Available online at www.sciencedirect.com
ScienceDirect
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
Journal of Computational and Applied Mathematics 199 (2007) 397–402
www.elsevier.com/locate/cam

Proving the existence of zeros using the topological degree and interval arithmetic

A. Frommer[a,*], F. Hoxha[b], B. Lang[a]

[a] Bergische Universität Wuppertal, Applied Computer Science and Scientific Computing, D-42097 Wuppertal, Germany
[b] Department of Mathematics, Faculty of Natural Sciences, University of Tirana, Bld. Zogu I, Tirana, Albania

Received 21 December 2004

Abstract

The invariance of the topological degree under certain homotopies is used to computationally prove the existence of zeros of nonlinear mappings in $\mathbb{R}^n$. These existence tests use interval arithmetic to enclose the range of a function over a box. We show that our test is more general than a well-known test based on Miranda's theorem, and we show by a numerical example that the new test can be successful on substantially larger boxes.

Keywords: Nonlinear systems; Topological degree; Rigorous solution; Interval computations; Existence tests

1. Introduction

Let $\Omega \subseteq \mathbb{R}^n$ be open and bounded and let $F : \overline{\Omega} \to \mathbb{R}^n$ be a continuous mapping. Let $\partial\Omega$ denote the boundary of Ω and assume that $y \in \mathbb{R}^n$ is such that $F(x) \neq y$ for all $x \in \partial\Omega$. Then the topological degree $d(\Omega, F, y)$, an integer, is defined, and it gives us information on the number of solutions $x \in \Omega$ of $F(x) = y$. We refer to [2,11] for a thorough introduction to the topological degree. For the purpose of the present paper, out of the many attractive properties of the topological degree, only the following will be used.

Theorem 1 (*See [2,11]*).

(i) *If F is (Fréchet-)differentiable and the Jacobian $F'(x)$ is non-singular at all points $x \in \Omega$ where $F(x) = y$, then*

$$d(\Omega, F, y) = \sum_{x\in\Omega,\, F(x)=y} \operatorname{sign}(\det F'(x)).$$

(*Note that $\{x \in \Omega,\ F(x) = y\}$ is necessarily finite.*)

* Corresponding author.
E-mail addresses: frommer@math.uni-wuppertal.de (A. Frommer), hoxhaf@fshn.edu.al (F. Hoxha), lang@math.uni-wuppertal.de (B. Lang).

doi:10.1016/j.cam.2005.07.030

(ii) *If* $H : [0,1] \times \overline{\Omega} \to \mathbb{R}^n$ *is continuous and such that* $H(t,x) \neq y$ *for all* $t \in [0,1]$ *and for all* $x \in \partial\Omega$, *then* $d(\Omega, H(t,\cdot), y)$ *does not depend on* t.
(iii) *If* $d(\Omega, F, y) \neq 0$, *then there exists an* $x \in \Omega$ *such that* $F(x) = y$.

In this paper we want to computationally prove the existence of a zero of f within a box in $\mathbb{R}^n$, i.e. we have $y = 0$ and $\Omega = \mathrm{int}([x])$, the topological interior of an interval vector $[x]$ in $\mathbb{R}^n$, see Section 2. We will use the fact that by (i) of Theorem 1 the mapping $I_{\check{x}} : x \mapsto x - \check{x}$ satisfies

$$d(\mathrm{int}([x]), I_{\check{x}}, 0) = 1 \quad \text{whenever } \check{x} \in \mathrm{int}([x]).$$

We will then consider the homotopy

$$H(t,x) = t \cdot A \cdot F(x) + (1-t) \cdot (x - \check{x}),$$

where $A \in \mathbb{R}^{n \times n}$ is a suitable, non-singular matrix. By the use of interval arithmetic we will be able to computationally check sufficient conditions such that $H(t,x) = 0$ has no solution on $\partial[x]$ for all $t \in [0,1]$. By (ii) and (iii) of Theorem 1 we then have shown that $A \cdot F$ and thus F has a zero in the box $[x]$, since then $d(\Omega, A \cdot F, 0) = d(\Omega, I_{\check{x}}, 0) = 1$.

