SIAM J. SCI. COMPUT.
Vol. 15, No. 4, pp. 959–976, July 1994

A SHADOWING LEMMA APPROACH TO GLOBAL ERROR ANALYSIS FOR INITIAL VALUE ODES*

SHUI-NEE CHOW† AND ERIK S. VAN VLECK‡

Abstract. The authors show that for dynamical systems that possess a type of piecewise hyperbolicity in which there is no decrease in the number of stable modes, the global error in a numerical approximation may be obtained as a reasonable magnification of the local error. In particular, under certain conditions the authors prove the existence of a trajectory on an infinite time interval of the given ordinary differential equation uniformly close to a given numerically computed orbit of the same differential equation by allowing for different initial conditions. For finite time intervals a general result is proved for obtaining a posteriori bounds on the global error based on computable quantities and on finding and bounding the norm of a right inverse of a particular matrix. Two methods for finding and bounding/estimating the norm of a right inverse are considered. One method is based upon the choice of the pseudo or generalized inverse. The other method is based upon solving multipoint boundary value problems (BVPs) with the choice of boundary conditions motivated by the piecewise hyperbolicity concept. Numerical results are presented for the logistic equation, the forced pendulum equation, and the space discretized Chafee–Infante equation.

Key words. numerical initial value ODEs, global error analysis, piecewise hyperbolicity

AMS subject classification. 65L

1. Introduction.

In this paper we consider initial value ordinary differential equations and their discretization. We investigate both theoretically and numerically the global error in computing numerical approximations. It is well known (see [Ge], [S1]) that the global error between the numerical approximation and the actual trajectory with the same initial condition may become large. In fact, for a numerical method of order p with a fixed stepsize of h and a differential equation with Lipschitz constant L, classical error estimates are of the form $\exp(Lt)h^p$ at time t. On the other hand, local error control provides a tight bound in many instances. For example, the local error is a good estimate of the global error for many stiff initial value problems in which one-sided Lipschitz constants appear. Our contribution is to show that global errors stay reasonably bounded for the wider class of initial value problems that are piecewise hyperbolic with no decrease in the number of stable modes where in fact the number of stable modes may increase. These are not stable initial value problems in the classical sense, and the global error may not be of the order of the local error. We will show that if the initial condition for the discrete approximation is allowed to differ from that of the continuous trajectory, then for a large class of problems the global error may be represented as a reasonable magnification of the local error. This is important when one is employing numerical simulations to study qualitative as well as quantitative features of dynamical systems. In this paper we reserve the word orbit to denote a discrete sequence of points and the word trajectory to denote a continuous function of time.

We consider the class of problems in which there is no decrease in the number of stable modes of the linear variational equation along solution paths. This is reminiscent of the case in which there are several hyperbolic fixed points (i.e., saddle points) and a trajectory that passes near these fixed points in which the dimension of the stable manifold of these fixed points is not decreasing. This situation arises in certain space discretized parabolic partial differential equations that occur as models of chemical, biological, and other physical systems.

*Received by the editors September 14, 1992; accepted for publication (in revised form) May 25, 1993.

†School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332 (chow@math.gatech.edu). The work of this author was supported in part under National Science Foundation grant DMS-9005420.

‡Department of Mathematics and Statistics, Simon Fraser University, Burnaby, British Columbia V5A 1S6 Canada. Current address, Department of Mathematical and Computer Sciences, Colorado School of Mines, Golden, Colorado 80401 (erikvv@lyapunov.mines.colorado.edu). The work of this author was supported in part under Natural Sciences and Engineering Research Council of Canada grant OGP0121873.

Our method is based upon showing that there exists a nearby trajectory in which there is no local error and is somewhat similar to defect correction (see [S2]). Our approach is different from defect correction in that we do not attempt to provide a correction, but instead estimate the magnification of the local error that gives the global error. In [Be1] it is shown that in a neighborhood of a single hyperbolic fixed point, the discrete stable and unstable manifolds of the map defined by the numerical method converge to the stable and unstable manifolds of the fixed point of the continuous problem. Aspects of backward error analysis for initial value ordinary differential equations are studied in [E2] and under certain conditions it is shown that even on an infinite time interval there exists a nearby equation that is solved exactly by the numerical result.

The idea of showing that near an orbit with a small local error there exists an orbit with no local error is formalized in the dynamical systems community in terms of the shadowing lemma. Results in this direction were first given by Anosov [A] and Bowen [Bo] for uniformly hyperbolic maps on a differential manifold. These results were generalized, and recently an analytic proof of the shadowing lemma has been given in [CLP] under the assumption of exponential dichotomy. The infinite time result presented in this paper is proven under the assumption of piecewise exponential dichotomy in which the rank of the projection onto the stable subspace is, under certain assumptions, allowed to increase with time.

Numerical methods for computing the global error for maps, where the local error is comprised solely of roundoff error (as opposed to roundoff error and discretization error when one is solving differential equations numerically), were initially given in [HYG1] and [HYG2] for the logistic map and the Henon map. These mappings are not uniformly hyperbolic in the sense of Anosov and Bowen, but are on average hyperbolic. The methods used in [HYG1] and [HYG2] are based upon interval arithmetic to provide a sequence of intervals containing both the orbit with the small local error and the orbit with no local error. Other numerical methods for shadowing of maps have been given in [CP1], [CP2], and [CVV2]. In [SY] a new proof of the shadowing lemma is given and numerical methods are presented that apply to both mappings and ordinary differential equations. The methods in [SY] are based upon performing Newton's method to find a zero of a certain function. They answer a slightly different question than we do, by showing that there exists a noisy discrete approximation with some unknown initial condition near a trajectory of the same problem with a given initial condition. The methods in [SY] provide a rigorous verification that there exists a nearby trajectory and rigorous bounds on the distance from the trajectory to the discrete approximation using Taylor series methods to integrate numerically. Taylor series methods are used to obtain explicit bounds on the local errors although the methods in [SY] may be applied with only estimates of the local errors. The methods presented here do not provide rigorous bounds, but instead provide estimates using existing initial value software and local error estimates provided by the initial value problem (IVP) software. Our purpose is to derive numerical methods for obtaining a posteriori global error estimates that are compatible with existing numerical integration software.

In §2 we present a notion of piecewise hyperbolicity due to Pliss [Pl1] and present results that give sufficient conditions for the existence of a trajectory on the positive real line uniformly close to a discrete numerical approximation when the linear variational equation has this type of piecewise hyperbolicity. We show that there exists a trajectory nearby by showing that there exists a zero of a certain mapping, F. Under certain assumptions, Newton's method will converge to a zero of F given a numerically computed orbit as an initial guess. We do not actually perform Newton's method, but find a bound on the norm of a right inverse of the linearized function DF to prove the existence of a zero of F in a neighborhood of our initial guess. The main result of this section (Theorem 2.3) is essentially a shadowing lemma for the case in which the linear variational equation is piecewise hyperbolic in the sense of Pliss.

This includes the case in which the linear variational equation is exponentially dichotomic. In §3 we state a result for a numerically approximating orbit of finite length and give a simple proof in terms of quantities that are numerically computable. The result is based upon having a right inverse for an approximation of DF and a bound on the norm of this right inverse. The challenge to obtain accurate estimates of the global error is to minimize the norm of the right inverse over the set of all right inverses. Section 4 is devoted to developing numerical methods to estimate quantities necessary to apply the result in §3. Most of our efforts are in finding a suitable right inverse and in providing a bound or estimate on the norm of this right inverse. Two methods are developed. One is based on the choice of the pseudo or generalized inverse as our right inverse, and the second method is based on ideas related to the well conditioning of multipoint BVPs. These multipoint BVPs correspond to right inverses and, motivated by the concept of piecewise hyperbolicity the interior boundary conditions, are chosen to occur at points in which there is an increase in the number of stable modes. To provide estimates of the global error using existing numerical ordinary differential equation (ODE) software, all of our global error estimates are in terms of the supremum norm. Numerical examples are presented in §5. Our methods are applied to the logistic equation, the forced pendulum equation, and the space-discretized Chafee–Infante equation. Conclusions and references are presented in §§6 and 7, respectively.

