ZAMM · Z. Angew. Math. u. Mech. 65 (1985) 12, 49 – 55

NEUMAIER, A.

Existence Regions and Error Bounds for Implicit and Inverse Functions

Dedicated to Prof. KARL NICKEL on the occasion of his sixtieth birthday.

Es werden verschiedene Sätze bewiesen, die die Existenz implizit definierter bzw. inverser Funktionen in vorgeschriebenen Gebieten garantieren. Als Anwendungen werden Fehlerschranken für ungenaue Lösungsabbildungen impliziter Gleichungssysteme und für ungenäherte Nullstellen für Gleichungssysteme abgeleitet.

Several theorems are proved which guarantee the existence of implicitly defined functions and inverse functions, respectively, in prescribed domains. As applications, error bounds for inaccurate solution maps of implicit systems of equations and for approximate zeros of systems of equations are derived.

Доказываются некоторые теоремы обеспечивающие существование неявно определенных функций и обратных функций, соответственно, в предписанных областях. В качестве примеров выводятся границы погрешности для неточных преобразований решения неявных систем уравнений и для приближенных корней систем уравнений.

Introduction

Motivated by the problem of numerically constructing error bounds for approximate zeros of systems of equations which may or may not involve free parameters we investigate in this paper constructive conditions which guarantee that a given implicit system of equations $F(x, y) = 0$ allows for all y in a given region $E \subseteq \mathbb{R}^m$ a solution function $x = H(y)$ with values in another given region $D \subseteq \mathbb{R}^n$. The map H may or may not be required to be continuous. As a special case, conditions are sought which guarantee that a given n-dimensional function has an inverse function in a given region $E \subseteq \mathbb{R}^n$ with values in $D \subseteq \mathbb{R}^n$.

In contrast to the well-known local results — the inverse and implicit function theorem — which specify conditions for which there are sufficiently small neighbourhoods D, E of points such that the above holds, there are only a few results concerning larger regions. Namely, a number of conditions, like bounded inverse of the derivative ("Hadamard's theorem"), norm coercivity, or uniform monotonicity guarantee global invertibility of a differentiable function (cf. ORTEGA and RHEINBOLDT [7]) — the case $E = \mathbb{R}^n$, and hidden in exercises of DIEUDONNÉ [2] (Ex. X. 2.6 -|- 7) some global results are derived for the case of implicit functions. But little seems to have been done in the semilocal case, where D and E are bounded regions of $\mathbb{R}^n$. The present paper is designed to fill this gap.

The organization of the material is as follows. In Section 1 we prove as a basic result a semilocal implicit function theorem. Nonsingular (Fréchet-) derivative and a simple boundary condition suffice to ensure the existence of the solution function in the prescribed domain (Theorem 1). As in DIEUDONNÉ's treatment of the global case, the local implicit function theorem is combined with continuation techniques to obtain the result. A simple transformation then allows to extend the theorem to the case when an initial solution point is not known (Theorem 2). As an application, a computable, realistic error bound for approximate solution functions is derived (see the remarks following Proposition 5).

In Section 2 particular cases of the fixpoint theorems of Brouwer and Leray-Schauder are derived from Theorem 1, indicating that there might be a more general theorem. Indeed, using the methods of degree theory, the differentiability assumption can be considerably weakened, resulting in a theorem (Theorem 3) which contains these fixpoint theorems as particular cases. The price paid is that the resulting solution map need no longer be continuous.

Section 3 is concerned with applications to inverse functions. A semilocal inverse function theorem (Theorem 5) is supplemented by two uniqueness criteria (Propositions 6 and 7). The power of the semilocal inverse function theorem is demonstrated by the derivation of the (global) norm coercivity theorem (Corollary 3) and several criteria for the existence of a zero in a prescribed domain. Suggestions how to verify the boundary conditions involved conclude this section.

The terminology and notation is close to that of ORTEGA and RHEINBOLDT [7]. We refer to this book for background results; e.g. reference to result 5.2.1 of [7] is given as OR 5.2.1. Another useful general reference is SCHWETLICK [10].

1. A semilocal implicit function theorem

In this section our fundamental result, the semilocal implicit function theorem, will be proved and discussed. We give a simple and an extended version of the theorem; the extended version has the special feature that it does not require an initial solution point of the implicit system of equations. Finally, a numerical test is given to verify the validity of the boundary condition. It is shown that this test can be used to bound the error in approximate solution maps of implicit functions.

Theorem 1 (Semilocal implicit function theorem): *Let D be an open and bounded subset of $\mathbb{R}^n$, and let E be a simply connected Hausdorff space. Let $F: \bar{D} \times E \to \mathbb{R}^n$ be continuous, and suppose that the following two conditions hold:*

(H1) $\partial_1 F(x, y)$ *exists for* $x \in D$, $y \in E$, *is continuous and nonsingular.*

(H2) $F(x, y) \neq 0$ *for* $x \in \partial D$, $y \in E$.

Then for given $(x^0, y^0) \in D \times E$ *with* $F(x^0, y^0) = 0$, *there is a unique continuous map* $H: E \to D$ *with*

$$Hy^0 = x^0\,, \qquad F(Hy, y) = 0 \quad \textit{for} \quad y \in E\,. \tag{1), (2)}$$

Here we call the space E *simply connected* if (i) any two points $y^0, y^1 \in E$ can be joined by a (continuous) path $q: [0, 1] \to E$ with $q(0) = y^0$, $q(1) = y^1$, and (ii) if $q = q_0$ and $q = q_1$ are two such paths then there exists a continuous (deformation) map $q: [0, 1] \times [0, 1] \to E$ such that $q(s, 0) = y^0$, $q(s, 1) = y^1$, $q(0, t) = q_0(t)$, $q(1, t) = q_1(t)$ for all $s, t \in [0,1]$. In applications, E often is an interval or a subset of $\mathbb{R}^n$ with the induced topology.

