Shadowing pseudo-orbits and gradient descent noise reduction

Kevin Judd
School of Mathematics and Statistics
The University of Western Australia

March 15, 2007

Keywords: nonlinear dynamical systems, shadowing, noise reduction
Short title: Shadowing pseudo-orbits

Abstract

Shadowing trajectories are one of the most powerful ideas of modern dynamical systems theory, providing a tool for proving some central theorems and a means to assess the relevance of models and numerically computed trajectories of chaotic systems. Shadowing has also been seen to have a role in state estimation and forecasting of nonlinear systems. Shadowing trajectories are guaranteed to exist in hyperbolic systems, but this is not true of non-hyperbolic systems, indeed it can be shown there are systems that cannot have long shadowing trajectories. In this paper we consider what might be called *shadowing pseudo-orbits*. These are pseudo-orbits that remain close to a given pseudo-orbit, but have smaller *mismatches* between forecast state and verifying state. Shadowing pseudo-orbits play a useful role in the understanding and analysis of gradient descent noise reduction, state estimation and forecasting nonlinear systems, because their existence can be ensured for a wide class of non-hyperbolic systems. New theoretical results are presented that that extend classical shadowing theorems to shadowing pseudo-orbits. These new results provide some insight into the convergence behaviour of gradient descent noise reduction methods. The paper also discusses, in the light of the new results, some recent numerical results for an operational weather forecasting model when *gradient descent noise reduction* was employed.

1 Introduction

From a mathematical perspective this paper can be viewed as a continuation of a long chain of results concerning the existence of shadowing trajectories, but the results presented here are not intended to be just statements of mathematical existence; they are intended to provide a tool to better understand noise reduction, state estimation and forecasting in nonlinear systems.

Shadowing trajectories play a central role in dynamical systems theory with many deeply profound results about the nature of dynamical systems having been obtained from them [19, 23, 24]. These results begin with the work of Anosov [1] and Bowen [3] in the 1960's for hyperbolic systems, then Pesin [27] expanded the shadowing results to non-uniformly hyperbolic systems using ergodic methods, and recently Chow and Palmer [5], and Diamond, Kloeden and others [9, 10, 20] have considered shadowing in more general settings. The work presented here owes a lot to these recent developments, although the content is complementary rather than an extension of it.

Of particular interest to us is the numerical computation, or approximation, of shadowing trajectories. There have been developed algorithms for approximating shadowing trajectories, most notably gradient descent methods [4, 7, 12, 13]. Recent proofs show that these methods converge to the true trajectory of a hyperbolic system, given observations with sufficiently small noise and a perfect model [4, 28]. When noise levels are larger, or the system is non-hyperbolic, then it has been described how near tangencies of invariant stable and unstable manifolds can cause the failure of numerical methods for finding shadowing trajectories [28]. The algorithms fail to shadow the true trajectory because they convergence to nearby trajectories that are homoclinic to the true trajectory [28], a phenomenon sometimes referred to as *glitches* [29]. There are other problems. It has been shown that there exist non-hyperbolic systems for which shadowing trajectories cannot not be obtained [8, 21, 22, 31]; these systems have local Lyapunov exponents that fluctuate about zero over time. On the other hand, Pilyugin and others have shown that shadowing is a *generic* property of homeomorphisms in the C^0-topology [26].

The results of this paper provide a counter point to the problems known to hinder numerical computation of shadowing trajectories. It is argued that a non-hyperbolic system may behave sufficiently like a hyperbolic system to enable useful noise reduction, state estimation and forecasting.

The real motivation, and intended application, of the results of this paper is in the realm of forecasting with imperfect observations and imperfect models. From the author's point of view this paper is a piece of a larger puzzle. For some time there has been growing interest in using shadowing ideas and methods as a diagnostic of model error [11, 30, 25] and as a basis for ensemble forecasting in nonlinear systems [16, 17]. (Preliminary experiments have been made with operational weather forecasting models [15].) There is already evidence that shadowing-based methods provide better state estimation for nonlinear systems than traditional Kalman filter methods [14], but a more general argument is being developed along the lines that the increasingly widely used sequential-update Bayesian methods (ensemble Kalman filters and particle filters) have fundamental limitations that shadowing methods avoid. This line of argument will not be developed here because the purpose of this paper is to present technical results.

