SIAM J. MATH. ANAL.
Vol. 40, No. 1, pp. 413–441
© 2008 Society for Industrial and Applied Mathematics

NORMAL FORMS, QUASI-INVARIANT MANIFOLDS, AND BIFURCATIONS OF NONLINEAR DIFFERENCE-ALGEBRAIC EQUATIONS*

R. BEARDMORE† AND K. WEBSTER†

Abstract. We study the existence of quasi-invariant manifolds in a neighborhood of a fixed point of the *difference*-algebraic equation (ΔAE) $F(z_n, z_{n+1}) = 0$, where $F : \mathbb{R}^{2m} \to \mathbb{R}^m$ is a smooth map satisfying $F(0, 0) = 0$. We demonstrate the existence of quasi-invariant manifolds on which one can define forward and backward orbits of the ΔAE under mild assumptions on its linearization at the fixed point $z = 0$. Indeed, by assuming this linearization to be a regular matrix pencil, one obtains a functional equation satisfied by invariant manifolds which can be solved using an extension of the contraction mapping to spaces that satisfy an interpolation property. If the ΔAE under study is permitted to depend smoothly on a parameter, we then obtain a Neimark–Sacker bifurcation theorem as a corollary that can be deduced from the existence of a normal form for nonlinear ΔAEs.

Key words. invariant manifolds, bifurcations, difference-algebraic equations

AMS subject classifications. 39B72, 39A11, 37G99

DOI. 10.1137/050638618

1. Introduction.

The purpose of this paper is to provide an analysis of the invariant manifolds and bifurcations found in a class of *difference*-algebraic equations (ΔAEs) of the form

$$F(z_n, z_{n+1}) = 0; \tag{1.1}$$

the nomenclature and chosen acronym for (1.1) have been taken from [4, 22]. We assume that $F(= F(z, \bar{z})) : \mathbb{R}^{2m} \to \mathbb{R}^m$ is a smooth map satisfying $F(0, 0) = 0$ and say that (1.1) is *singular* because the partial derivative $d_{\bar{z}}F(0, 0)$ is not an isomorphism from $\mathbb{R}^m$ to itself. The purpose of the first part of this paper is to provide conditions under which (1.1) has suitably defined invariant manifolds that contain the fixed point $z = 0$, where the main difficulty to overcome in this analysis is the fact that forward orbits of (1.1) are not necessarily uniquely defined in a neighborhood of the fixed point.

The second part of the paper utilizes the existence of the aforementioned invariant manifolds to investigate the presence of bifurcations in ΔAEs in the sense that by extending F to be a C^k-mapping of the form $F : \mathbb{R}^{2m} \times \mathbb{R} \to \mathbb{R}^{2m}$, where $k \geq 5$, we examine the structure of invariant sets in the family of ΔAEs

$$F(z_n, z_{n+1}, \mu) = 0. \tag{1.2}$$

Our rationale is taken from bifurcation theory for maps which leads to the following question. If $F(0, 0, \mu) = 0$ for all μ in some interval and the one-parameter family of matrix pencils

$$\mathcal{P}(\mu) := (A(\mu), B(\mu)) := (d_{\bar{z}}F(0, 0, \mu), d_z F(0, 0, \mu)),$$

*Received by the editors August 22, 2005; accepted for publication (in revised form) August 28, 2007; published electronically May 7, 2008.
http://www.siam.org/journals/sima/40-1/63861.html

†Department of Mathematics, Imperial College, 180 Queen's Gate, London, SW7 2AZ, United Kingdom (r.beardmore@ic.ac.uk, k.webster@ic.ac.uk). The first author was supported by Nuffield Foundation grant NAL/00511/G. The second author was supported by EPSRC grant GR/S/17215/01.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

where F has $(z, \overline{z}, \mu)$ as its argument, is such that $\mathcal{P}(\mu_0)$ has a finite eigenvalue of unit modulus in the complex plane, does an invariant set of (1.2) bifurcate from the fixed point $z = 0$ at $\mu = \mu_0$?

Due to the lack of forward uniqueness we modify what we mean by the term *invariant*, which we do by using the prefixed *quasi-invariant*, and say that a set $\mathcal{Q} \subset F^{-1}\{0\}$ is quasi-invariant for (1.1) if, for $(z, \overline{z}) \in \mathcal{Q}$, there is a subsequent iterate $(\overline{z}, \overline{\overline{z}}) \in \mathcal{Q}$ for some $\overline{\overline{z}} \in \mathbb{R}^m$.

The paper is organized in the following way: The remainder of section 1 briefly covers the linear prerequisites for (1.1). Section 2 then provides some motivating applications. Section 3 presents the basic definitions of how (1.1) defines a local dynamical system and gives the invariant manifold equation of fixed points of (1.1). Section 4 provides a reformulation of the invariant manifold equation from section 3 as a nonlinear fixed-point problem in suitable Banach spaces which then is shown to have a solution in section 4.3. Section 5 gives a normal form for (1.1) with and without the presence of a bifurcation parameter. This section concludes with theorems that can be deduced using these normal forms, giving bifurcation results for (1.1) when that parameter is included. Finally, section 6 finishes the paper with a series of examples.

1.1. The linear case: Kronecker normal form. As a precursor to the analysis of the nonlinear problem (1.1), consider the linear case

$$Bz_n + Az_{n+1} = 0, \tag{1.3}$$

where $A, B : \mathbb{R}^m \to \mathbb{R}^m$ are linear maps and A is singular. In order to discuss the behavior of (1.3) we first introduce a normal form for matrix pencils.

When A is singular, the matrix pencil (A, B) is said to be *regular* if there is an $\omega \in \mathbb{C}$ such that $\det(\omega A + B) \neq 0$. The following result is well known for regular matrix pencils (see [7, 3]): There are complementary subspaces $K_1 \simeq \mathbb{R}^p, K_2 \simeq \mathbb{R}^q \subset \mathbb{R}^m$ such that $p + q = m$ and nonsingular linear mappings P, Q on $\mathbb{R}^m$, $L : K_1 \to K_1$, and $N : K_2 \to K_2$ such that

$$PAQ = \begin{pmatrix} I_p & 0 \\ 0 & N \end{pmatrix}, \quad PBQ = \begin{pmatrix} L & 0 \\ 0 & I_q \end{pmatrix}; \tag{1.4}$$

I_p and I_q are identities on K_1 and K_2, respectively. Moreover, there is a $\nu \geq 1$ such that $N^\nu = 0$, and ν is said to be the Kronecker index of (A, B).

The Kronecker normal form (KNF) in (1.4) can be used to rewrite (1.3) as a coupled system of difference equations

$$Lu_n + u_{n+1} = 0, \quad v_n + Nv_{n+1} = 0, \tag{1.5}$$

which has the solution $u_n = (-L)^n u_0$ and $v_n \equiv 0$ for all n, and thus (1.3) *has* a quasi-invariant subspace that arises from the quasi-invariant space $\{(u, v) : v = 0\}$ associated with (1.5). It is the presence of the former that we shall exploit in the remainder of the paper to study nonlinear perturbations of (1.5) that arise from a consideration of problems of the form (1.1).

1.2. Notation. If we define the spectrum of a matrix pencil to be

$$\sigma(A, B) = \{\lambda \in \mathbb{C} : \det(\lambda A + B) = 0\},$$

then $\sigma(A, B) = -\sigma(L)$ (*note the minus sign*), and p as defined within the KNF above coincides with the number of finite eigenvalues of (A, B), where eigenvalues are

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

counted according to their algebraic multiplicity. The matrix pencil (A, B) is said to be *hyperbolic* if $\sigma(A, B)$ is nonempty and contains no elements of unit modulus; otherwise, it is said to be *elliptic*. We shall also write $\rho(A, B) = \sup\{|\lambda| : \lambda \in \sigma(A, B)\}$ and denote the spectral radius of any linear mapping L by $\rho(L)$. Throughout we shall use # to denote the cardinality of a set of eigenvalues, counted according to algebraic multiplicity.

We shall use $BL(X, Y)$ to denote the space of continuous linear maps from one normed linear space X to another Y, even when X and Y are finite-dimensional. We shall use $B_\epsilon(x)$ for the open ball of radius ϵ about x, and $B_\epsilon(x; X)$ will specify that this ball is contained in the space X. If $L \in BL(X, Y)$, we shall denote the usual operator norm by $\|L\|_{BL(X,Y)}$, which is given by $\sup\{\|Lx\|_Y : x \in X, \|x\|_X = 1\}$. If the context is clear, we shall simply write $\|L\|$, and $BL(X)$ is also used for $BL(X, X)$. Throughout, if $F : X \to Y$ is a nonlinear mapping, then $dF(x) \in BL(X, Y)$ shall denote the Fréchet derivative, and when acting on $h \in X$ it will be written with square brackets, as in $dF(x)[h]$. Similarly, $d^2F(x)[h, k]$ denotes the second derivative, and this is bilinear in $[h, k]$.

If n is a positive integer, we shall use $\mathcal{O}_n(x)$ on occasion to denote any mapping, H, say, with the property that $\lim_{x\to 0} \|H(x)\|/\|x\|^n$ exists.

2. Motivation.

There are several problems from control theory and numerical analysis that lead to discrete systems where the relationship between the current and future states of a system are not explicit; see [12, 10, 6, 14] for examples.

2.1. Discretized differential-algebraic equations.

In [11] the authors apply a Runge–Kutta method to solve a differential-algebraic boundary-value problem arising from an optimal control problem, yielding a nonlinear *difference-algebraic equation* where the control plays the role of an implicit variable. For example, using a forward-Euler method to discretize the differential-algebraic equation (DAE)

$$\dot{x} = f(x, y), 0 = g(x, y) \qquad ((x(0), y(0)) \text{ given})$$

yields the ΔAE

$$x_{n+1} = x_n + hf(x_n, y_n), 0 = g(x_n, y_n) \qquad ((x_0, y_0) \text{ given}),$$

where h is a small parameter. A singularity in this context occurs when the partial derivative $d_y g(x, y)$ is singular on some subset of $g^{-1}\{0\}$.

Over the past decade a great deal of attention has been devoted to singular DAEs

$$F(z, \dot{z}) = 0, \tag{2.1}$$

where $d_{\dot{z}}F(z, \dot{z})$ changes rank on some set, and our study of (1.1) can be viewed as an extension of the work undertaken on (2.1) to the discrete-time case.

It is well known that (2.1) supports a range of singular and regular behavior, including *impasse points* and *pseudoequilibria* [16, 19, 17, 20]. However, not a great deal of the current literature is devoted to the study of *bifurcations* of DAEs nor to the unfolding of singularities in DAEs, such as the image and kernel singularities defined in [23]. One reason for this is the difficulty of proving a suitable center manifold theorem that can cope with the kind of singularities peculiar to DAE. We do note, however, that a Hopf bifurcation theorem is presented in [9] for systems of the form $\dot{x} = f(x, \dot{x}, \alpha)$, where α is a bifurcation parameter, $x \in \mathbb{R}^m$, and f is *nonexpansive* with respect to $\dot{x}$, a case that does include certain DAE singularities; a Hopf bifurcation

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

theorem for regular DAEs can be found in [15]. We also note that there are results in the DAE literature that yield the existence of an invariant manifold containing an equilibrium point; for regular DAEs see [18], and for singular DAEs see [24, 2].

2.2. Output-nulling control. Take a discrete dynamical system of the form

$$x_{n+1} = f(x_n, u_n), \; y_n = g(x_n, u_n), \tag{2.2}$$

where (x_n) is a sequence of states, (u_n) are controls, and (y_n) is a sequence of outputs. One may ask whether there is an admissible control that nullifies or fixes the outputs: Given (y_n), does there exist a sequence pair $((x_n), (u_n))$ satisfying (2.2)? The resulting equation is an infinite-dimensional system of equations that has a structure reminiscent of the semiexplicit, index-1 DAE (for terminology, see [3]).

2.3. Optimal control. The preprint [14] is relevant to the present work as it presents an invariant manifold result which leads to the existence of a control for the following variational problem with an infinite horizon:

$$\min_{(u_k)} \left\{ \sum_{k=0}^{\infty} \ell(x_k, u_k) : x_{k+1} = f(x_k, u_k), x_0 \in \mathbb{R}^m, u_k \in B_\epsilon(0; \mathbb{R}^p) \right\}$$

such that $f(0, 0) = 0$ and $\ell(0, 0) = 0$. An optimal orbit satisfies the first-order optimality conditions given by the quasi-linear, *implicit* difference equation

$$x_{k+1} = f(x_k, u_k), \tag{2.3}$$

$$\frac{\partial H}{\partial x}(x_k, u_k, \lambda_{k+1}) = \lambda_k, \tag{2.4}$$

$$0 = \frac{\partial H}{\partial u}(x_k, u_k, \lambda_{k+1}), \tag{2.5}$$

where H is the Hamiltonian $H(x, u, \lambda) = \lambda^T f(x, u) + \ell(x, u)$. In [14] the author demonstrates the existence of a *stable manifold* associated with (2.3)–(2.5) which has a dimension that coincides with the number of eigenvalues of the linearization about its fixed point, and the existence of this stable manifold then provides the necessary optimal control. The obstacle treated in [14] is the existence of a zero closed-loop eigenvalue which is analogous to the type of singularity treated in this paper. One can see the resemblance of (2.3)–(2.5) to (1.5) in that state variables (x_k) propagate forwards in time in (2.3)–(2.5), whereas adjoint variables (λ_k) propagate backwards, a property shared by (1.5).

3. A functional equation for quasi-invariant manifolds. Let us now define in what sense we expect (1.1) to induce a dynamical system. An element $z \in \mathbb{R}^m$ is said to be a *fixed point* of (1.1) if $F(z, z) = 0$. If z denotes the first argument of F and $\overline{z}$ the second, as in $F(z, \overline{z})$, then we define the following conditions:

(A1) $z = 0$ is a fixed point of (1.1): $F(0, 0) = 0$,

(A2) $\det(d_{\overline{z}} F(0, 0)) = 0$, and

(A3) there is a $\xi \in \mathbb{C}$ such that $\det(d_z F(0, 0) + \xi d_{\overline{z}} F(0, 0)) \neq 0$.