The proof of Theorem 1 is split into several propositions. Proposition 1 is the well-known local implicit function theorem; Proposition 2 is a slight extension for the case when there are several solutions of $F(x, y) = 0$ at a point $y = y^0$. Proposition 3 contains the core of the argument, essentially treating the case where E is the unit square. As an immediate consequence, continuation along arbitrary paths is possible (Proposition 4). Finally, the simple connectedness of E allows to combine Propositions 3 and 4 to prove the theorem.

In the statements of the following propositions we always assume that the hypothesis of Theorem 1 holds.

Proposition 1: *If* $(x^1, y^1) \in D \times E$ *satisfies* $F(x^1, y^1) = 0$ *then there are open neighbourhoods* $D_1 \subseteq D$ *of* x^1 *and* $E_1 \subseteq E$ *of* y^1 *such that the equation* $F(x, y) = 0$ *has for all* $y \in E_1$ *a unique solution* $x = H_1 y \in D_1$. *Moreover, the map* $H_1: E_1 \to D_1$ *defined in this way is continuous.*

Proof: The proof of OR 5.2.4 applies with only trivial modifications. □

Proposition 2: *For* $y^1 \in E$, *the set*

$$X := \{x \in \bar{D} \mid F(x, y^1) = 0\}$$

is finite, and there are open, disjoint neighbourhoods $D(x) \subseteq D$ *of* $x \in X$ *and a closed neighbourhood* $E_1 \subseteq E$ *of* y^1 *such that the equation* $F(\xi, y) = 0$ *has for all* $y \in E_1$ *and all* $x \in X$ *a unique solution* $\xi = H_x y \in D(x)$. *Moreover, the maps* $H_x: E_1 \to D(x)$ *defined in this way are continuous.*

Proof. By (H2), X is a subset of D; hence by Proposition 1 there are (for $x \in X$) open neighbourhoods $D_1(x) \subseteq D$ of x and $E_1(x) \subseteq E$ of y^1 such that the equation $F(\xi, y) = 0$ has for all $y \in E_1(x)$ a unique solution $\xi = H_x y \in D_1(x)$, and $H_x: E_1(x) \to D_1(x)$ is continuous. In particular, for $x \in X$, the set $D_1(x)$ contains no point of X distinct from x. Therefore, X has no accumulation points and since X is closed (F continuous) and bounded (D bounded), X is finite. Hence $\mu := \min\{\|x - x'\| \mid x, x' \in X, x \neq x'\}$ is a positive number, the sets $D(x) := \{\xi \in D_1(x) \mid \|\xi - x\| < \mu/2\}$ are open neighbourhoods of $x \in X$, and $D(x) \cap D(x') = \emptyset$ for $x, x' \in X$, $x \neq x'$. Since X is finite and the H_x are continuous, the intersection E_1' of the sets $\{y \in E_1(x) \mid H_x y \in D(x)\}$, $x \in X$, is a neighbourhood of y^1. If we restrict H_x to a closed neighbourhood E_1 of y^1 contained in E_1', the assertion follows. □

Proposition 3: *If* $(x^0, y^0) \in D \times E$ *satisfies* $F(x^0, y^0) = 0$, *then for every continuous map* $q: [0, 1] \times [0,1] \to E$ *with*

$$q(s, 0) = y^0 \quad \textit{for} \quad s \in [0, 1]\,, \tag{3}$$

there is a unique continuous map $p: [0, 1] \times [0, 1] \to D$ *with*

$$p(s, 0) = x^0\,, \qquad F(p(s, t), q(s, t)) = 0 \tag{4), (5)}$$

for all $s, t \in [0, 1]$.

Proof: We write $J = [0, 1]$, and denote by Δ a subset of $J \times J$, maximal with respect to the properties

(i) $(s, t) \in \Delta\,, \qquad 0 \leqq u \leqq t \Rightarrow (s, u) \in \Delta\,,$

(ii) there is a unique continuous map $p: \Delta \to D$ such that (4) and (5) hold for all $(s, t) \in \Delta$.

Since (3) holds, $(s, 0) \in \Delta$ for all $s \in J$. In order to show that $\Delta = J \times J$ we define for $s \in J$ the number

$$t_s := \sup\{t \in J \mid (s, t) \in \Delta\}\,.$$

Since J is compact there is a number $\sigma \in J$ such that $\tau = t_\sigma$ is minimal. We now put

$$X = \{x \in \bar{D} \mid F(x, q(\sigma, \tau)) = 0\}\,.$$

By Proposition 2 there are disjoint, open neighbourhoods $D(x) \subseteq D$ of $x \in X$ and a closed neighbourhood $E_1 \subseteq E$ of $q(\sigma, \tau)$ such that the equation $F(\xi, y) = 0$ has for all $y \in E_1$ a unique solution $\xi = H_x y \in D(x)$, and the maps H_x are continuous. By continuity of p, the sets

$$\Delta(x) := \{(s, t) \in \Delta \mid p(s, t) \in D(x)\} \tag{6}$$

are disjoint and open in Δ. We now put

$$U(\varepsilon) := \{(s, t) \in J \times J \mid |s - \sigma| \leqq \varepsilon, |t - \tau| \leqq \varepsilon\}\,,$$

and show that there is $\delta > 0$ with

$$\Delta \cap U(\delta) \subseteq \Delta_0 := \bigcup_{x \in X} \Delta(x)\,. \tag{7}$$

For if this is not the case, there is a sequence $(s_i, t_i) \in \Delta \cap U(1/i)$, $i = 1, 2, \ldots$, such that $(s_i, t_i) \notin \Delta_0$, and by construction of $U(\varepsilon)$ we have $\lim_{i \to \infty} (s_i, t_i) = (\sigma, \tau)$. Since D is bounded, the sequence $p(s_i, t_i)$ has at least one accumulation point $x \in \bar{D}$. By continuity of F and q we have $F(x, q(\sigma, \tau)) = 0$, hence $x \in X$. By continuity of p, this implies $p(s_i, t_i) \in \Delta(x)$, hence $(s_i, t_i) \in \Delta(x) \subseteq \Delta_0$ for infinitely many values of i, a contradiction. Therefore, (7) holds for some $\delta > 0$.