2. Theoretical aspects. Throughout this paper we consider both sequences and continuous functions. We reserve the notation $\tilde{x}(t)$ for functions and the notation $x := \{x_n\}$ for sequences. Given a function $\tilde{x}(t)$ defined on some possibly infinite real interval and a sequence $\{t_n\}$ with values in this interval, we will write the restriction of $\tilde{x}(t)$ to $\{t_n\}$ as $x := \{x_n\}$ where $x_n := x(t_n)$ for all n. Unless otherwise stated $||\tilde{x}|| = \sup_t ||\tilde{x}(t)||$ and $||x|| = \sup_n ||x_n||$. In this section $||y||$ denotes the Euclidean norm for $y \in \mathbb{R}^N$.

Consider the initial value problem

$$\begin{aligned} \dot{\tilde{x}} &= f(\tilde{x}, t), \\ \tilde{x}(t_0) &= x_0, \end{aligned} \tag{2.1}$$

where $t_0 \in \mathbb{R}$, $\tilde{x}(t) \in \mathbb{R}^N$, $\dot{\tilde{x}} = \frac{d\tilde{x}}{dt}$, and $f \in C^k(\mathbb{R}^N, \mathbb{R}; \mathbb{R}^N)$ for some $k \geq 2$. Let $\phi : \mathbb{R}^N \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^N$ be the associated solution operator so that $\phi(x_0, t_0, t_0) = x_0$ and $\phi(x_0, t_0, t)$ is the solution at time t with initial condition x_0 at t_0. To solve this equation numerically, we consider one-step methods of the form

$$\begin{aligned} x_{n+1} &= x_n + h_n \Theta(f, x_n, t_n, h_n), \\ x_n &\qquad \text{given} \end{aligned} \tag{2.2}$$

to advance the solution from t_n to $t_{n+1} := t_n + h_n$.

Given an orbit $x := \{x_n\}_0^\infty$ produced by a one-step method, we define a corresponding piecewise discontinuous function that is double valued at t_n, $n = 1, \ldots, \infty$, called a *pseudo solution* $\{\tilde{x}_n(t)\}_0^\infty$ as

$$\tilde{x}_n(t) = \phi(x_n, t_n, t) \quad \text{for } t_n \leq t \leq t_{n+1} \tag{2.3}$$

$n = 0, \ldots, \infty$ so that $\tilde{x}_n(t_n) = x_n$. Let $\delta_n = \tilde{x}_{n+1}(t_{n+1}) - \tilde{x}_n(t_{n+1})$, $n = 0, \ldots, \infty$ denote the *local error* at the nth iterate, and let $\tilde{x}(t) = \tilde{x}_n(t)$ for $t_n \leq t < t_{n+1}$ so that $\tilde{x}(t)$ is well defined.

Let $l^\infty(\mathbb{N})$ denote the sequences $w = \{w_n\}_0^\infty$ with $w_n \in \mathbb{R}^N$ for all n and $\sup_n ||w_n|| < \infty$. Consider the operator $F : l^\infty(\mathbb{N}) \to l^\infty(\mathbb{N})$, where the nth iterate $(F(w))_n$ is defined for any

$w \in l^\infty(\mathbb{N})$ to be

$$(F(w))_n = w_{n+1} - \phi(w_n, t_n, t_{n+1}) \quad \text{for } n = 0, \ldots, \infty \tag{2.4}$$

so that F measures the local error at each iterate. We wish to find a solution $w \in l^\infty(\mathbb{N})$ of $F(w) = 0$, i.e., a solution of the original IVP (2.1).

Consider the first variation $DF(x) : l^\infty(\mathbb{N}) \to l^\infty(\mathbb{N})$ of $F(x)$ defined by

$$(DF(x)u)_n = u_{n+1} - \phi_x(x_n, t_n, t_{n+1})u_n, \tag{2.5}$$

where $\phi_x := \partial\phi/\partial x_n$.

Given a pseudo solution $\tilde{x}(t)$, we construct (as in [Pa]) a corresponding continuous function $\tilde{z}(t)$ defined by

$$\tilde{z}(t) = \begin{cases} \tilde{x}(t) + (t_{n+1} - t_n)^{-1}(t - s_n)\delta_n, & t_n \le t \le s_n, \\ \tilde{x}(t) + (t_{n+1} - t_n)^{-1}(t - s_n)\delta_{n+1}, & s_n \le t \le t_{n+1}, \end{cases} \tag{2.6}$$

where $s_n = (t_{n+1} + t_n)/2$ so that $\delta = \sup_n \delta_n$ implies $||\tilde{x} - \tilde{z}|| \le \frac{\delta}{2}$.

Let Φ denote the principal matrix solution for the linear variational equation about $\tilde{z}(t)$ so that

$$\partial_t \Phi(t, \tau) = Df(\tilde{z}(t), t)\Phi(t, \tau), \qquad \Phi(\tau, \tau) = I \tag{2.7}$$

for $t \ge \tau$ where $Df := \frac{\partial f}{\partial \tilde{z}}$.

DEFINITION 1. *For positive constants K, λ, the system (2.7) is said to be (K, λ)-hyperbolic on the interval $[a, b]$ if for given $s \in [a, b]$ there exists linear subspaces $S(s)$ and $U(s)$ of dimension k and $N - k$, respectively, such that, if $y_0 \in S(s)$,*

$$||\Phi(t, a)\Phi^{-1}(s, a)y_0|| \le Ke^{-\lambda(t-s)}||y_0||$$

for $t \ge s$ and $s, t \in [a, b]$, and, if $y_0 \in U(s)$,

$$||\Phi(t, a)\Phi^{-1}(s, a)y_0|| \le Ke^{-\lambda(s-t)}||y_0||$$

for $s \ge t$ and $s, t \in [a, b]$.

This is identical to the definition of exponential dichotomy on an interval that may be found in [Co] and [AMR], where $S(s)$ and $U(s)$ represent the decaying and growing solution components, respectively.

Given subspaces L, M in $\mathbb{R}^N$ we say that these subspaces intersect transversally if

$$\dim L + \dim M = \dim(L \cap M) + N.$$

Define $\angle(L, M)$, the angle between subspaces L, M, where $0 \le \angle(L, M) \le \pi/2$ by

$$\cos(\angle(L, M)) = \max_{||y_1||=||y_2||=1} |y_1^T y_2|,$$

where $y_1 \in L$, $y_2 \in M$ and $y_i \perp L \cap M$ for $i = 1, 2$.

It is assumed that (2.7) satisfies the following three hypotheses.

(H1) There is a mesh made up of what we will call the switching points, $\beta_0 < \beta_1 < \cdots < \beta_m < \beta_{m+1}$, where $\beta_0 = t_0$ and $\beta_{m+1} = +\infty$ such that (2.7) is (K, λ)-hyperbolic with decaying/growing solution spaces $S_j(t)$ and $U_j(t)$ on each of the intervals $[\beta_j, \beta_{j+1}]$, for $j = 0, \ldots, m$.

(H2) The inequalities

$$\dim U_j(\beta_{j+1}) > \dim U_{j+1}(\beta_{j+1})$$

are satisfied for $j = 0, \ldots, m-1$, and the subspaces $U_j(\beta_{j+1})$ and $S_{j+1}(\beta_{j+1})$ intersect transversally for $j = 0, \ldots, m-1$.

(H3) There is an $\alpha > 0$ such that

$$\angle(U_j(\beta_{j+1}), S_{j+1}(\beta_{j+1})) > \alpha$$

for $j = 0, \ldots, m-1$.

Remarks.

(i) Note that (H2) implies that necessarily $m \leq N$; i.e, the number of switching points is less than or equal to the dimension of the problem.

(ii) It is shown in [Pl2] that (H1)–(H3) are satisfied for differential equations of the form (2.1) that are periodic in t, hyperbolic on the nonwandering set and satisfy the strict transversality condition (see [Ro1] and [Ro2]).

(iii) The switching points denote points in time where there is a decrease in the number of unstable modes.

Consider now the inhomogeneous linear equation

$$\dot{\tilde{u}}(t) = Df(\tilde{z}(t), t)\tilde{u}(t) + \tilde{g}(t). \tag{2.8}$$

The following theorem shows that (H1)–(H3) are sufficient to imply the existence of uniformly bounded solutions of (2.8).