Throughout we make use of the matrix pencil (A, B), where

$$A := d_{\overline{z}} F(0, 0) \quad \text{and} \quad B := d_z F(0, 0), \tag{3.1}$$

and (A3) is the condition that (A, B) is regular. We shall assume that $F \in C^k(\mathbb{R}^{2m}, \mathbb{R}^m)$ for $k > 3$ and seek *local* and *global orbits* of the ΔAE (1.1) in the following sense.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

DEFINITION 1. *A sequence* $(z_n)_{j=0}^{J}$ *is said to be a* J-orbit *of* (1.1) *for* $J \in \mathbb{N}$ *if*

$$F(z_j, z_{j+1}) = 0 \quad \textit{for} \quad 0 \leq j \leq J-1 \quad \textit{and} \quad 2 \leq J < \infty,$$

and a local orbit *is a J-orbit for some $J \geq 2$. If there is a $z_2 \in \mathbb{R}^m$ such that $(z_0, z_1; z_2)$ is a local 2-orbit, then we say that the initial condition (z_0, z_1)* supports *this orbit. A sequence* $(z_n)_{j=0}^{\infty}$ *is said to be a* global orbit *of* (1.1) *as it is a J-orbit for each $J \geq 2$.*

Following the terminology used for DAEs, a pair (z_0, z_1) such that $F(z_0, z_1) = 0$ is said to be *consistent*, and if this pair supports some orbit, then it is said to be a consistent initial condition. We could also have analogously defined backward orbits for $J \leq -2$, but we omit this for brevity. Note that initial conditions lie in $\mathbb{R}^{2m}$ and not $\mathbb{R}^m$, a property that is analogous to DAEs whereby initial positions *and* certain initial derivatives must be provided in order to obtain the existence of solutions.

We now give the definition which stipulates how we expect (1.1) to induce a dynamical system.

DEFINITION 2. (1.1) *induces a local dynamical system on a manifold $\mathcal{M} \subset F^{-1}\{0\} \subset \mathbb{R}^{2m}$ which contains the origin of $\mathbb{R}^{2m}$ if there is a ball $\mathcal{M}_r := \mathcal{M} \cap B_r(0; \mathbb{R}^{2m})$ such that for each $(z, \overline{z}) \in \mathcal{M}_r$ there is a* unique *$(\overline{z}, \overline{\overline{z}}) \in \mathcal{M}$. If these conditions hold, $\mathcal{M}$ is said to be* a solution manifold *of* (1.1)*.*

This definition ensures that every point $(z, \overline{z}) \in \mathcal{M}_r$ supports the nontrivial 3-orbit $(z, \overline{z}; \overline{\overline{z}})$ and that the point $\overline{\overline{z}} \in \mathbb{R}^m$ is uniquely determined if we are to impose the requirement that $(\overline{z}, \overline{\overline{z}}) \in \mathcal{M}$.

DEFINITION 3. *A set $\mathcal{Q} \subset F^{-1}\{0\} \subset \mathbb{R}^{2m}$ is said to be* quasi-*invariant if, for each $(z, \overline{z}) \in \mathcal{Q}$,* there exists *a $\overline{\overline{z}} \in \mathbb{R}^m$ such that $(\overline{z}, \overline{\overline{z}}) \in \mathcal{Q}$.*

As an aside, note that (1.1) induces a trivial dynamical system on the quasi-invariant set $\{(0,0)\}$ by virtue of (A1), even if assumption (A3) fails. Note also that a solution manifold $\mathcal{M}$ is not necessarily unique; it is the local orbit within $\mathcal{M}$ that must be uniquely determined. Indeed, there may well be many possible choices for $\overline{\overline{z}}$ in order to keep the orbit on $F^{-1}\{0\}$, many of which may not be elements of $\mathcal{M}$.

3.1. The functional equation. Our strategy for locating quasi-invariant manifolds of (1.1) is to study a functional equation obtained in an analogous manner to the center-manifold equation from the theory of invariant manifolds for maps. The solution of this equation then provides the manifold $\mathcal{M}$ needed to form a local dynamical system for (1.1). We shall show in Theorem 1 that one can find a linear space $K_1 \subset \mathbb{R}^{2m}$ with an associated locally defined, differentiable map $\varphi : K_1 \to K_1$, a manifold $\mathcal{M} \subset F^{-1}\{0\}$, and a local diffeomorphism $\theta : K_1 \to \mathcal{M}$ such that

$$F(\theta(u)) = 0 \Longrightarrow F(\theta(\varphi(u))) = 0.$$

As a result, we will be able to ensure that (1.1) induces a local dynamical system on $\mathcal{M}$ essentially by iterating the map φ. This simply means that if $(z, \overline{z}) = \theta(u)$ is a consistent initial condition, then $(\overline{z}, \overline{\overline{z}}) = \theta(\varphi(u))$ and $(\overline{\overline{z}}, \overline{\overline{\overline{z}}}) = \theta(\varphi(\varphi(u)))$ provide subsequent iterates of (1.1).

Returning to (1.1), let us change the form of the problem by setting $w_n = z_{n+1}$, so that along an orbit of (1.1) we have

$$z_{n+1} = w_n, \tag{3.2}$$

$$0 = Bz_n + Aw_n + \Phi(w_n, z_n), \tag{3.3}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

where Φ is the C^k function which satisfies $\Phi(0,0) = 0, d\Phi(0,0) = 0$ and which is defined by

$$F(z,w) - Bz - Aw := \Phi(w,z).$$

The problem of finding an initial condition which is consistent, (z_0, w_0), say, is of an algebraic nature, whereas the problem of finding an orbit which is supported by this initial condition is a *dynamic* problem. This means that the problem of finding a manifold of orbits of a ΔAE will lead not to an algebraic equation that one could tackle using an elementary version of the implicit function theorem but instead to a functional equation.

Let us now obtain this functional equation. By applying the condition that (A, B) is a regular matrix pencil (condition (A3)), it follows that

$$(\mathcal{A}, \mathcal{B}) := \left(\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & I \\ B & A \end{pmatrix} \right) \tag{3.4}$$

is also a regular matrix pencil. If we define the vector $W_n = (z_n, w_n) \in \mathbb{R}^{2m}$, using (3.2)–(3.3), (1.1) can be written in the semilinear form

$$\mathcal{A}W_{n+1} = \mathcal{B}W_n + \Psi(W_n),$$

where Ψ is the C^k-mapping $\Psi(W) := (0, \Phi(W))$, so that $\Psi(0) = 0$ and $d\Psi(0) = 0$. There are mappings P and Q that put $(\mathcal{A}, \mathcal{B})$ in Kronecker normal form, and, by setting $W_n = QX_n$, we may write (3.2)–(3.3) in the form

$$[P\mathcal{A}Q]\, X_{n+1} = [P\mathcal{B}Q]\, X_n + P\Psi(QX_n), \tag{3.5}$$

where the terms in square brackets are in normal form:

$$\begin{bmatrix} I & 0 \\ 0 & N \end{bmatrix} X_{n+1} = \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix} X_n + P\Psi(QX_n).$$

Consequently, there are linear spaces $K_1 \simeq R^p$ and $K_2 \simeq \mathbb{R}^q$ such that $K_1 \oplus K_2 \simeq \mathbb{R}^{2m}$ and $X_n = (u_n, v_n) \in K_1 \oplus K_2$, where (u_n, v_n) satisfies the difference equation in normal form

$$\begin{cases} u_{n+1} &= Cu_n + f(u_n, v_n), \\ Nv_{n+1} &= \quad v_n - g(u_n, v_n). \end{cases} \tag{NF}$$

We now ask that there is a manifold given by the graph of some function h on which one can solve (NF) uniquely in a neighborhood of the fixed point $(u, v) = (0, 0)$ in the sense that $v_n = h(u_n)$ holds along orbits. This imposes the two conditions

$$u_{n+1} = Cu_n + f(u_n, h(u_n)) \quad \text{and} \quad Nh(u_{n+1}) = h(u_n) - g(u_n, h(u_n))$$

on h, and it follows that the local orbit (u_n, v_n) of (NF) can be found if h satisfies the functional equation

$$h(u) = Nh(Cu + f(u, h(u))) + g(u, h(u)), \qquad h(0) = 0,\ dh(0) = 0, \tag{3.6}$$

for all u in some neighborhood of the origin in $\mathbb{R}^p$. The boundary conditions in (3.6) ask first that the fixed point $(u, v) = (0, 0)$ of (NF) lies on the graph of h and then that this graph is tangent to the quasi-invariant subspace obtained on setting $f = 0$ and $g = 0$ in (NF).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

3.2. Further preliminaries.

3.2.1. Perturbation of eigenvalues. For completeness we have included the following two preliminary results regarding the spectra of one-parameter families of matrix pencils, which are mappings of the form

$$\mathcal{P} : (-1, 1) \to BL(\mathbb{R}^m) \times BL(\mathbb{R}^m); \mu \mapsto (A(\mu), B(\mu)).$$

If we define the family of analytic functions

$$f_\mu(\omega) = \det(\omega A(\mu) + B(\mu)),$$

the multiplicity of an eigenvalue of a matrix pencil is then the multiplicity of the corresponding zero of $f_\mu(\cdot)$, which is at most m. The identity

$$\frac{d}{d\omega} f_\mu(\omega) = f_\mu(\omega)\mathrm{tr}[(\omega A(\mu) + B(\mu))^{-1} A(\mu)], \tag{3.7}$$

whenever this inverse is defined, can be used to obtain the following two lemmas.

Lemma 1 (C^1-dependence of eigenvalues). *Suppose that $\mathcal{P}(\mu) := (A(\mu), B(\mu))$ is a C^1-parameterized family of real matrix pencils, with $\mu \in (-1, 1)$, such that $\mathcal{P}(0)$ is a regular matrix pencil. An element $\lambda_0 \in \sigma(\mathcal{P}(0))$ is said to be an* algebraically simple eigenvalue *of $\mathcal{P}(0)$ if*

$$\ker(\lambda_0 A(0) + B(0)) = \langle x_0 \rangle \quad \text{and} \quad x_0 \notin \mathrm{ran}(\lambda_0 A(0) + B(0)).$$

If λ_0 is an algebraically simple eigenvalue of $\mathcal{P}(0)$, then there is a C^1-parameter family of algebraically simple eigenvalues $\lambda(\mu) \in \mathbb{C}$ of $\mathcal{P}(\mu)$ such that $\lambda(0) = \lambda_0$, with a corresponding C^1 family of unit eigenvectors $x(\mu)$, with $x(0) = x_0$.

Proof. This follows from the implicit function theorem applied to the system $F(\lambda, x, \mu) = (0, 0)$, where $F(\lambda, x, \mu) := \left[(\lambda A(\mu) + B(\mu))x, \|x\|_2^2 - 1\right]$. ∎

Lemma 2 (C^0-dependence of eigenvalues). *Suppose that $\mathcal{P}(\mu) := (A(\mu), B(\mu))$ is a C^0-parameterized family of real matrix pencils, with $\mu \in (-1, 1)$, such that $\mathcal{P}(0)$ is regular. If λ_0 is an eigenvalue of $\mathcal{P}(0)$ of algebraic multiplicity l, then it is isolated in the complex plane, and for each $\epsilon > 0$ there is a $\delta > 0$ such that if $|\mu| < \delta$, then $\mathcal{P}(\mu)$ has l eigenvalues (counted according to algebraic multiplicity) in the disk $D(\lambda_0, \epsilon)$.*

Proof. As $f_0(\cdot)$ does not vanish identically because $\mathcal{P}(0)$ is regular by assumption, neither can $f_\mu(\cdot)$ for sufficiently small μ. The *isolatedness* of eigenvalues of $\mathcal{P}(\mu)$ is a consequence of the fact that analytic functions have isolated zeros. Now by using (3.7) we integrate around a closed circle in the complex plane with center $\omega = \lambda_0$ and radius ϵ, from where

$$\#\{\sigma(\mathcal{P}(\mu)) \cap D(\lambda_0, \epsilon)\} = \frac{1}{2\pi i} \oint_{\partial D(\lambda_0, \epsilon)} \mathrm{tr}[(\omega A(\mu) + B(\mu))^{-1} A(\mu)] d\omega,$$

where $D(\lambda_0, \epsilon)$ is an open disk of radius ϵ about λ_0 in the complex plane. This quantity is integer-valued and depends continuously on μ, and the result now follows. ∎

3.2.2. Notation. From this point we shall identify the linear space K_1 from the KNF with $\mathbb{R}^p$ and K_2 with $\mathbb{R}^q$; now let $|\cdot|_p$ and $|\cdot|_q$ denote norms on $\mathbb{R}^p$ and $\mathbb{R}^q$, and let $\Omega_\delta = \{u \in \mathbb{R}^p : |u|_p < \delta\}$. We also assume that the unit sphere $\partial\Omega_1$ is a C^∞ manifold.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Let $C^0(\overline{\Omega}_\delta, \mathbb{R}^q)$ be the Banach space of continuous maps on $\overline{\Omega}_\delta$ with norm $\|h\|_{C^0} = \sup_{u\in\overline{\Omega}_\delta} |h(u)|_q$. Similarly, let $C^j(\overline{\Omega}_\delta, \mathbb{R}^q)$ be the space of all j-times continuously differentiable functions on $\overline{\Omega}_\delta$ with norm $\|h\|_{C^j} = \max_{0\le i\le j} \sup_{u\in\overline{\Omega}_\delta} \|d^i h(u)\|_{C^0}$, where d^i denotes the ith (Fréchet) derivative, so that $d^j h(u)$ is a j-linear form which we denote $[k_1, \dots, k_j] \to d^j h(u)[k_1, \dots, k_j]$. Consequently, we have the norm of a higher derivative given by the formula

$$\|d^j h\|_{C^0} = \sup_{u\in\overline{\Omega}_\delta} \sup_{|k_i|_p\le 1} |d^j h(u)[k_1, \dots, k_j]|_q. \tag{3.8}$$

If M is any multilinear form on a linear space Z and $z \in Z$, then $M[z]^{(k)}$ is shorthand for $M[z, z, \dots, z]$.

From the smoothness of the unit sphere $\partial\Omega_1$, it follows that the embedding of $C^{j+1}(\overline{\Omega}_\delta)$ into $C^j(\overline{\Omega}_\delta)$ is compact, so that if $(h_n) \subset C^{j+1}(\overline{\Omega}_\delta)$ is bounded in the norm of the latter space, there is a subsequence (h_{n_k}) which converges in $C^j(\overline{\Omega}_\delta)$ to some element of $C^j(\overline{\Omega}_\delta)$. We shall also make limited use of the Hölder spaces, which we denote by $C^{j+\alpha}(\overline{\Omega}_\delta)$ whenever j is an integer and $0 < \alpha < 1$, recalling the compact embedding $C^{j+\alpha}(\overline{\Omega}_\delta) \subset C^{j+\beta}(\overline{\Omega}_\delta)$ if $\alpha > \beta$.

