SIAM J. NUMER. ANAL.
Vol. 41, No. 5, pp. 1948–1973

RIGOROUS SHADOWING OF NUMERICAL SOLUTIONS OF ORDINARY DIFFERENTIAL EQUATIONS BY CONTAINMENT[*]

WAYNE B. HAYES[†] AND KENNETH R. JACKSON[†]

Abstract. An exact trajectory of a dynamical system lying close to a numerical trajectory is called a *shadow*. We present a general-purpose method for proving the existence of finite-time shadows of numerical ODE integrations of arbitrary dimension in which some measure of hyperbolicity is present and there are either 0 or 1 expanding modes, or 0 or 1 contracting modes. Much of the rigor is provided automatically by interval arithmetic and validated ODE integration software that is freely available. The method is a generalization of a previously published *containment* process that was applicable only to two-dimensional maps. We extend it to handle maps of arbitrary dimension with the above restrictions, and finally to ODEs. The method involves building n-cubes around each point of the discrete numerical trajectory through which the shadow is guaranteed to pass at appropriate times. The proof consists of two steps: first, the rigorous computational verification of a simple geometric property, which we call the *inductive containment property*, and second, a simple geometric argument showing that this property implies the existence of a shadow. The computational step is almost entirely automated and easily adaptable to any ODE problem. The method allows for the rescaling of time, which is a necessary ingredient for successfully shadowing ODEs. Finally, the method is local, in the sense that it builds the shadow inductively, requiring information only from the most recent integration step, rather than more global information typical of several other methods. The method produces shadows of comparable length and distance to all currently published results. Finally, we conjecture that the inductive containment property implies the existence of a shadow without restriction on the number of expanding and contracting modes, although proof currently eludes us.

Key words. error analysis, ordinary differential equations, dynamical systems, chaos, shadowing, computational techniques, interval arithmetic

AMS subject classifications. 37M05, 65C20, 68U20, 70K99, 81T80

DOI. 10.1137/S0036142901399100

1. Introduction.

Consider the *initial value problem* (IVP) for an autonomous *ordinary differential equation* (ODE)

$$\mathbf{y}'(t) = \mathbf{f}(\mathbf{y}(t)), \tag{1.1}$$

$$\mathbf{y}(t_0) = \mathbf{y}_0, \tag{1.2}$$

where the ODE (1.1) is called the *defining equation*, (1.2) is called the *initial condition*, $\mathbf{y}$ is an n-dimensional vector, and f is an n-dimensional vector-valued function. Standard forward error analysis (e.g., Dahlquist and Björck (1974) or Kahaner, Moler, and Nash (1989)) tells us that, for a large class of ODEs, it is impossible in fixed-precision arithmetic to produce a numerical solution to an IVP which remains uniformly close to the exact solution for a long time. As a result, forward error bounds are impractical in such cases. However, one can often guarantee that a numerical solution remains uniformly close not to the solution starting at the initial condition specified, but instead to the exact solution to the ODE (1.1) starting at a nearby initial condition. In other words, if one allows the initial condition to have a nonzero error, just as one is satisfied with a nonzero error at all other times (Murdock (1995)), then it may be possible to guarantee that the numerical solution remains uniformly close to *some* exact

[*]Received by the editors December 3, 2001; accepted for publication (in revised form) March 17, 2003; published electronically November 19, 2003. This work was supported by the NSERC.
http://www.siam.org/journals/sinum/41-5/39910.html

[†]Department of Computer Science, University of Toronto, Toronto, ON, M5S 2E4, Canada (wayne@cs.toronto.edu, krj@cs.toronto.edu).

solution for a long time. Such an exact solution is called a *shadow* of the numerical solution.

Backward error analysis is a general term applied to methods of error analysis that relate a numerical solution to the exact solution of a "nearby" problem (Corless (1994), for example). In the context of IVPs for ODEs, "nearby" has at least two interpretations: we can either perturb the defining equation, or we can perturb the initial condition. Defect-based and other backward error analyses allow a time-dependent perturbation to the defining equation while leaving the initial condition untouched. In contrast, shadowing perturbs only the initial condition. For many physical systems which are modelled using ODEs, the governing equations are well defined, and virtually all error is introduced by imprecise knowledge of initial conditions and/or by numerical error in the computation of the solution. In these contexts, shadowing may a be more appropriate method of error analysis than defect-based methods. On the other hand, rigorous shadowing as presented in this paper and elsewhere is extremely expensive. Whereas nonrigorous defect-controlled methods are of roughly equal expense compared to more traditional integration methods, rigorous shadowing requires validated ODE integration, which at present tends to be several orders of magnitude more expensive in both time and memory than nonvalidated methods, even for low-dimensional problems. Thus, the goal of shadowing should not be to validate every numerical solution computed, but instead to study under what conditions we can expect a numerical solution to have a shadow.

Procedures for finding shadows usually involve some sort of fixed-point method. These include nonrigorous numerical methods akin to Newton's method (Grebogi et al. (1990); Quinlan and Tremaine (1992); Hayes (1995)) and methods that employ a theorem to prove the existence of a shadow, usually relying on Brouwer's fixed-point theorem or the Newton–Kantorovich theorem (Sauer and Yorke 1991; Chow and Palmer 1991, 1992; Chow and Van Vleck 1994).

An important advance has been the realization that ODEs differ fundamentally from maps in that they have errors in *time* as well as in space. By employing a *rescaling of time*, shadow lengths for ODEs can be increased by several orders of magnitude (Coomes, Koçak, and Palmer (1994b), (1995a), (1995b); Van Vleck 1995), even allowing the proof of existence of periodic trajectories near periodic pseudotrajectories (Coomes, Koçak, and Palmer (1994a); Coomes, Koçak, and Palmer (1997)). Shadowing has also been used to demonstrate that conservative integrations that approximately satisfy a first integral can have shadows that exactly satisfy it (Coomes (1997)), and that more explicit control of the numerical error in the stable versus unstable subspaces can lead to better shadowing results (Van Vleck (2000)). An interesting application has been to prove that a chaotic trajectory exists near an apparently chaotic pseudotrajectory (Stoffer and Palmer 1999). Hayes (2001) provides a more detailed survey of ODE shadowing results.

This paper extends the work of Grebogi, Hammel, Yorke, and Sauer (1990)(hereafter GHYS), who introduced an elegant geometrical method called *containment* for proving the existence of shadows. Their proof is valid for iterated maps in two dimensions, and is also practical for two-dimensional ODE problems that do not require a rescaling of time. We extend their results to maps of arbitrary dimension in which some measure of hyperbolicity is present and there are either 0 or 1 expanding modes, or 0 or 1 contracting modes. Although we firmly believe that containment can work with an arbitrary number of expanding and contracting directions, proving the general case is a work in progress. We also introduce a new method complementary to containment that facilitates a rescaling of time. In contrast to the above methods that use a fixed-point result, containment, including our new rescaling of time, uses an entirely geometrical argument. We rigorously verify the conditions of our theorems using validated ODE integration (Nedialkov (1999); Nedialkov, Jackson, and Corliss (1999)) and demonstrate that containment is capable of proving the existence of shadows of IVPs for ODEs that are of comparable quality to any currently in the literature. We also demonstrate how containment can reproduce the proof of chaos given by Stoffer and Palmer (1999).

The outline of the paper is as follows. Section 2 presents the ideas for the proofs of containment in an informal, geometrical setting. We present the actual proofs in section 3. Formally, these proofs break into two steps. First, we must prove that the numerical trajectory satisfies a certain property called the *inductive containment property* (ICP, for short). The ICP can be proven computationally to hold, using a validated ODE integrator; we defer discussion of how this is done until section 4. Second, we must show that a numerical trajectory that satisfies the ICP has a shadow. We prove this for maps in n dimensions for the cases in which there is either one expanding or one contracting direction while all the others do the opposite (3.1), or all directions either expand or contract (3.2). The method to rescale time is presented in section 5. Section 6 presents experimental results and comparisons with previous work, followed in section 7 by our conclusions.

2. Informal description of containment.

Although containment was the first method introduced for proving the existence of finite-time shadows of numerical orbits, it has not, to our knowledge, been pursued beyond its initial conception. In this paper we demonstrate that, at least in the restricted cases discussed, containment is about as strong as any method currently in the literature.

2.1. Definitions.

In this paper, an *orbit* is a discrete sequence of points, a *solution* is a continuous curve, and a trajectory more generally refers to either an orbit or a solution, depending upon the context. The prefix *pseudo-* will be used to denote an approximate orbit, solution, or trajectory, although sometimes it will be omitted if the meaning is clear from the context.

Shadowing of numerical orbits was first applied to iterated maps.

DEFINITION 2.1. *An* orbit *of an* iterated map *consists of a sequence of points* $\mathbf{x}_i$ *generated by the recurrence* $\mathbf{x}_{i+1} = \varphi(\mathbf{x}_i)$ *for some map* φ.

DEFINITION 2.2. *A* homeomorphism *is a map which is continuous, one-to-one, and onto.*

For our purposes, we restrict φ to being a homeomorphism.

DEFINITION 2.3. *A* pseudo-orbit, *or* noisy *orbit, for* φ *satisfies* $\mathbf{y}_{i+1} = \varphi(\mathbf{y}_i) + \delta_i$, *where* δ_i *is the noise introduced at step* i. *If* $\|\delta_i\| < \delta$ *for all* i, *then it is called a* δ-pseudo-orbit *for* φ.

DEFINITION 2.4. *The exact orbit* $\{\mathbf{x}_i\}_{i=0}^{N}$ *is an* ε*-shadow of the pseudo-orbit* $\{\mathbf{y}_i\}_{i=0}^{N}$ *if* $\|\mathbf{y}_i - \mathbf{x}_i\| < \varepsilon$ *for* $i = 0, \ldots, N$.

Numerical solutions to ODEs can often be viewed as iterated maps by defining $\mathbf{x}_{i+1} = \varphi_{h_i}(\mathbf{x}_i)$, where φ_{h_i} is the *time-*h_i solution operator for the IVP (1.1), (1.2). The time-h_i solution operator is a homeomorphism as long as $\mathbf{f}$ in (1.1) is bounded and Lipschitz continuous over the domain of interest (Ascher, Mattheij, and Russell (1988)). For small h_i, a *one-step numerical method* approximates φ_{h_i} by $\tilde{\varphi}_{h_i}$ and then computes a sequence of discrete points $\mathbf{y}_{i+1} = \tilde{\varphi}_{h_i}(\mathbf{y}_i)$ representing approximations to $\mathbf{y}(t_{i+1})$, where $t_{i+1} = t_i + h_i$. We will term such a discrete sequence of points a *pseudotrajectory.* If the pseudotrajectory satisfies a *local error tolerance* of δ such that

$\|\mathbf{y}_{i+1} - \varphi_{h_i}(\mathbf{y}_i)\| \le \delta$, then we call it a *$\delta$-pseudotrajectory.* If h_i is constant, we can drop it as a subscript and treat the pseudotrajectory as a pseudo-orbit of the iterated map $\varphi \equiv \varphi_h$.

2.1.1. Hyperbolicity and pseudohyperbolicity.

One of the most important concepts in shadowing is that of *hyperbolicity.* Essentially, a system of ODEs is hyperbolic if the variational equation along a solution $\mathbf{y}(t)$ displays *exponential dichotomy* (Palmer 1988). This means that a perturbation δ to the solution $\mathbf{y}(t)$ at time $t = t_0$, $\mathbf{z}(t_0) = \mathbf{y}(t_0) + \delta$ produces a new solution $\mathbf{z}(t)$ with one of two properties: if δ lies in the *stable subspace* of $\mathbf{y}(t)$, then $\mathbf{z}(t)$ converges exponentially to $\mathbf{y}(t)$ as t increases; if δ lies in the *unstable subspace*, then $\mathbf{z}(t)$ diverges exponentially away from $\mathbf{y}(t)$ as t increases. More details can be found in Hayes (2001) or Palmer (1988). If a system is hyperbolic, then the angle between the stable and unstable subspaces is always bounded away from 0 (see GHYS).

This paper deals not with hyperbolic systems, but with systems whose pseudotrajectories are shadowable for finite but nontrivial lengths of time even though they are not hyperbolic. For this to occur, a system must display pseudohyperbolicity. We say that a system is *pseudohyperbolic* if a small perturbation to a trajectory $\mathbf{y}(t)$ produces a new solution $\mathbf{z}(t)$ which falls into one of two classes: those which *tend to* diverge exponentially away from $\mathbf{y}(t)$ as t increases, and those that *tend to* converge exponentially towards $\mathbf{y}(t)$ as t increases. In addition, $\mathbf{z}(t)$ should behave in this manner over nontrivial periods of time. In short, a pseudohyperbolic system should "mimic" the behavior of a hyperbolic system over finite but nontrivial periods of time. This notion could be quantified by, for example, attempting to find the stable and unstable subspaces using *refinement* (GHYS; Quinlan and Tremaine (1992); Hayes (1995), (2001)), and then performing least-squares fits to exponential curves of the growth and decay of these subspaces.