As we will explain, this approach might be seen as a generalization of the existence test of Moore and Kioustelidis [9], which is based on Miranda's theorem [7]. This test, in turn, may essentially be regarded as a generalization of a well-known existence test based on the Krawczyk operator [8] as was recently investigated in [1,5,13]. The new test also relates to the tests based on Borsuk's theorem presented in [3,4].

2. A new existence test

We first introduce some additional notation. A real compact interval is indicated by square brackets, e.g. $[a]$. Its lower and upper bound are $\underline{a}, \overline{a}$, respectively, so that we have $[a] = [\underline{a}, \overline{a}]$.

The set of all compact real intervals is denoted $\mathbb{IR}$. A *box* in $\mathbb{R}^n$ is an interval vector $[x] = ([x_1], [x_2], \ldots, [x_n])^{\mathrm{T}}$, where $[x_i] \in \mathbb{IR}$. The topological boundary of an interval vector $[x]$ is given by its n pairs of opposite facets

$$\left.\begin{array}{l} [x]^{i,+} = ([x_1], \ldots, [x_{i-1}], \overline{x}_i, [x_{i+1}], \ldots, [x_n])^{\mathrm{T}} \\ [x]^{i,-} = ([x_1], \ldots, [x_{i-1}], \underline{x}_i, [x_{i+1}], \ldots, [x_n])^{\mathrm{T}} \end{array}\right\}, \quad i = 1, \ldots, n.$$

Given $F : [x] \to \mathbb{R}^n$ we denote the components of F by F_j, i.e. $F(x) = (F_1(x), \ldots, F_n(x))^{\mathrm{T}}$. Actually, instead of F we will mainly be concerned with an affinely scaled version

$$G(x) = A \cdot F(x), \quad A \in \mathbb{R}^{n \times n} \text{ non-singular},$$

where A is chosen such that G resembles $I_{\check{x}}$ on $[x]$.

The range of G or of one of its components over some set $S \subseteq [x]$ is denoted by

$$\mathrm{range}(G, S) = \{G(x), x \in S\}.$$

If S is an interval vector $[y]$ then, by continuity, $\mathrm{range}(G_j, [y])$ is an interval for each component G_j. We assume that we are able to compute enclosing interval vectors $[r([y])]$ for $\mathrm{range}(G, [y])$, i.e. we have for $j = 1, \ldots, n$

$$[r([y])_j] \supseteq \mathrm{range}(G_j, [y]). \tag{1}$$

The new existence test is based on the following theorem in which the arithmetic operations are to be understood as standard interval operations.

Theorem 2. *Let* $[x]$ *be an interval vector in* $\mathbb{R}^n$ *and* $\check{x} \in \mathrm{int}([x])$ *be a point vector. Let*

$$\partial[x] = \bigcup_{k=1}^{m} [y^k] \quad \textit{and} \quad [0,1] = \bigcup_{\ell=1}^{p} [t^\ell]$$

be coverings of $\partial[x]$ *and* $[0, 1]$ *by interval vectors and intervals, respectively. Then, if for all* $k = 1, \ldots, m$ *and for all* $\ell = 1, \ldots, p$ *there exists a* $j = j(k, \ell) \in \{1, \ldots, n\}$ *such that*

$$0 \notin [t^\ell] \cdot [r_j([y^k])] + (1 - [t^\ell]) \cdot ([y^k] - \check{x})_j \tag{2}$$

with $[r([y^k])_j]$ *according to* (1), *the function* G (*and thus* F) *has a zero* x^* *in* $\mathrm{int}([x])$.