THEOREM 2.1 ([Pl1]). *There exist constants $T(K, \lambda, \alpha)$ and $\eta(K, \lambda, \alpha)$ such that if* (2.7) *satisfies* (H1)–(H3) *for some switching points $\{\beta_j\}_0^{m+1}$, with*

$$\beta_{j+1} - \beta_j > T(K, \lambda, \alpha)$$

for $j = 1, \ldots, m-1$, then for any continuous function $\tilde{g}(t)$ with

$$||\tilde{g}(t)|| \leq \eta(K, \lambda, \alpha),$$

the system (2.8) *has a solution $\tilde{u}(t)$ satisfying*

$$||\tilde{u}(t)|| \leq 1 \quad \textit{for all } t.$$

The following theorem is an approximate implicit function theorem (basically Newton's method) that we use to find a zero of the function F defined in (2.4) given a sufficiently small local error and a bounded right inverse for $DF(x)$ defined in (2.5). Theorem 2.2 is easily generalized to the case of doubly infinite sequences and finite sequences without change in the proof.

THEOREM 2.2. *Let $F : l^\infty(\mathbb{N}) \to l^\infty(\mathbb{N})$ be a C^2 map. Let x be a point in $l^\infty(\mathbb{N})$ such that $DF(x)$ has a bounded right inverse $DF(x)^\dagger$ and let $\epsilon_0 > 0$ be chosen so that*

$$||DF(x) - DF(w)|| \leq 1/(2||DF(x)^\dagger||) \tag{2.9}$$

for $||w - x|| \leq \epsilon_0$. If $0 < \epsilon \leq \epsilon_0$ and

$$||F(x)|| \leq \epsilon/(2||DF(x)^\dagger||), \tag{2.10}$$

then the equation $F(w) = 0$ has a solution w such that $||w - x|| \leq \epsilon$.

Proof. For the proof see [CVV1]. □

The following theorem gives sufficient conditions for the existence of a trajectory near an orbit $x := \{x_n\}_0^\infty$ with $||F(x)|| \leq \delta$. The result is based upon the uniform boundedness result of Pliss (Theorem 2.1) and an application of Theorem 2.2.

THEOREM 2.3. *Consider the* IVP (2.1) *and assume that it is solved using a numerical method to produce an orbit* $x := \{x_n\}_0^\infty$ *with local error uniformly bounded by* $\delta > 0$. *Let* $\tilde{x}(t)$ *denote the corresponding pseudo solution and* $\tilde{z}(t)$ *the corresponding continuous function constructed as in* (2.6). *Let* $h_{\max} = \sup_n\{h_n\}$ *and* $h_{\min} = \inf_n\{h_n\}$ *denote the maximum and minimum stepsize, respectively. Let* $L_{Df}(\tilde{x}, \gamma)$ *denote a bound on the Lipschitz constant for* Df *in a* γ*-neighborhood of* $\tilde{x}(t)$ *and let* $B_{Df}(\tilde{x}, \gamma)$ *denote a bound on* Df *in a* γ*-neighborhood of* $\tilde{x}(t)$. *Assume that*

(i) *the linear system* (2.7) *satisfies* (H1)–(H3) *with* $\eta(K, \lambda, \alpha)$ *and* $T(K, \lambda, \alpha)$ *defined as in Theorem* 2.1 *with* $\beta_{j+1} - \beta_j > T(K, \lambda, \alpha)$ *for* $j = 1, \ldots, m-1$;

(ii) *the inequality* $c > r$ *is satisfied where* $r = h_{\max} L_{Df}(\tilde{x}, \delta/2)\delta/2 + h_{\max}^2 B_{Df}^2(\tilde{x}, \delta/2)/2$ *and* $c = \eta(K, \lambda, \alpha)h_{\min}(1-\rho)$ *for an arbitrary* ρ *such that* $1 \gg \rho > 0$;

(iii) *the inequality*

$$h_{\max} L_{Df}(\tilde{x}, \epsilon)\epsilon \leq \left(\frac{c+s}{c}\right)^{-1} \frac{c-r}{2}$$

is satisfied for $s := h_{\max}^2 B_{Df}(\tilde{x}, \delta/2)\eta(K, \lambda, \alpha)/2$ *and* $\epsilon := 2\delta(c-r)^{-1}$.

Then there exists a solution $\tilde{w}(t)$ *of the* IVP (2.1) *such that* $||\tilde{w}(t_n) - x_n|| \leq \epsilon$ *for* $n = 0, \ldots, \infty$.

Proof. Since (i) holds, the Pliss Theorem implies that the inhomogeneous equation (2.8) has a solution $\tilde{u}(t)$ satisfying $||\tilde{u}|| \leq 1$ for all continuous functions $\tilde{g}(t)$ such that $||\tilde{g}|| \leq \eta(K, \lambda, \alpha)$. If $g_n := -\int_{t_n}^{t_{n+1}} \tilde{g}(s)ds$ and $u_n := \tilde{u}(t_n)$, then

$$\begin{aligned} u_{n+1} &= u_n + \int_{t_n}^{t_{n+1}} Df(\tilde{z}(s), s)\tilde{u}(s)ds - g_n \\ &= \phi_x(x_n, t_n, t_{n+1})u_n - g_n + r_n, \end{aligned} \tag{2.11}$$

where ϕ_x is defined in (2.5) and

$$\begin{aligned} r_n &= \int_{t_n}^{t_{n+1}} [Df(\tilde{z}(s), s)\tilde{u}(s) - Df(\tilde{x}(s), s)u_n]ds \\ &= \int_{t_n}^{t_{n+1}} [Df(\tilde{z}(s), s) - Df(\tilde{x}(s), s)]u_n ds \\ &\quad + \int_{t_n}^{t_{n+1}} Df(\tilde{z}(s), s)\left[\int_{t_n}^{s} Df(\tilde{z}(\tau), \tau)\tilde{u}(\tau) + \tilde{g}(\tau)d\tau\right]ds. \end{aligned}$$

Then we have $||r_n|| \leq r||u_n|| + h_{\max}^2 B_{Df}(\tilde{x}, \delta/2)\eta(K, \lambda, \alpha)/2 \equiv r||u_n|| + s$.

Let $z_n = z(t_n)$ for $n = 0, \ldots, \infty$ so that by (2.11) we have

$$(DF(x)u)_n = (G(z)u)_n - r_n, \quad \text{for } n = 0, \ldots, \infty,$$

where the linear operator $G(z)$ is defined so that $G(z)u = g$ implies

$$(G(z)u)_n \equiv u_{n+1} - u_n - \int_{t_n}^{t_{n+1}} Df(\tilde{z}(s), s)\tilde{u}(s)ds = g_n \quad \text{for all } n.$$

Given a sequence $g := \{g_n\}_0^\infty$ with $||g_n|| \le c$ there exists a continuous function $\tilde{g}(t)$ such that $||\tilde{g}|| \le \eta(K, \lambda, \alpha)$. By the Pliss Theorem, $G(z)$ is onto; i.e., given any sequence $g \in l^\infty(\mathbb{N})$ there exists a sequence $u \in l^\infty(\mathbb{N})$ such that $G(z)u = g$. Therefore, $G(z)$ has a right inverse $G(z)^\dagger$ with

$$||G(z)^\dagger|| = \sup_{||g|| \le c} \frac{||G(z)^\dagger g||}{c} \le \sup_{||g|| \le c} \frac{\{||u|| : u = G(z)^\dagger g\}}{c} \le c^{-1}.$$

In general, if $u = \{u_n\}_0^\infty$ satisfies $(G(z)u)_n = g_n + r_n$, then $(DF(x)u)_n = g_n$ and $||u|| \le ||G(z)^\dagger||(c + r||u|| + s)$ so that (ii) implies $(1 - r/c)||u|| \le ||G(z)^\dagger||(c + s)$. Thus, if (ii) is satisfied,

$$||DF(x)^\dagger|| = \sup_{||g|| \le c} \frac{||DF(x)^\dagger g||}{c} \le \sup_{||g|| \le c} \frac{\{||u|| : u = DF(x)^\dagger g\}}{c}$$
$$\le ||G(z)^\dagger|| \left(\frac{c+s}{c}\right)\left(\frac{1-r}{c}\right)^{-1} \le \left(\frac{c+s}{c}\right)(c-r)^{-1}.$$

Now apply the Fixed Point Theorem to F with $||F(x)|| \le \delta$ and $||DF(x)^\dagger|| \le (\frac{c+s}{c})(c - r)^{-1}$. For $\epsilon_0 := \epsilon$ and $||x - w|| \le \epsilon_0$ and using (iii), we have (2.9) satisfied since

$$||DF(x) - DF(w)|| \le h_{\max} L_{Df}(\tilde{x}, \epsilon)||x - w|| \le \left(\frac{c+s}{c}\right)^{-1} \frac{c-r}{2} \le (2||DF(x)^\dagger||)^{-1}.$$

Thus, there exists a solution $w \in l^\infty(\mathbb{N})$ of $F(w) = 0$ such that $||w_n - x_n|| \le \epsilon$. Define $\tilde{w}(t) = \phi(w_n, t_n, t)$ for $t_n \le t < t_{n+1}$ and $n = 0, \ldots, \infty$ to complete the proof. □

3. Numerical aspects. We now consider the case in which we have produced a finite orbit using a numerical method. We would like to know whether there is a trajectory satisfying the same differential equation but with a nearby initial condition such that the trajectory is close to the numerically computed orbit at the mesh points. Our intent is to provide verifiable assumptions given certain numerically computable quantities so that we may apply the theorem and obtain an a posteriori bound on the global error.