It can be somewhat notationally cumbersome to include all of the references to the underlying spaces in all of the norms that we use, so we shall limit their use and expect that the precise meaning can be taken from context.

4. Solving the fixed-point problem (3.6). It is not (NF) that we shall seek to solve directly, but we make the substitution

$$u = \epsilon\tilde{u}, v = \epsilon\tilde{v}$$

in (NF) to give (after removal of the tildes for clarity)

$$\left\{ \begin{aligned} u_{n+1} &= Cu_n + \epsilon^{-1} f(\epsilon u_n, \epsilon v_n), \\ N v_{n+1} &= \quad v_n - \epsilon^{-1} g(\epsilon u_n, \epsilon v_n). \end{aligned} \right. \tag{NF$_\epsilon$}$$

As the functions f and g are higher than linear order at the origin, $(\mathrm{NF})_\epsilon$ is in fact smooth with respect to variations in ϵ.

Let us define the one-parameter family of C^k functions f_ϵ and g_ϵ (with C^{k-1} dependence on ϵ) by

$$\mathsf{f}_\epsilon(u, v) = \epsilon^{-1} f(\epsilon u, \epsilon v) \quad \text{and} \quad \mathsf{g}_\epsilon(u, v) = \epsilon^{-1} g(\epsilon u, \epsilon v),$$

respectively. For $j \in \mathbb{N}$ we also have

$$d^j \mathsf{f}_\epsilon(u, v) = \epsilon^{j-1} d^j f(\epsilon u, \epsilon v) \quad \text{and} \quad d^j \mathsf{g}_\epsilon(u, v) = \epsilon^{j-1} d^j g(\epsilon u, \epsilon v), \tag{4.1}$$

whenever these derivatives are defined.

Now seek an invariant manifold $\mathcal{M}$ of $(\mathrm{NF})_\epsilon$ given by a graph on which

$$v_n = h(u_n),$$

and then $\mathcal{M}$ can be realized as such a graph if there is a solution of the nonlinear functional equation

$$\left\{ \begin{array}{l} h(u) = Nh(Cu + \mathsf{f}_\epsilon(u, h(u))) + \mathsf{g}_\epsilon(u, h(u)), \\ h(0) = 0, dh(0) = 0. \end{array} \right. \tag{4.2}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

4.1. Preliminary estimates. The following are simple but essential estimates on the derivatives of f and g. By the mean-value inequality and the fact that the mapping $(u,v) \mapsto (f(u,v), g(u,v))$ and its derivative vanish at $(u,v) = (0,0)$, there exists an ℓ $(= \ell(\delta, r)) > 0$ such that

$$|f(u,v)|_p \le \ell \|(u,v)\|^2, \quad |g(u,v)|_q \le \ell \|(u,v)\|^2, \tag{4.3}$$

and

$$\|df(u,v)\| \le \ell \|(u,v)\|, \quad \|dg(u,v)\| \le \ell \|(u,v)\|, \tag{4.4}$$

whenever $|u|_p \le \delta, |v|_q \le r$, where here and throughout we use the norm

$$\|(u,v)\| = \max(|u|_p, |v|_q) \qquad (\forall (u,v) \in \mathbb{R}^p \times \mathbb{R}^q).$$

Here $\|df(u,v)\|$ and $\|dg(u,v)\|$ both refer to induced operator norms, treating $df(u, v)$ and $dg(u,v)$ as linear mappings. By using the mean-value inequality we obtain

$$\|d^2 g(u,v) - d^2 g(0,0)\| \le \sup_{|u|_p \le \delta, |v|_q \le r} \|d^3 g(u,v)\| \|(u,v)\|,$$

and the triangle inequality gives

$$\|d^2 g(u,v)\| \le \|d^2 g(0,0)\| + \overline{\ell} \|(u,v)\| \qquad (|u|_p \le \delta, |v|_q \le r),$$

where $\overline{\ell}$ $(= \overline{\ell}(\delta, r)) = \sup_{|u|_p \le \delta, |v|_q \le r} \|d^3 g(u,v)\|$, whence

$$\|d^2 \mathsf{g}_\epsilon(u,v)\| \le \epsilon \left(\|d^2 g(0,0)\| + \epsilon \overline{\ell} \|(u,v)\| \right) \qquad (|u|_p \le \delta, |v|_q \le r). \tag{4.5}$$

An analogous inequality holds for f and f_ϵ:

$$\|d^2 \mathsf{f}_\epsilon(u,v)\| \le \epsilon \left(\|d^2 f(0,0)\| + \epsilon \overline{\ell} \|(u,v)\| \right) \qquad (|u|_p \le \delta, |v|_q \le r). \tag{4.6}$$

It is the $O(\epsilon)$ size of these quantities that will be important later.

4.2. Introducing a cutoff function. It is not (4.2) that we shall seek to solve directly, but we must employ a cutoff function to rewrite (4.2) in a fixed-point form that is amenable to a Picard iteration. This is not the case at present because if we were to define a nonlinear operator acting on h by the right-hand side of (4.2), there is no reason for it or its iterates to be well-defined on a suitable function space.

For any $\delta > 0$, there is a cutoff function $\psi \in C^\infty(\mathbb{R}^p)$ such that

$$\psi(u) = \begin{cases} u & \text{if} \quad |u|_p \le \delta/2, \\ 0 & \text{if} \quad |u|_p \ge 3\delta/2 \end{cases}$$

and such that $|\psi(u)|_p \le \delta$. By using this cutoff we define a Nemitskii operator π as follows:

$$\pi(h)(u) = \psi(Cu + \mathsf{f}_\epsilon(u, h(u))) \qquad (\forall u \in \overline{\Omega}_\delta, h : \overline{\Omega}_\delta \subset \mathbb{R}^p \to \mathbb{R}^q), \tag{4.7}$$

so that

$$\pi : C^0(\overline{\Omega}_\delta, \mathbb{R}^q) \to C^0(\overline{\Omega}_\delta, \mathbb{R}^p),$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

and the inequality $|\pi(h)(u)|_p \leq \delta$ holds pointwise. Moreover, by the C^k-regularity of f, it follows that π is itself a C^k-mapping for each $\epsilon > 0$ fixed, with Fréchet derivative

$$d\pi(h)[k](\cdot) = d\psi(C\cdot + \mathsf{f}_\epsilon(\cdot, h))[d_v\mathsf{f}_\epsilon(\cdot, h)[k]] \qquad (\forall h, k \in C^0(\overline{\Omega}_\delta, \mathbb{R}^q)). \tag{4.8}$$

Also note the following estimate, which will be important later.

LEMMA 3. *Suppose that* $h \in C^0(\overline{\Omega}_\delta, \mathbb{R}^q)$ *satisfies* $\|h\|_{C^0} \leq r$, *then*

$$\|d\pi(h)\|_{BL(C^0)} \leq \epsilon \ \cdot \|\psi\|_{C^1}\ell \max(\delta, r), \tag{4.9}$$

and there are constants $\kappa_1, \kappa_2 > 0$, *depending on* ℓ, δ, *and* r, *but not on* ϵ, *such that*

$$\|d^2\pi(h)\|_{BL(C^0)\times BL(C^0)} \leq \epsilon\|\psi\|_{C^2}\left(\kappa_1(\ell, \delta, r) + \epsilon\kappa_2(\ell, \delta, r)\right). \tag{4.10}$$

Proof. By using (4.8) we obtain

$$\begin{aligned}
\sup_{k\in C^0, \|k\|_{C^0}=1} \|d\pi(h)[k]\|_{C^0} &= \sup_{k\in C^0, \|k\|_{C^0}=1} \|d\psi(C\cdot + \mathsf{f}_\epsilon(\cdot, h))[d_v\mathsf{f}_\epsilon(\cdot, h)[k]]\|_{C^0} \\
&\leq \|\psi\|_{C^1} \cdot \sup_{k\in C^0, \|k\|_{C^0}=1} \|d_v\mathsf{f}_\epsilon(\cdot, h)[k]\|_{C^0} \\
&\leq \|\psi\|_{C^1} \cdot \sup_{|u|_p\leq\delta, |v|_q\leq r} \|d_v\mathsf{f}_\epsilon(u, v)\| \\
&\leq \|\psi\|_{C^1} \cdot \ell \cdot \sup_{|u|_p\leq\delta, |v|_q\leq r} \|(\epsilon u, \epsilon v)\| \text{ (by (4.1) and (4.3))},
\end{aligned}$$

and the first part follows. The second part follows from

$$d^2\pi(h)[k_1, k_2] = d^2\psi(C\cdot + \mathsf{f}_\epsilon(\cdot, h))[d_v\mathsf{f}_\epsilon[k_1], d_v\mathsf{f}_\epsilon[k_2]] + d\psi[d^2_{vv}\mathsf{f}_\epsilon(\cdot, h)[k_1, k_2]],$$

so that, for $i = 1, 2$,

$$\begin{aligned}
\sup_{k_i\in C^0, \|k_i\|_{C^0}=1} \|d^2\pi(h)[k_1, k_2]\|_{C^0} &\leq \sup_{u\in\mathbb{R}^p} \|d\psi(u)\| \cdot \|d^2_{vv}\mathsf{f}_\epsilon(\cdot, h)\| \\
&\qquad + \sup_{u\in\mathbb{R}^p} \|d^2\psi(u)\| \cdot \|d_v\mathsf{f}_\epsilon(\cdot, h)\|^2 \\
\leq \|\psi\|_{C^2}\left(\sup_{|u|_p\leq\delta, |v|_q\leq r} \|d^2_{vv}\mathsf{f}_\epsilon(u, v)\| + \sup_{|u|_p\leq\delta, |v|_q\leq r} \|d_v\mathsf{f}_\epsilon(u, v)\|^2\right) \\
\leq \|\psi\|_{C^2}\left(\epsilon\|d^2f(0,0)\| + \epsilon^2\overline{\ell}\max(\delta, r) + \sup_{|u|_p\leq\delta, |v|_q\leq r} \|d_vf(\epsilon u, \epsilon v)\|^2\right) \quad \text{(by (4.5))} \\
\leq \|\psi\|_{C^2}\left(\epsilon\|d^2f(0,0)\| + \epsilon^2\overline{\ell}\max(\delta, r) + \sup_{|u|_p\leq\delta, |v|_q\leq r} \|(\epsilon u, \epsilon v)\|^2\ell^2\right) \quad \text{(by (4.3))},
\end{aligned}$$

and the result follows directly from here. ∎

In order to solve (4.2), we now tackle the following nonlinear fixed-point problem:

$$h = Nh(\pi(h)) + \mathsf{G}_\epsilon(h), \tag{4.11}$$

where $\mathsf{G}_\epsilon : C^0(\overline{\Omega}_\delta, \mathbb{R}^q) \to C^0(\overline{\Omega}_\delta, \mathbb{R}^q)$ is the Nemitskii operator

$$\mathsf{G}_\epsilon(h)(u) := \mathsf{g}_\epsilon(u, h(u)).$$

Notice that the cutoff ψ has been used in (4.11), but, because ψ coincides with the identity on some balls around the origin, solutions of (4.11) will satisfy (4.2) on this ball.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

This construction ensures that $h(\pi(h)) \in C^0(\overline{\Omega}_\delta, \mathbb{R}^q)$ whenever $h \in C^0(\overline{\Omega}_\delta, \mathbb{R}^q)$, and we may, as a result, define the operator

$$T : C^0(\overline{\Omega}_\delta, \mathbb{R}^q) \to C^0(\overline{\Omega}_\delta, \mathbb{R}^q); \quad h \mapsto h(\pi(h)),$$

noting that the operators $T : C^1(\overline{\Omega}_\delta, \mathbb{R}^q) \to C^1(\overline{\Omega}_\delta, \mathbb{R}^q)$ and $T : C^2(\overline{\Omega}_\delta, \mathbb{R}^q) \to C^2(\overline{\Omega}_\delta, \mathbb{R}^q)$ are also well-defined as the restrictions of T to various subspaces of $C^0(\overline{\Omega}_\delta, \mathbb{R}^q)$ as f and g are C^3 functions.

However, it is not (4.11) that we shall solve, but we exploit the nilpotency of N to bring the *functional part* of (4.2) and (4.11), that is, $Nh(\pi(h))$, into a higher-order contribution to the problem. However, because $g(u, v)$ is a second-, or possibly higher-order function, the operator G_ϵ contains no linear terms, and this will help us to obtain a contractive sequence by iterating T.

By way of example, let us suppose that $N \neq 0$ but $N^2 = 0$. If there exists a solution of $h = Nh(\pi(h)) + \mathsf{G}_\epsilon(h)$, there results $Nh = N\mathsf{G}_\epsilon(h)$, and therefore $Nh(\pi(h)) = N\mathsf{G}_\epsilon(h(\pi(h)))$. In this case we find that h must satisfy

$$h = N\mathsf{G}_\epsilon(h(\pi(h))) + \mathsf{G}_\epsilon(h), \tag{4.12}$$

and one observes that the *functional part* of the equation (that is, $h(\pi(h))$) now sits inside a higher-order term (and not a linear one as in (4.11)). Conversely, if h satisfies (4.12), then $Nh = N\mathsf{G}_\epsilon(h)$, so that $N\mathsf{G}_\epsilon(h(\pi(h))) = Nh(\pi(h))$ because $N^2 = 0$, and $h = Nh(\pi(h)) + \mathsf{G}_\epsilon(h)$ follows.

When the nilpotency index of the map N is arbitrary we extend this idea in the following lemma, where here and in the remainder we shall write

$$\mathcal{G}_\epsilon(h) = \sum_{j=0}^{\nu-1} N^j \mathsf{G}_\epsilon(T^j(h)),$$

and

$$T^{j+1} = T(T^j), \qquad \text{where} \qquad T^0 = I,$$

and the latter denotes the identity on $C^0(\overline{\Omega}_\delta, \mathbb{R}^q)$.

