1}\right| \operatorname{wid} J_f\left(\hat{x}+\boldsymbol{x}\right) \operatorname{wid} \boldsymbol{x}\,[-1,1] \subseteq S_{\mathrm{R}}\left(\boldsymbol{x}, \hat{x}\right) \\ &= \frac{1}{4}\left|J_f\left(\hat{x}\right)^{-1}\right| \operatorname{wid} J_f\left(\hat{x}+\boldsymbol{x}\right) \operatorname{wid} \boldsymbol{x}\,[-1,1] + \boldsymbol{y}, \end{aligned}$$

where $\boldsymbol{y} = \frac{1}{2}\left|J_f\left(\hat{x}\right)^{-1}\left(J_f\left(\hat{x}\right) - \operatorname{mid} J_f\left(\hat{x}+\boldsymbol{x}\right)\right)\right| \operatorname{wid} \boldsymbol{x}\,[-1,1]$.

Lemma 3.2. *Let A, $B \in \mathbb{R}^{n\times n}$ be nonsingular, and $A \geq B$, $A^{-1} \geq 0$, $B^{-1} \geq 0$. Then*

$$B^{-1} \geq A^{-1}.$$

Proof. Because $A - B \geq 0$ ($A \geq B$), $A^{-1} \geq 0$ and $B^{-1} \geq 0$, simple calculation yields

$$B^{-1} - A^{-1} = A^{-1}(A - B)B^{-1} \geq 0,$$

i.e.,

$$B^{-1} \geq A^{-1}.$$

Corollary 3.3. *Let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable functions and $\hat{x} \in \mathbb{R}^n$ be a simple zero of f, $\boldsymbol{x} \in \mathbb{IR}^n$ be symmetric with $\hat{x} + \boldsymbol{x} \subseteq D$ such that any matrix in $J_f\left(\hat{x}+\boldsymbol{x}\right)$ is nonsingular. If $J_f\left(\hat{x}\right)^{-1} \geq 0$, $\left(\operatorname{mid} J_f\left(\hat{x}+\boldsymbol{x}\right)\right)^{-1} \geq 0$ and $\operatorname{mid} J_f\left(\hat{x}+\boldsymbol{x}\right) \geq J_f\left(\hat{x}\right)$, then the inclusion (7) is valid.*

Proof. According to Lemma 3.2, we get

$$J_f\left(\hat{x}\right)^{-1} \geq \left(\operatorname{mid} J_f\left(\hat{x}+\boldsymbol{x}\right)\right)^{-1}$$

from $J_f\left(\hat{x}\right)^{-1} \geq 0$, $\left(\operatorname{mid} J_f\left(\hat{x}+\boldsymbol{x}\right)\right)^{-1} \geq 0$ and $\operatorname{mid} J_f\left(\hat{x}+\boldsymbol{x}\right) \geq J_f\left(\hat{x}\right)$. And we have

$$\left|J_f\left(\hat{x}\right)^{-1}\right| \geq \left|\left(\operatorname{mid} J_f\left(\hat{x}+\boldsymbol{x}\right)\right)^{-1}\right|$$

because of $J_f\left(\hat{x}\right)^{-1} \geq 0$ and $\left(\operatorname{mid} J_f\left(\hat{x}+\boldsymbol{x}\right)\right)^{-1} \geq 0$, which, by Theorem 3.1, implies

$$S_{\mathrm{H}}\left(\boldsymbol{x}, \hat{x}\right) \subseteq S_{\mathrm{R}}\left(\boldsymbol{x}, \hat{x}\right).$$

In practice there exists functions f that satisfy conditions $J_f\left(\hat{x}\right)^{-1} \geq 0$, $\left(\text{mid}\, J_f\left(\hat{x}+\boldsymbol{x}\right)\right)^{-1} \geq 0$ and $\text{mid}\, J_f\left(\hat{x}+\boldsymbol{x}\right) \geq J_f\left(\hat{x}\right)$ presented in Corollary 3.3, for example, the inverse isotone mappings f with convex-like derivatives. Moreover, numerical experiments presented in Section 5 have shown that the inclusion $S_{\mathrm{H}}\left(\boldsymbol{x}, \hat{x}\right) \subseteq S_{\mathrm{R}}\left(\boldsymbol{x}, \hat{x}\right)$ holds true for a wider class of functions.