In this section, let $||y||$ denote the supremum norm for $y \in \mathbb{R}^N$. Consider the IVP (2.1) and for some finite positive integer M consider the orbit $\{x_n\}_0^M$ produced using (2.2). Let $\tilde{x}_n(t)$ denote the pseudo solution defined as in (2.3) and let $\tilde{x}(t) := \tilde{x}_n(t)$ for $t_n \le t < t_{n+1}$. Let δ denote a bound on the local error. Consider the linearized problem about the pseudo solution

$$\dot{u} = Df(\tilde{x}(t), t)u. \tag{3.1}$$

Define the operator $F : l^\infty(\{0, \ldots, M\}) \to l^\infty(\{0, \ldots, M-1\})$ by

$$(F(x))_n = x_{n+1} - \phi(x_n, t_n, t_{n+1}) \tag{3.2}$$

and its first variation $DF(x) : l^\infty(\{0, \ldots, M\}) \to l^\infty(\{0, \ldots, M-1\})$ by

$$(DF(x)u)_n = u_{n+1} - \phi_x(x_n, t_n, t_{n+1})u_n, \tag{3.3}$$

where ϕ_x is defined as in (2.5).

Let $\tilde{z}(t)$ be a function with the property that $||\tilde{x}(t) - \tilde{z}(t)|| \leq \delta$ for all t. For one-step methods with an associated Taylor polynomial, an obvious choice for the function $\tilde{z}(t)$ is the interpolant defined locally by the Taylor polynomial corresponding to the numerical method. In particular, for the Runge–Kutta Felberg integrator RKF45 used for the examples in §5 we employ the associated fifth-order interpolant. The function $\tilde{z}(t)$ may be discontinuous at the mesh points. Consider the linearization about $\tilde{z}$,

$$\dot{u} = Df(\tilde{z}(t), t)u. \tag{3.4}$$

We define $G(z) : l^\infty(\{0, \ldots, M\}) \to l^\infty(\{0, \ldots, M-1\})$ by

$$(G(z)u)_n = u_{n+1} - \Phi(t_{n+1}, t_n)u_n, \tag{3.5}$$

where Φ is defined as in (2.7).

Let A_n denote a quadrature formula used to approximate $\Phi(t_{n+1}, t_n)$. Let s denote a bound on the relative error in the quadrature approximation, i.e.,

$$||\Phi(t_{n+1}, t_n)u_n - A_n u_n|| \leq s||u_n||. \tag{3.6}$$

Define the operator $H(A) : l^\infty(\{0, \ldots, M\}) \to l^\infty(\{0, \ldots, M-1\})$ by

$$(H(A)u)_n = u_{n+1} - A_n u_n. \tag{3.7}$$

We now state the following theorem similar to Theorem 2.3, but with assumptions that may be easily verified computationally.

THEOREM 3.1. *Consider the* IVP (2.1) *and assume that it is solved numerically producing an orbit* $x := \{x_n\}_0^M$ *with local error uniformly bounded by* $\delta > 0$. *Let* $\tilde{x}(t)$ *denote the corresponding pseudo solution and let* $\tilde{z}(t)$ *denote a function with the property that* $||\tilde{x}(t) - \tilde{z}(t)|| < \delta$ *for all* t. *Let* A_n *denote the quadrature formula used to approximate the linear variational equation about* $\tilde{z}(t)$ *from* t_n *to* t_{n+1}. *Let* $h_{\max} = \sup_n\{h_n\}$ *denote the maximum stepsize. Let* $L_{Df}(\tilde{x}, \gamma)$ *denote a bound on the Lipschitz constant for* Df *in an* γ*-neighborhood of* $\tilde{x}(t)$. *Assume that*

(i) *the inequality* $c > r + s$ *is satisfied where* $||H(A)^\dagger|| \leq c^{-1}$ *and* $H(A)$ *is defined in* (3.7), $H(A)^\dagger$ *is a right inverse of* $H(A)$, $r = h_{\max} L_{Df}(\tilde{x}, \delta)\delta$, *and* s *is defined in* (3.6);

(ii) *the inequality*

$$h_{\max} L_{Df}(\tilde{x}, \epsilon)\epsilon \leq (c - r - s)/2$$

is satisfied for $\epsilon = 2\delta(c - r - s)^{-1}$.

Then there exists a solution $\tilde{w}(t)$ *of the* IVP (2.1) *such that* $||\tilde{w}(t_n) - x_n|| \leq \epsilon$ *for* $n = 0, \ldots, M$.

Proof. Given a sequence $g := \{g_n\}_0^{M-1}$, we have that $u := \{u_n\}_0^M$ is a solution of $(DF(x)u)_n = g_n$ if $(H(A)u)_n = g_n + r_n + s_n$, where s_n is defined to be

$$s_n := [\Phi(t_{n+1}, t_n) - A_n]u_n$$

and

$$r_n := [\phi_x(x_n, t_n, t_{n+1}) - \Phi(t_{n+1}, t_n)]u_n = \int_{t_n}^{t_{n+1}} [Df(\tilde{x}(s), s) - Df(\tilde{z}(s), s)]u_n ds.$$

By (3.6) we have that $||s_n|| \leq s||u_n||$ and we have that

$$||r_n|| \leq h_{\max} L_{Df}(\tilde{x}, \delta)\delta||u_n||,$$

so that $||r_n|| \leq r||u_n||$. For $||g_n|| \leq 1$, we have

$$||u|| \leq ||H(A)^\dagger|| \sup_n \{||g_n|| + ||r_n|| + ||s_n||\} \leq c^{-1}(1 + r||u|| + s||u||)$$

so that by (i)

$$||u|| \leq (c - r - s)^{-1}.$$

Thus,

$$\begin{aligned} ||DF(x)^\dagger|| &= \sup_{||g||=1} ||DF(x)^\dagger g|| \leq \sup_{||g||=1} \{||u|| : u = DF(x)^\dagger g\} \\ &\leq (c - r - s)^{-1}. \end{aligned}$$

Now apply the Fixed Point Theorem to F with $||F(x)|| \leq \delta$ and $||DF(x)^\dagger|| \leq (c - r - s)^{-1}$. For $\epsilon_0 := \epsilon$ and $||x - w|| \leq \epsilon_0$ and using (ii), we have (2.9) satisfied since

$$||DF(x) - DF(w)|| \leq h_{\max} L_{Df}(\tilde{x}, \epsilon)||x - w|| \leq (c - r - s)/2 \leq (2||DF(x)^\dagger||)^{-1}.$$

Thus, there exists a solution $w \in l^\infty(\{0, \ldots, M\})$ of $F(w) = 0$ such that $||w_n - x_n|| \leq \epsilon$. Define $\tilde{w}(t) = \phi(w_n, t_n, t)$ for $t_n \leq t < t_{n+1}$ and $n = 0, \ldots, M - 1$ to complete the proof. □

Remarks. (i) Note that no explicit bounds on the inverses of f or Df or on higher-order derivatives are required.

(ii) The theorem may be applied to maps by simply using δ as a bound on the roundoff error and by setting $r = s = 0$.