LEMMA 4. *Suppose that $N^\nu = 0$ but $N^{\nu-1} \neq 0$; then h is a solution of* (4.11) *if it is a solution of the fixed-point problem*

$$h = \mathcal{G}_\epsilon(h). \tag{4.13}$$

Proof. Let us suppose that h is a solution of (4.13); then

$$Nh = N\mathcal{G}_\epsilon(h) = N\left[\sum_{j=0}^{\nu-1} N^j \mathsf{G}_\epsilon(T^j(h))\right] = \sum_{j=0}^{\nu-1} N^{j+1}\mathsf{G}_\epsilon(T^j(h)),$$

and so, because $N^\nu = 0$, we obtain

$$\begin{aligned} Nh(\pi(h)) &= \sum_{j=0}^{\nu-1} N^{j+1}\mathsf{G}_\epsilon(T^j(h(\pi(h)))) \\ &= \sum_{j=0}^{\nu-1} N^{j+1}\mathsf{G}_\epsilon(T^{j+1}(h)) = \sum_{j=1}^{\nu-1} N^j \mathsf{G}_\epsilon(T^j(h)) = \mathcal{G}_\epsilon(h) - \mathsf{G}_\epsilon(h) = h - \mathsf{G}_\epsilon(h), \end{aligned}$$

which therefore provides a solution of (4.11) as required. ∎

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

4.3. The main result. Our strategy for solving (4.13), and hence (4.2), is to show that $\mathcal{G}_\epsilon$ satisfies a *refined* Banach contraction theorem of the type given in [26, p. 286]. The idea that we employ several times is encapsulated in the following idea. Consider Banach spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ such that $x_0 \in Y$ and $Y \subset X$, and moreover $\mathcal{T}$ is a mapping satisfying $\mathcal{T} : Y \to Y$ and $\mathcal{T} : X \to X$. Now, if there is an $r > 0$ and a $\kappa \in [0, 1)$ such that

1. $\mathcal{T} : \overline{B}_r(x_0; X) \to \overline{B}_r(x_0; X)$,
2. $\|\mathcal{T}(y) - \mathcal{T}(y')\|_X \leq \kappa\|y - y'\|_X$ for all $y, y' \in Y \cap \overline{B}_r(x_0; X)$,

then $\mathcal{T}$ has a fixed point $y^* \in \overline{B}_r(x_0; X)$. In addition, if

3. $\mathcal{T} : \overline{B}_\rho(x_0; Y) \to \overline{B}_\rho(x_0; Y)$ and
4. there is an interpolating Banach space Z such that $Y \subset Z$ with compact embedding and $Z \subset X$ with continuous embedding,

then $y^* \in Z$.

The point here is that we cannot ensure that $y^* \in Y$, although one still obtains a fixed point in some space from the standard iteration scheme. By using this idea one can prove an existence and regularity result for (4.2), where we have in mind $x_0 = 0, \mathcal{T} = \mathcal{G}_\epsilon, Y = C^{k+1}, Z = C^{k+\alpha}$, and $X = C^k$, where $\alpha \in (0, 1)$. We begin by providing the details to cover the cases $k = 0$ and $k = 1$.

The following theorem is the main result of this paper from which the invariant manifold and bifurcation theorems are deduced.

THEOREM 1. *Let $\alpha \in [0, 1)$. There exists an $\epsilon_0 > 0$ such that, for each $\epsilon \in (0, \epsilon_0)$,* (4.13) *has a solution $h \in C^{1+\alpha}(\overline{\Omega}_\delta, \mathbb{R}^q)$; moreover $h(0) = 0$ and $dh(0) = 0$.*

Proof. Let X_r be the C^0-closed ball of radius r about zero in $C^0(\overline{\Omega}_\delta, \mathbb{R}^q)$ and Y_r the C^1-closed ball of radius r about zero in $C^1(\overline{\Omega}_\delta, \mathbb{R}^q)$. Let $h_0 \in Y_r$, and define a sequence

$$h_{n+1} := \mathcal{G}_\epsilon(h_n).$$

We shall show that we can choose ϵ such that (h_n) is well-defined, contractive, and hence Cauchy in X_r, and it therefore converges. Throughout the remainder of the proof we shall use the positive constant

$$n_* := \sum_{j=0}^{\nu-1} \|N\|^j_{BL(\mathbb{R}^q)}.$$

We now give a proof of Theorem 1 in four short steps, each placing a stronger restriction on ϵ relative to the fixed choice of δ and r to ensure that $\mathcal{G}_\epsilon$ contracts when acting on C^1 functions, measured in the C^0 norm.

Step 1. If $\epsilon < r(\ell n_* \max(\delta, r)^2)^{-1} =: \epsilon_1$, then $\mathcal{G}_\epsilon : X_r \to X_r$.

Proof. Suppose that $h \in X_r$, and then $T(h) = h(\pi(h))$ is continuous as h is. Moreover if $\|h\|_{C^0} \leq r$, then $\|T(h)\|_{C^0} = \|h(\pi(h))\|_{C^0} \leq r$; similarly $\|T^j(h)\|_{C^0} \leq r$ for all $0 \leq j \leq \nu - 1$. By definition,

$$\begin{aligned}\|\mathcal{G}_\epsilon(h)\|_{C^0} &\leq \sum_{j=0}^{\nu-1} \|N\|^j \|\mathsf{G}_\epsilon(T^j(h))\|_{C^0} \leq \sum_{j=0}^{\nu-1} \|N\|^j \sup_{\|\overline{h}\|_{C^0} \leq r} \|\mathsf{G}_\epsilon(\overline{h})\|_{C^0} \\ &\leq n_* \sup_{|u|_p \leq \delta, |v|_q \leq r} |\mathsf{g}_\epsilon(u, v)|_q = n_* \sup_{|u|_p \leq \delta, |v|_q \leq r} \epsilon^{-1}|g(\epsilon u, \epsilon v)|_q.\end{aligned}$$

From (4.3) it follows that

$$\|\mathcal{G}_\epsilon(h)\|_{C^0} \leq \epsilon\ \ell n_* \sup_{|u|_p \leq \delta, |v|_q \leq r} \|(u, v)\|^2 = \epsilon\ \ell n_* \max(\delta, r)^2.$$

By assumption, $\epsilon \ell n_* \max(\delta, r)^2 < r$, and Step 1 is complete. $\square$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Step 2. If $\epsilon < \min\{\epsilon_1, [r\ell \max(\delta, r)\|\psi\|_{C^1}]^{-1}, r[n_*(1+r)\ell \max(\delta, r)]^{-1}\} =: \epsilon_2$, then $\mathcal{G}_\epsilon : Y_r \to Y_r$.

Proof. Suppose that $h \in Y_r$ so that $h \in X_r$, then $T^j(h)$ is differentiable on $\overline{\Omega}_\delta$ as h is; moreover throughout the remainder of the proof of Step 2 we shall write $H := \mathcal{G}_\epsilon(h)$ for brevity. It then follows that

$$dH(u) = \sum_{j=0}^{\nu-1} N^j \left[d_u \mathsf{g}_\epsilon(u, T^j(h)(u)) + d_v \mathsf{g}_\epsilon(u, T^j(h)(u))[d_u(T^j(h))(u)]\right], \tag{4.14}$$

and we now need to estimate $\|dH(u)\|_{C^0}$. By using the fact that $\|h\|_{C^0} \leq r$, since $T : X_r \to X_r$ from Step 1, we obtain $|T^j(h)(u)|_q = |h(\pi(h(\ldots))\ldots)|_q \leq r$, and therefore

$$\begin{aligned}
\sup_{|u|_p \leq \delta} \|dH(u)\|_{BL(\mathbb{R}^p,\mathbb{R}^q)} &\leq \sum_{j=0}^{\nu-1} \|N\|^j \big(\|d_u \mathsf{g}_\epsilon(u, T^j(h)(u))\| \\
&\qquad + \|d_v \mathsf{g}_\epsilon(u, T^j(h)(u))\| \|d_u(T^j(h))(u)\|\big) \\
&\leq n_* \sup_{0 \leq j \leq \nu-1} \big(\|d_u \mathsf{g}_\epsilon(u, T^j(h)(u))\| + \|d_v \mathsf{g}_\epsilon(u, T^j(h)(u))\| \|d_u(T^j(h))(u)\|\big) \\
&\leq n_* \sup_{\substack{0 \leq j \leq \nu-1 \\ |u|_p \leq \delta, |v|_q \leq r}} \big(\|d_u \mathsf{g}_\epsilon(u, v)\| + \|d_v \mathsf{g}_\epsilon(u, v)\| \|d_u(T^j(h))(u)\|\big) \\
&\leq \epsilon\, n_* \ell \max(r, \delta) \left(1 + \sup_{0 \leq j \leq \nu-1, |u|_p \leq \delta} \|d_u(T^j(h))(u)\|\right).
\end{aligned}$$

Now we estimate the final bracketed term in the latter expression. The linear mapping obtained from differentiating $T^j(h)(u)$ with respect to u is

$$d_u(T^j(h))(u) = d_u(h(\pi(h(\ldots \pi(h) \ldots)))),$$

which can be written as the recurrence

$$d_u(T^j(h))(u) = dh(\pi(T^{j-1}(h))(u)) d\pi(T^{j-1}(h)(u)) \cdot d_u(T^{j-1}(h)(u)), \tag{4.15}$$

where, by definition, $d_u(T^0(h))(u) = d_u(h)(u) = dh(u)$. By taking C^0 norms and setting

$$\xi_j := \|d_u(T^j(h))(u)\|_{BL(\mathbb{R}^p,\mathbb{R}^q)},$$

we obtain the relation

$$\xi_j \leq \|dh\|_{C^0} \sup_{\|\overline{h}\|_{C^0} \leq r} \|d\pi(\overline{h})\|_{BL(C^0)} \cdot \xi_{j-1},$$

where $\xi_0 = \|dh\|_{C^0} \leq \|h\|_{C^1} \leq r$. From (4.9) of Lemma 3, we find that

$$\xi_j \leq \epsilon \ell r \max(\delta, r) \|\psi\|_{C^1} \cdot \xi_{j-1} \leq (\epsilon \ell r \max(\delta, r) \|\psi\|_{C^1})^j\, \xi_0 \leq r$$

because $\epsilon \ell r \max(\delta, r) \|\psi\|_{C^1} < 1$ by assumption. As a result, the inequality

$$\begin{aligned}
\sup_{|u|_p \leq \delta} \|dH(u)\|_{BL(\mathbb{R}^p,\mathbb{R}^q)} &\leq \epsilon\, n_* \ell \max(r, \delta) \left(1 + \sup_{0 \leq j \leq \nu-1} \xi_j\right) \\
&\leq \epsilon n_* \ell \max(r, \delta)(1 + r) \leq r
\end{aligned}$$

also now follows from the assumption of the claim, and we have proven that $h \in Y_r$, as required. ∎

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Step 3. If $\epsilon < \epsilon_2$ and we define the $O(\epsilon)$ quantity $\kappa(\epsilon)$ by

$$\kappa(\epsilon) := \epsilon \cdot \ell \max(\delta, r) \sum_{i=0}^{\nu-1} (r\epsilon \cdot \|\psi\|_{C^1} \ell \max(\delta, r))^i,$$

then $\|h_{n+1} - h_n\|_{C^0} \le \kappa(\epsilon)\|h_n - h_{n-1}\|_{C^0}$. As a result, (h_n) is Cauchy in X_r if ϵ is further restricted so that $\kappa(\epsilon) < 1$, and so (h_n) converges in X_r.

Proof. For brevity, let us write h in place of h_{n+1} and k for h_n after setting $h_0 = 0 \in Y_r$ and $h_{n+1} = \mathcal{G}_\epsilon(h_n)$. Now

$$\begin{aligned} \|\mathcal{G}_\epsilon(h) - \mathcal{G}_\epsilon(k)\|_{C^0} &\le \sum_{j=0}^{\nu-1} \|N\|^j \|\mathsf{G}_\epsilon(T^j(h)) - \mathsf{G}_\epsilon(T^j(k))\|_{C^0} \\ &\le n_* \sup_{0 \le j \le \nu-1} \|\mathsf{g}_\epsilon(u, T^j(h)(u)) - \mathsf{g}_\epsilon(u, T^j(k)(u))\|_{C^0}. \end{aligned}$$

The mean-value inequality yields

$$\begin{aligned} &|\mathsf{g}_\epsilon(u, T^j(h)(u)) - \mathsf{g}_\epsilon(u, T^j(k)(u))|_q \\ &\qquad \le \sup_{z \in [T^j(h)(u), T^j(k)(u)]} \|d_v \mathsf{g}_\epsilon(u, z)\| \|T^j(h) - T^j(k)\|_{C^0} \\ &\qquad \le \sup_{|u|_p \le \delta, |v|_q \le r} \|d_v \mathsf{g}_\epsilon(u, v)\| \|T^j(h) - T^j(k)\|_{C^0} \qquad \text{(using Step 1)} \\ &\qquad \le \epsilon\ell \max(r, \delta) \cdot \|T^j(h) - T^j(k)\|_{C^0} \qquad \text{(by (4.3))}, \end{aligned} \tag{4.16}$$

where, for any $z_1, z_2 \in \mathbb{R}^p$, the generalized interval from z_1 to z_2 is given by

$$[z_1, z_2] := \{\lambda z_1 + (1 - \lambda) z_2 : 0 \le \lambda \le 1\}.$$

So let us define $\chi_j := \|T^j(h) - T^j(k)\|_{C^0}$, and we estimate χ_j as follows:

$$\begin{aligned} |T^j(h)(u) - T^j(k)(u)|_q &= |h(\pi(T^{j-1}(h))) - k(\pi(T^{j-1}(k)))|_q \\ &\le |h(\pi(T^{j-1}(h))) - k(\pi(T^{j-1}(h)))|_q \\ &\qquad + |k(\pi(T^{j-1}(h))) - k(\pi(T^{j-1}(k)))|_q \\ \implies \qquad \chi_j &\le \|h - k\|_{C^0} + \|k(\pi(T^{j-1}(h))) - k(\pi(T^{j-1}(k)))\|_{C^0}. \end{aligned}$$

However, from Step 2 we have $\|k\|_{C^1} \le r$, and therefore

$$|k(u) - k(u')|_q \le r|u - u'|_p \qquad \forall u, u' \in \overline{\Omega}_\delta.$$

Using the fact that π is a Fréchet differentiable mapping on $C^0(\overline{\Omega}_\delta, \mathbb{R}^q)$ with norm bounded according to (4.9), an application of the mean-value inequality gives

$$\chi_j \le \|h - k\|_{C^0} + r \sup_{\|\overline{h}\|_{C^1} \le r} \|d\pi(\overline{h})\|_{BL(C^0)} \cdot \chi_{j-1},$$

so that $\chi_j \le \|h - k\|_{C^0} + \epsilon \cdot r\|\psi\|_{C^1} \ell \max(\delta, r)\chi_{j-1}$, where $\chi_0 = \|h - k\|_{C^0}$. The discrete Gronwall inequality now gives

$$\chi_j \le \|h - k\|_{C^0} \sum_{i=0}^{\nu-1} \left(r\epsilon\|\psi\|_{C^1} \ell \max(\delta, r)\right)^i,$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

and with (4.16) we have the desired inequality

$$\|\mathcal{G}_\epsilon(h) - \mathcal{G}_\epsilon(k)\|_{C^0} \le \kappa(\epsilon) \cdot \|h - k\|_{C^0}.$$

The standard contraction argument now shows that $(h_n) \subset Y_r$ is Cauchy in X_r as claimed, and there therefore exists an $h \in X_r$ such that $h_n \xrightarrow{C^0} h$ as $n \to \infty$. ◻

Since $(h_n) \subset Y_r$, it follows that there is a subsequence (n_j) such that $h_{n_j} \xrightarrow{C^\alpha} \overline{h}$ as $j \to \infty$ for some $\overline{h} \in X_r \cap C^\alpha(\overline{\Omega}_\delta)$, and we deduce that $\overline{h} = h \in C^\alpha(\overline{\Omega}_\delta)$. By restricting ϵ further we can actually ensure that h is differentiable, as follows.