Proof. The condition (2) implies that for all k and ℓ we have

$$0 \notin [t^\ell] \cdot [r([y^k])] + (1 - [t^\ell]) \cdot ([y^k] - \check{x})$$

and thus

$$\begin{aligned} 0 &\notin \bigcup_{k=1}^{m} \bigcup_{\ell=1}^{p} ([t^\ell] \cdot [r([y^k])] + (1 - [t^\ell]) \cdot ([y^k] - \check{x})) \\ &\supseteq \{t \cdot G(x) + (1 - t) \cdot (x - \check{x}),\ t \in [0, 1],\ x \in \partial[x]\}, \end{aligned}$$

the last relation holding due to (1) and the inclusion property of interval arithmetic. Therefore, the homotopy $H(x, t) = t \cdot G(x) + (1 - t) \cdot (x - \check{x})$ does not go through a zero on $\partial[x]$ so that $d(\Omega, G, 0) = d(\Omega, I_{\check{x}}, 0) = 1$. □

Enclosures $[r([y^k])_j]$ for $\mathrm{range}(G_j, [y^k])$ can be obtained in several different ways. Of these, the mean value forms yield reasonably tight enclosures at affordable cost. To describe them, let $\check{y}^k$ be a point in $[y^k]$, usually the midpoint, and assume $[Y^k]$ is an interval slope matrix satisfying

$$F(y) - F(\check{y}^k) \in [Y^k](y - \check{y}^k) \quad \text{for all } y \in [y^k].$$

Then

$$AF(y) \in AF(\check{y}^k) + (A[Y^k]) \cdot (y - \check{y}^k) \quad \text{for all } y \in [y^k]$$

and therefore

$$\mathrm{range}(G, [y^k]) \subseteq AF(\check{y}^k) + (A[Y^k]) \cdot ([y] - \check{y}^k) =: [\mathrm{rm}(G, [y^k])]. \tag{3}$$

If F is differentiable, one can take $[Y^k]$ to be any interval matrix with

$$[Y^k] \supseteq \{F'(y), y \in [y^k]\}.$$

One such $[Y^k]$ can be obtained, for example, as an interval arithmetic evaluation of the Jacobian F'. We refer to [6,10] for other means to computationally obtain slope matrices.

Note that the test of Theorem 2 cannot be successful if we do not subdivide the interval [0, 1] for t. Indeed, if $p = 1$ and, consequently, $[t^1] = [0, 1] = 1 - [t^1]$, we necessarily have

$$0 \in [t^1] \cdot [r([y^k])] + (1 - [t^1]) \cdot ([y^k] - \check{x})$$

for all k and for all components j.

As soon as we take $p \geqslant 2$, however, the test implicated by Theorem 2 seems to bear quite a big potential. For example, we have the following proposition where we chose the $[y^k]$ as the facets $[x]^{i,\pm}$ of $[x]$.

Proposition 1. *Assume that we have the covering*

$$\partial[x] = \bigcup_{i=1}^{n} [x]^{i,+} \cup [x]^{i,-}$$

and that for $i = 1, \ldots, n$

$$[r_i([x]^{i,+})] > 0, \quad [r_i([x]^{i,-})] < 0 \tag{4}$$

holds, meaning that all elements of these intervals satisfy the respective inequality. Then, if we take $p = 2$ *and* $[t^1] = [0, 0.5]$, $[t^2] = [0.5, 1]$, *we have that* (2) *is fulfilled for* $\ell = 1, 2$ *and all facets* $[x]^{i,\pm}$ *with* $j = j(i, \ell) = i$.

Proof. Since we have

$$([x]^{i,+} - \check{x})_i > 0 \quad \text{and} \quad ([x]^{i,-} - \check{x})_i < 0$$

we see that

$$[t^1] \cdot [r_i([x]^{i,+})] + (1 - [t^1]) \cdot ([x]^{i,+} - \check{x})_i > 0,$$
$$[t^2] \cdot [r_i([x]^{i,+})] + (1 - [t^2]) \cdot ([x]^{i,+} - \check{x})_i > 0,$$
$$[t^1] \cdot [r_i([x]^{i,-})] + (1 - [t^1]) \cdot ([x]^{i,-} - \check{x})_i < 0,$$
$$[t^2] \cdot [r_i([x]^{i,-})] + (1 - [t^2]) \cdot ([x]^{i,-} - \check{x})_i < 0,$$

which shows that (2) holds. □

Testing for (4) has been proposed by Moore and Kioustelidis in [9], where, using a theorem of Miranda [7], it was actually shown that we can allow for non-strict inequalites in (4). With the exception of this slightly larger generality, we see that the Moore–Kioustelidis test may be regarded as a special case of our new test.