4. Algorithms. In this section we present algorithms to estimate quantities needed to apply Theorem 3.1. To apply Theorem 3.1 we must supply a bound on a right inverse of $H(A)$. Since the norm of the right inverse of $H(A)$ measures to a large degree the magnification of the local error that gives the global error, it is advantageous to find, if possible, a right inverse of $H(A)$ that has small norm. Our philosophy is to use existing ODE solvers and other existing software to develop methods for obtaining an accurate estimate of the global error. As such, our error estimates will be in terms of the supremum norm, $||\cdot|| := ||\cdot||_\infty$, since most ODE solvers provide error estimates in this norm. We will use the absolute and relative local error tolerances that are provided by most standard solvers. Our intent is to provide practical estimates but not necessarily rigorous bounds on the global error. Most of our effort will be devoted to finding a suitable right inverse $H(A)^\dagger$ and to obtaining a bound or estimate of $||H(A)^\dagger||_\infty$.

An obvious choice for the right inverse is the pseudo or generalized inverse. In this case finding the right inverse is trivial, but estimating or bounding its norm may be difficult. If we consider $H(A)$ as a matrix and write $H(A)$ in terms of its singular value decomposition, then $H(A) = U\Sigma V^T$ where U, V are orthogonal and Σ is a nonnegative diagonal matrix. If $H(A)$ is full rank, then $||H(A)^\dagger||_2 = 1/\sigma_1$ where σ_1 is the smallest singular value of $H(A)$. The difficulty with using the pseudo inverse is that although the pseudo inverse is optimal in the 2-norm sense it is not necessarily optimal in the ∞-norm sense. In fact, in general, $||H(A)^\dagger||_\infty \leq \sqrt{NM}||H(A)^\dagger||_2$ for an ODE in $\mathbb{R}^N$ and an orbit of length M.

For explicit one-step methods, $H(A)$ has the matrix form

$$H(A) = \begin{pmatrix} -A_0 & I_N & & \\ & \ddots & \ddots & \\ & & -A_{M-1} & I_N \end{pmatrix}, \tag{4.1}$$

where $H(A)$ is an $M \cdot N \times (M+1) \cdot N$ matrix, A_i is an $N \times N$ matrix, and I_N is the $N \times N$ identity matrix. The matrix A_i advances the discrete solution of the linear variational equation from t_i to t_{i+1}. Note that $H(A)H(A)^T$ is a symmetric block tridiagonal matrix. We will restrict attention to explicit one-step methods, although similar results will apply for implicit one-step methods and multistep methods. The matrix $H(A)$ has the form of a multiple shooting matrix for a linear boundary value problem, but without the N additional rows that are used to specify the boundary conditions. The next approach will be to outline strategies for adding boundary conditions and thus specifying a right inverse for $H(A)$.

Our second approach to finding a right inverse involves appending boundary conditions at multiple points to obtain a well-conditioned BVP (see [dHM2] and [Ma2]). In this way we obtain a linear multipoint BVP. Our challenge is to find boundary conditions, possibly at more than the initial and terminal times, so that a BVP has a uniformly bounded solution. When appending boundary conditions we look for switching points to dynamically change the number of stable and unstable components. In particular, if we have k stable directions initially, then up to a suitable orthogonal change of variables (see [MS]), we adjoin the boundary condition

$$\begin{pmatrix} 0 & 0 \\ 0 & I_k \end{pmatrix} u(\beta_0) = 0,$$

where I_k is the $k \times k$ identity matrix. Similarly, if at the terminal time there are l unstable directions, then we adjoin the boundary condition

$$\begin{pmatrix} I_l & 0 \\ 0 & 0 \end{pmatrix} u(\beta_{m+1}) = 0.$$

The intermediate switching points β_j for $j = 1, \ldots, m$ produce boundary conditions of the form

$$\begin{pmatrix} 0_{N-k-j} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0_{k+j-1} \end{pmatrix} u(\beta_j) = 0.$$

These are linear boundary conditions so that appending boundary conditions is equivalent to adding N rows to $H(A)$ in (4.1).

To apply Theorem 3.1, we must have estimates of the following quantities:

1. δ, the local absolute error for the original problem (2.1);
2. $h_{\max}$, the maximum stepsize;
3. L_{Df}, a bound on the Lipschitz constant of Df;
4. c^{-1}, a bound on $||H(A)^\dagger||_\infty$;
5. s, the local relative error for the quadrature formula of the linearized problem.

Our basic algorithm is as follows.

ALGORITHM.

Step 1. Integrate simultaneously

$$\dot{x} = f(x,t),$$
$$\dot{u} = Df(x(t),t)u$$

from t_j to t_{j+1} with initial data x_j and $u_j = I$ for $j = 0, \ldots, M-1$, where $h_j = t_{j+1} - t_j$ is the stepsize chosen by the integrator with given absolute and relative error tolerances to obtain estimates for δ, s, and $h_{\max}$. Thus, we obtain the x_j and A_j for all j.

Step 2. Find a bound or estimate c^{-1} for $||H(A)^\dagger||_\infty$, where $H(A)^\dagger$ is either the pseudo inverse or a right inverse formed by adjoining boundary conditions.

Step 3. Compute a posteriori bounds for L_{Df}.

Step 4. If (i) and (ii) in Theorem 3.1 are satisfied, then apply the theorem to obtain the global error $\epsilon := 2\delta(c - s - r)^{-1}$.

Remarks. (i) By integrating both the original equation and the linear variational equation simultaneously, we obtain an approximation of the linear variational equation about the interpolant defined by the numerical method.

(ii) The local and relative error tolerances provide bounds for δ and s, respectively.

(iii) The a posteriori bounds for L_{Df} may be obtained as in [SY] using Gronwall's inequality or using coarse a priori bounds as we have done for the examples in §5.

(iv) The quantity r in Theorem 3.1 may be computed in terms of δ, $h_{\max}$, L_{Df}.

We now present the details of implementations for finding a right inverse and an ∞-norm bound or estimate on this right inverse in Step 2. The first method we consider is based on finding the smallest singular value of the matrix $H(A)$ in (4.1). The second method is based on adjoining boundary conditions and solving a suitable linear inhomogeneous BVP.

Instead of directly finding the smallest singular value σ_1 of the matrix $H(A)$, we will find the smallest eigenvalue λ_1 of the symmetric block tridiagonal matrix $H(A)H(A)^T$ and then set $\sigma_1 = \sqrt{\lambda_1}$. We find λ_1 by the Lanczos process (see [GvL]). The Lanczos process is an iterative method for finding the extremal eigenvalues of a matrix. The method generates a sequence of tridiagonal matrices whose eigenvalues are progressively better estimates of the extremal eigenvalues of the original matrix that is typically large and sparse. The convergence to the extremal eigenvalues is rapid provided the relative spacing between these eigenvalues is large (see [GvL]). We have made modifications to the software package LAS2 (see [Ber]) to apply the Lanczos procedure to symmetric block tridiagonal matrices $B := H(A)H(A)^T$, where $H(A)$ is of the form (4.1). We employ the error estimation procedure that is provided as part of the software. Although there is no guarantee that we have actually found the smallest eigenvalue, we were able to confirm that for smaller examples the Lanczos process did in fact provide accurate estimates of the smallest eigenvalue.

Remarks. (i) The use of the pseudo inverse has the advantage that information about the number of stable and unstable modes is not necessary.

(ii) The use of the Lanczos method to take advantage of the sparse structure of $H(A)$ may be implemented in a memory efficient or a time efficient manner by either recalculating the A_i, $i = 1, \ldots, M$ or by storing the A_i, respectively.

(iii) As was remarked above, we obtain a supremum norm estimate of $\sqrt{NM}\sigma_1^{-1}$ as an estimate of the norm of the pseudo inverse of $H(A)$.

The boundary value problem approach is based on considering the difference equation

$$u_{n+1} = A_n u_n + g_n \tag{4.2}$$

along with the appended boundary conditions where $g = \{g_n\}_0^M$ is an arbitrary sequence with $||g||_\infty = 1$. We have that

$$\begin{aligned} ||H(A)^\dagger||_\infty &= \sup_{||g||_\infty = 1} ||H(A)^\dagger g||_\infty \\ &= \sup_{||g||_\infty = 1} \{||u||_\infty : u \quad \text{satisfies} \quad (4.2)\}. \end{aligned} \tag{4.3}$$

Our task now is to replace the problem (4.2) for an arbitrary sequence g of norm one with a problem for a fixed sequence and obtain a bound on $||H(A)^{\dagger}||_{\infty}$. This is similar to the situation when one is attempting to estimate the norm of the inverse of a matrix for condition number estimation. We will replace the arbitrary sequence g with a sequence in which every element is of absolute value one.