Step 4. There is an $\epsilon_3 > 0$ such that $h \in C^{1+\alpha}(\overline{\Omega}_\delta)$ whenever $\epsilon < \min(\epsilon_2, \epsilon_3)$ and $\kappa(\epsilon) < 1$.

Proof. Let $h \in C^2(\overline{\Omega}_\delta)$ satisfy $\|h\|_{C^2} \le r$, and recall that $H := \mathcal{G}_\epsilon(h)$. From (4.14) we obtain

$$\begin{aligned} d^2H(u) = \sum_{j=0}^{\nu-1} N^j \{ & d^2_{uu}\mathsf{g}_\epsilon(u, T^j(h)(u)) + 2d^2_{uv}\mathsf{g}_\epsilon(u, T^j(h)(u))[I, d_u(T^j(h)(u))] \\ & + d^2_{vv}\mathsf{g}_\epsilon(u, T^j(h)(u))[d_u(T^j(h)(u)), d_u(T^j(h)(u))] \\ & + d_v\mathsf{g}_\epsilon(u, T^j(h)(u))[d^2_{uu}(T^j(h)(u))]\}. \end{aligned}$$

In seeking a bound on $\|H\|_{C^2}$, we now examine the term $d^2_{uu}(T^j(h)(u))$ more closely as bounds on the remaining elements of d^2H can be obtained from Steps 1 and 2. By applying the chain rule (4.15) we obtain the recurrence

$$\begin{aligned} d^2_{uu}(T^j(h)(u)) &= d_u \left\{ dh(\pi(T^{j-1}(h))(u)) d\pi(T^{j-1}(h)(u)) \cdot d_u(T^{j-1}(h)(u)) \right\} \\ &= d^2h(\pi(T^{j-1}(h))(u))[d\pi(T^{j-1}(h)(u)) \cdot d_u(T^{j-1}(h)(u))]^{(2)} \\ &\quad + dh(\pi(T^{j-1}(h))(u))[d^2\pi(T^{j-1}(h)(u))[d_u(T^{j-1}(h)(u))]^{(2)}] \\ &\quad + dh(\pi(T^{j-1}(h))(u))[d\pi(T^{j-1}(h)(u))[d^2_{uu}(T^{j-1}(h)(u))]], \end{aligned}$$

and taking norms gives

$$\begin{aligned} \|d^2_{uu}(T^j(h)(u))\|_{C^0} \le\ & \|d^2h\|_{C^0}\|d\pi(h)\|^2\|d_u(T^j(h)(u))\|_{C^0} \\ & + \|dh\|_{C^0}\|d^2\pi(h)\|\|d_u(T^j(h)(u))\|^2_{C^0} \\ & + \|dh\|_{C^0}\|d\pi(h)\|\|d^2_{uu}(T^j(h)(u))\|_{C^0}. \end{aligned}$$

If we write

$$\eta_j := \|d^2_{uu}(T^j(h)(u))\|_{C^0},$$

and use the fact that $\xi_j = \|d_u(T^j(h))(u)\|_{BL(\mathbb{R}^p,\mathbb{R}^q)} \le r$ for all $0 \le j \le \nu - 1$, which was established in Step 2, we obtain the difference inequality

$$\eta_j \le r^3\|d\pi(h)\|^2_{C^0} + r^2\|d^2\pi(h)\|_{C^0} + r\|d\pi\|_{C^0} \cdot \eta_{j-1}$$

such that $\eta_0 \le r$ by definition. There results, for $j \ge 1$,

$$\eta_j \le (r\|d\pi(h)\|_{C^0})^j\, \eta_0 + \left(r^3\|d\pi(h)\|^2_{C^0} + r^2\|d^2\pi(h)\|_{C^0}\right) \sum_{i=0}^{j-1}(r\|d\pi(h)\|_{C^0})^i,$$

and the bounds (4.9) and (4.10) show that $\eta := \max_{0\le j\le \nu-1} \eta_j$ has an $O(1)$ dependence on ϵ in the sense that there is an $M > 0$ such that $\eta \le M$ whenever $0 \le \epsilon \le 1$. (In fact, one can choose M to be less than r if ϵ is sufficiently small.)

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

We now obtain

$$\begin{aligned}\|d^2H\|_{C^0} &\le n_* \sup_{0\le j\le \nu-1} \left\{\|d^2_{uu}\mathsf{g}_\epsilon(u,T^j(h)(u))\|_{C^0} + 2\xi_j\|d^2_{uv}\mathsf{g}_\epsilon(u,T^j(h)(u))\|_{C^0}\right.\\ &\qquad\left. + \xi_j^2\|d^2_{vv}\mathsf{g}_\epsilon(u,T^j(h)(u))\|_{C^0} + \eta_j\|d_v\mathsf{g}_\epsilon(u,T^j(h)(u))\|_{C^0}\right\}\\ &\le n_* \sup_{\|\overline{h}\|_{C^0}\le r, 0\le j\le\nu-1} \left\{\|d^2_{uu}\mathsf{g}_\epsilon(u,\overline{h})\|_{C^0} + 2r\|d^2_{uv}\mathsf{g}_\epsilon(u,\overline{h})\|_{C^0}\right.\\ &\qquad\left. + r^2\|d^2_{vv}\mathsf{g}_\epsilon(u,\overline{h})\|_{C^0} + \eta\|d_v\mathsf{g}_\epsilon(u,\overline{h})\|_{C^0}\right\},\end{aligned} \tag{4.17}$$

where η has been used to bound the last term in (4.17). By using (4.5) to estimate the second derivative terms we find that

$$\begin{aligned}\|d^2H\|_{C^0} \le \epsilon\cdot n_* \sup_{0\le j\le\nu-1} &\left\{\|d^2g(0,0)\| + \epsilon\overline{\ell}\max(r,\delta)\right.\\ &+ 2r(\|d^2g(0,0)\| + \epsilon r\overline{\ell}\max(r,\delta))\\ &\left.+ r^2(\|d^2g(0,0)\| + \epsilon\overline{\ell}\max(r,\delta)) + \eta\ell\max(r,\delta)\right\}.\end{aligned} \tag{4.18}$$

It is immediate from (4.18) and Steps 1 and 2 that a suitably small choice of ϵ ensures that $\|H\|_{C^2} \le r$ whenever $\|h\|_{C^2} \le r$. As a result, if we impose the following restriction on the initial guess for a fixed point of $\mathcal{G}_\epsilon$:

$$h_0 \in Z_r := \{h \in C^2(\overline{\Omega}_\delta, \mathbb{R}^q) : \|h\|_{C^2} \le r\},$$

then the C^0-convergent sequence from Step 3 also satisfies $(h_n) \subset Z_r$. This means that a $C^{1+\alpha}$-convergent subsequence can now be extracted from (h_n), so that the convergence of h_n to h actually occurs in $C^{1+\alpha}$ and h therefore lies in this smoother space. □

We have shown that there is a differentiable solution of (4.2) on a sufficiently small ball around the origin, but there remains to prove the last part of Theorem 1 regarding the behavior of h at the origin. So let h be a C^1 solution of (4.2) on some ball $\overline{\Omega}_\delta$, and put $\zeta = h(0)$. It follows that ζ is a solution of the algebraic equation

$$-\zeta + Nh(\mathsf{f}_\epsilon(0,\zeta)) + \mathsf{g}_\epsilon(0,\zeta) = 0, \tag{4.19}$$

and (4.19) has solution $\zeta = 0$. As the linearization of the left-hand side of (4.19) at $\zeta = 0$ is a multiple of the identity, the inverse function theorem ensures that $\zeta = 0$ is the only solution of (4.19) in some neighborhood of zero, and this ensures that $h(0) = 0$. Since any solution of (4.13) provides one of (4.2), we can differentiate (4.2) with respect to u and set $u = 0$; this gives

$$dh(0) = Ndh(0)[C],$$

but then we can continue in an inductive manner to deduce that

$$Ndh(0)[C] = N^2dh(0)[C^2] = \cdots = N^\nu dh(0)[C^\nu] = 0,$$

as N is nilpotent. We find that $dh(0) = 0$, and this concludes the proof of Theorem 1. □

The question of maximal regularity of a solution of (4.2), or equivalently (4.13), is not addressed, although the method of proof used in Theorem 1 can be continued by restricting ϵ further as required to show that $\mathcal{G}_\epsilon$ maps the ball of radius r in $C^j(\overline{\Omega}_\delta, \mathbb{R}^p)$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

into itself. This ensures that the sequence (h_n) constructed in the proof of Theorem 1 converges to h in as strong a C^j norm as we like, provided that f and g are sufficiently smooth. We cannot, however, be sure that the resulting solution h lies in $C^\infty(\overline{\Omega}_\delta, \mathbb{R}^p)$ because the interval in which ϵ must reside so as to obtain a C^j fixed point of $\mathcal{G}_\epsilon$ could shrink indefinitely as j grows.

Theorem 1 does not ensure the existence of a continuous fixed point of (4.2) which is not C^α for some $0 < \alpha < 1$. If we take a continuous initial guess for a solution, $h_0 \in C^0(\overline{\Omega}_\delta, \mathbb{R}^q)$, say, then, although there is a sequence of continuous functions defined by $h_{n+1} = \mathcal{G}_\epsilon(h_n)$, there is no reason to suspect that the sequence of iterates (h_n) will satisfy the property of being a C^0-contractive sequence. On the other hand, a suitably small C^1 initial guess will lead to C^α convergence of the resulting iterates for any $\alpha \in [0, 1)$.

4.4. Stability implies uniqueness. There are some simple cases where uniqueness and smoothness can be easily established. The most obvious is where $N = 0$, and then Theorem 1 can be proven using the elementary implicit function theorem. Another occurs when the matrix denoted C that arises from the Kronecker normal form of $(\mathcal{A}, \mathcal{B})$ in (3.5) satisfies $\|C\| < 1$ in the norm induced by $|\cdot|_p$. In this case the cutoff function ψ used above is not needed in order to obtain a well-defined operator π. If we define the Nemitskii operator

$$\pi(h)(u) = Cu + \mathsf{f}_\epsilon(u, h(u)),$$

and $h \in C^0(\overline{\Omega}_\delta, \mathbb{R}^q)$ satisfies $\|h\|_{C^0} \leq r$, then $|Cu + \mathsf{f}_\epsilon(u, h(u))|_p \leq \|C\||u|_p + |\mathsf{f}_\epsilon(u, h(u))|_p \leq \|C\|\delta + \epsilon \cdot \ell \max(\delta, r)^2$. In order for $h(\pi(h))$ to be well-defined, we now need only to choose ϵ such that $\|C\|\delta + \epsilon \cdot \ell \max(\delta, r)^2 \leq \delta$, which can be done. The proof of Theorem 1 then goes through with this minor modification, and the resulting fixed point of $\mathcal{G}_\epsilon$ is unique in the space of continuously differentiable functions.

4.5. Polynomial approximation. Let us remove the dependence of (4.2) on ϵ for clarity and return to the fixed-point problem (3.6) directly, which we recall defines a nonlinear operator G via

$$\begin{aligned} h(u) = Nh(Cu + f(u, h(u))) + g(u, h(u)) =: (Gh)(u), \\ \text{where } h(0) = 0, dh(0) = 0. \end{aligned} \tag{4.20}$$

PROPOSITION 1. *Suppose that $h \in C^k(\overline{\Omega}_\delta, \mathbb{R}^q)$ is a solution of* (4.20), *and suppose that $\overline{h} \in C^k(\overline{\Omega}_\delta, \mathbb{R}^q)$ satisfies*

$$\overline{h}(u) - G(\overline{h})(u) = e(u), \quad \overline{h}(0) = 0,$$

where e is a given C^k function that satisfies $e(0) = 0, de(0) = 0, \ldots, d^k e(0) = 0$. Then

$$|h(u) - \overline{h}(u)|_q = o(|u|_p^k) \text{ as } u \to 0.$$

Proof. If we define the function $\Delta := h - \overline{h}$, then we have to prove that

$$\Delta(0) = 0, d\Delta(0) = 0, \ldots, d^k\Delta(0) = 0,$$

and the result then follows from the basic properties of the derivative. Clearly $\Delta(0) = 0$, and differentiating (4.20) with respect to u gives

$$\begin{aligned} dh(u) = Ndh(Cu + f(u, h(u)))[C + d_u f(u, h(u)) + d_v f(u, h(u))[dh(u)]] \\ + d_u g(u, h(u)) + d_v g(u, h(u))[dh(u)], \end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

and similarly,

$$d\overline{h}(u) = Nd\overline{h}(Cu + f(u, \overline{h}(u)))[C + d_u f(u, \overline{h}(u)) + d_v f(u, \overline{h}(u))[d\overline{h}(u)]] \\ + d_u g(u, \overline{h}(u)) + d_v g(u, \overline{h}(u))[d\overline{h}(u)] + de(u).$$

As a result, because $h(0) = \overline{h}(0) = 0$, we find that

$$d\Delta(0) = Nd\Delta(0)[C] + de(0) = Nd\Delta(0)[C],$$

by the assumption that $de(0) = 0$, so that

$$d\Delta(0) = Nd\Delta(0)[C] = N^2 d\Delta(0)[C^2] = \cdots = N^\nu d\Delta(0)[C^\nu] = 0.$$

Continuing in a similar vein, we obtain

$$d^2h(u) = Nd^2h(Cu + f(u, h(u)))[C + d_u f(u, h(u)) \\ + d_v f(u, h(u))[dh(u)]]^{(2)} + d^2_{uu} g(u, h(u)) + 2d^2_{uv} g(u, h(u))[I, dh(u)] \\ + d^2_{vv} g(u, h(u))[dh(u), dh(u)] + d_v g(u, h(u))[d^2h(u)],$$

with a similar expression for $d^2\overline{h}(u)$, with the additional presence of the term $d^2e(u)$. We find that

$$d^2h(0) = Nd^2h(0)[C, C] + d^2_{uu} g(0, 0), \tag{4.21}$$

and, because $d^2e(0) = 0$, (4.21) also holds with $h(0)$ replaced by $\overline{h}(0)$. We deduce that the bilinear form $d^2\Delta(0)$ satisfies

$$d^2\Delta(0)[X, Y] = Nd^2\Delta(0)[CX, CY] \qquad (\forall X, Y \in \mathbb{R}^p),$$

so that

$$d^2\Delta(0)[X, Y] = N^\nu d^2\Delta(0)[C^\nu X, C^\nu Y] = 0 \qquad (\forall X, Y \in \mathbb{R}^p).$$

We omit the details, but by continuing inductively and assuming $d^j e(0) = 0$, one obtains the result that the j-linear form $d^j\Delta(0)$ satisfies

$$d^j\Delta(0)[X_1, X_2, \ldots, X_j] = Nd^j\Delta(0)[CX_1, CX_2, \ldots, CX_j]$$

for all $X_1, \ldots, X_j \in \mathbb{R}^p$. The nilpotency of N now ensures that the latter quantity is zero. ∎

A simple corollary to Proposition 1 is that if (4.20) has two infinitely differentiable solutions h and $\overline{h}$ defined on some neighborhood of zero, then they agree beyond all orders at zero:

$$\lim_{u \to 0} \frac{|h(u) - \overline{h}(u)|_q}{|u|_p^k} = 0 \qquad (\forall k \geq 1).$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

5. Applications.

5.1. Nonlinear normal form. The first application of Theorem 1 is the following result, which says that (1.1) induces a local dynamical system on a manifold in the sense of Definition 2. Recall the definition of the matrix pencil (A, B) via

$$(A, B) := (d_{\overline{z}}F(0,0), d_z F(0,0)),$$

where F has $(z, \overline{z})$ as its argument, and note that the following matrix pencil defined on $\mathbb{R}^{2m}$:

$$(\mathcal{A}, \mathcal{B}) := \left(\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & I \\ B & A \end{pmatrix} \right),$$

satisfies $\sigma(\mathcal{A}, \mathcal{B}) = -\sigma(A, B)$.