By minimality of τ, $\Delta \cap U(\delta)$ is connected (two points (s, t), (s', t') are joined by line segments joining $(s, t) - (s, t_0) - (s', t_0) - (s', t')$, where $t_0 = \max(0, \tau - \delta)$, and this path is in $\Delta \cap U(\delta)$ if its endpoints are). Since the $\Delta(x)$ are disjoint and open in Δ, $\Delta \cap U(\delta)$ is contained in one of the $\Delta(x)$, say, in $\Delta(\bar{x})$, $\bar{x} \in X$. By (6),

$$p(s, t) \in D(\bar{x}) \quad \text{for} \quad (s, t) \in \Delta \cap U(\delta)\,. \tag{8}$$

Denote by $\bar{\tau}$ the largest $t \in J$ with $q(\sigma, t) \in E_1$. By continuity of q,

$$\text{either} \quad \bar{\tau} > \tau \,, \quad \text{or} \quad \bar{\tau} = \tau = 1 \,.$$

By (8) and the construction of $D(\bar{x})$, E_1, and $H_{\bar{x}}$, the unique continuous continuation of p along the line segment $s = \sigma, \tau \leqq t \leqq \bar{\tau}$ is given by $p(s, t) := H_{\bar{x}} q(s, t)$. Hence the maximality of Δ implies $\bar{\tau} = \tau = 1$, $(\sigma, 1) = (\sigma, \tau) \in \Delta$. Therefore, $\Delta = J \times J$, and the assertion is proved. □

Proposition 4: *If $(x^0, y^0) \in D \times E$ satisfies $F(x^0, y^0) = 0$ then for every continuous map $q_0 : [0, 1] \to E$ with $q_0(0) = y^0$ there is a unique continuous map $p_0 : [0, 1] \to D$ satisfying*

$$p_0(0) = x^0 \,, \qquad F(p_0(t), q_0(t)) = 0 \quad \text{for all} \quad t \in [0, 1] \,. \tag{4a), (5a)}$$

Proof: We apply Proposition 3 to the map q defined by $q(s, t) := q_0(t)$ for $s, t \in [0, 1]$. Then (4a) and (5a) hold for the map p_0 defined by $p_0(t) = p(0, t)$ for $t \in [0, 1]$. Conversely, if (4a) and (5a) hold then the map p' defined by $p'(s, t) = p_0(t)$ for $s, t \in [0.1]$ is continuous and satisfies (4) and (5) for $s, t \in [0, 1]$ (with p replaced by p'). Hence $p' = p$; in particular $p_0(t) = p'(0, t) = p(0, t)$ for $t \in [0, 1]$ and p_0 is unique. □

Proof of Theorem 1: Write again $J = [0, 1]$. Denote for arbitrary paths $q_0 : J \to E$ with $q_0(0) = y^0$ by $x(q_0)$ the endpoint $p_0(1)$ of the path p_0 defined by Proposition 4. We show that $x(q_0)$ depends only on the endpoint $y^1 = q_0(1)$ of q_0.

For, suppose that $q_1 : J \to E$ is another path with $q_1(0) = y^0$, $q_1(1) = y^1$. Since E is simply connected there is a continuous map $q : J \times J \to E$ such that

$$q(s, 0) = y^0 \,, \qquad q(s, 1) = y^1 \,, \quad \text{for} \quad s \in J \,; \qquad q(0, t) = q_0(t) \,, \qquad q(1, t) = q_1(t) \quad \text{for} \quad t \in J \,.$$

By Proposition 3 there is a unique continuous map $p : J \times J \to D$ satisfying (4) and (5). By Proposition 4, $p_0(t)$ coincides with $p(0, t)$; in particular,

$$x^1 := x(q_0) = p_0(1) = p(0, 1) \,.$$

Now the set $\Delta := \{s \in J \mid p(s, 1) = x^1\}$ is closed and nonempty; hence $\sigma := \sup \Delta \in \Delta$. By Proposition 1 there is a neighbourhood D_1 of $p(\sigma, 1) = x^1$ such that $x = x^1$ is the only solution of $F(x, y^1) = 0$ in D_1. If $\sigma < 1$ the continuity of p implies the existence of $s \in (\sigma, 1)$ such that $p(s, 1) \in D_1$. Now $F(p(s, 1), y^1) = F(p(s, 1), q(s, 1)) = 0$, hence $p(s, 1) = x^1$ by construction of D_1. But then $s \in \Delta$, contradicting the maximality of σ. Therefore $\sigma = 1$; in particular, $p(1, 1) = x^1 = x(q_0)$. But by Proposition 4, $p_1(t)$ coincides with $p(1, t)$; hence $x(q_1) = p_1(1) = p(1, 1) = x(q_0)$. Therefore, $x(q_0)$ indeed only depends on the endpoint of q_0.

This allows us to define a map $H : E \to D$ by the rule $Hy := x(q_0)$ where $q_0 : J \to E$ is an arbitrary path with $q_0(0) = y^0$, $q_0(1) = y$. This map satisfies (1) and (2), and is continuous by Proposition 1. Conversely, if $H' : E \to D$ is a continuous map satisfying (1) and (2) (with H replaced by H'), and if $q_0 : J \to E$ is a path with $q_0(0) = y^0$, $q_0(1) = y$ then the path p_0 defined by $p_0(t) := H' q_0(t)$ for $t \in J$ is (by Proposition 4) the unique path satisfying (4a) and (5a). Therefore $H' y = H' q_0(1) = p_0(1) = x(q_0) = Hy$, and H is unique. □

Remarks:

1. A similar result for the case when $D = \mathbb{R}^n$ is contained in DIEUDONNÉ [2], Ex. X.2.6, with a similar proof. For our applications to error bounds, however, it is relevant that D can be chosen as a small bounded region.