The first sections of this paper state, develop, and prove the main results, then in section 5 we use the results to better understand and assess iterative methods for finding shadowing trajectories, for example, gradient descent algorithms. We discuss the main result in the context of the expected convergence properties of gradient descent shadowing algorithms when applied to non-hyperbolic models, including an operational weather forecasting model.

2 Main results

Let $f\colon \mathbb{U} \to \mathbb{U}$ be a diffeomorphism on a metric space $\mathbb{U}$. Of interest is a situation where f represents a model of some dynamical system (some aspect of reality) and there is a given sequence of states $x \in \mathbb{U}^n$, which may have been estimated from observations. These states are not a trajectory of f, that is, there is a *mismatch* between x_i and $f(x_{i-1})$. The question asked is whether there exists another sequence of states $y \in \mathbb{U}^n$ that is both close to x and has significantly smaller mismatches. As already alluded to, the existence of such a y has implications for noise reduction, state estimation and forecasting. If f were globally hyperbolic, then a shadowing trajectory exists for all time, which has zero mismatch, and the gradient descent algorithm will converge to such a trajectory if the observational noise is sufficiently small [28]. This cannot be said of a non-hyperbolic system [8, 21, 22, 31]. Of course, if a non-hyperbolic system happens to be locally hyperbolic in the region of x and y, then some of the results pertaining to (globally) hyperbolic systems may apply, but this does not help when the system is non-hyperbolic in this region. On the other hand, there are practical difficulties even in the (locally) hyperbolic case. It can be shown that convergence of the gradient descent algorithm is limited by the smallest Lyapunov exponent [28], so convergence, even for a hyperbolic system, could be very slow, and in practice no method ever obtains an actual trajectory.

The core idea of the results presented here is that if one only wants to obtain from x another sequence of states y that has smaller mismatches than x, then one does not need the model f to be hyperbolic, f only has to have dynamics that are *partially hyperbolic* in the sense we now define.

We will say that f is *partially hyperbolic* if there exists an interval $(\lambda_0, \lambda_1) \subset (0, 1]$, such that for all $\lambda \in (\lambda_0, \lambda_1)$ there is a splitting $T_z\mathbb{U} = E^{(-1)}_\lambda(z) \oplus E^{(0)}_\lambda(z) \oplus E^{(+1)}_\lambda(z)$, continuous with $z \in \mathbb{U}$, with $df(z)(E^\kappa_\lambda(z)) = E^\kappa_\lambda(f(z))$ for $\kappa \in \{-1, 0, +1\}$, where $E^\kappa_\lambda(z) \neq 0$ for $\kappa \in \{-1, +1\}$ and

$$\forall v \in E^\kappa_\lambda(z), \quad \|df(z)^\kappa v\| \leq \lambda \|v\|, \quad \kappa \in \{-1, +1\}.$$

In this definition $E^\kappa_\lambda(z)$ for $\kappa \in \{-1, +1\}$ describe the principle expanding and contracting directions, where as $E^0_\lambda(z)$ represents the other directions, sometimes called "(nearly) neutral modes", which may have near zero growth exponents, or have local exponents that fluctuate about zero over time. If $\lambda_1 = 1$ then f will be said to be *weakly hyperbolic*. The more familiar notion of *uniformly hyperbolic* occurs when there exists $\lambda \in (\lambda_0, \lambda_1)$ such that $E^{(0)}_\lambda(z) = 0$. Such a λ is called a *uniform bound*.

It will become apparent later that a lower bound on the angle $\theta(z)$ between the $E^\kappa_\lambda(z)$ for $\kappa \in \{-1, +1\}$ is of importance. Let $\theta = \inf_z \theta(z)$, which will be called the *angle bound*. It will be necessary for our purposes that $\theta > 0$, which will be seen to be easily assured by having $\sup_z \|df(z)\|_\infty$ bounded.