2.2. Containment in two dimensions.

The first studies of shadows of pseudohyperbolic systems with both expanding and contracting directions appear to be Beyn (1987) and Hammel, Yorke, and Grebogi (1987). Hammel, Yorke, and Grebogi (1988) and GHYS provide the first proof of the existence of a shadow for a nonhyperbolic system over a nontrivial length of time. Their method consists of two steps. First, they *refine* a noisy trajectory using an iterative method that produces a nearby trajectory with less noise. When refinement converges to the point that the noise is of the order of the machine precision, they invoke *containment*, which can prove the existence of a nearby exact trajectory. Their containment method, which we now describe, is directly applicable only to two-dimensional maps.

Let $\{\mathbf{y}_i\}_{i=0}^{N} \subset \mathbf{R}^2$ be a two-dimensional δ-pseudo-orbit of φ. As i increases, orbits separated from each other by a small distance along the expanding direction diverge on average away from each other, while orbits separated by a small distance along the contracting direction approach each other on average. The containment process consists of building a parallelogram M_i around each point $\mathbf{y}_i$ of the pseudo-orbit such that two sides $C_i^{\pm 1}$ are approximately normal to, and separated from each other along, the contracting direction, while the other two sides $E_i^{\pm 1}$ are approximately normal to, and separated from each other along, the expanding direction.[1] The diameter of

[1] Note that this naming convention is exactly opposite to that of GHYS, because in two dimensions they emphasized the direction to which the sides of M_i were *parallel.* In higher dimensions, the faces of an n-cube are not parallel to a unique direction, and it is the direction along which a face is separated from the center of the n-cube that matters. We change the naming convention now to avoid confusion later.

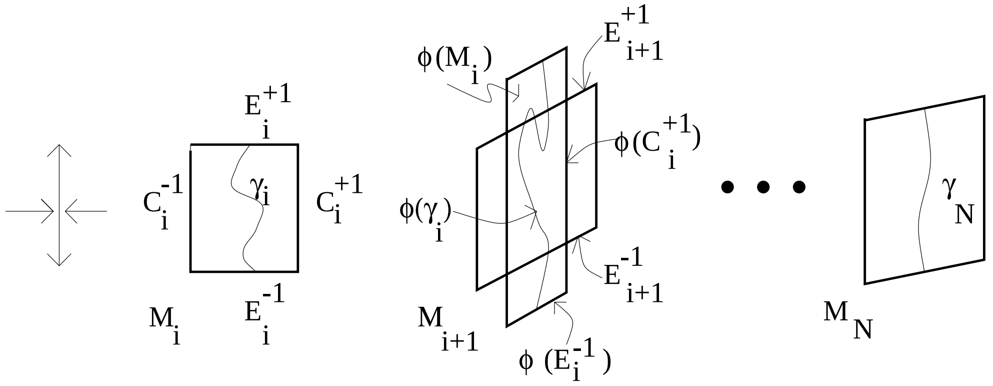

FIG. 2.1. *Containment in two dimensions, reproduced from GHYS. The horizontal direction is contracting, and the vertical direction is expanding.*

M_i will bound the distance from the pseudo-orbit to the shadow. In order to prove the existence of a shadow, the image of M_i under φ must intersect M_{i+1} such that $\varphi(M_i)$ makes a "plus sign" with M_{i+1} (Figure 2.1). To ensure that this property holds, GHYS require a bound on the second derivative of φ, and the amounts of expansion and contraction need to be resolvable to within the machine precision. The proof of the existence of an exact orbit then relies on the following argument. For any $i \in \{0, 1, \ldots, N-1\}$ let γ_i be a continuous curve in M_i connecting the expanding sides E_i^{-1} and E_i^{+1}. Its image $\varphi(\gamma_i)$ is then stretched such that there is a section of $\varphi(\gamma_i)$ lying wholly within M_{i+1}, and in particular $\varphi(\gamma_i)$ leaves M_{i+1} through the expanding sides $E_{i+1}^{\pm 1}$ at both ends. Let γ_{i+1} be a continuous subsection of $\varphi(\gamma_i)$ lying wholly within M_{i+1} connecting the expanding sides $E_{i+1}^{\pm 1}$. Repeating this process along the orbit produces γ_N lying wholly within the final parallelogram M_N. Then any point $\mathbf{x}_N \in \gamma_N$ traced backwards via φ^{-1} yields a point $\mathbf{x}_i \in \gamma_i \subset M_i$, $i = N-1, \ldots, 1, 0$. Note that $\{\mathbf{x}_i\}_{i=0}^N$ is an exact orbit. Moreover, since $\mathbf{x}_i, \mathbf{y}_i \in M_i$, we infer that $\|\mathbf{x}_i - \mathbf{y}_i\| \le \varepsilon$, where ε bounds the diameter of M_i, $i = 0, \ldots, N$. Thus, $\{\mathbf{x}_i\}_{i=0}^N$ is an ε-shadow of $\{\mathbf{y}_i\}_{i=0}^N$. We make the intuitive argument described here rigorous in section 3.

With this picture in mind, there is a nice geometric interpretation of the requirement that the angle between the stable and unstable directions be bounded away from 0: if the angle gets too small, then the parallelogram essentially loses a dimension, and $\varphi(M_i)$ can not make a "plus sign" with M_{i+1}. Practically speaking, this occurs when the angle becomes comparable to the noise amplitude of the pseudo-orbit. Hence, the more accurate the orbit, the longer it can be shadowed (GHYS, Quinlan and Tremaine (1992)).

2.3. Containment in three dimensions. The process described by GHYS is not directly applicable to systems with more than two dimensions, and GHYS provided no indication of how it could be extended beyond two dimensions. We describe how the method can be extended to three dimensions, in which there are precisely two interesting cases:

(i) One expanding direction and two contracting (Figure 2.2). Assume that the z direction is expanding, while the x and y directions are contracting. (We assume, for simplicity of exposition and for ease of drawing, that these three directions are roughly orthogonal, although in practice they need only

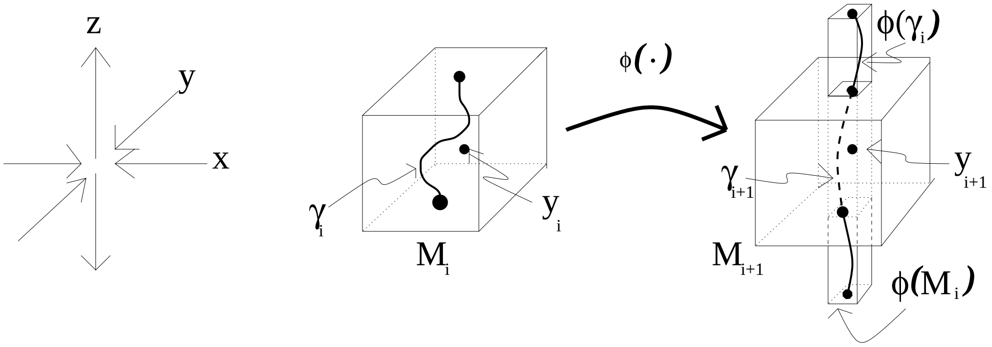

FIG. 2.2. *Containment in three-dimensions, case* (i)*: one expanding direction and two contracting.*

be resolvable from each other.) Then, analogously to the two-dimensional argument, assume we can draw a *cube* M_i of diameter no larger than ε around each noisy point $\mathbf{y}_i$, and, for $i = 0, 1, \ldots, N-1$, assume we can verify that $\varphi(M_i)$ maps over M_{i+1} so that φ stretches M_i into a long, thin tube, a segment of which lies wholly in M_{i+1}. Then, precisely as in the two-dimensional case, we can prove that an ε-shadow of $\{\mathbf{y}_i\}_{i=0}^N$ exists as follows. We introduce a curve γ_i that runs approximately along the expanding (vertical) direction from any point on the top of M_i to its bottom. If $\varphi(M_i)$ maps over M_{i+1} as in Figure 2.2, then we are guaranteed that a contiguous section of $\varphi(\gamma_i)$ lies inside M_{i+1}, connecting its top and bottom along the expanding direction. This segment of $\varphi(\gamma_i)$ becomes γ_{i+1}. Any point $\mathbf{x}_N \in \gamma_N \subset M_N$ can be traced backwards via φ^{-1} to a point $\mathbf{x}_i \in \gamma_i \subset M_i$ for $i = 0, 1, \ldots, N-1$. As in the two-dimensional case, $\{\mathbf{x}_i\}_{i=0}^N$ is an ε-shadow of $\{\mathbf{y}_i\}_{i=0}^N$.

(ii) Two expanding and one contracting direction. We note that if time is reversed in such a system, then expanding and contracting directions reverse their roles. Thus, we simply look at the pseudotrajectory in reverse and apply the above argument. That is, we set $\mathbf{z}_i = \mathbf{y}_{N-i}, i = 0, \ldots, N$, and apply the above argument to the noisy trajectory $\{\mathbf{z}_i\}_{i=0}^N$.

3. Containment theorems and proofs.

3.1. Containment in *n* dimensions with one expanding direction. Here we provide a proof of what we call the $(n, 1)$-inductive containment theorem: the n-dimensional case in which precisely one direction is expanding, while all the others contract. Previous proofs of containment required explicit a priori bounds on spatial derivatives, whereas our proof requires no such bounds.[2]

Let M_i be a parallelepiped in $\mathbf{R}^n$ with faces F_i^j, for $i = 0, \ldots, N$ and $j = \pm 1, \ldots, \pm n$, with opposite signs in the superscript representing opposite faces of a parallelepiped (see Figure 3.1). Without loss of generality, we assume that the first direction is the "expanding" one. We will denote the union of a set of faces by listing all of them in the superscript; for example, $F_i^{\pm 2, \ldots, \pm n}$ represents the set of all the faces of M_i except F_i^{-1} and F_i^{+1}. Let $\partial_E M_i \equiv F_i^{-1} \cup F_i^{+1} \equiv F_i^{\pm 1}$ and $\partial_C M_i \equiv \bigcup_{j=2}^n F_i^{-j} \cup F_i^{+j} \equiv F_i^{\pm 2, \ldots, \pm n}$. Let $\varphi : \mathbf{R}^n \to \mathbf{R}^n$ be a homeomorphism. Let

[2] Of course, our validated ODE integration (Nedialkov (1999)) must compute bounds on derivatives in order to compute enclosures, but these bounds are not a priori; they are computed on-the-fly, and if a bounds check fails, we can always try a smaller timestep to compensate.

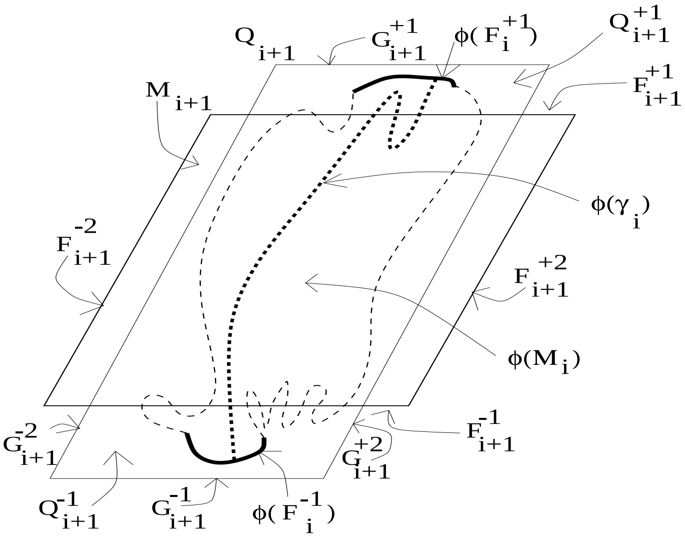

FIG. 3.1. *The image $\varphi(M_i)$ and M_{i+1} for two dimensions. The dark curves at the bottom and top are $\varphi(F_i^{\pm 1})$. The dashed curves at the left and right are $\varphi(F_i^{\pm 2})$.*

int X represent the interior of X. Then M_i and M_{i+1} satisfy the $(n, 1)$-ICP if

(1) $\varphi(F_i^{\pm 1}) \cap M_{i+1} = \emptyset$, and $\varphi(F_i^{-1})$ and $\varphi(F_i^{+1})$ are on opposite sides of the infinite slab between the two hyperplanes containing F_{i+1}^{-1} and F_{i+1}^{+1}.

(2) $\exists Q_{i+1}$, a parallelepiped in $\mathbf{R}^n$ with each face G_{i+1}^{j} parallel to the corresponding face F_{i+1}^{j} of M_{i+1} for $j = \pm 1, \ldots, \pm n$, such that
   (a) $\varphi(M_i) \subset \text{int } Q_{i+1}$,
   (b) $F_{i+1}^{\pm 2, \ldots, \pm n} \cap Q_{i+1} = \emptyset$, and $\forall j \in \{2, \ldots, n\}$, F_{i+1}^{-j} and F_{i+1}^{+j} are on opposite sides of the infinite slab between the two hyperplanes containing G_{i+1}^{-j} and G_{i+1}^{+j}.