2. The difficulty in the proof of Theorem 1 is to show the existence of a continuous map H solving (1) and (2). To show the existence of some solution $x \in D$ with $F(x, y) = 0$ only one can proceed very simply:

Suppose $y \in E$. Since E is path connected, there is a continuous map $p : [0, 1] \to E$ such that $p(0) = y^0$, $p(1) = y$. The set

$$\Delta := \{t \in [0, 1] \mid F(x, p(t)) = 0 \quad \text{for some} \quad x \in D\}$$

contains zero; hence $\tau := \sup \Delta$ is defined and there is a sequence of numbers $t_i \in \Delta$ with limit τ. By definition of Δ, there are $x_i \in D$ with $F(x_i, p(t_i)) = 0$. Since D is bounded, there is a convergent subsequence x_{i_k}, $k = 1, 2, \ldots$, of the sequence x_i. The limits $x^1 = \lim x_{i_k}$, $y^1 = \lim p(t_{i_k}) = p(\tau)$ satisfy $F(x^1, y^1) = 0$ and by (H2), $x^1 \in D$.

If $\tau < 1$ then by the local implicit function theorem (Proposition 1) the equation $F(x, p(t))$ has a solution x for $p(t)$ in a neighbourhood of $y^1 = p(\tau)$, hence (by continuity of p) for some $t \in (\tau, 1)$, contradicting the maximality of τ. Therefore $\tau = 1$, $y^1 = p(\tau) = y$, and $x = x^1$ is a solution of $F(x, y) = 0$.

3. The differentiability condition (H1) is only used to prove Proposition 1. Therefore, (H1) may be replaced by the statement of Proposition 1.

4. If Theorem 1 shall be applied numerically, two problems arise. Namely, due to finite precision arithmetic, it is often the case that in place of an initial pair (x^0, y^0) with $f(x_0, y_0) = 0$ only a pair with $F(x^0, y^0)$ "small" is available. Moreover, a practical way is needed to verify the boundary condition (H2). In the remainder of this section we show how to handle both problems.

Theorem 2 (Semilocal implicit function theorem — extended form): *Let D be an open and bounded subset of $\mathbb{R}^n$, and let E be a simply connected Hausdorff space. Let $F : \overline{D} \times E \to \mathbb{R}^n$ be continuous, and suppose that for some point $(x^0, y^0) \in D \times E$ and some continuous map $\Delta : E \to \mathbb{R}^n$ with $\Delta y^0 = F(x^0, y^0)$, the following two conditions hold:*

(H1) *$\partial_1 F(x, y)$ exists for $x \in D$, $y \in E$, is continuous and nonsingular.*

(H2*) *$F(x, y) \neq t \, \Delta y$ for $x \in \partial D$, $y \in E$, $t \in [0, 1]$.*

Then there is a continuous map $H : E \to D$ with $F(Hy, y) = 0$ for all $y \in E$.

Proof: Define $E^* := E \times [0, 1]$. Then E^* is again simply connected, the map $F^* : \overline{D} \times E^* \to \mathbb{R}^n$ defined by $F^*(x, y, t) := F(x, y) - t \, \Delta y$ is continuous, and $\partial_1 F^*(x, y, t) = \partial_1 F(x, y)$ for $(x, y, t) \in D \times E^*$. Moreover $F^*(x^0, y^0, 1) = 0$, and $F^*(x, y, t) \neq 0$ for $x \in \partial D$, $(y, t) \in E^*$ by (H2*). Hence by Theorem 1, there is a continuous map $H^* : E^* \to D$ such that $H^*(y^0, 1) = x^0$ and $F^*(H^*(y, t), y, t) = 0$ for all $(y, t) \in E^*$. Therefore the map $H : E \to D$ defined by $Hy := H^*(y, 0)$ is continuous and satisfies $F(Hy, y) = F^*(Hy, y, 0) = F^*(H^*(y, 0), y, 0) = 0$ for all $y \in E$. □

Note that the conclusion has no reference to (x^0, y^0) whence uniqueness cannot be guaranteed.

The following computable form of the boundary condition (H2*) is similar in spirit to an existence condition for zeros of equations based on Brouwer's fixpoint theorem given in NEUMAIER [5]. In that paper it is also discussed how the required constants can be computed efficiently. The norm in the following may be any monotone vector norm; absolute value and inequalities are meant componentwise.

Proposition 5: *Let D be an open bounded subset of $\mathbb{R}^n$, let E be a connected Hausdorff space, and let $F: \bar{D} \times E \to \mathbb{R}^n$ and $H_0: E \to D$ be continuous functions. For some nonsingular $n \times n$-matrix A, define*

$$\delta_y := A^{-1}(\Delta y) \quad \text{for} \quad y \in E \tag{9}$$

and suppose that there is a number $\varkappa > 1$ such that for all $y \in E$,

$$D_y := \{x \in \mathbb{R}^n \mid \|x - H_0 y\| \leqq \varkappa \|\delta_y\|\} \subseteq D .$$

Let c be a nonnegative vector such that

$$|F(x, y) - \Delta y - A(x - H_0 y)| \leqq c \quad \text{for} \quad (x, y) \in \bar{D} \times E . \tag{10}$$

If the vector

$$b := |A^{-1}| c$$

satisfies the inequality

$$\|b\| < (\varkappa - 1) \|\delta_y\| \quad \text{for all} \quad y \in E$$

then the boundary condition (H2*) *is satisfied.*

Proof: Suppose that $F(x, y) = t\,\Delta y$ for some $x \in \bar{D}$, $y \in E$, $t \in [0, 1]$. With $x^0 := H_0 y$ we then have by (9) and (10),

$$|x - x^0| \leqq |x - x^0 - (t - 1)\,\delta_y| + (1 - t)\,|\delta_y| \leqq |A^{-1}|\,|(t - 1)\,A\delta_y - A(x - x^0)| + |\delta_y|$$
$$\leqq |A^{-1}|\,|F(x, y) - \Delta y - A(x - x^0)| + |\delta_y| \leqq |A^{-1}|\,c + |\delta_y| = b + |\delta_y|$$

whence $\|x - x^0\| \leqq \|b\| + \|\delta_y\| \leqq \varkappa \|\delta_y\|$. Therefore $x \in D_y \subseteq D$, and the boundary condition (H2*) follows. □

Remarks:

1. In order to make the vector c small we may proceed as follows: For A we take an approximation of $\partial_1 F(x^0, y^0)$. For H_0 we take an approximation of the expected solution map; if none is known, a suitable substitute may be

$$H_0 y := x^0 = \text{const}, \quad \text{or} \quad H_0 y := x^0 - A^{-1} F(x^0, y) .$$

Then Δ may be chosen as

$$\Delta y := F(H_0 y, y), \quad \text{or} \quad \Delta y := F(x^0, y^0) = \text{const}.$$

2. The proof of Proposition 5 (case $t = 0$) shows that the solution map H satisfies

$$Hy \in D_y \quad \text{for all} \quad y \in E .$$

Therefore, Theorem 2 combined with Proposition 5 and the preceding remarks amounts to an error bound for approximate solution maps of $F(x, y) = 0$, $y \in E$. The discussion in Neumaier [5] shows that for $\varkappa = 1.5$ and $\Delta y = F(H_0 y, y)$, the bounds obtained will overestimate the true errors by a factor of at most three, and hence be quite realistic.