The Moore–Kioustelidis test can be successful only if for $i = 1, \ldots, n$ the component G_i has constant non-negative sign on $[x]^{i,+}$ and constant non-positive sign on $[x]^{i,-}$. The new test, however, can still be successful in situations where this is not true as the following numerical example will show.

3. A numerical example

As in [3], we consider the simple two-dimensional mapping

$$F(u, v) = \begin{pmatrix} 4 - 2(u-1)^2 \\ (2 - (u+1)^2) \cdot (2 - (v-1)^2) \end{pmatrix},$$

which has a unique zero $x^* = (u^*, v^*) = (1 - \sqrt{2}, 1 - \sqrt{2})$ in the box $[x] = [u] \times [v] = [-1, 1] \times [-1, 1]$. As was shown in [3], any test based on Miranda's theorem for the box $[x]$ must fail for this function, since for any non-singular matrix A the function $G(x) = A \cdot F(x)$ does not have constant sign on the respective facet in the respective component. We performed our new test for this function on boxes

$$[x]^q = x^* + 2^{q/8} \cdot ([-0.5, 0.5], [-0.5, 0.5]), \quad q = 0, 1, 2, \ldots.$$

Thus the width of the test boxes increases geometrically with rather small steps. The enclosures for the range of G were obtained using the mean value form (3) with the slope matrix obtained by the 'naive' interval arithmetic evaluation of the Jacobian of F. The matrix A in $G = AF$ was either taken to be the identity or it was the inverse of the midpoint matrix of the naive interval arithmetic evaluation of the Jacobian over the whole box $[x]^q$.

The t interval [0, 1] was subdivided into $p = 2$ or $p = 10$ equally sized subintervals. The boxes $[y]^k$ were obtained by the following strategy: as a first try we started with the set of all facets of $[x]^q$. Those facets where (2) was not fulfilled were then bisected along their longest edge and we tried the test on each of these parts. If this is still not successful, we recursively bisect further until a given limit on the total number of boxes to be considered is reached. This bisection process is, of course, costly, so that a successful test for a small number of bisections is better than a successful one with more bisections.

Figs. 1 and 2 summarize our numerical results which were obtained using Rump's INTLAB toolbox for Matlab [12]. The horizontal axis gives the number of bisections we allowed for each facet. The vertical axis plots the diameter

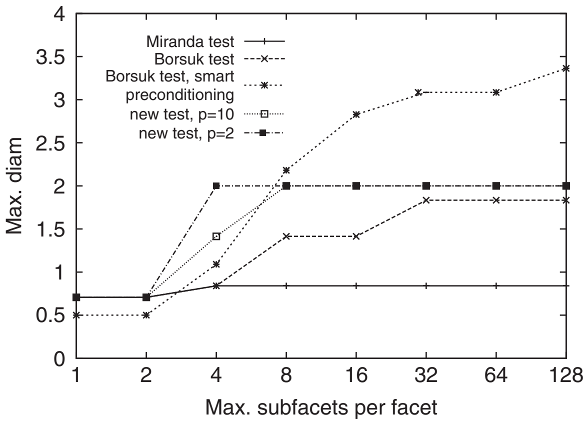

Fig. 1. Results for different tests, inverse midpoint preconditioning.

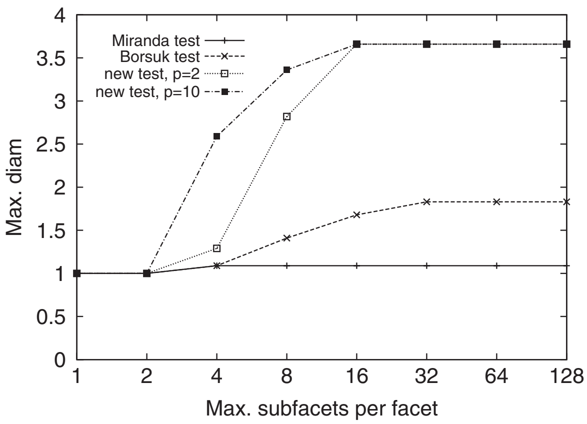

Fig. 2. Results for different tests, no preconditioning.

of the largest box $[x]^q$ for which the test was successful. In addition to our new test, we also give results for the Miranda test using bisection (checking for constant sign of G_i on all parts of the facets $[x]^{i,\pm}$) and for the Borsuk test from [3,4].