THEOREM 2 (nonlinear Kronecker normal form). *Let $\alpha \in [0, 1)$, and suppose that* (A1)–(A3) *hold; then there is a linear space K_1 and a $C^{1+\alpha}$-manifold $\mathcal{M}$ modeled on K_1 such that* $\dim(K_1) = \#\sigma(A, B)$ *and*

1. *for all $(z, w) \in \mathcal{M}$ there results $F(z, w) = 0$,*
2. *there is an r' such that for each $(z, w) \in \mathcal{M}$, with $\|(z, w)\| < r'$, there is a unique $(\overline{z}, \overline{w}) \in \mathcal{M}$ such that $w = \overline{z}$, and*
3. *there is a $C^{1+\alpha}$-diffeomorphism $\theta : K_1 \to \mathcal{M}$ such that if $(z, w) = \theta(u)$ for some $u \in K_1$, then $(\overline{z}, \overline{w}) = \theta(\varphi(u))$, where*

$$\varphi(u) = Cu + p(u)$$

for some linear map $C : K_1 \to K_1$ that satisfies $\sigma(C) = \sigma(A, B)$. Moreover, $p : B_\delta(0; K_1) \to K_1$ satisfies $p(0) = 0$ and $d_u p(0) = 0$.

Proof. From Theorem 1, first identify the linear space K_1 from the KNF of $(\mathcal{A}, \mathcal{B})$ with $\mathbb{R}^p$ and K_2 with $\mathbb{R}^q$. Now define the $C^{1+\alpha}$-graph

$$\hat{\mathcal{M}} := \{(u, h(u)) \in \mathbb{R}^p \oplus \mathbb{R}^q : u \in \Omega_\delta\},$$

and let the r-neighborhood of zero in $\hat{\mathcal{M}}$ be $\hat{\mathcal{M}}_r := \{(u, h(u)) \in \mathbb{R}^p \oplus \mathbb{R}^q : |u|_p < r\}$ whenever $r < \delta$.

As a consequence of Theorem 1, there is an $r > 0$ such that for each $(u, v) = (u, h(u)) \in \hat{\mathcal{M}}_r$ we can find a pair $(\overline{u}, \overline{v})$ given by $(\overline{u}, h(\overline{u})) \in \hat{\mathcal{M}}$ such that (NF) is satisfied:

$$\overline{u} = Cu + f(u, v), \quad N\overline{v} = v - g(u, v),$$

where C is obtained from the KNF of $(\mathcal{A}, \mathcal{B})$ so that $\sigma(C) = -\sigma(\mathcal{A}, \mathcal{B}) = \sigma(A, B)$.

Let $\varphi(u) := Cu + f(u, h(u))$, and for each $u \in \mathbb{R}^p$ of sufficiently small norm set

$$\theta(u) := Q(u, h(u)) \qquad \text{and} \qquad \mathcal{M} \equiv Q(\hat{\mathcal{M}}), \quad \mathcal{M}_r \equiv Q(\hat{\mathcal{M}}_r),$$

where the linear map Q is taken from the KNF of $(\mathcal{A}, \mathcal{B})$ from the discussion that immediately follows (3.4). The map θ then provides a local diffeomorphism between Ω_δ and $\mathcal{M}$; moreover both

$$F(\theta(u)) = 0 \qquad \text{and} \qquad F(\theta(\varphi(u))) = 0$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

follow by the construction of h. The following diagram illustrates how a map is induced on $\mathcal{M}$ in this way:

$$\begin{array}{ccc} \mathcal{M}_r & \longrightarrow & \mathcal{M} \\ \theta^{-1}\downarrow & & \uparrow\theta \\ \Omega_r & \xrightarrow{\varphi} & \Omega_\delta \end{array},$$

where $\theta^{-1}(Q(u,h(u))) = u$, and this concludes the proof. ∎

The following immediate corollaries of Theorem 1 provide some information regarding the stable and unstable behavior of orbits in a neighborhood of a fixed point of (1.1).

COROLLARY 1. *Suppose that F satisfies* (A1)–(A3) *and (A,B) is a regular matrix pencil with $\rho(A,B) < 1$; then there is a $C^{1+\alpha}$-solution manifold $\mathcal{M}$ of* (1.1) *containing* 0 *such that each $(z,\overline{z}) \in \mathcal{M}$ supports a global orbit $(z_n)_{n=0}^{\infty}$ with $(z_n, z_{n+1}) \in \mathcal{M}$, $z_0 = z$, $z_1 = \overline{z}$, and $\lim_{n\to\infty} z_n = 0$.*

Proof. The orbit is constructed by iterating the map $\varphi(u) = Cu + p(u)$ given in part 3 of Theorem 2: Because $\rho(A,B) < 1$ we have $\rho(C) < 1$ so that φ is a contraction near the origin in some norm, and the result follows. ∎

The following are the natural definitions of stable and unstable sets associated with fixed points of (1.1); note that they are subsets of $\mathbb{R}^{2m}$ and not $\mathbb{R}^m$.

DEFINITION 4. *The set*

$$W^s_{\mathrm{loc}}(0) := \{(z,\overline{z}) \in \mathbb{R}^{2m} : \exists \text{ global orbit } (z_n)_{n=0}^{\infty}, z_0 = z, z_1 = \overline{z}, \lim_{n\to\infty} z_n = 0\}$$

is the local stable set associated with the zero fixed point of (1.1), *and*

$$W^u_{\mathrm{loc}}(0) := \{(z,\overline{z}) \in \mathbb{R}^{2m} : \exists \text{ global orbit } (z_n)_{n=0}^{-\infty}, z_{-1} = z, z_0 = \overline{z}, \lim_{n\to-\infty} z_n = 0\}$$

is the local unstable set.

In case $\sigma(A,B)$ contains elements outside the unit disk, one can apply the stable manifold theorem for maps to φ in Theorem 2 to give the following result.

COROLLARY 2. *Suppose that* (A1)–(A3) *are satisfied and (A,B) possesses n_s eigenvalues in the open unit disk; then* (1.1) *possesses a subset of the local stable set which is a differentiable manifold of dimension n_s.*

There is an analogous corollary to show that the unstable set is nonempty and contains a manifold of dimension n_u, where n_u is the number of elements of $\sigma(A,B)$ lying outside the closed unit disk. This result is obtained by applying Corollary 2 to (1.1) but with time running *backwards*.

COROLLARY 3. *Suppose that* (A1)–(A3) *are satisfied and (B,A) possesses n_u eigenvalues in the open unit disk; then* (1.1) *possesses a subset of the local unstable set which is a differentiable manifold of dimension n_u.*

Proof. Let us rewrite (1.1) in the form

$$F(z_{n-1}, z_n) = 0 \tag{5.1}$$

to emphasize the fact that we are seeking an orbit that propagates backwards in time, with (z_{-1}, z_0) given. The linearization of (5.1) is of the form

$$Bz_{n-1} + Az_n$$

and if $\det B \neq 0$, then we may locally solve (5.1) for $z_{n-1} = f(z_n)$, and then one can apply the stable manifold theorem for maps to this.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

On the other hand if $\det B = 0$, then conditions (A1)–(A3), appropriately modified by exchanging the roles of z and $\overline{z}$ because time is flowing backwards, still apply to (5.1) because (B, A) is a regular matrix pencil due to the fact that (A, B) is regular. The result then follows from Corollary 2. ∎

As A is a singular mapping it follows that the finite spectrum of (B, A) satisfies

$$\sigma(B, A) = (\sigma(A, B)\backslash\{0\})^{-1} \cup \{0\},$$

whether or not B is singular, and hence $\sigma(B, A)$ contains zero so that $n_u \geq 1$. An unstable manifold therefore always exists for (1.1) under conditions (A1)–(A3).

5.2. Bifurcation theorems. We now consider a C^k-mapping $F : \mathbb{R}^{2m} \times \mathbb{R} \to \mathbb{R}^{2m}$, where $k \geq 5$, and examine the family of difference equations

$$F(z_n, z_{n+1}, \mu) = 0. \tag{5.2}$$

Define the one-parameter family of matrix pencils

$$\mathcal{P}(\mu) := (A(\mu), B(\mu)) := (d_{\overline{z}}F(0, 0, \mu), d_z F(0, 0, \mu)),$$

where F has $(z, \overline{z}, \mu)$ as its argument.

THEOREM 3 (parameterized nonlinear KNF). *Suppose that $(0, 0)$ is a fixed point of (5.2) for all $\mu \in \mathbb{R}$ and that $\mathcal{P}(0)$ is a regular matrix pencil. Then there is a linear space K_1 and a $C^{1+\alpha}$-parameter family of $C^{1+\alpha}$-manifolds $\mathcal{M}_\mu$ modeled on K_1 such that $\dim(K_1) = \#\sigma(\mathcal{P}(0))$ and*

1. *for all $(z, w) \in \mathcal{M}_\mu$ there results $F(z, w, \mu) = 0$,*
2. *there is an r' (independent of μ) such that for each $(z, w) \in \mathcal{M}_\mu$, with $\|(z, w)\| < r'$, there is a unique $(\overline{z}, \overline{w}) \in \mathcal{M}_\mu$ such that $w = \overline{z}$, and*
3. *there is a $C^{1+\alpha}$-parameter family of $C^{1+\alpha}$-diffeomorphisms $\theta_\mu : K_1 \to \mathcal{M}_\mu$ such that if $(z, w) = \theta_\mu(u)$ for some $u \in K_1$, then $(\overline{z}, \overline{w}) = \theta_\mu(\varphi(u, \mu))$, where*

   $$\varphi(u, \mu) = C(\mu)u + p(u, \mu).$$

   Moreover, $C(\cdot)$ is a $C^{1+\alpha}$-parameter family of maps in $BL(K_1)$ and, for some $\delta > 0$, $p : B_\delta(0; K_1) \times B_\delta(0; \mathbb{R}) \to K_1$ satisfies

   $$p(0, \mu) \equiv 0, \ \ d_u p(0, \mu) \equiv 0.$$

4. *If $\lambda : [-\delta, \delta] \to \mathbb{C}$ is a continuous (and so bounded) curve, then $\lambda(\mu) \in \sigma(\mathcal{P}(\mu))$ for all $\mu \in [-\delta, \delta]$ if and only if $\lambda(\mu) \in \sigma(C(\mu))$ for all $\mu \in [-\delta, \delta]$.*

Proof. Consider the suspended difference equation

$$\begin{cases} \mu_{n+1} = \mu_n, \\ z_{n+1} = w_n, \\ \ \ \ 0 \ \ \ = F(z_n, w_n, \mu_n). \end{cases} \tag{S}$$

Let us write

$$F(z, w, \mu) = A(\mu)w + B(\mu)z + \mathcal{F}(z, w, \mu),$$

where $F(0, 0, \mu) = 0$, $d_z F(0, 0, \mu) = B(\mu)$, and $d_w F(0, 0, \mu) = A(\mu)$, and then consider the following matrix pencil on $\mathbb{R}^{2m}$:

$$(\mathcal{A}, \mathcal{B}(\mu)) := \left(\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & I \\ B(\mu) & A(\mu) \end{pmatrix} \right).$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

This satisfies $\sigma(\mathcal{A},\mathcal{B}(\mu)) = -\sigma(\mathcal{P}(\mu))$ and is a regular matrix pencil when $\mu = 0$, and we can exploit this fact using the resulting Kronecker normal form to put (S) into a normal form. If we set $\mathbf{z} = (z, w)$, then we may write (S) as

$$\begin{aligned}\overline{\mu} &= \mu,\\ \mathcal{A}\overline{\mathbf{z}} &= \mathcal{B}(\mu)\mathbf{z} + \mathcal{F}(\mathbf{z},\mu),\end{aligned}$$

where an overbar is used to denote a forward iterate. Now there is a matrix pair (P, Q) such that $P\mathcal{A}Q = \left(\begin{smallmatrix} I_{K_1} & 0\\ 0 & N\end{smallmatrix}\right)$ and $P\mathcal{B}(0)Q = \left(\begin{smallmatrix} C & 0\\ 0 & I_{K_2}\end{smallmatrix}\right)$, where $Q : \mathbb{R}^{2m} \to K_1 \oplus K_2 = \mathbb{R}^{p+q}$ and N is nilpotent. With $(u, v) := \mathbf{w} = Q\mathbf{z} \in K_1 \oplus K_2$, we obtain

$$\overline{\mu} = \mu, \tag{5.3}$$
$$\overline{u} = \alpha(\mu)u + \beta(\mu)v + \mathcal{G}_1(u, v, \mu), \tag{5.4}$$
$$N\overline{v} = \gamma(\mu)u + \delta(\mu)v + \mathcal{G}_2(u, v, \mu), \tag{5.5}$$

where

$$\alpha : K_1 \to K_1, \quad \beta : K_2 \to K_1, \quad \gamma : K_1 \to K_2, \quad \text{and} \quad \delta : K_2 \to K_2$$

are differentiable linear maps in μ, and

$$\alpha(0) = C, \quad \beta(0) = 0, \quad \gamma(0) = 0, \quad \text{and} \quad \delta(0) = I_{K_2},$$

where $C \in BL(K_1)$ is provided by the KNF of $\mathcal{P}(0)$ and $\sigma(C) = -\sigma(\mathcal{A},\mathcal{B}(0)) = \sigma(\mathcal{P}(0))$. Moreover, $\mathcal{G}_1$ and $\mathcal{G}_2$ represent $\mathcal{O}_2(u, v)$-functions parameterized by μ.