2. Special cases and a weaker hypothesis

We begin this section with two corollaries of Theorem 1, namely special cases of the fixpoint theorems of Brouwer and Leray-Schauder.

Corollary 1: *Let D be a nonempty, convex, open, and bounded subset of $\mathbb{R}^n$, and let $\Phi: \bar{D} \to D$ be a continuously differentiable map. If for all $x \in D$ none of the real eigenvalues of $\Phi'(x)$ exceeds 1 then Φ has a fixpoint in $\bar{D}$.*

Proof: Put $E := [0, 1)$, $y^0 := 0$, and, for a fixed $x^0 \in D$, define

$$F(x, y) := y(\Phi x - x^0) + x^0 - x, \qquad x \in \bar{D}, \qquad y \in [0, 1) .$$

Then $\partial_1 F(x, y) = y\Phi'(x) - I$ is continuous and nonsingular and $F(x^0, y^0) = 0$. Moreover, $F(x, y) = 0$ implies that $x = y(\Phi x) + (1 - y)\,x^0$ is a convex combination of $x^0 \in D$ and $\Phi x \in \bar{D}$, and since $y \neq 1$ we have $x \in D$. Therefore (H2) is satisfied. By Theorem 1 there is now a continuous map $H: [0, 1) \to D$ such that $F(Hy, y) = 0$ for all $y \in [0, 1)$. Clearly, for $y \to 1$, Hy has at least one accumulation point $x^* \in \bar{D}$, and we have $F(x^*, 1) = 0$, i.e. $\Phi x^* = x^*$. □

Corollary 2: *Let D be an open and bounded subset of $\mathbb{R}^n$, and let $\Phi: D \to \mathbb{R}^n$ be a continuously differentiable map.* If $0 \in D$,

$$\Phi x \neq \lambda x \quad \text{for all} \quad x \in \partial D, \lambda > 1 ,$$

and if for all $x \in D$ none of the real eigenvalues of $\Phi'(x)$ exceeds 1 then Φ has a fixpoint in $\bar{D}$.

Proof: Put $E := [0, 1)$, $y^0 := 0$, and define

$$F(x, y) := y(\Phi x) - x, \qquad x \in \bar{D}, \qquad y \in [0, 1) .$$

Then $\partial_1 F(x, y) = y\Phi'(x) - I$ is continuous and nonsingular, and $F(x^0, y^0) = 0$. Moreover, $F(x, y) = 0$ implies $x = 0$ (if $y = 0$) and $\Phi(x) = \lambda x$, $\lambda = y^{-1} > 1$ (if $y > 0$). Hence by assumption, $x \notin \partial D$. Therefore, (H2) holds, and as before, a fixpoint of Φ exists. □

In fact, both the eigenvalue assumption and the differentiability assumption can be dropped; both results hold for any continuous Φ. This is the contents of the famous fixpoint theorems of Brouwer (cf. OR6. 3.2) and Leray-Schauder (cf. OR6.3.3). This suggests that by weakening the differentiability assumptions of Theorem 1 a stronger existence theorem can be proved which contains the Brouwer and Leray-Schauder fixpoint theorems as special cases. Indeed, this is possible, but the deeper methods of degree theory are required.

Theorem 3: *Let D be an open bounded subset of $\mathbb{R}^n$, and let E be a connected Hausdorff space. Let $F: \bar{D} \times E \to \mathbb{R}^n$ be continuous, and suppose that for some pair $(x^0, y^0) \in D \times E$ with $F(x^0, y^0) = 0$, the following two conditions hold.*

(H1*) $F_0(x) := F(x, y^0)$ *has in D a continuous, nonsingular derivative .*

(H2) $F(x, y) \neq 0$ *for* $x \in \partial D$, $y \in E$.

Then the equation $F(x, y) = 0$ has for all $y \in E$ at least one solution $x \in D$.

Proof: We shall use degree theory as discussed in Chapter 6 of ORTEGA and RHEINBOLDT [7]. The map $F_0: x \to F(x, y^0)$ satisfies the assumption of OR6.1.2 (with obvious change of notation); hence by OR6.1.5,

$$\deg(F_0, D, 0) = \sum \operatorname{sgn} \det F_0'(x) ,$$

where the sum extends over all zeros $x \in D$ of F_0. Since $F_0(x^0) = F(x^0, y^0) = 0$, the sum is not empty, and since $\det F_0'(x)$ is continuous and nonzero in D it has constant sign. Therefore $\deg(F_0, D, 0) \neq 0$. Now for arbitrary $y \in E$, there is a path $q: [0, 1] \to E$ such that $q(0) = y^0$, $q(1) = y$, and the map $H: \bar{D} \times [0, 1] \to \mathbb{R}^n$ defined by $H(x, t) := F(x, q(t))$ satisfies the assumption of OR6.2.2. Hence

$$\deg(F(\cdot, y), D, 0) = \deg(H(\cdot, 1), D, 0) = \deg(H(\cdot, 0), D, 0) = \deg(F_0, D, 0) \neq 0 ,$$

and by OR6.3.1, the equation $F(x, y) = 0$ has a solution $x \in D$. □

Remark: If we now repeat the arguments leading to Corollary 1 and 2 but use Theorem 3 instead of Theorem 1 we just obtain the fixpoint theorems of Brouwer and Leray-Schauder; the differentiability assumption in (H1*) becomes trivial since $F_0(x)$ has now the form $x^0 - x$ and $-x$, respectively.