In Fig. 1, in all but one test we took A as the computed inverse of the midpoint of $[F'([x]^q)]$, the so-called inverse midpoint preconditioning. The exception is one of the tests based on Borsuk's theorem, where we used a more advanced preconditioning technique which is done individually for each pair of facets ('smart' preconditioning), see [4] for details. Fig. 2 gives the results in the case where we do not apply any preconditioning, i.e. when $A = I$.

From Fig. 1 we see that our new test is better if we allow for only a small number of subdivisions. We also see that the new test is always superior to the Miranda test and to the Borsuk test using midpoint preconditioning. The number p of subdivisions of the interval [0, 1] for t has a remarkably little influence on the results. Note that making p larger does not introduce substantial additional cost, since the number of (interval) function and derivative evaluations remains unchanged. The Borsuk test with 'smart' preconditioning performs better than the new test if we allow for a larger number of subdivisions. In this context it should be noted that the 'smart' preconditioning technique requires additional cost, see [4].

The results in Fig. 2 are somewhat surprising: a comparison with Fig. 1 shows that the midpoint-inverse preconditioned methods actually perform worse than their non-preconditioned counterparts. We attribute this to the large size of the boxes considered. In the unpreconditioned case, the new test is always best, and it even beats the Borsuk test with 'smart' preconditioning.

Acknowledgements

F. Hoxha gratefully acknowledges support by DAAD through its exchange visit program in 2004. B. Lang was partially supported by the VolkswagenStiftung within the project “Konstruktive Methoden der Nichtlinearen Dynamik zum Entwurf verfahrenstechnischer Prozesse”, Geschäftszeichen I/79 288.

References

[1] G. Alefeld, F. Potra, Z. Shen, On the existence theorems of Kantorovich, Moore and Miranda, Comput. Suppl. 15 (2001) 21–28.
[2] K. Deimling, Nonlinear Functional Analysis, Springer, Berlin, Heidelberg, New York, 1985.
[3] A. Frommer, B. Lang, Existence tests for solutions of nonlinear equations using Borsuk’s theorem, SIAM J. Numer. Anal. 43 (2005) 1348–1361.
[4] A. Frommer, B. Lang, On preconditioners for the Borsuk existence test, Proc. Applied Math. Mech. 4 (2004) 638–639.
[5] A. Frommer, B. Lang, M. Schnurr, A comparison of the Moore and Miranda existence tests, Computing 72 (2004) 349–354.
[6] E.R. Hansen, Global Optimization Using Interval Analysis, Marcel Dekker, New York, 1992.
[7] C. Miranda, Un’ osservazione su un teorema di Brouwer, Bolletino Unione Mathematica Italiana (1940) 5–7.
[8] R.E. Moore, A test for existence of solutions to nonlinear systems, SIAM J. Numer. Anal. 14 (1972) 611–615.
[9] R. Moore, J. Kioustelidis, A simple test for accuracy of approximate solutions to nonlinear (or linear) systems, SIAM J. Numer. Anal. 17 (1980) 521–529.
[10] A. Neumaier, Interval Methods for Systems of Equations, Cambridge University Press, Cambridge, 1990.
[11] J. Ortega, W. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, 1970.
[12] S.M. Rump, INTLAB—interval laboratory, in: T. Csendes (Ed.), Developments in Reliable Computing, Kluwer, New York, 1999, pp. 77–104.
[13] M. Schnurr, On the proofs of some statements concerning the theorems of Kantorovich, Moore and Miranda, Reliab. Comp. 11 (2005) 77–85.