Let $\gamma_0 \subset M_0$ be a simple curve joining F_0^{-1} to F_0^{+1}, but otherwise remaining in the interior of M_0. That is,

$$\gamma_0 \cap F_0^{-1} \neq \emptyset \;\; \wedge \;\; \gamma_0 \cap F_0^{+1} \neq \emptyset \;\; \wedge \;\; \text{int } \gamma_0 \subset \text{int } M_0.$$

THEOREM 3.1 ($(n, 1)$-inductive containment theorem). *If M_i and M_{i+1} satisfy the $(n, 1)$-ICP $\forall i = 0, \ldots, N-1$, then $\forall i = 0, \ldots, N$*

$$\exists \textit{ simple curve } \gamma_i \subseteq \varphi^i(\gamma_0) \textit{ s.t. } \gamma_i \cap F_i^{-1} \neq \emptyset \;\; \wedge \;\; \gamma_i \cap F_i^{+1} \neq \emptyset \;\; \wedge \;\; \text{int } \gamma_i \subset \text{int } M_i. \tag{3.1}$$

That is, γ_i touches the boundary of M_i in precisely two places, connecting F_i^{-1} to F_i^{+1}, but otherwise remains entirely inside M_i.

Proof. We proceed by induction on i. The proof of the base case $i = 0$ is immediate, by the definition of γ_0. For the inductive case, assume $\exists$ a simple curve $\gamma_i \subseteq \varphi^i(\gamma_0)$ such that $\gamma_i \cap F_i^{-1} \neq \emptyset \;\wedge\; \gamma_i \cap F_i^{+1} \neq \emptyset \;\wedge\; \text{int } \gamma_i \subset \text{int } M_i$. From ICP(1), $\varphi(F_i^{\pm 1}) \subset Q_{i+1}$, and the fact that Q_{i+1} is convex, we know that Q_{i+1} intersects both F_{i+1}^{-1} and F_{i+1}^{+1}; and from ICP(2(b)), Q_{i+1} does not intersect $F_{i+1}^{\pm 2, \ldots, \pm n}$. Thus, since Q_{i+1} is convex, $Q_{i+1} - M_{i+1}$ is disconnected by the slab defined in ICP(1) into two

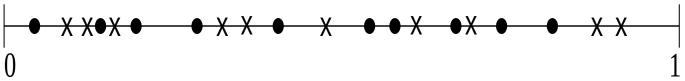

FIG. 3.2. *Schematic representation of the sets $\gamma^{-1}(s^{-1})$ (dots) and $\gamma^{-1}(s^{+1})$ ($\times$'s).*

disjoint components,[3] say Q_{i+1}^{-1} and Q_{i+1}^{+1}, each containing one of $\varphi(F_i^{\pm 1})$, by ICP(1). Without loss of generality, assume $\varphi(F_i^j) \subset Q_{i+1}^j$, $j = \pm 1$. Now, consider one of the components, say Q_{i+1}^{-1}. It contains one of the two endpoints of $\varphi(\gamma_i)$, since one endpoint is in $\varphi(F_i^{-1}) \subset Q_{i+1}^{-1}$, while the other endpoint of $\varphi(\gamma_i)$ is in $\varphi(F_i^{+1}) \subset Q_{i+1}^{+1}$. Since γ_i is a simple curve and φ is a homeomorphism, $\varphi(\gamma_i)$ is a simple curve. Now, $Q_{i+1}^{-1} \cap Q_{i+1}^{+1} = \emptyset$, and $\varphi(\gamma_i)$ connects the two. Thus, $\varphi(\gamma_i)$ must cross the boundary of Q_{i+1}^{-1}. This boundary consists of exactly two mutually exclusive patches, one of which is a subset of ∂Q_{i+1}, the other a subset of F_{i+1}^{-1}. Since $\varphi(\gamma_i) \subset \varphi(M_i) \subset \text{int } Q_{i+1}$, we infer that $\varphi(\gamma_i) \cap \partial Q_{i+1} = \emptyset$, and so $\varphi(\gamma_i)$ leaves Q_{i+1}^{-1} through F_{i+1}^{-1}. A similar argument shows that $\varphi(\gamma_i)$ leaves Q_{i+1}^{+1} through F_{i+1}^{+1}. Thus, $\varphi(\gamma_i) \cap F_{i+1}^j \neq \emptyset$, $j = \pm 1$. It remains to show that there exists a segment γ_{i+1} of $\varphi(\gamma_i)$ which is a simple curve and maintains the property defined in (3.1).

Since $\varphi(\gamma_i)$ is a simple curve, there exists a parameterization $\gamma(t)$ for $t \in [0,1]$ such that $\gamma([0,1]) = \varphi(\gamma_i)$ and $\gamma(t)$ is a homeomorphism (Munkres (1975)). Let $s^j = \varphi(\gamma_i) \cap F_{i+1}^j$, $j = \pm 1$. Now, s^{-1} and s^{+1} are disjoint sets since $F_{i+1}^{-1} \cap F_{i+1}^{+1} = \emptyset$, and they are compact because (1) F_{i+1}^j for $j = \pm 1$ are compact; (2) γ_i is compact, φ is a homeomorphism, and so $\varphi(\gamma_i)$ is compact; and (3) the intersection of two compact sets in $\mathbf{R}^n$ is compact. Finally, $\gamma^{-1}(s^{\pm 1})$ are compact because γ is a homeomorphism. To prove that there exists a simple curve $\gamma_{i+1} \subset \varphi(\gamma_i)$ such that $\gamma_{i+1} \cap F_{i+1}^{-1} \neq \emptyset$, $\gamma_{i+1} \cap F_{i+1}^{+1} \neq \emptyset$, and int $\gamma_{i+1} \subset$ int M_{i+1}, we need to show that there exist two points in $[0,1]$, one each from $\gamma^{-1}(s^{-1})$ and $\gamma^{-1}(s^{+1})$, such that no points from either set are between them (see Figure 3.2). This will prove that there exists a simple curve, which is a section of $\varphi(\gamma_i)$, that connects F_{i+1}^{-1} to F_{i+1}^{+1} without otherwise intersecting ∂M_{i+1}. To this end, let $G = \gamma^{-1}(s^{-1})$ and $R = \gamma^{-1}(s^{+1})$, and note that G and R are compact, disjoint, nonempty subsets of $[0,1]$. The following lemma completes the proof. □

LEMMA 3.2. *Let G and R be (possibly infinite) disjoint, compact, nonempty subsets of $[0,1]$. Then $\exists g \in G$, $r \in R$ such that $(g,r) \cap (G \cup R) = \emptyset$, where we have assumed without loss of generality than $g < r$.*

Proof. Consider the function $f(x,y) = |x - y|$ over the subset $G \times R$ of the plane. Since f is continuous and $G \times R$ is compact, f attains its minimum at some point $(g,r) \in G \times R$. That is, $|g - r| \leq |g' - r'|$ for any other $g' \in G$, $r' \in R$. Thus, there is no element of either set G or R between g and r, and so the open interval (g,r) is disjoint from $G \cup R$. □

THEOREM 3.3 (shadowing containment theorem). *Let φ be a homeomorphism. Let $\{M_i\}_{i=0}^N$ be a sequence of parallelepipeds enclosing a pseudotrajectory $\{\mathbf{y}_i\}_{i=0}^N$. Let ε be the maximum diameter of M_i for $i = 0, 1, \ldots, N$. Let $\gamma_i \subset M_i, \gamma_i \neq \emptyset$, $i = 0, \ldots, N$, and let $\gamma_{i+1} \subseteq \varphi(\gamma_i)$, $i = 0, \ldots, N-1$. Then $\exists$ an ε-shadow $\{\mathbf{x}_i\}_{i=0}^N$ of $\{\mathbf{y}_i\}_{i=0}^N$. That is, there is an exact trajectory $\{\mathbf{x}_i\}_{i=0}^N$ of φ such that $\|\mathbf{x}_i - \mathbf{y}_i\| < \varepsilon$, $i = 0, \ldots, N$.*

[3] This is because F_{i+1}^{-1} and F_{i+1}^{+1} are each patches of an $(n-1)$-dimensional hyperplane residing in n dimensions, and so they each disconnect any convex set they intersect, as long as that convex set does not intersect their boundaries ∂F_{i+1}^{-1} and ∂F_{i+1}^{+1}, respectively.

Proof. Since φ is a homeomorphism, φ^{-1} is a well-defined function. Pick any point $\mathbf{x}_N \in \gamma_N$, and recursively define $\mathbf{x}_i = \varphi^{-1}(\mathbf{x}_{i+1}),\ i = N-1, N-2, \dots, 0$. Since $\gamma_{i+1} \subseteq \varphi(\gamma_i)$, $\varphi^{-1}(\gamma_{i+1}) \subseteq \gamma_i$, and so by induction $\mathbf{x}_i \in \gamma_i$ for $i = N, N-1, \dots, 0$. Since $\mathbf{y}_i \in M_i$ and $\mathbf{x}_i \in \gamma_i \subset M_i$, $\|\mathbf{y}_i - \mathbf{x}_i\| \leq \operatorname{diam}(M_i) \leq \varepsilon,\ i = 0, \dots, N$. □

Thus, applying Theorem 3.3 to an orbit satisfying the $(n,1)$-ICP implies the existence of a shadow.

Remark 3.1. Note that Theorem 3.3 is independent of the number of dimensions n, and of the number of expanding and contracting directions, because the only parts of the inductive containment theorem that are used are the conclusions that $\gamma_{i+1} \subseteq \varphi(\gamma_i)$ for $i = 0, 1, \dots, N-1$ and $\emptyset \neq \gamma_i \subset M_i$ for $i = 0, \dots, N$. The 0-expanding and 0-contracting directions are handled separately. We conjecture that the general (n,k)-inductive containment theorem (work in progress) will also assert this property, so that the above shadowing containment theorem is applicable to the general (n,k) case, in which k directions are expanding and $n-k$ are contracting.

As mentioned previously, the case with one *contracting* dimension while the other $n-1$ directions expand can be handled simply by reversing the arrow of time and applying the above argument. We call this the $(n, n-1)$ case. Another proof, which is more likely to be generalizable to an arbitrary number of expanding and contracting directions, is presented in Hayes (2001).

3.2. Containment with zero contracting or zero expanding directions. For completeness, we mention the trivial cases in which all directions are contracting, or all directions are expanding. We call these the $(n,0)$ and (n,n) cases, respectively. The former case is entirely trivial, because the problem is stable: if $\varphi(M_i) \subset M_{i+1}$ for all i, then clearly any exact solution starting in M_0 will be in M_i for all $i > 0$. Similarly, if all directions are expanding, then we apply the same argument in the reverse direction: if $\varphi^{-1}(M_{i+1}) \subset M_i$ for all i, then any exact solution *finishing* in M_N, traced backwards, lies in M_i for $i = N-1, N-2, \dots, 0$.

3.3. Discussion. The four cases $(n,0), (n,1), (n,n-1)$, and (n,n) cover *all* cases for $n = 1, 2, 3$. That is, the theorems in this paper can prove the existence of shadows for any n-dimensional system, $n \leq 3$, in which some measure of pseudohyperbolicity is present. Furthermore, although the proofs, for simplicity, deal only with a single function φ, the induction argument could just as easily use a *different* function φ_i at each step. In particular, φ_i could be the ODE time-h_i solution operator φ_{h_i}. Thus, modulo a rescaling of time (which we discuss below), the above proofs can be used to find shadows of noisy trajectories of ODE systems, as well as maps, with up to three dependent variables. They can also be used in the case of n dependent variables, with the restriction that solutions have either one expanding and $n-1$ contracting directions, or one contracting and $n-1$ expanding directions.

Finally, we believe that a generalized (n,k)-ICP implies the existence of a shadow (work in progress; more discussion in Hayes (2001)).

3.4. Proving the existence of chaotic orbits. Following the analysis of Stoffer and Palmer (1999), we describe how to use containment to prove the existence of chaotic orbits. We quote directly from their introduction.

> The idea is to construct two periodic pseudo-orbits which happen to be close to each other at some point. We call this the branching point. Then it is possible to construct an infinite number of pseudo-orbits as follows. You follow one or the other of the periodic pseudo-orbits. When you reach the branching point, you either stay on your periodic orbit for at least one more loop, or else you switch to the other periodic orbit. Each time you

> arrive at the branching point you can again choose to stay or to switch, ad infinitum. Assume that for each such pseudo-orbit there is a unique *orbit* of the system which is close to the pseudo-orbit. Then the dynamical system indeed behaves chaotically, at least in a certain neighbourhood of the two periodic pseudo-orbits. (Stoffer and Palmer 1999)

To use this approach together with containment to prove the existence of a chaotic orbit, assume that the first orbit has a sequence of N parallelepipeds M_i satisfying the (n,k)-ICP with $M_N = M_0$. Then the (n,k)-inductive containment theorem can be invoked ad infinitum around this periodic pseudo-orbit and proves the existence of infinitely long exact orbits that remain in the vicinity of this pseudo-orbit.[4] Similarly, assume that the second orbit has a sequence of P parallelepipeds Q_i satisfying the (n,k)-ICP with $Q_P = Q_0$. Assume further that $M_i = Q_j$ for some i, j. Then $M_i = Q_j$ is the branching point, the (n,k)-inductive containment theorem can be invoked ad infinitum around *both* of these pseudo-orbits, and each time we pass $M_i = Q_j$ we can choose which pseudo-orbit to follow. The (n,k)-inductive containment theorem proves that a shadow follows us as we go.