To free ourselves from the apriori knowledge of an initial pair (x^0, y^0) with $F(x^0, y^0) = 0$ we may proceed as in Theorem 2 and obtain (the details are left to the reader):

Theorem 4: *Let D be an open and bounded subset of $\mathbb{R}^n$, and let E be a connected Hausdorff space. Let $F: \bar{D} \times E \to \mathbb{R}^n$ be continuous, and suppose that for some pair $(x^0, y^0) \in D \times E$ and some continuous map $\Delta: E \to \mathbb{R}^n$ with $\Delta y^0 = F(x^0, y^0)$, the following two conditions hold:*

(H1*) *$F_0(x) := F(x, y^0)$ has in D a continuous, nonsingular derivative,*

(H2*) *$F(x, y) \neq t\,\Delta y$ for $x \in \partial D$, $y \in E$, $t \in [0, 1]$.*

Then the equation $F(x, y) = 0$ has for all $y \in E$ at least one solution $x \in D$. □

Remarks:

1. Since Proposition 5 has no differentiability assumption, it can be used even in this more general situation to verify the boundary condition (H2*).

2. Under the assumptions of Theorem 3 or Theorem 4, the set-valued map $H: E \to \mathfrak{P}(D)$ with

$$\{Hy := x \in D \mid F(x, y) = 0\}$$

is upper semicontinuous; i.e. if $(x^i, y^i) \in D \times E$, $i = 1, 2, \ldots$, is a convergent sequence with limit $(x, y) \in D \times E$ then

$$x^i \in H(y^i) , \quad i = 1, 2, \ldots \Rightarrow x \in H(y) .$$

3. Inverse functions

In this section we apply the preceding results to the problem of inverting a map $F: \bar{D} \subseteq \mathbb{R}^n \to \mathbb{R}^n$. Moreover, we point our several existence tests for zeros of such a map some of which can be easily verified computationally.

Theorem 5 (semilocal inverse function theorem): *Let D be an open and bounded subset of $\mathbb{R}^n$, and let $F: \bar{D} \to \mathbb{R}^n$ be a continuous function which has in D a nonsingular, continuous derivative. Let E be a simply connected subset of $\mathbb{R}^n$ such that*

$$F(x) \notin E \quad \text{for} \quad x \in \partial D .$$

Then for given $x^0 \in D$ with $F(x^0) \in E$ there is a unique continuous map $F^{-1}: E \to D$ such that

$$F^{-1}(F(x^0)) = x^0 , \qquad F(F^{-1}(y)) = y \quad \text{for} \quad y \in E .$$

Moreover, F^{-1} is continuously differentiable in E, and for $y \in E$ we have

$$x = F^{-1}(y) \Rightarrow (F^{-1})'(y) = F'(x)^{-1} .$$

Proof: Existence and uniqueness follows from Theorem 1, applied with $y^0 := F(x^0)$ to $F(x) - y$ in place of $F(x, y)$. Differentiability and the formula for the derivative follows from OR5.2.1. □

Remark: For a similar result, in which the assumption "E simply connected" is replaced by the assumption "$E = F(D)$", see POURCIAU [9].

In general, the uniqueness of F^{-1} with $F^{-1}(F(x^0)) = x^0$ need not imply that $F(x) = y \in E$ has a unique solution $x \in D$; cf. OR, E5.3.1. But since we know already the existence of some continuous inverse function, two uniqueness criteria are available; both use the same hypothesis

(I) *Let D^* be a connected subset of $\mathbb{R}^n$, and let $F: D^* \to \mathbb{R}^n$ be a continuous function which has in D^* a nonsingular, continuous derivative. Suppose that on a subset E^* of $\mathbb{R}^n$, F has a continuous inverse $F^{-1}: E^* \to D^*$.*

Proposition 6: *Under the hypothesis* (I), *suppose that*

$$F(x) \in E^* \quad \text{for all} \quad x \in D^* .$$

Then for all $x \in D^*$ we have

$$F(x) = y \Leftrightarrow x = F^{-1}(y) .$$

Proof: Let $x^1 \in D^*$ and put $y^1 := F(x^1)$. Since D^* is connected there is a path $x: [0, 1] \to D^*$ such that $x(0) = x^0$, $x(1) = x^1$. Put $\bar{x}(t) := F^{-1}(F(x(t)))$, and define $\Delta := \{t \in [0, 1] \mid x(t) = \bar{x}(t)\}$. Then Δ is closed and $0 \in \Delta$; in particular $\bar{t} := \sup \Delta \in \Delta$. By Proposition 1 (applied with $F(x) - y$ in place of $F(x, y)$) there are neighbourhoods $D_0 \subseteq D^*$ of $x(\bar{t})$ and $E_0 \subseteq E^*$ of $F(x, \bar{t})$ such that

for $y \in E_0$, the equation $F(x) = y$ has a unique solution $x \in D_0$. If $\bar{t} < 1$ then by continuity of x, $\bar{x}$, and F, there is a $t' \in (\bar{t}, 1)$ such that $x(t')$, $\bar{x}(t') \in D_0$ and $F(x(t')) = F(\bar{x}(t')) \in E_0$; hence $x(t') = \bar{x}(t')$, $t' \in \Delta$, $t' = \sup \Delta \geqq t'$, contradiction. Therefore $\bar{t} = 1$, whence $x^1 = x(1) = \bar{x}(1) = F^{-1}(F(x^1)) = F^{-1}(y^1)$. □