Seeking an invariant manifold on which $v = h(u, \mu)$, we put (5.3)–(5.5) into the form

$$\overline{\mu} = \mu, \tag{5.6}$$
$$\overline{u} = Cu + \mathcal{O}_2(u, v, \mu), \tag{5.7}$$
$$N\overline{v} = v + \mathcal{O}_2(u, v, \mu), \tag{5.8}$$

where $\mathcal{O}_2(u, v, \mu)$ denotes a function of (u, v, μ) which vanishes to second or higher order at the origin. From Theorem 1 we obtain a local invariant manifold of (5.6)–(5.7) on which $v = h(u, \mu)$. Moreover $h(0, 0) = 0, dh(0, 0) = 0$, and we may assume that h is C^1.

It follows that $h(u, \mu)$ satisfies the functional equation

$$Nh(\alpha u + \beta h(u, \mu) + \mathcal{G}_1(u, h(u, \mu), \mu)) = \gamma(\mu)u + \delta h(u, \mu) + \mathcal{G}_2(u, h(u, \mu), \mu),$$

and if we set $x = h(0, \mu)$, then x satisfies the equation

$$Nh(\beta x + \mathcal{G}_1(0, x, \mu), \mu) = \delta(\mu)x + \mathcal{G}_2(0, x, \mu).$$

The latter is an algebraic equation for x which we denote $a(x, \mu) = 0$; moreover $a(0, \mu) = 0$ holds for all μ near 0, whence $h(0, \mu) \equiv 0$. In addition, a short calculation shows that

$$d_x a(0, \mu) = \delta(\mu) - Nd_u h(0, \mu)[\beta(\mu)],$$

which is an identity mapping when $\mu = 0$. The implicit function theorem now ensures that $x = 0$ is the only solution of $a(x, \mu) = 0$ for all μ near 0.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

The functional equation satisfied by $d_u h(u,\mu)$ is then

$$\begin{aligned}-Nd_u h(\alpha u+\beta h+\mathcal{G}_1(u,h,\mu))[\alpha+\beta d_u h+d_u\mathcal{G}_1+d_v\mathcal{G}_1\cdot d_u h]\\ +\gamma+\delta d_u h+d_u\mathcal{G}_2+d_v\mathcal{G}_2\cdot d_u h=0,\end{aligned}$$

where various arguments have been omitted for brevity. If we write τ for the linear map $d_u h(0,\mu)$, then

$$\gamma(\mu)+\delta(\mu)\tau=N\tau\cdot[\alpha(\mu)+\beta(\mu)\tau].$$

This is a Riccati equation for τ that can be solved near $\mu=0$ for τ as a function of μ using the implicit function theorem and the properties of α,β,γ, and δ. The result that $\tau(0)=0$ then follows because N is nilpotent and $\tau(0)=N\tau(0)[C]$.

We are now in a position to define the one-parameter family of matrices, denoted by $C(\mu)$ in the statement of the theorem, namely,

$$C(\mu):=\alpha(\mu)+\beta(\mu)\tau(\mu),$$

so that $C(0)=C$. If $C(\mu)$ has an eigenvalue λ, say, then

$$(\alpha+\beta\tau)w=\lambda w\implies\gamma+\delta\tau=N\tau\cdot[\lambda w],$$

whence

$$\begin{bmatrix}\alpha&\beta\\ \gamma&\delta\end{bmatrix}\begin{bmatrix}w\\ \tau w\end{bmatrix}=\lambda\begin{bmatrix}w\\ N\cdot\tau w\end{bmatrix},$$

and therefore

$$-\lambda\in\sigma\left(\begin{bmatrix}I&0\\ 0&N\end{bmatrix},\begin{bmatrix}\alpha&\beta\\ \gamma&\delta\end{bmatrix}\right)=\sigma(\mathcal{A},\mathcal{B}(\mu))=-\sigma(\mathcal{P}(\mu)).$$

We have deduced that

$$\sigma(\alpha(\mu)+\beta(\mu)\tau(\mu))\subseteq\sigma(\mathcal{P}(\mu)),$$

but the left-hand side of this inclusion has $\dim(K_1)$ elements counted according to algebraic multiplicity, whereas the right-hand side may have more unless, that is, $\mu=0$, in which case the inclusion is replaced by an equality because $\sigma(\alpha(0))=\sigma(C)=\sigma(\mathcal{P}(0))$.

As a result, if $\lambda(\mu)\in C([-\delta,\delta],\mathbb{C})$ is a continuous path of eigenvalues of $\mathcal{P}(\mu)$, then by virtue of the fact that $\lambda(0)\in\sigma(\alpha(\mu)+\beta(\mu)\tau(\mu)|_{\mu=0})$, it follows by the continuous dependence of eigenvalues on μ and by counting their location in the complex plane that $\lambda(\mu)\in\sigma(\alpha(\mu)+\beta(\mu)\tau(\mu)|_{\mu\neq0})$. $\square$

Conclusion 4 of Theorem 3 is not equivalent to saying that a locus of eigenvalues of $\mathcal{P}(\mu)$ is necessarily a locus of eigenvalues of $C(\mu)$. This is because $\mathcal{P}(\mu)$ may have other eigenvalues for small μ which become unbounded as μ tends to zero, and such curves cannot correspond to eigenvalues of $C(\mu)$. This happens, for instance, in the singularity-induced bifurcation theorem from [1] because an eigenvalue has a pole with respect to the bifurcation parameter.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

5.2.1. Existence of bifurcations for (1.1). One can easily prove a period-doubling bifurcation theorem for (1.1) without having recourse to the nonlinear normal form given in Theorem 2, but we now include this result for completeness.

THEOREM 4. *Suppose that* (5.2) *satisfies the following hypotheses:*

1. $1 \notin \sigma(\mathcal{P}(0))$ *and* $F(0,0,\mu) \equiv 0$,
2. $\ker(-A(0)+B(0)) = \langle k \rangle$, *so that* $-1 \in \sigma(\mathcal{P}(0))$, *and*
3. $B'(0)k \notin \mathrm{im}(-\mathrm{A}(0)+\mathrm{B}(0))$.

Then $\mu = 0$ *is a period-doubling bifurcation point for* (5.2) *from the trivial solution* $z = 0$.

Proof. Consider the algebraic equation $G(z,w,\mu) = 0$, where $G : \mathbb{R}^{2m+1} \to \mathbb{R}^{2m}$ is given by

$$G(z,w,\mu) := \left[\begin{array}{c} F(z,w,\mu) \\ F(w,z,\mu) \end{array} \right],$$

and moreover G has the trivial solution branch. Also define $\overline{G}(z,\mu) := F(z,z,\mu)$, so that

$$d_{(z,w)}G(0,0,\mu) = \left[\begin{array}{cc} A(\mu) & B(\mu) \\ B(\mu) & A(\mu) \end{array} \right],$$

and $d_z\overline{G}(0,\mu) = A(\mu)+B(\mu)$. By assumption, $A(0)+B(0)$ is invertible, and therefore $d_z\overline{G}(0,\mu)$ is an invertible map for small $|\mu|$, so that if $G(z,w,\mu) = 0$ near $\mu = 0$, then $z \neq w$ unless $z = w = 0$. The theorem now follows from the simple eigenvalue bifurcation theorem applied to G at $\mu = 0$, noting that the kernel of $d_{(z,w)}G(0,0,0)$ is $(k,-k)^T$. ◻

One can of course formulate a similarly straightforward fold bifurcation for (1.1) in an entirely analogous fashion. However, the following theorem relies on Theorem 3 in a nontrivial way.

THEOREM 5 (Neimark–Sacker bifurcation). *Suppose that* (5.2) *has the fixed point* $z = 0$ *for all* $\mu \in \mathbb{R}$ *and that* $\lambda(\mu) \in \sigma(\mathcal{P}(\mu))$ *is a curve which satisfies the following:*

1. $\mathcal{P}(0)$ *is a regular matrix pencil;*
2. $|\lambda(0)| = 1$, *and* $\lambda(0)$ *is an algebraically simply eigenvalue of* $\mathcal{P}(0)$*;*
3. $\lambda(0)^n \neq 1$ *for* $n \in \{1,2,3,4\}$*;*
4. $\frac{d}{d\mu}|\lambda(\mu)|\big|_{\mu=0} \neq 0$.

Then modulo a further nonresonance condition[1] *there is a half-interval* $J \subset \mathbb{R}$ *containing* 0 *in its closure such that* (5.2) *possesses a quasi-invariant circle* $\Gamma_\mu \subset \mathbb{R}^{2m}$ *for all* $\mu \in J$. *Moreover, if* $\mathrm{diam}(\Gamma_\mu) = \sup\{\|z-w\| : z,w \in \Gamma_\mu\}$, *then* $\lim_{\mu\to 0} \mathrm{diam}(\Gamma_\mu) = 0$.

Proof. Theorem 5 follows immediately from the Neimark–Sacker bifurcation for maps applied to $\varphi(u,\mu)$ from Theorem 3 (part 3). ◻

6. Examples. *Example* 1 (output-nulling control problem). The results in this paper give sufficient conditions for a positive answer to the following question:

(Q) *Given* $f(= f(x,u)) : \mathbb{R}^{n+m} \to \mathbb{R}^n, g(= g(x)) : \mathbb{R}^n \to \mathbb{R}^m$ *such that* $f(0,0) = 0, g(0) = 0$, *does there exist a sequence of states* (x_n) *given by the iterates of* f *and controls* (u_n) *such that* $g(x_n) \equiv 0$ *and* $x_n \to 0$ *as* $n \to \infty$*?*

Thus, we seek a global orbit of the ΔAE

$$x_{n+1} = f(x_n,u_n),\ g(x_n) = 0.$$

[1]See [21] or [25, p. 376], where the open condition "$a \neq 0$" is given, and this requires the computation of third-order terms in the normal form for this bifurcation.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

A necessary condition for the existence of such a solution can be obtained by substituting the dynamic part of the problem into the constraint, to give the hidden constraint

$$g(f(x_n, u_n)) = 0 \qquad (\forall n \geq 1).$$

Hence, provided the function $g(f(x,u))$ has an invertible partial u-derivative at $(x,u) = (0,0)$, by the stable manifold theorem the response to (**Q**) is affirmative if the spectrum of the x-derivative of $f(x,U(x))$, also evaluated at $(x,u) = (0,0)$, contains an element of the open unit disk. Here $U(x)$ denotes the solution of the equation $g(f(x,U)) = 0$ given locally by the implicit function theorem. Note for a moment that if the stated u-derivative $dg(0)d_uf(0,0)$ is invertible, it follows that the matrix pencil

$$(A,B) := \left(\begin{pmatrix} I_x & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} d_xf & d_uf \\ dg & 0 \end{pmatrix} \right)\Bigg|_{(x,u)=(0,0)}$$

is regular, and thus (A1)–(A3) hold for this problem.

However, by using Theorem 1 one can dispense with the condition that $dg(0)d_uf(0, 0)$ is invertible. In fact, let us assume that $K := \ker(dg(0)d_uf(0,0)) \neq \{0\}$. In this case, the matrix $\lambda A + B$ is invertible if $-\lambda \notin \sigma(d_xf(0,0))$ and $dg(\lambda I_x + d_xf)^{-1}d_uf$ is also invertible at $(x,u) = (0,0)$. However, for λ large we have

$$\begin{aligned} \lambda^2 dg(\lambda I_x + d_xf)^{-1}d_uf &= \lambda dg(I_x + \lambda^{-1}d_xf)^{-1}d_uf \\ &= \lambda dg(I_x - \lambda^{-1}d_xf + O(\lambda^{-2}))d_uf \\ &= \lambda\ dg \cdot d_uf - dg \cdot d_xf \cdot d_uf + O(\lambda^{-1}), \end{aligned}$$

evaluating all of the stated derivatives at $(x,u) = (0,0)$. As a result, if the weaker condition holds that the pencil $(dg \cdot d_uf, dg \cdot d_xf \cdot d_uf)$ is regular, then (A,B) is regular and (A1)–(A3) still apply. The response to (Q) is again affirmative if $\sigma(A,B)$ contains at least one member of the open unit disk.

In fact one can show that output-nulling control problems are well-posed in sequence spaces as follows. First consider the linear problem

$$Az_{n+1} + Bz_n = \Gamma_n, \tag{6.1}$$

where $n \in \mathbb{N}$ and (Γ_n) is a given sequence in

$$\ell^\infty_{\mathbb{N}}(\mathbb{R}^m) = \left\{ (z_n)_{n\in\mathbb{N}} : z_n \in \mathbb{R}^m, \sup_{n\in\mathbb{N}} \|z_n\|_{\mathbb{R}^m} < \infty \right\},$$

but $\det A = 0$. If (A,B) is a regular matrix pencil with index ν, the KNF allows us to write (6.1) in the form

$$u_{n+1} = Cu_n + \alpha_n, \tag{6.2}$$

$$Nv_{n+1} = v_n + \beta_n, \tag{6.3}$$

where $(u_n, v_n) \in \mathbb{R}^{p+q}$. In order to solve (6.1), let us consider the linear operator $I - N\sigma$ on a space of sequences $\ell^\infty_{\mathbb{N}}(\mathbb{R}^q)$. We take linear maps $T \in BL(\mathbb{R}^q)$ to act pointwise on $\ell^\infty_{\mathbb{N}}(\mathbb{R}^q)$, so

$$T(w_n)_{n\in\mathbb{N}} = (Tw_n)_{n\in\mathbb{N}} \qquad (\forall (w_n)_{n\in\mathbb{N}} \in \ell^\infty_{\mathbb{N}}(\mathbb{R}^q)),$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

and we define σ the forward-shift map by

$$\sigma(w_n)_{n\in\mathbb{N}} = (w_{n+1})_{n\in\mathbb{N}} \qquad (\forall (w_n)_{n\in\mathbb{N}} \in \ell^\infty_{\mathbb{N}}(\mathbb{R}^q)).$$

Any such T will commute with σ, and one can see by a direct multiplication that

$$(I - N\sigma)\sum_{i=0}^{\nu-1} N^i\sigma^i = I,$$

where I denotes the identity on $\ell^\infty_{\mathbb{N}}(\mathbb{R}^q)$.