4. Verifying the inductive containment property.

We present one method of verifying that the general (n,k)-ICP holds for a given pseudotrajectory derived from the numerical solution of an ODE. (Three more methods for verifying the ICP are presented in Hayes (2001).) We note in passing that this scheme (as well as the other three discussed in Hayes (2001)) could easily be adapted to the simpler problem of maps. We require the use of validated interval arithmetic, or a validated ODE integrator if φ derives from an ODE. The validated ODE integrator that we use is called VNODE (Nedialkov (1999); Nedialkov, Jackson, and Corliss (1999)). VNODE works with n-dimensional parallelepipeds and satisfies the following property: given an n-dimensional parallelepiped A and a timestep h, VNODE will return an n-dimensional parallelepiped B such that $\varphi_h(A) \subset B$, where φ_h is the time-h solution operator. For the purposes of this description, we will denote the output B by $\bar{\varphi}_h(A)$. Thus,

$$\varphi_h(A) \subset \bar{\varphi}_h(A).$$

We will usually omit the timestep parameter h; we will talk only of φ, keeping in mind that, in the induction, φ can be different for each step.

We verify the ICP using an iterative method that we have found empirically to require about 3–4 validated integrations per step on average, independent of n. This method rigorously verifies the ICP in the cases for which we have proven the inductive containment theorem and is the method we actually used to produce our numerical results. Three noniterative, deterministic methods are presented in Hayes (2001); however, we found this method to be the most efficient with the validated ODE solver we used (Nedialkov (1999)).

We first look at the simple two-dimensional case in which one of the directions is expanding, while the other is contracting. To begin, assume that the only information provided by our validated ODE integration is an outer bound $\bar{\varphi}(M_i)$ on $\varphi(M_i)$. Then, it is *not* possible to verify the $(2,1)$-ICP with only one validated integration, because this information can only prove contraction, not expansion, as shown in Figure 4.1. In both Figures 4.1(a) and 4.1(b), $\bar{\varphi}(M_i)$ is a valid enclosure of $\varphi(M_i)$. In both figures, $\bar{\varphi}(M_i)$ can be used to prove that $\varphi(M_i)$ has contracted in the horizontal

[4] Slightly more is required to prove the existence of periodic orbits or to prove uniqueness.

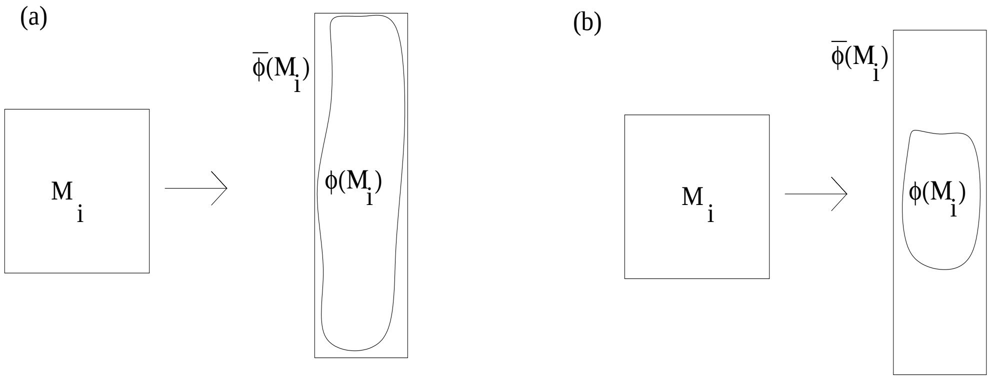

FIG. 4.1. *Enclosure methods can prove contraction but not expansion.*

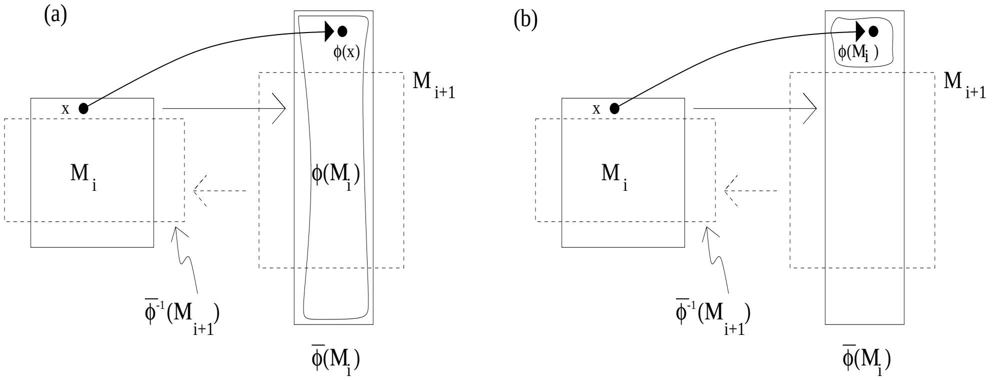

FIG. 4.2. (a) *The two validated integrations required to prove the* $(2,1)$*-ICP.* (b) *A potential problem, which is solved by doing a (cheap) point integration of one point on each expanding face, to verify that there are points of* $\varphi(M_i)$ *on both side of* M_{i+1}.

direction. However, enclosure methods cannot directly prove expansion, as Figure 4.1(b) illustrates: although $\bar{\varphi}(M_i)$ is a valid enclosure of $\varphi(M_i)$, it is not a very good one, because the actual image $\varphi(M_i)$ of M_i has not expanded in any direction. To solve this problem, we perform two validated integrations; refer to Figure 4.2(a). The first integration (solid rectangles) is a forward integration that provides $\bar{\varphi}(M_i)$, which in turn gives us a bound on the size of $\varphi(M_i)$ in the contracting directions (depicted as the horizontal direction in the figure). Now, assume we can find an M_{i+1} which satisfies the ICP not with $\varphi(M_i)$, but with $\bar{\varphi}(M_i)$. (If we cannot find such an M_{i+1}, then our method fails and we cannot prove the existence of a shadow beyond step i.) A validated integration *backwards* (dashed rectangles) is then performed on M_{i+1}, giving $\bar{\varphi}^{-1}(M_{i+1})$. If $\bar{\varphi}^{-1}(M_{i+1})$ proves that *contraction* has occurred in the nominally expanding directions when moving back from M_{i+1} to M_i, then we argue that expansion in forward time has occurred, as follows. Choose any $\mathbf{x} \in M_i - \bar{\varphi}^{-1}(M_{i+1})$. Since $\mathbf{x} \notin \bar{\varphi}^{-1}(M_{i+1}) \supset \varphi^{-1}(M_{i+1})$, this implies $\varphi(\mathbf{x}) \in \varphi(M_i) - M_{i+1}$. Since $F_i^{\pm 1} \subset M_i - \bar{\varphi}^{-1}(M_{i+1})$, this tells us that $\varphi(F_i^{\pm 1}) \cap M_{i+1} = \emptyset$. This is insufficient to prove ICP(1), as illustrated in Figure 4.2(b): perhaps $\bar{\varphi}(M_i)$ is a loose enclosure of $\varphi(M_i)$, and all of $\varphi(M_i)$ is actually on one side of M_{i+1}. To verify that this is not the case, we pick one point on each of F_i^{+1} and F_i^{-1} and perform a validated point

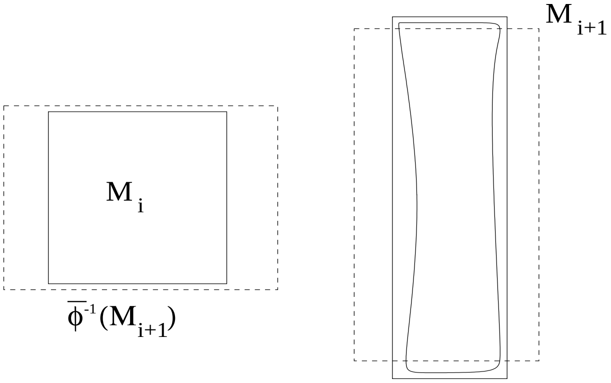

FIG. 4.3. *Shortcomings of the two-integration method: sometimes it can not prove expansion even if the M_{i+1} is valid.*

integration of each (which can be done cheaply) to verify that they land on opposite sides of M_{i+1}.[5] Since there is exactly one expanding direction, M_{i+1} cuts $\bar{\varphi}(M_i)$ into two disjoint sets, and a simple continuity argument shows that the two faces in their entirety land on opposite sides of M_{i+1}, thus verifying ICP(1). A similar argument in reverse time shows that the chosen M_{i+1} also verifies ICP(2(b)).

The argument of the previous paragraph clearly applies just as well in n dimensions when there is one expanding direction and $n-1$ contracting directions, for the same reasons that the two-dimensional proof of containment is easily transformed into Theorem 3.1 (Hayes (2001)). To prove that it also works when there is one contracting direction and $n-1$ expanding directions, note that there is a precise symmetry between the two cases (one expanding vs. one contracting): if we simultaneously reverse the order of $\{M_i\}_{i=0}^N$, giving $L_i = M_{N-i}$, and let $\psi = \varphi^{-1}$, then the above argument applies to the sequence $\{L_i\}_{i=0}^N$ using ψ as the homeomorphism. Thus, by symmetry, this method is also rigorous in the case when there is one contracting direction and $n-1$ expanding ones.

Figure 4.3 illustrates that it is possible to choose an M_{i+1} that satisfies the ICP but for which we cannot *verify* that the ICP holds. This occurs when M_{i+1} is chosen to be "almost as large" as $\bar{\varphi}(M_i)$ in the expanding directions; then, the excess when computing $\bar{\varphi}^{-1}(M_{i+1})$ swamps the contraction that occurs when integrating the expanding direction backwards in time. We solve this problem by iteratively shrinking M_{i+1} in the nominally expanding directions until $\bar{\varphi}^{-1}(M_{i+1})$ fits inside M_i in those directions. If we shrink M_{i+1} to size zero in the expanding direction without being able to integrate it backwards to fit inside M_i, then the method fails, and we cannot prove the existence of a shadow beyond step i. We have found empirically that, when the algorithm is succeeding, no more than 2 to 3 backwards integrations are usually required, independent of n. The number of backwards integrations is occasionally significantly larger, when the system encounters areas of nonhyperbolicity.

If the system were hyperbolic, then the nominally expanding directions would always expand, and the nominally contracting directions would always contract. However, in systems that are only pseudohyperbolic, the nominally expanding directions

[5] We have found empirically that this problem must be very rare, because it has not happened even once during our experiments. We suspect that it may be possible to prove the ICP without this extra point integration, but we have not devoted much thought to this matter.

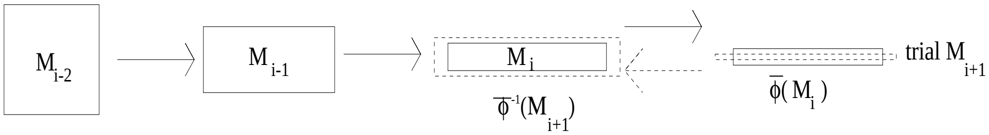

FIG. 4.4. *Example of the nominally expanding direction contracting too much for our integrator to prove contraction in the backwards direction.*

may expand most of the time, but not always, and vice versa for the contracting directions. One of the reasons our shadowing method can fail is if a nominally expanding direction contracts too much or for too long a time (Figure 4.4). Then, the expanding dimensions of M_i can become so small that no backwards integration from M_{i+1} can fit inside M_i in the nominally expanding directions.

4.1. Implementation issues and discussion. In the original paper that described containment, Grebogi et al. (GHYS) appear to have used boxes M_i of fixed size and found that smaller boxes seemed to work better. In contrast, our method dynamically grows and shrinks the M_i as i progresses, in an effort to maintain the ICP. In fact, we find it advantageous to choose the expanding dimension of M_i to be fairly large, to allow us to "absorb" possible future nonexpansion, in an effort to avoid the situation depicted in Figure 4.4. Similarly, we choose the contracting dimensions to be relatively small, to avoid the opposite effect (allowing us to "absorb" noncontraction without the nominal contracting dimensions becoming too large). Practically, we find that our "boxes" can be extremely long and thin: typically, they are of length 10^{-3} to 10^{-6} in the expanding dimensions, and as small as 10^{-12} to 10^{-14} in the contracting dimensions.