Remark 3.4. Although we can only prove that $S_{\mathrm{H}}\left(\boldsymbol{x}, \hat{x}\right) \subseteq S_{\mathrm{R}}\left(\boldsymbol{x}, \hat{x}\right)$, in fact, the inclusion $S_{\mathrm{H}}\left(\boldsymbol{x}, \tilde{x}\right) \subseteq S_{\mathrm{R}}\left(\boldsymbol{x}, \tilde{x}\right)$ is also observed provided that $\left|J_f\left(\tilde{x}\right)^{-1}\right| \geq \left|\left(\text{mid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)^{-1}\right|$ or $J_f\left(\tilde{x}\right)^{-1} \geq 0$, $\left(\text{mid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)^{-1} \geq 0$ and $\text{mid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right) \geq J_f\left(\tilde{x}\right)$ for $\tilde{x}$ with $f(\tilde{x}) \approx 0$. Indeed, all the numerical examples that we have encountered support this observation.

4. An improved verification algorithm

In this section we describe a new verification algorithm for simple solutions to nonlinear systems of equations based on the interval operator $S_{\mathrm{H}}\left(\boldsymbol{x}, \tilde{x}\right)$(6) and Theorem 2.4. Like algorithm 2.5, the new verification algorithm is presented as executable MATLAB/INTLAB codes. In the algorithm we take R as an approximate inverse of $\text{mid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right)$ calculated in floating-point arithmetic.

Algorithm 4.1

```
function XX = ImpVerifyNonLinSys(f,xs)
XX = NaN;
Y0 = f(gradientinit(intval(xs)));    % inclusion of f(xs)
M = mid(Y0.dx);
R = inv(M);
Y = -R*Y0.x;
X = hull(Y*infsup(0.9,1.1) + 1e-20*infsup(-1,1) , 0 );
                                                  % interval vector X satisfying requirements
YY = f(gradientinit(xs+X));
M = mid(YY.dx);
R = inv(M);    % approximate inverse of m(J_f(xs+X))
M = abs(R);
x = inf(X);  y = sup(X);  z = max(abs(x),abs(y));
Y = -R*Y0.x + 0.5*M*diam(YY.dx)*z*infsup(-1,1);    % X ⊆ z*infsup(-1,1)
if all(in0(Y,X))
XX = xs+Y;
end
```

5. Numerical experiments

The following experiments are done using MATLAB R2012a and INTLAB V6 under Windows 7 on a Lenovo PC (1.70 GHz Intel(R) Core(TM) i5-3317U processor, 4 GB of memory).

Example 5.1. Consider

$$f(x) = \begin{pmatrix} 3x_1 - cos(x_2x_3) - \dfrac{1}{2} \\ x_1^2 - 81(x_2+0.1)^2 + sinx_3 + 1.06 \\ e^{-x_1x_2} + 20x_3 + \dfrac{10\pi - 3}{3} \end{pmatrix} = 0. \tag{8}$$

We use algorithm 4.1 and algorithm 2.5, with only one verification step executed, to compute the inclusion of the solution to (8) near some numerical approximation $\tilde{x} = \begin{pmatrix} 0.500000002581808 \\ -0.000028492129453 \\ -0.523599487583918 \end{pmatrix} \in \mathbb{R}^3$, where the numerical solution $\tilde{x}$ is obtained by Newton's methods. Table 5.1 displays the computational results and computing times (second) of the algorithms.

In the following numerical examples, the numerical solution $\tilde{x}$ is first obtained by Newton's methods with the given initial approximation for each dimension and then the inclusion of the solution near $\tilde{x}$ is computed by algorithm 4.1 and algorithm 2.5. We use `XX1`, `XX2` to denote the inclusions provided by algorithm 2.5 with only one verification step executed, algorithm 4.1, respectively in Examples 5.2–5.4. From the relevant computation results, it can be seen that although the functions f presented in (9), (10) and (11) do not have the properties $\left|J_f\left(\tilde{x}\right)^{-1}\right| \geq \left|\left(\text{mid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)^{-1}\right|$ or $J_f\left(\tilde{x}\right)^{-1} \geq 0$, $\left(\text{mid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right)\right)^{-1} \geq 0$ and $\text{mid}\, J_f\left(\tilde{x}+\boldsymbol{x}\right) \geq J_f\left(\tilde{x}\right)$, there is still `XX1` = `XX2` for each dimension. Because of the large number of data, we use the maximum relative error to describe the relationship between `XX1` and `XX2`.

Table 5.1
Results and computing times for (8).