One can solve (6.2) in a suitably weighted sequence space if no restrictions are to be placed on the spectrum of (A, B). Equation (6.3) can also be solved:

$$\mathbf{v} = -(I - N\sigma)^{-1}\beta = -\sum_{i=0}^{\nu-1} N^i\sigma^i\beta,$$

where $\mathbf{v} = (v_i)_{i\in\mathbb{N}}$ and $\beta = (\beta_i)_{i\in\mathbb{N}}$, whence

$$v_n = -\sum_{i=0}^{\nu-1} N^i\beta_{n+i}.$$

So from a temporal point of view the current values of the state depend on future values of the input, but (6.1) is still well-posed in a sequence space as $\mathbf{v}$ depends continuously on β.

This means that a second- or higher-order nonlinear perturbation of (6.1),

$$Az_{n+1} + Bz_n + \mathcal{F}(z_n) = \Gamma_n, \tag{6.4}$$

say, where $\mathcal{F}(0) = 0, d\mathcal{F}(0) = 0$, can be written as an infinite-dimensional problem

$$(A\sigma + B)\mathbf{z} + \mathcal{F}(\mathbf{z}) = \mathbf{\Gamma} \tag{6.5}$$

in $\ell^\infty_{\mathbb{N}}(\mathbb{R}^m)$, and one can apply the inverse function theorem to solve locally for small-norm solutions of the form $\mathbf{z} = \mathbf{z}(\mathbf{\Gamma})$, where $\mathbf{z}(\mathbf{0}) = \mathbf{0}$.

This solution can be found via the Picard iteration $\mathbf{z}(\mathbf{\Gamma}) = \lim_{n\to\infty} \mathbf{y}^{(n)}$, where $\mathbf{y}^{(0)} = \mathbf{0}$ and

$$\mathbf{y}^{(n+1)} = -(A\sigma + B)^{-1}[\mathcal{F}(\mathbf{y}^{(n)}) - \mathbf{\Gamma}].$$

As a result, writing the solution sequence $\mathbf{z}(\mathbf{\Gamma})$ as $(z_n(\mathbf{\Gamma}))_{n\in\mathbb{N}}$, it is clear that the nonlinear perturbation will have the effect of making each z_n depend on infinitely many elements of the sequence $\mathbf{\Gamma}$, unless ν happens to equal 1.

This effect has been observed before in the literature in the context of delay DAEs [5, 13], where it is noted in the former reference that linear systems of delay DAEs can act like advanced systems when their index is two or higher. The problem (6.5) is displaying exactly this behavior.

Example 2. This example serves to illustrate how we can use Theorem 1 to deduce qualitative similarities between a DAE and its discrete counterpart. Take the DAE

$$\dot{x} = f(x, y), \tag{6.6}$$

$$g(x, y) = 0, \tag{6.7}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

subject to $x \in \mathbb{R}^n, y \in \mathbb{R}^m$ with an equilibrium at the origin, so that $f(0,0) = 0, g(0,0) = 0$. Let us impose a singularity of the form

$$\ker(d_y g(0,0)) = \langle k \rangle \quad \text{such that} \quad d_y f(0,0) d_x g(0,0) k \notin \operatorname{im}(d_y g(0,0)), \tag{6.8}$$

and the equilibrium solution is isolated:

$$\det \begin{pmatrix} d_x f(0,0) & d_y f(0,0) \\ d_x g(0,0) & d_y g(0,0) \end{pmatrix} \neq 0.$$

From [2] it is known that (6.6)–(6.7) has an invariant manifold W of dimension $n-1$ that contains the origin and intersects the singularity in an $n-2$-dimensional manifold of pseudoequilibria.

Now consider the forward-Euler method in state-space form [8, p. 375] applied to (6.6)–(6.7), resulting in the difference equation

$$x_{i+1} - x_i = h f(x_i, y_i), \tag{6.9}$$

$$g(x_{i+1}, y_{i+1}) = 0. \tag{6.10}$$

Using Theorem 1, in order to to show that (6.9)–(6.10) possesses a quasi-invariant manifold of solutions W_h that contains the origin and has dimension $n-1$, we need only show that (A1)–(A3) hold, which entails showing that the derivative at the origin of (6.9)–(6.10) is a regular matrix pencil. Hence we seek a $\xi \in \mathbb{C}$ such that

$$\det \left(\xi \begin{bmatrix} I & 0 \\ d_x g & d_y g \end{bmatrix} + \begin{bmatrix} I + h d_x f & h d_y f \\ 0 & 0 \end{bmatrix} \right) \Bigg|_{(x,y)=(0,0)} \neq 0.$$

However, the conditions in (6.8) ensure the existence of such a ξ for any fixed $h \neq 0$, and the existence of W_h follows. The dimension of W_h comes from counting the number of finite eigenvalues of the linearization of (6.9)–(6.10) which is given in [1] as $n-1$.

Example 3. Consider again (6.6)–(6.7) but now with a parameter α

$$\dot{x} = f(x, y, \alpha), g(x, y, \alpha) = 0, \tag{6.11}$$

and suppose that $(x, y) = (0, 0)$ is an equilibrium locus for all $\alpha \in \mathbb{R}$. Now suppose that the conditions for a Hopf bifurcation are formally satisfied by (6.11) at $\alpha = \alpha_0$: $\omega(\alpha) \in \sigma(M, -L(\alpha))$ is a locus of algebraically simple eigenvalues, where

$$M = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}, \ L(\alpha) = \begin{bmatrix} d_x f(0,0,\alpha) & d_y f(0,0,\alpha) \\ d_x g(0,0,\alpha) & d_y g(0,0,\alpha) \end{bmatrix};$$

and $\omega(\alpha_0) = i\omega_0$ is an eigenvalue of $(M, -L(\alpha))$ such that $\frac{d}{d\alpha} \Re(\omega(\alpha))\big|_{\alpha=\alpha_0}$ is nonzero. Also, no other eigenvalues of the regular matrix pencil $(M, -L(\alpha_0))$ have a zero real part.

Then, modulo an open condition on the third-order derivatives of the smooth functions f and g, we can show that the backward-Euler method

$$x_{i+1} - x_i = h f(x_{i+1}, y_{i+1}), \tag{6.12}$$

$$-h \cdot g(x_{i+1}, y_{i+1}) = 0, \tag{6.13}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

satisfies the conditions of the Neimark–Sacker bifurcation theorem for all sufficiently small $h > 0$. Note that $(M - hL(\alpha), -M)$ is the linearization of (6.12)–(6.13) about the zero fixed point and is a regular matrix pencil for $h > 0$ and $\alpha \approx \alpha_0$ such that

$$(1 - h\omega(\alpha))^{-1} \in \sigma(M - hL(\alpha), -M) \qquad (\forall \alpha).$$

In order to apply Theorem 5 to (6.12)–(6.13) we first note that the eigenvalue locus $(1 - h\omega(\alpha))^{-1}$ has unit length if and only if $1 - h\omega(\alpha)$ has unit length. If we define functions $R(\alpha)$ and $I(\alpha)$ by $\omega(\alpha) = R(\alpha) + iI(\alpha)$, then $|1 - h\omega(\alpha)| = 1$ if and only if $(1 - hR(\alpha))^2 + h^2 I(\alpha)^2 = 1$, which holds when

$$h\left(-R(\alpha) + \frac{h}{2}(R(\alpha)^2 + I(\alpha)^2)\right) = 0. \tag{6.14}$$

As a result, we define the function $b(\alpha, h) := -R(\alpha) + \frac{h}{2}(R(\alpha)^2 + I(\alpha)^2)$ and note that $b(\alpha, h) = 0$ for $h > 0$ ensures that the linearization of (6.12)–(6.13) at $(0, 0)$ has an eigenvalue of unit modulus. Now $b(\alpha_0, 0) = 0$ and $\frac{\partial b}{\partial \alpha}(\alpha_0, 0) = -R'(\alpha_0) \neq 0$ by assumption, and, as a result, one may solve $b(\alpha, h) = 0$ locally using the implicit function theorem for $\alpha = \alpha(h)$ such that $\alpha(0) = \alpha_0$.

From this calculation one can show that the numerical scheme (6.12)–(6.13) has a quasi-invariant circle for α in some half-neighborhood of $\alpha(h) = \alpha_0 + \frac{h\omega_0^2}{R'(\alpha)} + O(h^2)$ provided $h > 0$ is sufficiently small.

Note that no assumption is made regarding the invertibility of $d_y g(0, 0, 0)$. If this mapping were invertible in addition to the formal conditions given above for Hopf bifurcation, then the existence of a locus of periodic solutions of (6.11) could be deduced. However, without the invertibility of $d_y g(0, 0, 0)$, it is not known whether a Hopf bifurcation occurs in (6.11), but (6.12)–(6.13) has a locus of invariant circles nevertheless.

6.1. Concluding remark. The results in this paper can be used to show that second-order problems of the form

$$F(z_n, z_{n+1}, z_{n+2}) = 0 \tag{6.15}$$

have quasi-invariant manifolds provided that the appropriate matrix pencil is regular simply by rewriting (6.15) as a first-order problem. However, the methods of this paper do not easily extend to the study of invariant manifolds of the system that one would like to study if (1.1) had a period-2 orbit $(z, w, z, w, \ldots)$, namely, the system

$$F(z_n, z_{n+1}) = 0, \tag{6.16}$$

$$F(z_{n+1}, z_{n+2}) = 0, \tag{6.17}$$

where $F(z, w) = 0$ and $F(w, z) = 0$, with $w \neq z$. Another approach is needed to answer the question of whether overdetermined systems of this type have any invariant manifolds associated with them.

REFERENCES

[1] R. E. Beardmore, *The singularity-induced bifurcation and its Kronecker normal form*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 126–137.

[2] R. E. Beardmore and R. Laister, *The flow of a DAE near a singular equilibrium*, SIAM J. Matrix Anal. Appl., 24 (2002), pp. 106–120.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[3] K. E. Brenan, S. L. Campbell, and L. R. Petzold, *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*, North-Holland, Amsterdam, 1989.
[4] R. Bru, C. Coll, and E. Sanchez, *Structural properties of positive linear time-invariant difference-algebraic equations*, Linear Algebra Appl., 349 (2002), pp. 1–10.
[5] S. L. Campbell, *Singular linear systems of differential equations with delays*, Appl. Anal., 11 (1980), pp. 129–136.
[6] T. Fliegner, Ü. Kotta, and H. Nijmeijer, *Solvability and right-inversion of implicit nonlinear discrete-time systems*, SIAM J. Control Optim., 34 (1996), pp. 2092–2115.
[7] F. R. Gantmacher, *Theory of Matrices*, Vol. 2, Chelsea, New York, 1977.
[8] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations*, 2nd revised ed., Springer Ser. Comput. Math. II, Springer, New York, 2002.
[9] T. Kaczynski and W. Krawcewicz, *A local Hopf bifurcation theorem for a certain class of implicit differential equations*, Canad. Math. Bull., 36 (1993), pp. 183–189.
[10] P. E. Kloeden and P. Marín-Rubio, *Weak pullback attractors of non-autonomous difference inclusions*, J. Difference Equ. Appl., 9 (2003), pp. 489–502.
[11] J. Laurent-Varin, F. Bonnans, N. Berend, C. Talbot, and M. Haddou, *On the refinement of discretization for optimal control problems*, in Proceedings of the 16th IFAC Symposium on Automatic Control in Aerospace, St. Petersburg, 2004.
[12] J.-Y. Lin and Z.-H. Yang, *A discrete optimal control for descriptor systems*, IEEE Trans. Automat. Control, 34 (1989), pp. 177–181.
[13] H. Logeman, *Destabilizing effects of small time delays on feedback-controlled descriptor systems*, Linear Algebra Appl., (1998), pp. 131–153.
[14] C. Navasca, *Local stable manifold for the bidirectional discrete-time dynamics*, SIAM J. Control Optim., submitted.
[15] P. J. Rabier, *The Hopf bifurcation theorem for quasilinear differential-algebraic equations*, Comput. Methods Appl. Mech. Engrg., 170 (1999), pp. 355–371.
[16] P. J. Rabier and W. C. Rheinboldt, *On impasse points of quasilinear differential-algebraic equations*, Math. Anal. Appl., 181 (1994), pp. 429–454.
[17] P. J. Rabier and W. C. Rheinboldt, *Theoretical and Numerical Analysis of Differential-Algebraic Equations*, Handbook of Numerical Analysis, Vol. VIII, Elsevier Science, New York, 2003, Part 4, pp. 183–540.
[18] S. Reich, *On the qualitative behaviour of DAEs*, Circuits Systems Signal Proc., 14 (1995), pp. 427–443.
[19] G. Reißig and H. Boche, *On singularities of autonomous implicit ordinary differential equations*, IEEE CAS I, 50 (2003), pp. 922–931.
[20] R. Riaza and P. J. Zufiria, *Stability of singular equilibria in quasilinear implicit differential equations*, J. Differential Equations, 171 (2001), pp. 24–53.
[21] R. J. Sacker, *On Invariant Surfaces and Bifurcation of Periodic Solutions of Ordinary Differential Equations*, Ph.D. thesis, New York University, Courant Institute of Mathematical Sciences, 1964.
[22] G. Söderlind, *Remarks on the stability of high-index DAEs with respect to parametric perturbations*, Computing, 49 (1992), pp. 303–314.
[23] J. Sotomayor and M. Zhitomirskii, *Impasse singularities of differential systems of the form* $A(x)\dot{x} = f(x)$, J. Differential Equations, 169 (2001), pp. 567–587.
[24] V. Venkatasubramanian, *Singularity induced bifurcation in the Van Der Pol oscillator*, IEEE Trans. Circuits Syst. I Fund. Theory Appl., 41 (1994), pp. 765–769.
[25] S. Wiggins, *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, Texts Appl. Math., Springer, New York, 1990.
[26] E. Zeidler, *Nonlinear Analysis: Linear Monotone Operators*, Vol. II, Springer, New York, 1990.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.