Referring once again to Figure 4.4, we note that when containment fails, the "expanding" dimension of M_i has often shrunk to almost the same size as the contracting dimension, and both can be quite small (say, 10^{-12}), whereas when containment is "working," the expanding dimension of M_i can be several orders of magnitude larger than the contracting dimension. It is interesting to note that this implies that the hardest parts of an orbit to shadow are the places where our bounds on the distance between the noisy and shadow orbits are *smallest*, i.e., where we can prove that they are unusually close together. This appears counterintuitive but may be related to the one-dimensional result of Chow and Palmer (1991), in which they proved that shadows must maintain a minimum distance from the noisy orbit.

5. Rescaling time.

5.1. Informal description. Containment as presented thus far has put no restrictions on φ other than that it is a homeomorphism. As has also been mentioned, all of our theorems and proofs have been based on a single application of φ, and there is no explicit connection between the φ used at one step and the one used on the next. Thus, everything said thus far is also applicable if we allow φ to change between steps. In particular, at each step we could use the time-h_i solution operator φ_{h_i}, with h_i being the length of the ODE integration timestep taken at step i. The resulting method for shadowing numerical ODE integrations has been dubbed the *map method* by Coomes, Koçak, and Palmer (1994b), (1995a), (1995b). However, ODE integrations suffer from errors in time. For systems in which the $\mathbf{y}'$ direction lacks even pseudohyperbolicity, errors in time (which manifest themselves in phase space as errors in the $\mathbf{y}'$ direction) can lead to short shadowing times that can be dramatically increased if time is *rescaled*. In this section, we describe how containment can be augmented to allow for the rescaling of time.

Our idea for rescaling time in containment was inspired in part by the rescaling of time developed by Coomes, Koçak, and Palmer (1994b), (1995a) (although our proofs are very different from theirs), and partly by the idea of the *Poincaré section*, also known as a *Poincaré map* or *return map*. There are several variations on this idea, but the one that concerns us is the following. Assume that the solution to an ODE is "almost periodic," in the sense that the solution passes through some fixed neighborhood of a given plane $\mathcal{H}$ approximately every T time units, where $\mathcal{H}$ is approximately perpendicular to the trajectory at the point where it crosses the plane. The Poincaré map generates the sequence of points at which the trajectory intersects $\mathcal{H}$. To accomplish the general rescaling of time, we modify this idea to remove the almost-periodic requirement of the orbit, and simply place a plane $\mathcal{H}_i$ in the vicinity of the solution at time t_i, placed so that $\mathcal{H}_i$ is approximately perpendicular to $\mathbf{y}'(t_i)$. Note that we do not compute $\mathcal{H}_i$; we only prove that it exists.

To facilitate containment, we must extend the idea of the Poincaré section to encompass a small ensemble of solutions. To that effect, we wish to take a set $M_{i-1} \subset \mathcal{H}_{i-1}$, where the diameter of M_{i-1} is small, and place an $(n-1)$-dimensional hyperplane $\mathcal{H}_i$ approximately normal to the flow in the vicinity of $\varphi_{h_{i-1}}(M_{i-1})$. Then we define the Poincaré section of the set $\varphi_{h_{i-1}}(M_{i-1})$ pointwise as follows. Let Δh_{i-1} bound the time interval over which the ensemble $\varphi_{h_{i-1}}(M_{i-1})$ crosses $\mathcal{H}_i$:

$$\forall \mathbf{x} \in M_{i-1} \quad \exists h \in [h_{i-1} - \Delta h_{i-1}/2, h_{i-1} + \Delta h_{i-1}/2] \quad \text{s.t. } \varphi_h(\mathbf{x}) \in \mathcal{H}_i,$$

where we assume that, for each $\mathbf{x}$, the h chosen is unique. That is, we take the point-by-point Poincaré section of the points in M_{i-1} with respect to the plane $\mathcal{H}_i$. We call this a *splash* operation, because we imagine that the points in M_{i-1}, evolving via φ_h for $h \in [h_{i-1} - \Delta h_{i-1}/2, h_{i-1} + \Delta h_{i-1}/2]$, "splash" through $\mathcal{H}_i$ approximately simultaneously, and we assume that each trajectory intersects $\mathcal{H}_i$ precisely once during that interval; see Figure 5.1.

Our intent is to build $(n-1)$-dimensional parallelepipeds M_i inside $\mathcal{H}_i$ and then show that the point-by-point Poincaré section at $\mathcal{H}_i$—the splash operation—is a homeomorphism. We can then directly apply the previously proven containment theorems to the $(n-1)$-dimensional M_i's, which are each contained in the $(n-1)$-dimensional hyperplane $\mathcal{H}_i$, for an ODE system of n equations.

We note that since rescaling time via the splash operation effectively deletes one dimension from the problem, and since our map containment theorems are rigorous in three dimensions, this means that the methods presented in this paper are capable of rigorously shadowing ODE solutions of up to four dimensions, as long as a rescaling of time is applied.

5.2. Theorem: Splash is a homeomorphism. Refer to Figure 5.2. Let Q_i be an n-dimensional parallelepiped. Let $F_i^{\pm 1}$ be the two opposing faces of Q_i that are approximately normal to $\mathbf{y}'$ inside Q_i, and let $\mathbf{v}_i$ be the unit normal vector to these two faces, with $\mathbf{v}_i$ pointing from F_i^{-1} to F_i^{+1}. That is, $\mathbf{v}_i$ is approximately parallel to $\mathbf{y}'$ inside Q_i. Let D be the distance between F_i^{-1} and F_i^{+1} along $\mathbf{v}_i$. Let the infinite hyperplanes containing F_i^{-1} and F_i^{+1} be H_i^{-1} and H_i^{+1}, respectively, and let Z_i be the closed infinite slab between them. Let B_i be a parallelepiped with faces parallel to Q_i satisfying $Q_i \subset B_i \subset Z_i$, with two of the faces of B_i contained in $H_i^{\pm 1}$. Let $\{\mathbf{f}(\mathbf{x}) \cdot \mathbf{v}_i \mid \mathbf{x} \in B_i\} \subset [v_0, v_1]$, and assume $0 < v_0 \leq v_1$.

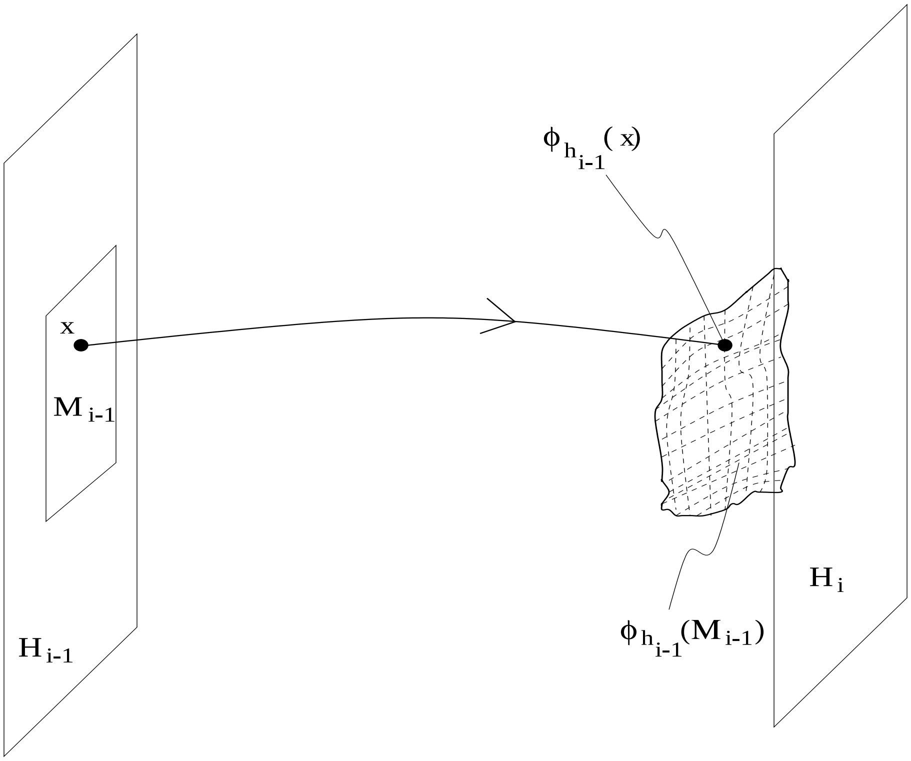

FIG. 5.1. *The "splash" operation depicted for a two-dimensional ensemble evolving in a three-dimensional configuration space. M_{i-1} is embedded in the plane $\mathcal{H}_{i-1}$ and evolves through one timestep to $\varphi_{h_{i-1}}(M_{i-1})$. As depicted, the ensemble is about to splash through $\mathcal{H}_i$.*

LEMMA 5.1. *If a trajectory remains in B_i while it is in Z_i, then it remains in Z_i for at least time $\underline{\varepsilon}_i^t \equiv D/v_1$ and at most $\bar{\varepsilon}_i^t \equiv D/v_0$.*

Proof. Let $\mathbf{y}(t)$ be a trajectory that remains in B_i while it is in Z_i. Let $z(t) = \mathbf{y}(t) \cdot \mathbf{v}_i$. Since $0 < v_0 \leq z'(t) \leq v_1$ and the width of B_i in the $\mathbf{v}_i$ direction is D, the maximum time to cross B_i is D/v_0, while the minimum time to cross is D/v_1. □

Let $\bar{\mathbf{f}}(B_i)$ be an enclosure of $\{\mathbf{f}(\mathbf{x}) \mid \mathbf{x} \in B_i\}$. Let S_i be a parallelepiped enclosure of $\{Z_i \cap (Q_i + h\bar{\mathbf{f}}(B_i)) \mid h \in [-\bar{\varepsilon}_i^t, \bar{\varepsilon}_i^t]\}$, and assume $S_i \subseteq B_i$.

Remark 5.1. S_i is intended to enclose the distance that a trajectory can drift from Q_i along the direction approximately perpendicular to $\mathbf{y}'$ as it travels across Z_i. This is required because a point in Q_i may not remain in Q_i when it is "splashed" onto $\mathcal{H}_i$. The following lemma formalizes this statement.

LEMMA 5.2. *Any trajectory intersecting Q_i remains in S_i while in Z_i, and thus remains in B_i as well.*

Proof. Since $S_i \subseteq B_i$, $\bar{\mathbf{f}}(B_i)$ bounds $\mathbf{y}' \equiv \mathbf{f}$ inside S_i. Since a trajectory remaining in B_i as it crosses Z_i does so in time $\leq \bar{\varepsilon}_i^t$, and since $S_i \subset B_i$, $\{h\bar{\mathbf{f}}(B_i) \mid h \in [-\bar{\varepsilon}_i^t, \bar{\varepsilon}_i^t]\}$ encloses the maximum possible distance from Q_i that a trajectory can travel in time $|\bar{\varepsilon}_i^t|$ while it remains in B_i. Thus, since $Q_i \subset S_i \subseteq B_i$, $\{Q_i + h\bar{\mathbf{f}}(B_i) \mid h \in [-\bar{\varepsilon}_i^t, \bar{\varepsilon}_i^t]\}$ encloses the position of any trajectory $\mathbf{y}(t)$ that is within time $\bar{\varepsilon}_i^t$ of intersecting Q_i, unless $\mathbf{y}(t)$ leaves Z_i during that time. Intersecting with Z_i completes the proof. □

Let $\mathcal{H}_i$ be any plane perpendicular to $\mathbf{v}_i$ which intersects the interior of Q_i. That

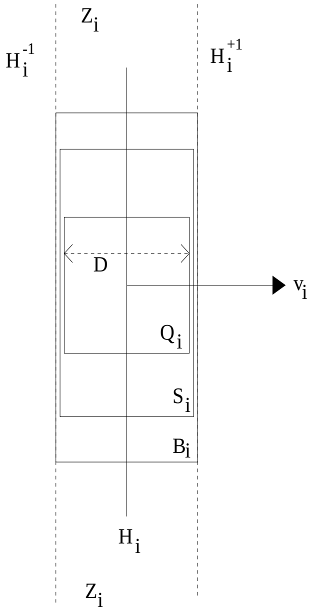

FIG. 5.2. *The objects used in Lemmas* 5.1–5.4. *Note that the left and right sides of Q_i, S_i, B_i, and Z_i are all in the planes H_i^{-1}, H_i^{+1}, respectively; they have been drawn as distinct for illustrative purposes only.*

is, $\mathcal{H}_i$ lies strictly between H_i^{-1} and H_i^{+1}.