Since our multipoint BVP may be thought as a several-coupled two-point BVP, it suffices to consider the difference equation

$$u_{n+1} = A_n u_n + g_n,$$
$$\begin{pmatrix} 0 & 0 \\ 0 & I_k \end{pmatrix} Q_0^T u_0 = \gamma_0, \begin{pmatrix} I_{N-k} & 0 \\ 0 & 0 \end{pmatrix} Q_M^T u_M = \gamma_M,$$

where for simplicity we let M denote the number of iterates between (possible intermediate) boundary points, and we let u_0 denote the value at the left boundary point. Here Q_0 is an appropriately chosen permutation matrix (see [MS], [dHM1]) and Q_M is an orthogonal matrix to be determined below.

To solve the BVP, we decouple using an orthogonal decoupling transformation. Using the modified Gram–Schmidt method (see [GvL]), we obtain the decomposition $Q_{n+1}R_n = A_n Q_n$ for $n = 0, .., M-1$, where R_n is upper triangular with positive diagonal elements and Q_{n+1} is orthogonal. Then the decoupling transformation is given by $v_n = Q_n^T u_n$ and the decoupled equation is

$$\begin{aligned} &v_{n+1} = R_n v_n + h_n, \\ &\begin{pmatrix} 0 & 0 \\ 0 & I_k \end{pmatrix} v_0 = \gamma_0, \qquad \begin{pmatrix} I_{N-k} & 0 \\ 0 & 0 \end{pmatrix} v_M = \gamma_M, \end{aligned} \tag{4.4}$$

where $h_n = Q_{n+1}^T g_n$. We note that for $l = 0$ or $l = M$ if l denotes the time of an initial or terminal boundary point, then γ_l is the vector of all zeros. If the left boundary point of our two-point BVP is an intermediate boundary point of the multipoint BVP, then the components of the vector γ_0 are given by

$$\gamma_0^{(j)} = \begin{cases} 0, j = 1, \ldots, N-k+1, \\ v_0^{(j)}, j = N-k+2, \ldots, N. \end{cases}$$

Similarly, if the right boundary point of our two-point BVP is an intermediate boundary point of the multipoint BVP, then the components of the vector γ_M are given by

$$\gamma_M^{(j)} = \begin{cases} v_M^{(j)}, j = 1, \ldots, N-k-1, \\ 0, j = N-k, \ldots, N. \end{cases}$$

We now write the recursion in block form by setting

$$v_n = \begin{pmatrix} v_n^{(1)} \\ v_n^{(2)} \end{pmatrix},$$

where $v_n^{(2)}$ is a k-vector and

$$R_n = \begin{pmatrix} R_n^{(11)} & R_n^{(12)} \\ 0 & R_n^{(22)} \end{pmatrix}.$$

Then for any integer $0 \le J \le M$ we have that

$$\begin{aligned} v_{n+J+1}^{(2)} = &\sum_{i=n}^{n+J-1} R_{n+J}^{(22)} \cdots R_{i+1}^{(22)} h_i^{(2)} \\ &+ h_{n+J} + R_{n+J}^{(22)} \cdots R_n^{(22)} v_n^{(2)} \end{aligned} \tag{4.5a}$$

and

$$v_n^{(1)} = \sum_{i=n}^{n+J-1} [R_i^{(11)} \dots R_n^{(11)}]^{-1} \{R_i^{(12)} v_i^{(2)} + h_i^{(1)}\} + [R_{n+J}^{(11)} \dots R_n^{(11)}]^{-1} \{v_{n+J+1}^{(1)} - h_{n+J}^{(1)} - R_{n+J}^{(12)} v_{n+J}^{(2)}\}. \tag{4.5b}$$

For a matrix C and a vector b, let $|C|$ denote the corresponding matrix and $|b|$ the corresponding vector whose elements consist of the absolute value of the elements of C and b, respectively. Let $\mathbf{1} := (1, \dots, 1)^T$ denote the vector with all elements equal to one. Consider now the sequence $w = \{w_n\}_0^M$ formed as follows:

$$w_{n+J+1}^{(2)} = \sum_{i=n}^{n+J-1} |R_{n+J}^{(22)} \dots R_{i+1}^{(22)}| \mathbf{1} + \mathbf{1} + |R_{n+J}^{(22)} \dots R_n^{(22)}| w_n^{(2)} \tag{4.6a}$$

and

$$w_n^{(1)} = \sum_{i=n}^{n+J-1} |[R_i^{(11)} \dots R_n^{(11)}]^{-1}| \{|R_i^{(12)}| w_i^{(2)} + \mathbf{1}\} + |[R_{n+J}^{(11)} \dots R_n^{(11)}]^{-1}| \{w_{n+J+1}^{(1)} + \mathbf{1} + |R_{n+J}^{(12)}| w_{n+J}^{(2)}\}. \tag{4.6b}$$

It is easy to see that we have the following lemma.

LEMMA 4.1. *If γ_0, γ_M in* (4.4) *are nonnegative vectors, then for v computed in* (4.5) *and w computed in* (4.6), *we have that $||v||_\infty \geq ||w||_\infty$ for all sequences h in* (4.5) *with $||h|| = 1$.*

As a consequence of this lemma, we have that $N||v||_\infty \geq ||H(A)^\dagger||_\infty$ since the supremum norm condition number of an $N \times N$ orthogonal matrix is bounded by N.

Remarks. (i) To determine the switching points (times where there is a change in stability), we monitor the diagonal elements of the upper triangular matrices R_n. If $|R_{m-1}^{(j,j)}| > 1$ and $|R_m^{(j,j)}| < 1$ for some j, $1 \leq j \leq N$, then the mth iterate is a candidate to be a switching point. Since the number of unstable modes cannot be increased, we only decrease the number of unstable modes when the magnitude of the diagonal element of R_n has magnitude less than one over several iterates.

(ii) We compute the sequence w in (4.5) to account for roundoff errors using the methods in [Wi].

(iii) Other decoupling transformations are possible besides discrete orthogonal decoupling transformation. In particular, the Riccati transformation (see [DOR1], [DOR2], [Me1]) allows us to integrate a subset of the N^2 variables in the linear variational equation and may be better conditioned in the supremum norm. Another choice for a decoupling transformation is the continuous orthogonal transformation or continuous orthonormalization [D], [Me2].

5. Numerical examples. In this section we apply the algorithms for estimating the global error to three example problems. All computations were performed on a Silicon Graphics workstation with 64 megabytes of memory in double-precision arithmetic (machine epsilon $\approx 2.2\text{E}-16$). All computations were done using the Runge–Kutta Fehlberg integrator RKF45 of Shampine and Watts [SWD]. We have chosen RKF45 for convenience since it is a widely used automatic integrator with absolute and relative error tolerances. In Tables 1 through 6 we use T or Time to denote the final value of the independent variable t, Iterates to denote the number of timesteps that were taken, and CPU Time is the CPU time recorded in seconds. For the BVP method c^{-1} is a bound on the infinity-norm of a right inverse, and for the singular

value decomposition (SVD) method c^{-1} is an estimate of the infinity-norm of the pseudo inverse. We use δ to denote the local error tolerance and ϵ is our global error estimate.

Example 5.1. The first problem we consider is the logistic equation

$$\dot{y} = y(1-y), \qquad y(0) = \xi, 1 \gg \xi > 0,$$

which was also considered in [Be1].

We let the switching point β_1 be the time t_n such that $y(t_n) \approx \frac{1}{2}$, where $n \in \{0, \ldots, M\}$. This choice for the switching point is consistent with our theoretical results. In fact, for the exact solution $x(t)$ of the logistic equation with $x(0) = \frac{1}{2}$, the fundamental matrix solution about $x(t)$ is (K, λ)-hyperbolic for $K = 4$ and $\lambda = 1$ with $S(t) = \mathbb{R}$ for $t \in [0, +\infty)$ and $U(t) = \mathbb{R}$ for $t \in (-\infty, 0]$. When employing the boundary value method as outlined in §4, we adjoin the condition $u(\beta_1) = 0$ to the linear variational equation.