	Inclusion	Time (s)
Algorithm 4.1	$\texttt{XX1} := \begin{pmatrix} [0.49999999995138,\ 0.50000000004119] \\ [-0.00000000533052,\ 0.00000000451574] \\ [-0.52359877573822,\ -0.52359877547977] \end{pmatrix}$	$\texttt{t}_1 = 0.049925$
Algorithm 2.5	$\texttt{XX2} := \begin{pmatrix} [0.49999999994728,\ 0.50000000008202] \\ [-0.00000000577963,\ 0.00000000899207] \\ [-0.52359877575001,\ -0.52359877536227] \end{pmatrix}$	$\texttt{t}_2 = 0.052508$

Where $\texttt{t}_2 : \texttt{t}_1 = 1.0517$ and it is easy to see that $\texttt{XX1} \subseteq \texttt{XX2}$.

Table 5.2
Solution of (9) using algorithm 2.5 and algorithm 4.1.

Dim	Algorithm 2.5 Computing time $\texttt{t}_1$ (s)	Algorithm 4.1 Computing time $\texttt{t}_2$ (s)	$\texttt{t}_1 : \texttt{t}_2$	$\texttt{mrelerr}(\boldsymbol{x})$
50	0.061858	0.060962	1.0147	$3.818 \cdot 10^{-16}$
100	0.078813	0.068845	1.1448	$3.826 \cdot 10^{-16}$
200	0.081962	0.061697	1.3285	$3.954 \cdot 10^{-16}$
500	0.113260	0.092930	1.2188	$4.246 \cdot 10^{-16}$
1000	0.249051	0.182910	1.3616	$4.593 \cdot 10^{-16}$
2000	0.753223	0.508220	1.4821	$4.438 \cdot 10^{-16}$

Where $\boldsymbol{x} = \texttt{XX1}$ or XX2, i.e., $\texttt{mrelerr}(\texttt{XX1}) = \texttt{mrelerr}(\texttt{XX2})$.

The relative error for $\boldsymbol{x} \in \mathbb{IR}$ denoted by $\texttt{relerr}(\boldsymbol{x})$ is defined in [8] by

$$\texttt{relerr}(\boldsymbol{x}) := \begin{cases} \left| \dfrac{\text{rad}\ \boldsymbol{x}}{\text{mid}\ \boldsymbol{x}} \right|, & \text{if } 0 \notin \boldsymbol{x} \\ \text{rad}\ \boldsymbol{x}, & \text{otherwise} \end{cases}.$$

And the maximum relative error for $\boldsymbol{x} \in \mathbb{IR}^n$ denoted by $\texttt{mrelerr}(\boldsymbol{x})$ is defined by

$$\texttt{mrelerr}(\boldsymbol{x}) = \max_i \{\texttt{relerr}(\boldsymbol{x}_i)\},$$

where $\boldsymbol{x}_i \in \mathbb{IR},\ i = 1, 2, \ldots, n$ is the ith entry of $\boldsymbol{x}$.

Example 5.2. Consider the two point boundary value problem

$$3\ddot{y}y + \dot{y}^2 = 0, \quad \text{with } y(0) = 0,\ y(1) = 20,$$

given by Abbott and Brent [29]. The true solution is $y = 20x^{0.75}$. The above equation can be discretized as

$$\begin{cases} f_k(y) \equiv 3y_k\,(y_{k+1} - 2y_k + y_{k-1}) + \left(\dfrac{y_{k+1} - y_{k-1}}{2}\right)^2 = 0,\ 1 \le k \le n \\ y_0 = 0,\ y_{n+1} = 20, \end{cases} \tag{9}$$

As initial approximation we use the values at the equally spaced points in $[0, 20]$. The results outputted by both algorithms for different dimensions, with only one verification step executed, are displayed in Table 5.2.

Example 5.3 ([30])**.** Consider the discretization of

$$u''(t) = \frac{1}{2}\,(u(t) + t + 1)^3,\ 0 < t < 1, \quad u(0) = u(1) = 0.$$

Denote $u_k = u\,(t_k)$, we have

$$\begin{cases} f_k(u) \equiv u_{k+1} - 2u_k + u_{k-1} - \dfrac{1}{2}h^2\,(u_k + t_k + 1)^3 = 0,\ 1 \le k \le n \\ u_0 = u_{n+1} = 0,\ t_k = k \cdot h;\ h = (n+1)^{-1}, \end{cases} \tag{10}$$

Take initial guess $u \equiv (\xi_i)$, $\xi_i = t_i\,(t_i - 1)$, $1 \le i \le n$. The results outputted by both algorithms for different dimensions, with only one verification step executed, are displayed in Table 5.3.