LEMMA 5.3. *Every trajectory intersecting Q_i intersects $\mathcal{H}_i$ at precisely one point while it crosses Z_i.*

Proof. Let $\mathbf{y}(t)$ be a trajectory that intersects Q_i. By Lemma 5.2, $\mathbf{y}(t)$ remains in $S_i \subseteq B_i$ while it crosses Z_i. Let $z(t) = \mathbf{y}(t) \cdot \mathbf{v}_i$. Let the z coordinates of $H_i^{-1}, \mathcal{H}_i, H_i^{+1}$ be z_{-1}, z_0, z_{+1}, respectively. While the trajectory remains in $S_i \subseteq B_i$, $z'(t) \geq v_0 > 0$, and, since $z(t)$ is continuous, it increases monotonically while $\mathbf{y}(t)$ remains in S_i, taking on each value between z_{-1} and z_{+1} precisely once, by the intermediate value theorem. In particular, it takes on the value z_0 precisely once and thus crosses $\mathcal{H}_i$ precisely once. □

Assume that Q_i is an enclosure of $\varphi_{h_{i-1}}(M_{i-1})$. For a point $\mathbf{x} \in M_{i-1}$, let $\varphi_{i-1}(\mathbf{x})$ be the unique point in $\mathcal{H}_i$ defined by Lemma 5.3. Let $\bar{M}_i = S_i \cap \mathcal{H}_i$. Clearly, $\bar{M}_i$ is an enclosure of $\varphi_{i-1}(M_{i-1})$. To show that φ_{i-1} applied to M_{i-1} is a homeomorphism, we need to show that it is continuous and one-to-one. We first prove it is one-to-one.

Let $\varepsilon^t > 0$ be given. Recall $\bar{\varepsilon}_i^t$ as defined in Lemma 5.1.

Assumption 1. Assume $\bar{\varepsilon}_i^t < \varepsilon^t$ and $\nexists$ distinct $\mathbf{x}, \mathbf{y} \in M_{i-1}$ such that $\mathbf{y} = \varphi_t(\mathbf{x})$ for $|t| < \varepsilon^t$.

Each of the assumptions introduced in this section is assumed to hold throughout the remainder of section, once it is introduced.

LEMMA 5.4. *Each point in $\varphi_{i-1}(M_{i-1})$ comes from only one point in M_{i-1}.*

Proof. Assume to the contrary that there exist distinct $\mathbf{x}, \mathbf{y} \in M_{i-1}$ such that $\varphi_{i-1}(\mathbf{x}) = \varphi_{i-1}(\mathbf{y}) = \mathbf{z} \in \bar{M}_i$. Since $\varphi_{h_{i-1}}(\mathbf{x}), \varphi_{h_{i-1}}(\mathbf{y})$ both splash to $\mathbf{z}$, they are on the same trajectory, and since they are both in Q_i, the time-shift between them is $\leq \bar{\varepsilon}_i^t$. Thus, $\exists t_1, t_2$ such that $\varphi_{t_1}(\mathbf{x}) = \mathbf{z} = \varphi_{t_2}(\mathbf{y})$ with $|t_1 - t_2| \leq \bar{\varepsilon}_i^t$. Then $\mathbf{y} = \varphi_{t_1 - t_2}(\mathbf{x})$, contradicting Assumption 1. □

THEOREM 5.5. *φ_{i-1} applied to M_{i-1} is one-to-one.*

Proof. Lemma 5.3 proves that $\varphi_{i-1}(M_{i-1})$ is many-to-one, and Lemma 5.4 proves it is one-to-many. Thus, it is one-to-one. □

We now prove that $\varphi_{i-1}(\mathbf{x})$ is continuous for all $\mathbf{x} \in M_{i-1}$.

Assumption 2. $\varphi_t(\mathbf{x})$ exists and is continuous in both t and $\mathbf{x}$ $\forall \mathbf{x} \in M_{i-1}$ and $\forall t$ such that $\varphi_t(\mathbf{x}) \in B_i$. Note that this is true as long as $\mathbf{f}$ is Lipschitz continuous (Stuart and Humphries (1996, Theorem 2.1.12)).

We will need the following theorem.

THEOREM 5.6. *If $\mathbf{y}$ and $\mathbf{z}$ each satisfy the differential equation $\mathbf{y}'(t) = \mathbf{f}(\mathbf{y}(t))$ on the interval $[t_0, t_1]$, and if $\mathbf{f}$ is Lipschitz continuous with constant L, then $\forall t \in [t_0, t_1]$,*

$$\|\mathbf{y}(t) - \mathbf{z}(t)\| \leq \|\mathbf{y}(t_0) - \mathbf{z}(t_0)\| e^{L(t-t_0)}.$$

Proof. See Theorem 112J of Butcher (1987). □

For a point $\mathbf{x} \in M_{i-1}$, let $h_{i-1}(\mathbf{x})$ be that unique timestep defined by $\varphi_{h_{i-1}(\mathbf{x})}(\mathbf{x}) \in \mathcal{H}_i$. That is, $\varphi_{i-1}(\mathbf{x})$ specifies *where* $\mathbf{x}$ goes, and $h_{i-1}(\mathbf{x})$ specifies *how long* it takes to get there.

LEMMA 5.7. *If $\mathbf{f}$ is Lipschitz continuous, then $\forall \mathbf{x}_0 \in M_{i-1}$, $h_{i-1}(\mathbf{x})$ is continuous at $\mathbf{x} = \mathbf{x}_0$.*

Proof. For simplicity, we will drop the subscript from $h_{i-1}(\mathbf{x})$ during this proof.

Let L be the Lipschitz constant for $\mathbf{f}$. Then by Theorem 5.6, for any $\mathbf{x}_0, \mathbf{x}$,

$$\|\varphi_{h(\mathbf{x}_0)}(\mathbf{x}) - \varphi_{h(\mathbf{x}_0)}(\mathbf{x}_0)\| \leq \|\mathbf{x} - \mathbf{x}_0\| e^{Lh(\mathbf{x}_0)} \equiv \delta_3(\mathbf{x}, \mathbf{x}_0).$$

Since we are interested only in the behavior of $h(\mathbf{x})$ in a neighborhood of $\mathbf{x}_0$, choose $\mathbf{x} \in M_{i-1}$ close enough to $\mathbf{x}_0$ so that $\varphi_{h(\mathbf{x}_0)}(\mathbf{x}) \in B_i$. Now, since $\varphi_{h(\mathbf{x}_0)}(\mathbf{x}_0) \in \mathcal{H}_i$, the distance from $\varphi_{h(\mathbf{x}_0)}(\mathbf{x})$ to $\mathcal{H}_i$ is also bounded above by $\delta_3(\mathbf{x}, \mathbf{x}_0)$. Since $\varphi_{h(\mathbf{x}_0)}(\mathbf{x}) \in B_i$, the maximum time to intersect $\mathcal{H}_i$ is $\delta_3(\mathbf{x}, \mathbf{x}_0)/v_0$. Thus, $h(\mathbf{x}) \in [h(\mathbf{x}_0) - \delta_3(\mathbf{x}, \mathbf{x}_0)/v_0, h(\mathbf{x}_0) + \delta_3(\mathbf{x}, \mathbf{x}_0)/v_0]$. The continuity of $h(\mathbf{x})$ at $\mathbf{x}_0$ follows by letting $\mathbf{x} \to \mathbf{x}_0$. □

LEMMA 5.8. *$\varphi_{i-1}(\mathbf{x})$ is continuous $\forall \mathbf{x} \in M_{i-1}$.*

Proof. By definition, $\varphi_{i-1}(\mathbf{x}) = \varphi_{h_{i-1}(\mathbf{x})}(\mathbf{x})$, and by construction, $\varphi_{h_{i-1}(\mathbf{x})}(\mathbf{x}) \in S_i \subseteq B_i$. Since the composition of two continuous functions is continuous and Lemma 5.7 asserts that $h_{i-1}(\mathbf{x})$ is continuous, Assumption 2 directly implies that $\varphi_{i-1}(\mathbf{x})$ is continuous. □

Thus, $\varphi_{i-1}(\mathbf{x}) \equiv \varphi_{h_{i-1}(\mathbf{x})}(\mathbf{x})$ is the unique splash point of $\mathbf{x}$ in $\mathcal{H}_i$.

Finally, the second part of Assumption 1 cannot be taken for granted. The following lemma is applied at step i to give us the second part of Assumption 1 at step $i + 1$.

Let W_i be an infinite slab with width $E > D$ in the $\mathbf{v}_i$ direction, parallel to Z_i such that $Z_i \subset W_i$. Let C_i be a parallelepiped with sides parallel to Q_i, also with a width of E in the $\mathbf{v}_i$ direction, satisfying $M_i \subset C_i \subset W_i$, where M_i is built inside $\mathcal{H}_i$ to satisfy the ICP with M_{i-1} under φ_{i-1}. Let $E_{+1} > 0$ be the distance from $\mathcal{H}_i$ in the $\mathbf{v}_i$ direction to the face of W_i, and let $E_{-1} > 0$ be the distance to the opposite face of W_i. Note that $E_{-1} + E_{+1} = E$. Let $\{\mathbf{f}(\mathbf{x}) \cdot \mathbf{v}_i \mid \mathbf{x} \in C_i\} \subset [u_0, u_1]$, and assume $0 < u_0 \leq u_1$. Let $\bar{\mathbf{f}}(C_i)$ be an enclosure of $\{\mathbf{f}(\mathbf{x}) \mid \mathbf{x} \in C_i\}$. Let T_i be a parallelepiped enclosure of $\{W_i \cap (M_i + h\bar{\mathbf{f}}(C_i)) \mid h \in [-\varepsilon^t, \varepsilon^t]\}$, and assume $T_i \subseteq C_i$.

Assumption 3. Assume $E/u_1 > \varepsilon^t$. That is, the minimum crossing time of C_i is greater than ε^t.

LEMMA 5.9. *$\nexists$ distinct $\mathbf{x}, \mathbf{y} \in M_i$ such that $\mathbf{y} = \varphi_t(\mathbf{x})$ for $|t| < \varepsilon^t$.*

Proof. Substituting M_i for Q_i, W_i for Z_i, T_i for S_i, and C_i for B_i in Lemmas 5.1–5.3, we see that

(1) If a trajectory remains in C_i while it is in W_i, then it remains in W_i for at least time E/u_1 and at most E/u_0. By a similar argument, the minimum and maximum times between such a trajectory's entering C_i and intersecting $\mathcal{H}_i$ are E_{-1}/u_1 and E_{-1}/u_0, respectively, and the corresponding times between such a trajectory's intersecting $\mathcal{H}_i$ and exiting C_i are E_{+1}/u_1 and E_{+1}/u_0.

(2) Any trajectory intersecting M_i remains in T_i while it is in W_i, and thus it remains in C_i.

(3) Every trajectory intersecting M_i intersects $\mathcal{H}_i$ at precisely one point while it remains in W_i, where $\mathcal{H}_i \subset W_i$ and $\mathcal{H}_i$ is parallel to the planes enclosing W_i.

Thus, by point (3), to intersect $\mathcal{H}_i$ more than once inside M_i, a trajectory must, at least, first traverse the distance from $\mathcal{H}_i$ to ∂C_i, exit and then reenter C_i, and traverse the distance from ∂C_i back to $\mathcal{H}_i$. By point (1), it takes time at least $E_{-1}/u_1 + E_{+1}/u_1 = E/u_1$ to do so. By Assumption 3, $E/u_1 > \varepsilon^t$. Thus, no trajectory can intersect M_i, exit T_i, and then reenter T_i to again intersect M_i in time less than ε^t. □

Remark 5.2. The base case of the induction is produced by substituting M_0 for M_i in Lemma 5.9, after building suitable W_0, C_0, and T_0.

5.3. Algorithmic details. Algorithmic verification of the requirements for the above theorems and lemmas is fairly straightforward: Q_i is simply the enclosure of $\varphi_{h_{i-1}}(M_{i-1})$ given to us by VNODE; the size of B_i is computed heuristically in an effort to ensure that $S_i \subseteq B_i$, and if our first guess is incorrect, we simply increase its size until $S_i \subseteq B_i$, or fail if increasing the size of B_i results in $0 \in \{\mathbf{f}(\mathbf{x}) \cdot \mathbf{v}_i \mid \mathbf{x} \in B_i\}$; ε^t, which is an upper bound on the time error introduced at each step by the rescaling of time, must currently be prechosen by trial and error, although we believe that good, simple heuristics for choosing it probably exist. The sole complication is in maintaining the property that Q_i has a pair of faces approximately normal to $\mathbf{y}'$ inside Q_i. Note that VNODE maintains a rotation matrix A_i, which represents the orientation of the parallelepiped Q_i. Let the columns of A_i be $\mathbf{a}_i^j, j = 1, \ldots, n$. We simply assign $\mathbf{a}_i^1$ to be parallel to our best estimate of $\mathbf{y}'(t_i)$. VNODE then ensures that $\mathbf{a}_{i+1}^1$ evolves via the variational equation to be approximately parallel to $\mathbf{y}'(t_{i+1})$. To account for the slow buildup of error that would allow $\mathbf{a}_i^1$ to drift away from $\mathbf{y}'(t_i)$, we reset $\mathbf{a}_i^1$ to be parallel to the computed $\mathbf{y}'(t_i)$ at each timestep. This corresponds to rotating Q_i about its center by a small angle θ, computed by solving

$$\cos(\theta) = \frac{\mathbf{a}_i^1 \cdot \mathbf{y}'(t_i)}{\|\mathbf{a}_i^1\| \, \|\mathbf{y}'(t_i)\|},$$

where $\mathbf{a}_i^1$ is the vector computed via evolution of the ODE from the previous timestep, and $\mathbf{y}'(t_i)$ is the value of $\mathbf{y}'$ computed directly from the right-hand side of the ODE at the current timestep. The largest distance a point in Q_i will move as a result of this rotation is $r\theta$, where r is the distance of the furthest corner in Q_i from its center. Thus, after rotating Q_i by θ, we increase its size by $r\theta$ in all directions, thus ensuring that it still encloses $\varphi_{h_{i-1}}(M_{i-1})$.