Proposition 7: *Under the hypothesis* (1), *suppose that* D^* *is convex, and suppose that* $\mathfrak{M}$ *is a convex set of nonsingular* $n \times n$*-matrices such that*

$$F'(x) \in \mathfrak{M} \quad \textit{for all} \quad x \in D^* .$$

Then

$$x, \bar{x} \in D^* , \qquad F(x) = F(\bar{x}) \Rightarrow x = \bar{x} .$$

Proof: Suppose that $x, \bar{x} \in D^*$ and $F(x) = F(\bar{x})$. Define $y(t) := F(x + t(\bar{x} - x))$. Then $y(0) = y(1)$ and $y'(t) = F'(x + t(\bar{x} - x))(\bar{x} - x)$. Therefore,

$$0 = y(1) - y(0) = \int_0^1 y'(t)\, dt = \int_0^1 F'(x + t(\bar{x} - x))(\bar{x} - x)\, dt = A(\bar{x} - x) ,$$

where

$$A = \int_0^1 F'(x + t(\bar{x} - x))\, dt \in \mathfrak{M}$$

since D^* and $\mathfrak{M}$ are convex. Hence A is nonsingular, and from $A(\bar{x} - x) = 0$ we find $x = \bar{x}$. □

Remark: For practical applications, suitable convex sets of non-singular matrices are the sets

$$\mathfrak{M}_1 := \{A \in \mathbb{R}^{n \times n} \mid ||I - CA|| \leqq \beta\} , \tag{2}$$

where $C \in \mathbb{R}^{n \times n}$ is a fixed nonsingular matrix and $\beta \in (0, 1)$, and

$$\mathfrak{M}_2 := \{A \in \mathbb{R}^{n \times n} \mid \langle A \rangle \geqq A_0\} , \tag{3}$$

where A_0 is an M-matrix and $\langle A \rangle$ is Ostrowski's comparison matrix of A, defined by (Ostrowski [8])

$$\langle A_{ii} \rangle := |A_{ii}| , \qquad \langle A \rangle_{ik} := -|A_{ik}| \quad \text{for } i \neq k .$$

Indeed, if $A \in \mathfrak{M}_1$ then by standard perturbation theory (OR2.3.2), CA and hence A is nonsingular, and

$$||(CA)^{-1}|| \leqq (1 - \beta)^{-1} , \tag{4}$$

whereas if $A \in \mathfrak{M}_2$ then by Ostrowski [8], Satz III(8), A is nonsingular, and

$$|A^{-1}| \leqq A_0^{-1} , \tag{5}$$

where the absolute value is taken component-wise.

As a simple consequence of the semilocal inverse function theorem we obtain the well-known norm coerciveness theorem (cf. OR5.3.8).

Corollary 3: *Let* D *be an open, connected subset of* $\mathbb{R}^n$, *and let* $F: D \to \mathbb{R}^n$ *be a continuous function which has a nonsingular continuous derivative* $F'(x)$. *Suppose that* F *is norm coercive on* D, *i.e. for each* $\gamma > 0$ *there is a closed and bounded set* $D_\gamma \subseteq D$ *such that* $||F(x)|| > \gamma$ *for all* $x \in D \setminus D_\gamma$. *Then* F *is a homeomorphism, i.e. for all* $y \in \mathbb{R}^n$ *there is a unique solution* $x = F^{-1}(y) \in D$ *of* $F(x) = y$ *depending continuously on* y.

Proof: Let E be a compact subset of $\mathbb{R}^n$. Since F is continuous, the number $\gamma := \sup \{||F(x)|| \mid x \in E\}$ is finite. With D_γ as in the hypothesis, there is an open and bounded set D_E with $D_\gamma \subseteq D_E \subseteq D$. By construction, $F(x) \notin E$ for $x \in \partial D_E$ so that Theorem 5 applies with D_E in place of D. Hence on each compact subset E of $\mathbb{R}^n$ containing a fixed $x^0 \in \mathbb{R}^n$, the map F has a unique continuous inverse F_E^{-1} with $F_E^{-1}(F(x^0)) = x^0$. Hence F has a unique continuous inverse $F^{-1}: \mathbb{R}^n \to D$ with $F^{-1}(F(x^0)) = x^0$. Proposition 6 now guarantees uniqueness. □

We now consider applications of the semilocal inverse function theorem to the existence problem for zeros of functions.

Theorem 6: *Let* D *be an open and bounded subset of* $\mathbb{R}^n$, *and let* $F: \bar{D} \to \mathbb{R}^n$ *be a continuous function which has in* D *a nonsingular, continuous derivative. If* $x^0 \in D$ *and*

$$F(x) \neq sF(x^0) \quad \textit{for all} \quad x \in \partial D , \qquad s \in [0, 1] \tag{6}$$

then there is a vector $x^* \in D$ *with* $F(x^*) = 0$. *Moreover, if* D *is convex and* $\mathfrak{M}$ *is a convex set of nonsingular* $n \times n$*-matrices such that*

$$F'(x) \in \mathfrak{M} \quad \textit{for all} \quad x \in D ,$$

then x^* *is the only zero of* F *in* D.

Proof: Apply Theorem 5 with $E := \{sF(x^0) \mid s \in [0, 1]\}$, and put $x^* := F^{-1}(0)$. The uniqueness assertion follows from Proposition 7. □

Corollary 4 (cf. Bus [1], Theorem 2.8): *Let* D *be an open and bounded subset of* $\mathbb{R}^n$, *and let* $F: \bar{D} \to \mathbb{R}^n$ *be a continuous function which has in* D *a nonsingular, continuous derivative. If* $x^0 \in D$ *and*

$$||F(x)|| > ||F(x^0)|| \quad \textit{for all} \quad x \in \partial D \tag{7}$$

then there is a vector $x^* \in D$ *with* $F(x^*) = 0$.

Proof: (7) implies (6). □

In the remainder of this section we discuss some practical ways to verify condition (6).