Example 5.4 ([30])**.**

$$\overline{u}(t) + \int_0^1 H(s, t)\,(\overline{u}(s) + s + 1)^3\,ds = 0,$$

Table 5.3
Solution of (10) using algorithm 2.5 and algorithm 4.1.

Dim	Algorithm 2.5 Computing time t_1 (s)	Algorithm 4.1 Computing time t_2 (s)	$t_1 : t_2$	mrelerr ($\boldsymbol{x}$)
50	0.683220	0.593479	1.1512	$1.450 \cdot 10^{-15}$
100	1.530358	1.187107	1.2892	$2.917 \cdot 10^{-15}$
200	3.037722	2.472819	1.2285	$8.018 \cdot 10^{-15}$
500	7.266109	6.235905	1.1652	$8.542 \cdot 10^{-15}$
1000	14.799065	12.831169	1.1534	$4.058 \cdot 10^{-14}$
2000	34.824131	27.656904	1.2592	$8.523 \cdot 10^{-14}$

Where $\boldsymbol{x} =$ XX1 or XX2, i.e., mrelerr (XX1)=mrelerr (XX2).

Table 5.4
Solution of (11) using algorithm 2.5 and algorithm 4.1.

Dim	Algorithm 2.5 Computing time t_1 (s)	Algorithm 4.1 Computing time t_2 (s)	$t_1 : t_2$	mrelerr ($\boldsymbol{x}$)
10	0.864388	0.774734	1.1157	$4.7 \cdot 10^{-15}$
20	3.520333	2.983376	1.1800	$1.7 \cdot 10^{-14}$
50	25.508979	18.800986	1.3568	$2.3 \cdot 10^{-13}$
100	105.427856	66.999370	1.5736	$9.1 \cdot 10^{-13}$

Where $\boldsymbol{x} =$ XX1 or XX2, i.e., mrelerr (XX1)=mrelerr (XX2).

where $H(s,t) = \begin{cases} s(1-t), & s \le t \\ t(1-s), & s > t \end{cases}$.

$$\begin{cases} f_k(u) \equiv u_k + \dfrac{1}{2}\left\{(1-t_k)\sum_{j=1}^{k} t_j \left(u_j + t_j + 1\right)^3 + t_k \sum_{j=k+1}^{n} \left(1-t_j\right)\left(u_j + t_j + 1\right)^3\right\} = 0 \\ u_0 = u_{n+1} = 0, \;\; t_j = j \cdot h; \;\; h = (n+1)^{-1} \end{cases}, \tag{11}$$

where $u_k = \overline{u}\,(t_k)\,, \;\; 1 \le k \le n$.

Take initial guess $u_i = t_i\,(t_i - 1)$, $1 \le i \le n$. The results outputted by both algorithms for different dimensions, with only one verification step executed, are displayed in Table 5.4.

6. Conclusion

In this paper, we provided an improved method for the verification algorithm of nonlinear systems. The new method is based on a modified version of Krawczyk operator, and is faster than the classic one and gives also narrower inclusion of the solution in some cases.

Because Krawczyk operator (2) is nothing but the centered form of the interval extension [10,26,27] of the mapping

$$\phi(x) = x - Rf(x)$$

from the function interval extension [9,10,26] point of view, in this direction our future research will be to develop more efficient and effective algorithms for verified solutions to nonlinear systems with the idea of bicentered interval extension of functions and/or the boundary Krawczyk operator first presented in [28].

Acknowledgments

The authors would like to thank the Editors and the Reviewers for the warm work earnestly and the valuable suggestions and comments.

References

[1] N. Yamamoto, Numerical verification method for solutions of boundary value problems with local uniqueness by Banach's fixed-point theorem, SIAM J. Numer. Anal. 35 (5) (1998) 2004–2013.
[2] P. Zgliczynski, K. Mischaikow, Rigorous numerics for partial differential equations: the Kuramoto-Sivashinsky equation, Found. Comput. Math. 1 (3) (2001) 255–288.
[3] Z. Galias, Interval methods for rigorous investigations of periodic orbits, Int. J. Bifurcation Chaos 11 (9) (2001) 2427–2450.
[4] S. Day, O. Junge, K. Mischaikow, A rigorous numerical method for the global analysis of infinite-dimensional discrete dynamical systems, SIAM J. Appl. Dyn. Syst. 3 (2) (2004) 117–160.
[5] H. Wang, D. Cao, S. Li, Interval entropy method for equality constrained multiobjective optimization problems, Siberian J. Numer. Math. 11 (1) (2008) 29–39.