A simple variable stepsize algorithm was used: whenever containment of a particular step succeeds, we increase the stepsize by a small factor; whenever it fails, we decrease the stepsize by a factor of 2. We do not explicitly fail due to small stepsize, because too small a stepsize results in failures in other parts of the method, for example, as depicted in Figure 4.4.

6. Results and discussion. In this section, we present results of our containment method for ODEs, compare our results to those of others, discuss some of the

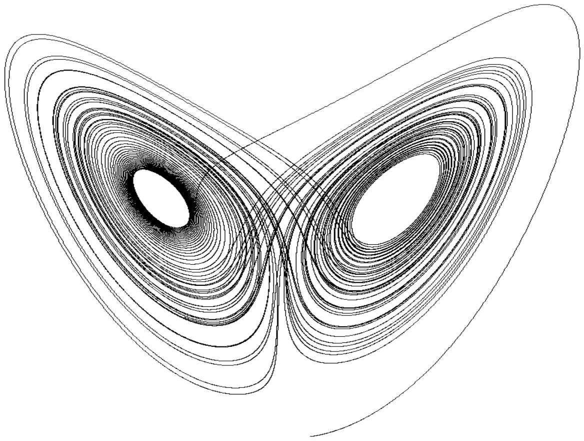

FIG. 6.1. *The "Lorenz butterfly."*

interesting implementation details of our method, and comment on observations of the behavior of our method, including how it sometimes fails.

6.1. Quantitative comparisons with other methods.

6.1.1. The Lorenz system of equations. The Lorenz equations (Lorenz (1963)),

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \sigma(y-x) \\ \rho x - y - xz \\ xy - \beta z \end{pmatrix}, \tag{6.1}$$

define a dissipative dynamical system (i.e., energy is not conserved), which was originally constructed to be a very simplified weather model. It can be shown (Coomes, Koçak, and Palmer (1995a)) that, under the Lorenz equations, the set

$$U = \{(x,y,z) : \rho x^2 + \sigma y^2 + \sigma(z-2\rho)^2 \le \sigma\rho^2\beta^2/(\beta-1)\}$$

is *forward invariant:* any solution that is in U at time t_0 remains in U for all time $t \ge t_0$. All the methods discussed in this section solve the Lorenz equations with the classical parameter values $\sigma = 10, \rho = 28, \beta = 8/3$ (Lorenz (1963)). It is easy to show that, for these parameter values, the cube $[0,15]^3$ lies in U, and so for our experiments we chose initial conditions randomly inside this cube. A set of initial conditions in this cube will invariably produce a solution whose three-dimensional shape has been dubbed the "Lorenz butterfly" (Figure 6.1). Schematically, the Lorenz butterfly consists of two two-dimensional disks in three-space with a "bridge" between them. The two disks together are termed a "chaotic attractor," because solutions tend to remain in the disks but jump chaotically from one to the other and back again. Solutions lack pseudohyperbolicity in the direction of the flow (Van Vleck (1995); Coomes, Koçak, and Palmer (1994b), (1995a)), and so a rescaling of time is required to shadow them effectively. As should be clear from Figure 6.1 and the above description, in addition to the $\mathbf{y}'$ direction, at any given point a solution has one contracting direction, which is perpendicular to the disk currently housing the solution, and one expanding direction, directed radially from the center of the disk. Provided a rescaling of time is employed, solutions to the Lorenz equations display

TABLE 6.1
Comparison of shadow lengths for the Lorenz system. VV=Van Vleck (1995)*; CKP = Coomes, Koçak, and Palmer* (1994b), (1995a).

Author	Local error	Global error	Map method	Rescaling time
VV	10^{-6}	10^{-5}	1–2	$10^2 \sim 10^4$
Hayes	10^{-6}	10^{-5}	$10 \sim 50$	$10^3 \sim 10^5$
CKP	10^{-13}	10^{-9}	$10 \sim 100$	$\geq 10^5$
Hayes	10^{-13}	10^{-9}	$10 \sim 1000$	$\geq 7.7\times10^5$

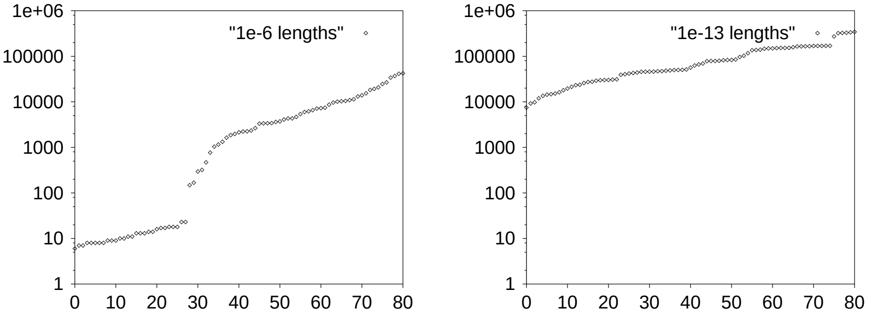

FIG. 6.2. *Distribution of shadow lengths computed by containment with a rescaling of time. Each panel shows a sorted list of shadow lengths for* 80 *simulations of the Lorenz equations. The horizontal axis is simply a label for each shadow; the vertical axis is its length. The magnitude of the noise (i.e., the local error) in the noisy orbits is about* 10^{-6} *in the left graph and* 10^{-13} *in the right.*

remarkable pseudohyperbolicity for extremely long periods of time. Thus, this system is a prime first candidate for testing shadowing methods.

We will compare our results to the only other published results on shadowing the Lorenz equations using a rescaling of time: Van Vleck (1995), whose results could be made rigorous but currently are not; and Coomes, Koçak, and Palmer (1994b), (1995a), whose results are completely rigorous.

First, with *no* rescaling of time (the "map method"), Van Vleck gives two examples of shadows with a local error[6] of about 10^{-5} lasting 1.04 and 1.38 time units; Coomes, Koçak, and Palmer have six examples with local error of about 10^{-13} lasting 9.7, 9.8, 9.9, 9.9, 86, and 126 time units. For this paper, we have simulated hundreds of shadows with various local errors. We have found that with local errors of about 10^{-5}, containment finds shadows that last between 1 and 30 time units, with a median and mean of about 20. With local errors of 10^{-13}, we find shadows lasting between 10 and 1000 time units, again with a mean and median about halfway through that range. Thus, it appears that, without a rescaling of time, the containment method is capable of finding shadows that are about an order of magnitude longer than other existing methods.

With a rescaling of time, Van Vleck gives many examples of shadows (with a local error of about 10^{-6}) ranging from 10^2 to 10^4 time units. Coomes, Koçak, and Palmer (with a local error of 10^{-13}) give six examples of shadows lasting at least 10^5 time units; they do not attempt to find longer shadows, so in fact their method

[6] The local errors used in the current paper were normalized to have comparable size per-unit-step to other methods, even though variable stepsize methods were used both for the validated ODE integration (Nedialkov (1999)) and for choosing the size of shadow steps.

may be capable of finding shadows longer than 10^5. The corresponding numbers for containment are 10^2 to 10^5 for local errors of 10^{-6}, and 10^2 to almost 10^6 for local errors of 10^{-13}. The results are summarized in Table 6.1.[7] It is clear that containment is at least as powerful as the other existing methods. It is worth noting that our results for local errors of 10^{-13} were produced using only a 17th-order Taylor series, whereas Coomes, Koçak, and Palmer used a Taylor series of 31st order.

Figure 6.2 shows two sets of results of shadow lengths, including the rescaling of time. The first is for eighty solutions with local error of approximately 10^{-6}, and the second for eighty solutions with local error of approximately 10^{-13}. The sharp increase in shadow lengths occurring just left of center in the first figure is probably due to the fact that, other than choosing $\mathbf{v}_0$ (cf. Figure 5.2 on page 1963) to be parallel to $\mathbf{y}'(t_0)$, the directions of the faces of M_0 are currently chosen at random. As a result, we sometimes choose nominally expanding and contracting directions that are not sufficiently close to the actual expanding and contracting directions. Thus, many shadows fail early due to this problem. However, if our nominally chosen directions are (by luck) close enough to the actual ones, then we get over this hump to find much longer shadows. There is probably a more clever way to choose the initial M_0, but we have not yet studied this problem closely. This problem becomes less pronounced as the local error decreases and is virtually absent in the right figure, which has local error $\delta = 10^{-13}$.

In addition, our shadowing distances (i.e., the maximum distance between the shadow and the numerical trajectory) are comparable to the methods of the above authors: for orbits with noise 10^{-6} and 10^{-13}, our method and those of Van Vleck and Coomes, Koçak, and Palmer find shadowing distances of approximately 10^{-5} and 10^{-9}, respectively. For containment, these sizes are based on ε^t and the maximum size of M_i over all i, which are at least in part user-controlled. For Van Vleck and Coomes, Koçak, and Palmer the shadowing distances are computed analytically based upon global bounds of various computed quantities.

6.1.2. Other systems of equations.

We have reproduced the shadowing experiments of several other authors, usually getting comparable results, as illustrated in Table 6.2. We discussed results for the Lorenz system in the previous section. In this section, we provide results for three other problems.

Forced damped pendulum. We first compare our results for the forced damped pendulum problem,

$$y'' + ay' + \sin y = b \cos t,$$

to those of GHYS, Sauer and Yorke (1991), and Chow and Van Vleck (1994). These authors use the values $a = 0.2, b = 2.4$ and $a = 1, b = 2.4$, with initial conditions $(y, y') = (0, 0)$, and mention that they get similar results with other pairs of values of a, b and initial conditions. We used the above two pairs of values for a, b and various random initial conditions in the unit square $[0, 1]^2$. We convert the second-order equation to two first-order equations by assigning $y_1 = y,\ y_2 = y'$, giving

$$y_1' = y_2,$$

[7] Our attempts to find the longest possible shadows for the latter case have been repeatedly confounded by having either workstation or disk crashes (independent of our code) while our simulations were running. The longest shadow we have observed is thus 7.7×10^5, even though, had our machines not crashed, the shadows might have been longer.

Table 6.2

Comparison of shadow lengths for four systems. For our results, the lengths shown are typical results after attempting many trials with the given local and global errors; the results of others are taken from their respective publications. Legend: δ = local error; ε = global space error; ε_t = global time error (if none is listed for our method, then we did not rescale time); L = shadow length; CKP = Coomes, Koçak, and Palmer (1994b), (1995a)*; SY = Sauer and Yorke* (1991)*; CVV = Chow and Van Vleck* (1994a)*; VV = Van Vleck* (1995)*; NR = not rigorous.*

System	Auth.	δ	ε	ε_t	L	Comment
Lorenz	VV	10^{-6}	10^{-5}		10^4	NR
	Hayes	10^{-6}	10^{-5}	2.5×10^{-5}	10^3–10^5	
	CKP	10^{-13}	10^{-9}		$\geq 10^5$	
	Hayes	10^{-13}	10^{-9}	$2.5{\times}10^{-9}$	$\geq 7.7 \times 10^5$	
Forced damped pendulum	SY	10^{-18}	10^{-9}		3×10^4	High machine precision
	Hayes	10^{-15}	10^{-6}	10^{-3}	10^3–3×10^4	
	CVV	10^{-6}	10^{-3}		10^4	NR
	Hayes	10^{-6}	10^{-5}	10^{-3}	10^3	
	CVV	10^{-11}	10^{-8}		10^3	NR
	Hayes	10^{-11}	10^{-8}	10^{-3}	10^3	
Forced van der Pol	VV	10^{-5}	10^{-4}		10^4	Periodic attractor NR
	Hayes	10^{-5}	10^{-6}	3×10^{-5}	$\geq 10^5$	
Logistic equation	CVV	10^{-7}	5×10^{-6}		9.22	$y_0 = 0.01$, fixed L, NR
	Hayes	10^{-7}	10^{-6}		9.22	
	CVV	10^{-7}	5×10^{-6}		18.46	$y_0 = 10^{-4}$, fixed L, NR
	Hayes	10^{-7}	10^{-6}		18.46	

$$y_2' = b \cos t - \sin y_1 - a y_2.$$

GHYS and Sauer and Yorke (1991) use extended precision arithmetic with a machine epsilon of 10^{-29} to generate a trajectory with local truncation error rigorously bounded by 10^{-18} per step, which allows them to find a shadow of length 3×10^4 and rigorous maximum distance 10^{-9} from their noisy trajectory. In comparison, we use standard IEEE754 floating-point numbers and arithmetic and obtain a local truncation error of about 10^{-15} at best, so our shadow distances are significantly less stringent at 10^{-6}, and tend to be shorter, although in a few instances we successfully found shadows of length $\sim 3 \times 10^4$. Given that Sauer and Yorke used higher precision, we are not surprised that our shadows tend to be shorter and not as close as theirs. Comparing our results to Chow and Van Vleck (1994), we see our method is capable of rigorously proving the existence of a shadow which is closer, but lasts for a shorter time, than their method does; on the other hand, our result is rigorous, whereas theirs is not, because they do not rigorously bound numerical errors before applying their theorem.