Proposition 8: *Let* D *be a convex, open, and bounded subset of* $\mathbb{R}^n$, *and let* $F: \bar{D} \to \mathbb{R}^n$ *be continuously differentiable. Suppose that* $\mathfrak{M}$ *is a convex set of nonsingular* $n \times n$*-matrices such that*

$$F'(x) \in \mathfrak{M} \quad \textit{for all} \quad x \in \bar{D} .$$

If $x^0 \in D$, and if the set

$$N := \{x^0 - A^{-1}F(x^0) \mid A \in \mathfrak{M}\} \tag{8}$$

is contained in D then F has a unique zero $x^* \in D$; moreover,

$$x^* \in N\,. \tag{9}$$

Proof: Suppose that for some $x \in \bar{D}$ and some $s \in [0, 1]$, $F(x) = sF(x^0)$. Define $y(t) := F(x^0 + t(x - x^0))$. Then, as in the proof of Proposition 7,

$$(s - 1)\,F(x^0) = y(1) - y(0) = A(x - x^0)$$

for a suitable $A \in \mathfrak{M}$. Hence

$$x = sx^0 + (1 - s)\,(x^0 - A^{-1}F(x^0)) \tag{10}$$

is a convex combination of $x^0 \in D$ and $x^0 - A^{-1}F(x^0) \in N \subseteq D$; hence $x \in D$ since D is convex. In particular, (6) holds, and by Theorem 6, F has a unique zero $x^* \in D$. But for $x = x^*$ we have $F(x) = sF(x^0)$ with $s = 0$, whence $x = x^0 - A^{-1}F(x^0) \in N$ by (10) and (8). □

Remarks:

1. The proof also shows (take $s = 0$) that even in case that $N \not\subseteq D$, if F has a zero $x^* \in D$ then x^* is unique (by Proposition 7) and lies in N.

2. If $\mathfrak{M} = \mathfrak{M}_1$ (as in (2)) then, with

$$x^1 := x^0 - CF(x^0)\,,$$

$x \in N$ implies

$$x - x^1 = \left(x^0 - A^{-1}F(x^0)\right) - \left(x^0 - CF(x^0)\right) = (CA)^{-1}\,(I - CA)\,(-CF(x^0)) =$$
$$= (CA)^{-1}\,(I - CA) = (x^1 - x^0)$$

for some $A \in \mathfrak{M}$, whence

$$\|x - x^1\| \leqq \|(CA)^{-1}\|\,\|I - CA\|\,\|x^1 - x^0\| \leqq \frac{\beta}{1 - \beta}\,\|x^1 - x^0\|$$

by (4). Therefore, N is a subset of

$$N_1 := \left\{x \in \mathbb{R}^n \mid \|x - x^1\| \leqq \frac{\beta}{1 - \beta}\,\|x^1 - x^0\|\right\},$$

and $N_1 \subseteq D$ implies the existence of a zero $x^* \in N_1$, unique in D.

3. If $\mathfrak{M} = \mathfrak{M}_2$ (as in (3)) then $x \in N$ implies that for suitable $A \in \mathfrak{M}$ we have

$$|x - x^0| = |A^{-1}F(x^0)| \leqq |A^{-1}|\,|F(x^0)| \leqq A_0^{-1}|F(x^0)|$$

by (5), whence N is a subset of

$$N_2 := \{x \in \mathbb{R}^n \mid |x - x^0| \leqq A_0^{-1}|F(x^0)|\}\,.$$

Again, if $N_2 \subseteq D$ then F has a zero $x^* \in N_2$, unique in D. Instead of solving a linear system to find $A_0^{-1}|F(x^0)|$, a simple upper bound can be constructed if $u, v > 0$ are known such that $A_0 u \geqq v$. Then there is a constant $\alpha > 0$ such that $|F(x^0)| \leqq \alpha u$, and since $A_0^{-1} \geqq 0$, this implies that $A_0^{-1}|F(x^0)| \leqq \alpha A_0^{-1}u \leqq \alpha v$.

4. If $\mathfrak{M} = \{A \in \mathbb{R}^{n\times n} \mid \underline{A} \leqq A \leqq \bar{A}\}$ is a matrix interval then interval arithmetic (cf. MOORE [4]), can be used in various ways to obtain vector intervals containing N. The resulting bounding methods for zeros of equations are often called interval Newton methods. The first existence test for an interval Newton method, a special case of the above proposition, was proved by NICKEL [6] for the special case that the enclosure of N is computed as

$$N_3 := x^0 - \mathfrak{B}F(x^0)\,,$$

where $\mathfrak{B}$ is a matrix interval containing all A^{-1} with $A \in \mathfrak{M}$. NICKEL's proof, however, is less elementary in that it uses Brouwer's fixpoint theorem. A number of other interval Newton methods have been proposed in the last few years; see e.g. KRAWCZYK [3] and the references there.

References

1 BUS, J. C. P., Numerical Solution of Systems of Nonlinear Equations, Math. Centre Tracts 122, Math. Centrum, Amsterdam 1980.
2 DIEUDONNÉ, J., Foundations of Modern Analysis, Enlarged and corrected printing, Academic Press, New York-London 1969.
3 KRAWCZYK, R., Intervalliterationsverfahren, Bericht Math.-Stat. Sect. Forschungszentrum Graz, **186**, 1982.
4 MOORE, R. E., Methods and Applications of Interval Analysis, SIAM Publications, Philadelphia 1979.
5 NEUMAIER, A., Simple bounds for zeros of systems of equations, in: Iterative Solution of Nonlinear Systems of Equations, Lecture Notes in Math. 953, Springer-Verlag, Berlin-Heidelberg-New York 1982, pp. 88—105.
6 NICKEL, K., On the Newton method in Interval Analysis, MRC Tech. Sum. Rep. (Univ. of Wisconsin, Madison) 1136, 1971.
7 ORTEGA, J. M.; RHEINBOLDT, W. C., Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York-London 1970.
8 OSTROWSKI, A. M., Über die Determinanten mit überwiegender Hauptdiagonale, Comment. Math. Helvet. **10** (1937), 69—96.
9 POURCIAU, B. H., Homeomorphisms and generalized derivatives, J. Math. Anal. Appl. **93** (1983), 338—343.
10 SCHWETLICK, H., Numerische Lösung nichtlinearer Gleichungen, VEB Deutscher Verlag der Wiss., Berlin 1978.

Received July 7, 1983, revised version December 1, 1983

Address: Dr. ARNOLD NEUMAIER, Institut für Angewandte Mathematik der Albert-Ludwigs-Universität, Hermann-Herder-Straße 10, D-7800 Freiburg i. Br., BRD