[6] M.D. Stuber, V. Kumar, P.I. Barton, Nonsmooth exclusion test for finding all solutions of nonlinear equations, BIT Numer. Math. 50 (4) (2010) 885–917.
[7] B. Hu, K. Xie, H. Tai, Inverse problem of power system reliability evaluation: analytical model and solution method, IEEE Trans. Power Syst. 33 (6) (2018) 6569–6578.
[8] S.M. Rump, Verification methods: Rigorous results using floating-point arithmetic, Acta Numer. 19 (2010) 287–449.
[9] T. Sunaga, Theory of an interval algebra and its application to numerical analysis, Res. Assoc. Appl. Geom. (RAAG) Mem. 2 (1958) 29–46.
[10] R.E. Moore, Interval Analysis, Prentice-Hall, Englewood Cliffs, NJ, 1966.
[11] R. Krawczyk, Newton-Algorithmen zur bestimmung von nullstellen mit fehlerschranken, Computing 4 (1969) 187–201.
[12] R.E. Moore, A test for existence of solutions to nonlinear systems, SIAM J. Numer. Anal. 14 (4) (1977) 611–615.
[13] S.M. Rump, Solving algebraic problems with high accuracy, habilitationsschrift, in: U.W. Kulisch, W.L. Miranker (Eds.), A New Approach to Scientific Computation, Academic Press, New York, 1983, pp. 51–120.
[14] S.M. Rump, INTLAB-Interval laboratory, in: Tibor Csendes (Ed.), Developments in Reliable Computing, Kluwer Academic Publishers, Dordrecht, 1999, pp. 77–104.
[15] Z. Yang, L. Zhi, Y. Zhu, Verified error bounds for real solutions of positive-dimensional polynomial systems, in: Proceedings of the 38th international symposium on symbolic and algebraic computation, ACM, 2013, pp. 371–378.
[16] X. Chen, R.S. Womersley, Existence of solutions to systems of underdetermined equations and spherical designs, SIAM J. Numer. Anal. 44 (6) (2006) 2326–2341.
[17] X. Chen, A. Frommer, B. Bruno Lang, Computational existence proofs for spherical t-designs, Numer. Math. 117 (2011) 289–305.
[18] S.M. Rump, S. Oishi, Verified error bounds for double roots of nonlinear equations, in: 2009 International Symposium on Nonlinear Theory and Its Applications, NOLTA'09, Sapporo, Japan, 2009.
[19] S.M. Rump, S. Graillat, Verified error bounds for multiple roots of systems of nonlinear equations, Numer. Algorithms 54 (2010) 359–377.
[20] N. Li, L. Zhi, Verified error bounds for isolated singular solutions of polynomial systems: case of breadth one, Theoret. Comput. Sci. 479 (2013) 163–173.
[21] N. Li, L. Zhi, Verified error bounds for isolated singular solutions of polynomial systems, SIAM J. Numer. Anal. 52 (4) (2014) 1623–1640.
[22] Z. Li, H. Sang, Verified error bounds for singular solutions of nonlinear systems, Numer. Algorithms 30 (2015) 309–331.
[23] A. Frommer, B. Hashemi, Verified computation of square roots of a matrix, SIAM J. Matrix Anal. Appl. 31 (3) (2009) 1279–1302.
[24] Tayyebe Haqiri, Federico Poloni, Methods for verified stabilizing solutions to continuous-time algebraic Riccati equations, J. Comput. Appl. Math. 313 (2017) 515–535.
[25] R.B. Kearfott, M. Nakao, A. Neumaier, S.M. Rump, S.P. Shary, P. van Hentenryck, Standardized notation in interval analysis, Comput. Technol. 15 (1) (2010) 7–13.
[26] A. Neumaier, Interval methods for systems of equations, Cambridge University Press, Cambridge, UK, 1990.
[27] R.B. Kearfott, Rigorous Global Search: Continuos Problems, Kluwer, Dordrecht, Netherlands, 1996.
[28] S.P. Shary, Krawczyk operator revised, in: Proceedings of International Conference on Computational Mathematics ICCM-2004. Workshops, ICM & MG Publisher, Novosibirsk, 2004, pp. 307–313.
[29] J.P. Abbott, R.P. Brent, Fast local convergence with single and multistep methods for nonlinear equations, Austr. Math. Soc. 19 (Series B) (1975) 173–199.
[30] J.J. Moré, M.Y. Cosnard, Numerical solution of non-linear equations, ACM Trans. Math. Softw. 5 (1) (1979) 64–85.