The primary problem with shadowing this system appears to be that it is nonautonomous. We currently handle a nonautonomous system by converting it to an autonomous system with one component of our solution, y_0, representing time: $y_0(0) = t_0,\ y_0'(t) = 1$. This has several drawbacks: (1) the new component is decidedly nonhyperbolic; (2) assuming we can solve the linear system $y' = 1$ exactly, the interval representing y_0 then accumulates roundoff error, and as time progresses, the error in

y_0 grows; (3) this is exacerbated by the minimum absolute error in y_0 increasing as $\varepsilon_{mach}t$, where ε_{mach} is the machine precision; (4) finally, the error in the computation of $\cos(y_0)$ adds to the error. These drawbacks, however, do not seem to adequately explain our poor shadowing results for this system. Perhaps the difficulties would vanish if a native procedure for validated integration of nonautonomous systems were used, or if we used higher precision, as did Sauer and Yorke (1991).

Forced van der Pol. The forced van der Pol equation,

$$x'' + \alpha(x^2 - 1)x' + x = \beta\cos(\omega t),$$

is studied by Van Vleck (1995). He defines the parameters implicitly with $\alpha = k = \sigma = 2/5$, where $k = \beta/(2\alpha)$ and $\sigma = (1 - \omega^2)/\alpha$, and uses the initial conditions $(x, x') = (0, 0)$. We try this initial condition, as well as others chosen randomly in the unit square $[0, 1]^2$, and we convert the second-order equation to two first-order equations by assigning $y_1 = x,\ y_2 = x'$, giving

$$y_1' = y_2,$$

$$y_2' = \beta\cos(\omega t) - (y_1^2 - 1)\alpha y_2 - y_1.$$

This equation has a hyperbolic periodic attractor, which all solutions approach asymptotically, and so this system is easy to shadow. With a local truncation error of 10^{-6}, Van Vleck found numerical shadows of length 10^4 and distance 10^{-4}, while we went significantly further, finding rigorous shadows lasting 10^5 and longer with a distance of 10^{-6}. Since solutions asymptotically approach a periodic solution that is hyperbolic, we conjecture that containment could be maintained indefinitely.

Logistic equation. Finally, the logistic equation,

$$y' = y(1 - y), \quad y(0) = \zeta,\ 0 < \zeta \ll 1,$$

was studied by Chow and Van Vleck (1994). In this problem, there is an unstable fixed point at $y = 0$ and a stable fixed point at $y = 1$. Chow and Van Vleck attempt shadowing two solutions, both starting at $y(0) = \zeta$ and integrating until $y(T) \approx 1 - \zeta$. If $\zeta = 10^{-2}$, then $T \approx 9.22$, and if $\zeta = 10^{-4}$, then $T \approx 18.46$. In both cases, we use a local truncation error of $\delta = 10^{-7}$. We find that we easily match their results, noting again that ours are rigorous, while theirs are not. In fact, we find that we can prove the existence of these shadows for $\varepsilon \approx 10\delta$ for δ down to about 10^{-14}.

6.2. Qualitative comparisons with other methods. First and foremost, our method has only been proven to work in a limited number of special cases. Generalizing the (n, k)-ICP to arbitrary (n, k) is straightforward Hayes (2001). Proving that it implies the existence of a shadow is more difficult, and is in progress. See Hayes (2001) for more discussion.

Although containment is rigorous, it appears to be less robust than nonrigorous methods. For example, in two examples out of three, the nonrigorous results of Chow and Van Vleck (1994) produced shadows that were about an order of magnitude longer than we could produce using containment. In addition, Hayes (1995) presented convincing evidence that the gravitational n-body problem is shadowable, and yet containment could prove the existence of shadows lasting only 1% as long as those found nonrigorously in Hayes (1995). Even worse, the VNODE package (Nedialkov (1999)) is capable of providing a validated enclosure of an IVP for the n-body problem which is about ten times as long as the containment-produced shadow! Clearly, if an enclosure of an IVP exists, then a shadow exists for the associated point solution for at least as long. Thus, at least for some problems, our implementation of containment is incapable of finding shadows even though they exist. This does not necessarily imply that the theorems proved in section 2.2 are deficient; it probably means that our implementation for verifying that the ICP holds can be improved, for example by reducing the excess of the validated numerical integrator.

Our method requires some a priori guesses; for example, the maximum and minimum sizes of the M_i, and the maximum time rescaling ε^t, need to be chosen before the algorithm runs. We typically had to choose these numbers by trial and error for each problem; if a certain ε^t did not work, for example, we often found that increasing it or decreasing the maximum size of M_i would allow us to find longer or closer shadows, respectively. Van Vleck's (1995) method also requires some a priori guesswork to make a rescaling of time work. Although Coomes, Koçak, and Palmer do not discuss their choice of parameters, it is likely that they require significant guesswork to find parameters that satisfy their theorems as well. Finally, *all* shadowing methods currently in the literature appear to require guesswork to discover the number of expanding and contracting dimensions and to choose a local error δ which is stringent enough to satisfy their respective theorems.

It is also not trivial to see how containment could be parallelized, since each M_i depends on M_{i-1}. Possibly an iterative method that guesses all the $\{M_i\}_{i=0}^{N}$ and then iteratively refines them in parallel could be constructed; this may also be related to two-point boundary value problems (Ascher, Mattheij, and Russell (1988)).

On the other hand, containment appears to have several advantages over other methods.

- We use an off-the-shelf validated integrator (Nedialkov (1999)) to verify that ICP holds; this integrator is almost as easy to use as any standard integrator, and thus getting the code "up and running" on a new problem usually takes only a few minutes. Another advantage of this simplicity is that it requires the user to have no deeper understanding of the system than knowing the defining equations.[8]
- Although the success of containment may depend, of course, upon global properties of the system, the method itself is local. By that we mean that it requires information only from the previous step to extend the length of the shadow. Several other methods require computing, storing, and updating global information such as the extent of nonhyperbolicity (cf. Chow and Palmer's p parameter 1991, 1992).

7. Conclusions.

We have extended the simple and elegant *containment* method of producing shadows from two-dimensional maps to maps of arbitrary dimension in which some measure of hyperbolicity is present and there is either 0 or 1 expanding modes, or 0 or 1 contracting modes, and added a rescaling of time to allow containment to work better for ODEs. We have demonstrated that this new method produces shadows of ODE integrations that are of comparable quality and length to any currently in the literature, and noted how it can be used to prove the existence of chaos.

[8] Some may consider this a disadvantage.

REFERENCES

U. M. Ascher, R. M. M. Mattheij, and R. D. Russell (1998), *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*, Prentice-Hall Series in Comput. Math., Prentice-Hall, Englewood Cliffs, NJ.

W.-J. Beyn (1987), *On invariant closed curves for one-step methods*, Numer. Math., 51, pp. 103–122.

J. C. Butcher (1987), *The Numerical Analysis of Ordinary Differential Equations*, Wiley, New York.

S.-N. Chow and K. J. Palmer (1991), *On the numerical computation of orbits of dynamical systems: The one-dimensional case*, J. Dynam. Differential Equations, 3, pp. 361–380.

S.-N. Chow and K. J. Palmer (1992), *On the numerical computation of orbits of dynamical systems: The higher dimensional case*, J. Complexity, 8, pp. 398–423.

S.-N. Chow and E. S. Van Vleck (1994), *A shadowing lemma approach to global error analysis for initial value ODEs*, SIAM J. Sci. Comput., 15, pp. 959–976.

B. A. Coomes (1997), *Shadowing orbits of ordinary differential equations on invariant submanifolds*, Trans. Amer. Math. Soc., 349, pp. 203–216.

B. A. Coomes, H. Koçak, and K. J. Palmer (1994a), *Periodic shadowing*, in Chaotic Numerics, P. Kloeden and K. Palmer, eds., Contemp. Math. 172, AMS, Providence, RI, pp. 115–130.

B. A. Coomes, H. Koçak, and K. J. Palmer (1994b), *Shadowing orbits of ordinary differential equations*, J. Comput. Appl. Math., 52, pp. 35–43.

B. A. Coomes, H. Koçak, and K. J. Palmer (1995a), *Rigorous computational shadowing of orbits of ordinary differential equations*, Numer. Math., 69, pp. 401–421.

B. A. Coomes, H. Koçak, and K. J. Palmer (1995b), *A shadowing theorem for ordinary differential equations*, Z. Angew. Math. Phys., 46, pp. 85–106.

B. A. Coomes, H. Koçak, and K. J. Palmer (1997), *Long periodic shadowing*, Numer. Algorithms, 14, pp. 55–78.

R. M. Corless (1994), *Error backward*, in Chaotic Numerics, P. Kloeden and K. Palmer, eds., Contemp. Math. 172, pp. 31–62.

G. Dahlquist and Å. Björck (1974), *Numerical Methods*, Prentice-Hall Series in Automatic Computation, Prentice-Hall, Englewood Cliffs, NJ.

C. Grebogi, S. M. Hammel, J. A. Yorke, and T. Sauer (1990), *Shadowing of physical trajectories in chaotic dynamics: Containment and refinement*, Phys. Rev. Lett., 65, pp. 1527–1530.

S. M. Hammel, J. A. Yorke, and C. Grebogi (1987), *Do numerical orbits of chaotic dynamical processes represent true orbits?*, J. Complexity, 3, pp. 136–145.

S. M. Hammel, J. A. Yorke, and C. Grebogi (1988), *Numerical orbits of chaotic dynamical processes represent true orbits*, Bull. Amer. Math. Soc., 19, pp. 465–470.

W. Hayes (1995), *Efficient Shadowing of High Dimensional Chaotic Systems with the Large Astrophysical n-Body Problem as an Example*, Master's thesis, Department of Computer Science, University of Toronto, Toronto.

W. B. Hayes (2001), *Rigorous Shadowing of Numerical Solutions of Ordinary Differential Equations by Containment*, Ph.D. thesis, Department of Computer Science, University of Toronto, Toronto; also available on the web at http://www.cs.toronto.edu/NA/reports.html#hayes-01-phd.

D. Kahaner, C. Moler, and S. Nash (1989), *Numerical Methods and Software*, Prentice-Hall Series in Comput. Math., Prentice-Hall, Englewood Cliff, NJ, 1989.

E. N. Lorenz (1963), *Deterministic nonperiodic flow*, J. Atmospheric Sci., 20, pp. 130–141. (Reprinted in Chaos, by H. Bai-Lin, World Scientific Publishing, Singapore, 1984.)

J. R. Munkres (1975), *Topology: A First Course*, Prentice-Hall, Englewood Cliffs, NJ.

J. Murdock (1995), *Shadowing multiple elbow orbits: An application of dynamical systems to perturbation theory*, J. Differential Equations, 119, pp. 224–247.

N. S. Nedialkov (1999), *Computing Rigorous Bounds on the Solution of an Initial Value Problem for an Ordinary Differential Equation*, Ph.D. thesis, Department of Computer Science, University of Toronto, Toronto.

N. S. Nedialkov, K. R. Jackson, and G. F. Corliss (1999), *Validated solutions of initial value problems for ordinary differential equations*, Appl. Math. Comput., 105, pp. 21–68.

K. J. Palmer (1988), *Exponential dichotomies, The shadowing lemma and transversal homoclinic points*, in Dynamics Reported, U. Kirchgraber and H. O. Walther, eds., Vol. 1, Wiley and Teubner.

G. D. Quinlan and S. Tremaine (1992), *On the reliability of gravitational N-body integrations*, Monthly Notices Roy. Astronom. Soc., 259, pp. 505–518.

T. Sauer and J. A. Yorke (1991), *Rigorous verification of trajectories for the computer simulation* *of dynamical systems*, Nonlinearity, 4, pp. 961–979.

D. Stoffer and K. J. Palmer (1999), *Rigorous verification of chaotic behaviour of maps using validated shadowing*, Nonlinearity, 12, pp. 1683–1698.

A. M. Stuart and A. R. Humphries (1996), *Dynamical Systems and Numerical Analysis*, Cambridge University Press, Cambridge, UK.

E. S. Van Vleck (1995), *Numerical shadowing near hyperbolic trajectories*, SIAM J. Sci. Comput., 16, pp. 1177–1189.

E. S. Van Vleck (2000), *Numerical shadowing using componentwise bounds and a sharper fixed point result*, SIAM J. Sci. Comput., 22, pp. 787–801.