For this problem we have $L_{Df} = 2$. Tables 1 and 2 show our results for approximate orbits between the fixed points $x = 0$ and $x = 1$. For the SVD method presented in §4 we have included the 2-norm bound for $||H(A)^{\dagger}||$ in parentheses. We compute our numerical orbit with initial data $y(0) = \xi \approx 0$ and numerically integrate from $t = 0$ to $t = T$ so that $y(T) \approx 1 - \xi$. In Table 1, $\xi = 1.E-2$ and $y(T) \approx 1 - \xi$, while in Table 2, $\xi = 1.E-4$ and $y(T) \approx 1 - \xi$.

TABLE 1

Example 5.1. ($T = 9.22$, Iterates $= 188$, $\delta = 1.E-07$)			
Method	CPU time	c^{-1}	ϵ
BVP	.04	24.31	4.86E–6
SVD	.61	309.71 (22.6)	6.20E–5

TABLE 2

Example 5.1. ($T = 18.46$, Iterates $= 390$, $\delta = 1.E-07$)			
Method	CPU time	c^{-1}	ϵ
BVP	.08	24.28	4.86E–6
SVD	1.26	459.15 (23.25)	9.20E–5

For the SVD method we attained convergence in less than 100 Lanczos iterations.

Example 5.2. The next example we consider is the forced pendulum equation

$$\ddot{y} + a\dot{y} + \sin y = b\cos t, \qquad \dot{y}(0) = y(0) = 0,$$

where $a = 0.2$ and $b = 2.4$. This equation was considered in [SY]. For this equation, we found that the optimal choice was to maintain one stable and one unstable mode throughout, although the linear variational equation is not uniformly hyperbolic. There are changes in stability, but there is not a monotone decrease in the number of unstable modes. Thus, we adjoin boundary conditions to obtain a two-point boundary value problem such that in the decoupled variables v, we have the boundary conditions

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} v(\beta_0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} v(\beta_1) = 0$$

when we integrate from $t = \beta_0 \equiv 0$ to $t = \beta_1 \equiv$ Time, where the values of Time are given in Tables 4 and 5.

For this problem we have $L_{Df} = 1$.

In our computations using the SVD method we attained convergence in the smallest eigenvalue of $H(A)H(A)^T$ to machine precision in less than 20 Lanczos iterations. For both the BVP and SVD method we were not able to compute for longer orbits due to memory

TABLE 3

Example 5.2. BVP method					
δ	Iterates	Time	CPU time	c^{-1}	ϵ
1.E–11	150,000	922	75	848,974	3.40E–05
1.E–12	300,000	1165	137	857,602	3.44E–06
1.E–12	500,000	1942	251	896,583	3.40E–06

TABLE 4

Example 5.2. SVD method					
δ	Iterates	Time	CPU time	c^{-1}	ϵ
1.E–11	150,000	922	472	808	1.62E–08
1.E–08	150,000	3677	469	802	1.61E–05
1.E-06	150,000	9229	634	799	1.60E–03

constraints, although less memory intensive implementations for both methods are possible by storing intermediate values of the orbit and then recomputing portions of the sequence $\{A_n\}_0^{M-1}$ as needed.

Although our methods serve different purposes than those considered in [SY] we now compare our results with the results obtained in [SY]. They obtain rigorous local error bounds through the error term of a fixed order, fixed stepsize Taylor series method. As such they obtain global error bounds. For the forced pendulum equation, Sauer and Yorke in [SY] were able to prove the existence of a trajectory within 1.E -9, a computer-generated orbit obtained using a seventh-order Taylor series method with a fixed step-size of $\Delta t \approx 3.\text{E} - 3$ to obtain local errors bounded by $1.\text{E} - 18$ for $0 \leq t \leq\approx 30,000$ on a machine with machine precision of $1.\text{E} - 28$. Our computations are performed using a standard fixed order, variable stepsize method, RKF45, in which we provide the local error tolerances and the integrator chooses the stepsize. In this way we obtain local error estimates, but not rigorous bounds on the local errors. For the BVP and the SVD methods presented in §4, our global error estimates are not rigorous bounds due to the local error estimation. Our use of the Lanczos method to estimate the smallest singular value of the matrix $H(A)$ defined in (4.1) seems to provide a reliable estimate for the norm of the pseudo inverse. For the forced pendulum equation, with the same parameter values and initial condition reported on in [SY], we were able to obtain a global error estimate of approximately $1.\text{E} - 3$ with a local error estimate of $1.\text{E} - 6$ for a trajectory of length $0 \leq t \leq\approx 10,000$ on a machine with machine precision approximately $1.\text{E} - 16$.

Example 5.3. The final example we consider is the space discretized Chafee–Infante equation with Neumann boundary conditions (see [Ch]). The Chafee–Infante equation is given by

$$\begin{aligned} v_t &= \xi^2 v_{xx} + f(v), \\ v_x(0) &= v_x(1) = 0, \\ v(x, t = 0) &\quad \text{given}, \end{aligned}$$

where $f(v) = v - v^3$. We consider the system of ODEs that is obtained after the above equation is discretized in its spatial variable. In particular, we consider the finite difference discretization

$$\begin{aligned} &\dot{v}_i = \left(\frac{\xi}{k}\right)^2 [v_{i+1} - 2v_i + v_{i-1}] + f(v_i), i = 1, \ldots, N, \\ &v_0 = v_1, v_{N+1} = v_N, \\ &v_i(0) \quad \text{given}, \end{aligned}$$

where $k = 1/(N-1)$.

TABLE 5

Example 5.3. BVP method					
δ	Iterates	Time	CPU time	c^{-1}	ϵ
1.E–10	5,000	6.34	451	57,698	1.15E–05
1.E–10	10,000	13.29	1490	259,811	5.20E–05
1.E–10	20,000	27.91	2220	492,733	9.86E–05

TABLE 6

Example 5.3. SVD method					
δ	Iterates	Time	CPU time	c^{-1}	ϵ
1.E–10	5,000	6.34	922	577	1.15E–07
1.E–08	5,000	15.31	3677	593	1.14E–05
1.E–06	5,000	36.51	9229	581	1.16E–03

The Chafee–Infante equation is a gradient system and its attracting set consists of the equilibrium solutions and the connections between the equilibrium solutions (see [H]). It is known (see [Ma1]) that the number of monotone pieces of the solution v, or lap number, is nonincreasing as a function of time. It is also known (see [BF]) that the dimension of the unstable subspace of an equilibrium solution is equal to the lap number of the equilibrium solution. We found that for a sufficiently fine discretization in space and sufficiently close to the attracting set, the number of unstable modes is nonincreasing along solution paths. It has recently been shown in [AD] and [LS] that for various discretizations of semilinear parabolic equations, the stable and unstable manifolds of the discretized problem converge to those of the continuous problem (see also [HLR]).

We set $N = 30$ and $\xi = 10^{-1}$ in our experiments and use the values $L_{Df} = 6$. We use the initial data $v_i(0) = \cos(3(i-1)\pi/(N-1))$ and monitor the eigenvalues of the matrix R_n to determine when we have attained the maximum number of unstable modes (during the initial transient there was an increase in the number of unstable modes). For all the experiments with the BVP method and for the SVD method with $\delta = 1.\mathrm{E}-10$, we only provide an error bound for the portion of the trajectory in which the number of unstable modes is nonincreasing. For the other examples in which the SVD method was used we include the initial transient phase. For all of our computations the number of unstable modes decreased from three to one.

In our computations with the BVP method we used $J = 0$ in (4.6(a), (4.6(b)). Somewhat better results were obtained using a larger value of J. For the SVD method, we obtained convergence to machine precision of the smallest eigenvalue of $H(A)H(A)^T$ within 20 Lanczos iterations.

6. Conclusions. In this paper we have shown that for a wide class of piecewise hyperbolic initial value ODEs, the global error in computing a discrete numerical approximation of a trajectory may be obtained as a reasonable magnification of the local error provided that we allow the true trajectory and the discrete approximation to have different initial conditions. The type of piecewise hyperbolicity we consider occurs in the case of several hyperbolic fixed points and certain space discretized parabolic partial differential equations.

From our numerical experiments it seems clear that although the SVD method was more expensive than the BVP method, we were able to obtain global error estimates with the SVD method for much smaller local error tolerances and longer time intervals. The SVD method gave better results than the BVP method but was more expensive in terms of both memory and time. It would be interesting to see if a more efficient Lanczos method could be developed specifically for the types of problems obtained when performing global error analysis. The cost of numerically integrating the linear variational equation may be decreased in the BVP

case by employing the Riccati transformation as a decoupling transformation in the case where there are very well-defined changes in the number of stable modes.

Our methods for providing global errors are not dependent on the particular integration method, although, in this paper, we restricted our attention to explicit one-step methods. In principle, any numerical integration scheme may be used, including implicit one-step methods and linear multistep methods. The amount of modification necessary depends on the particular implementation that one wishes to use. In order to use LSODE, for example, one would have to update the Nordsieck array before each step.

An interesting case for which we were not able to obtain good results is the case near a periodic orbit. This has been explored in [Be2], [E1], and [E2]. In particular, for the case in which $A_i = 1$ for all i, it is easy to see that the global error will grow linearly as a function of the length M of the orbit. This is due to the absence of hyperbolicity in solutions of systems of this type. Near a periodic orbit one does not expect to have hyperbolicity in the direction of the flow. It would be interesting to see if our methods could be applied to periodic systems to obtain global error estimates for those directions that are not in the direction of the flow.

Acknowledgments. We are grateful to Luca Dieci, Timo Eirola, Bob Russell, and the referees for helpful remarks on an earlier version of this paper.

REFERENCES

[AD] F. ALOUGES AND A. DEBUSSCHE, *On the qualitative behavior of the orbits of a parabolic partial differential equation and its discretization in the neighborhood of a hyperbolic fixed point*, Numer. Funct. Anal. Optim., 12 (1991), pp. 253–269.

[A] D. V. ANOSOV, *Geodesic Flows on Closed Riemannian Manifolds of Negative Curvature*, Trudy Math. Inst. Steklov, 90 (1967). (In Russian.)

[AMR] U. M. ASCHER, R. M. M. MATTHEIJ, AND R. D. RUSSELL, *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[BF] P. W. BATES AND P. C. FIFE, *Spectral comparison Principles for the Cahn–Hilliard and phase-field equations, and time scales for coarsening*, Phys. D, 43 (1990), pp. 335–348.

[Ber] M. W. BERRY, *SVDPACK: A Fortran-77 Software Library for the Sparse Singular Value Decomposition*, preprint.

[Be1] W.-J. BEYN, *On the numerical approximation of phase portraits near stationary points*, SIAM J. Numer. Anal., 24 (1987), pp. 1095–1113.

[Be2] ———, *On Invariant Closed Curves for One-Step Methods*, Numer. Math., 51 (1987), pp. 103–122.

[Bo] R. BOWEN, *ω-limit sets for Axiom A diffeomorphisms*, J. Differential Equations, 18 (1975), pp. 333–339.

[Ch] N. CHAFEE, *Asymptotic behavior for solutions of a one-dimensional parabolic equation with homogeneous Neumann boundary conditions*, J. Differential Equeations, 18 (1975), pp. 111–134.

[CLP] S. N. CHOW, X. B. LIN, AND K. J. PALMER, *A shadowing lemma with applications to semilinear parabolic equations*, SIAM J. Math. Anal., 20 (1989), pp. 547–557.

[CP1] S. N. CHOW AND K. J. PALMER, *On the numerical computation of orbits of dynamical systems: the one-dimensional case*, Dynam. Differential Equations, 3 (1991), pp. 361–380.

[CP2] S. N. CHOW AND K. J. PALMER, *On the numerical computation of orbits of dynamical systems: the higher dimensional case*, J. Complexity, 8 (1992), pp. 398–423.

[CVV1] S. N. CHOW and E. S. VAN VLECK, *A shadowing lemma for random diffeomorphisms*, Random Comput. Dynam., 1 (1992), pp. 197–218.

[CVV2] ———, *Shadowing of Lattice Maps*, manuscript.

[Co] W. A. COPPEL, *Dichotomies in Stability Theory*, *Lecture Notes in Mathematics* 629, Springer-Verlag, New York, 1978.

[D] A. DAVEY, *An automatic orthonormalization method for solving stiff* BVPs, J. Comput. Phys., 51 (1983), pp. 343–356.

[dHM1] F. R. DE HOOG AND R. M. M. MATTHEIJ, *An algorithm for solving multi-point boundary value problems*, Computing, 38 (1987), pp. 219–234.

[dHM2] ———, *On the conditioning of multipoint and integral boundary value problems*, SIAM J. Math. Anal., 20 (1989), pp. 200–214.

[DOR1] L. DIECI, M. R. OSBORNE, AND R. D. RUSSELL, *A Riccati transformation method for solving linear* BVPs. I: *Theoretical aspects*, SIAM J. Numer. Anal., 25 (1988), pp. 1055–1073.

[DOR2] ———, *A Riccati transformation method for solving linear* BVPs. II: *Computational aspects*, SIAM J. Numer. Anal., 25 (1988), pp. 1074–1092.

[E1] T. EIROLA, *Invariant curves of one-step methods*, BIT, 28 (1988), pp. 113–122.

[E2] ———, *Aspects of backward error analysis in numerical* ODEs, J. Comput. Appl. Math., 45 (1993), pp. 65–74.

[Ge] C. W. GEAR, *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1971.

[GvL] G. GOLUB AND C. VAN LOAN, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1983.

[HYG1] S. HAMMEL, J. A. YORKE, AND C. GREBOGI, *Do numerical orbits of chaotic dynamical processes represent true orbits?*, J. Complexity, 3 (1987), pp. 136–145.

[HYG2] ———, *Numerical orbits of chaotic processes represent true orbits*, Bull. Amer. Math. Soc., 19 (1988), pp. 465–470.

[HLR] J. K. HALE, X.-B. LIN, AND G. RAUGEL, *Upper semicontinuity of attracts of semigroups and partial differential equations*, Math. Comput., 50 (1988), pp. 89–123.

[H] D. HENRY, *Geometric Theory of Semilinear Parabolic Equations*, Lecture Notes in Mathematics 840, Springer-Verlag, New York, 1981.

[LS] S. LARSSON AND J.-M. SANZ-SERNA, *The Behavior of Finite Element Solutions of Semilinear Parabolic Problems Near Stationary Points*, preprint.

[Ma1] H. MATANO, *Nonincrease of the lap-number of a solution for a one-dimensional semilinear parabolic equation*, J. Fac. Sci. Univ of Tokyo IA Math., 29 (1982), pp. 401–441.

[Ma2] R. M. M. MATTHEIJ, *Decoupling and stability of algorithms for boundary value problems*, SIAM Rev., 27 (1985), pp. 1–44.

[MS] R. M. M. MATTHEIJ AND G. W. M. STAARINK, *An efficient algorithm for solving general linear two-point* BVP, SIAM J. Sci. Statist. Comput., 5 (1984), pp. 745–763.

[Me1] G. H. MEYER, *Initial Value Methods for Boundary Value Problems*, Academic Press, New York, 1973.

[Me2] ———, *Continuous orthonormalization for boundary value problems*, J. Comput. Phys., 62 (1986), pp. 248–266.

[Pa] K. J. PALMER, *Exponential dichotomies and transversal homoclinic points*, J. Differential Equations, 55 (1984), pp. 225–256.

[Pl1] V. A. PLISS, *Uniformly bounded solutions of linear systems of differential equations*, Differential Equations, 13 (1977), pp. 607–613.

[Pl2] ———, *Relationship between different conditions for structural stability*, Differential Equations, 17 (1981), pp. 545–550.

[Ro1] J. W. ROBBIN, *A structural stability theorem*, Ann. Math., 94 (1971), pp. 447–493.

[Ro2] R. C. ROBINSON, *Structural stability of C^1-diffeomorphisms*, J. Differential Equations, 22 (1976), pp. 28–73.

[SY] T. SAUER AND J. A. YORKE, *Rigorous verification of trajectories for the computer simulation of dynamical systems*, Nonlinearity, 4 (1991), pp. 961–979.

[SWD] L. F. SHAMPINE, H. A. WATTS, AND S. M. DAVENPORT, *Solving non-stiff* ODEs—*the state of the art*, SIAM Rev., 18 (1976), pp. 376–411.

[S1] H. J. STETTER, *Analysis of Discretization Methods for Ordinary Differential Equations*, Springer-Verlag, New York, 1973.

[S2] ———, *The defect correction principle and discretization methods*, Numer. Math., 29 (1978), pp. 425–443.

[Wi] J. H. WILKINSON, *Rounding Errors in Algebraic Processes*, Prentice-Hall, Englewood Cliffs, NJ, 1963.