It will be advantageous to assume that $\mathbb{U}$ is a finite dimensional Banach[1] space $\mathcal{B}$. Our aim is to establish the following proposition and discuss some of its consequences. The statement of this proposition is presented so as to give a clearer indication of results, rather than being a complete statement of all details. The statement is derived from approximation of stricter bounds given later.

In the following proposition there is a sequence of model states x derived from observations of the system. The proposition implies the existence of a y that shadows x and has smaller *mismatches*. The mismatches $x_{i+1} - f(x_i)$ represent the forecast errors of the model: these mismatches may involve errors resulting from errors in the observations, that is, as a consequence of assimilation of observational data into the model to obtain state "estimates" x_i and x_{i+1}; the mismatches may also result from model error. The mismatches $y_{i+1} - f(y_i)$ represent the *corrections* to the model states required to keep the model consistent with the observations. Consistency is judged in the proposition by the threshold ϵ, while β measures the relative success of the shadowing, that is, reduction of the mismatches.

Proposition 1 *Suppose $f\colon \mathcal{B} \to \mathcal{B}$ is a twice differentiable diffeomorphism on a finite dimensional Banach space $\mathcal{B}$, with bounded first and second derivatives, in particular let $J = \sup_{\mathcal{B}} \|d^2 f\|_\infty$. If $x \in \mathcal{B}^n$, with $\|x_{i+1} - f(x_i)\| \leq \delta$ and there exist $0 < \beta < 1$ and $0 < \gamma < 1$ so that $\frac{1}{n-1}\sum_{i=1}^{n-1}\|x_{i+1} - f(x_i)\| \leq \gamma\delta$ and $n^2\gamma^2 - 1 < 2n(n-1)\beta$, then there exists $y \in \mathcal{B}^n$ such that $\|x_i - y_i\| \leq \epsilon$, $\|y_{i+1} - f(y_i)\| \leq \beta\epsilon$, when*

$$\delta \leq \frac{1}{2K^2J}, \quad \epsilon = \frac{1 - \sqrt{1 - 2K^2J\delta}}{KJ} = \frac{K\delta}{1 + \sqrt{1 - 2K^2J\delta}},$$

and K is bounded as follows or is one, which ever is the larger. If f is uniformly hyperbolic with bound λ and angle bound θ, then

$$K \leq \frac{1}{2\sin\theta/2}\left(2\frac{1-\lambda^m}{1-\lambda} - 1 - \beta\frac{1-\lambda^n}{1-\lambda}\right) \quad \textit{where} \quad m = \frac{1}{2}(n\gamma + 1).$$

If f is partially hyperbolic on (λ_0, λ_1) with angle bound θ and γ small, then the above bound on K applies with

$$\lambda \geq \left(\frac{2-\beta}{\beta(n-1)}\right)^{\frac{1}{n}},$$

provided $\lambda \in (\lambda_0, \lambda_1)$, and

$$\beta\lambda_1^n - 2\lambda_1^m + \lambda_1 + 1 - \beta < 0.$$

[1] A complete normed linear space, of which, the Euclidean space $\mathbb{R}^d$. Since many complex models are constructed in this way this is a perfectly reasonable restriction. Weather forecasting models, for example, are essentially partial differential equations evolving vector fields, on the region bound by two concentric spheres. This function space is an infinite dimensional Banach space. In practice a model is constructed on a finite dimensional basis, such as the spherical harmonics, finite elements, or Galerkin approximation. Using such a finite dimensional basis reduces the problem to dynamics in $\mathbb{R}^d$, where for operational weather forecasting models d is currently several million.

The essential point of this proposition is that if the average mismatch of the initial (observation) sequence x is sufficiently small relative to the maximum mismatch, that is, γ is sufficiently small, then there is a y with smaller mismatches, $\beta\epsilon < \delta$. If γ is not sufficiently small, then the mismatches of the initial sequence x may not necessarily be significantly improved upon. The role and interplay of β, γ and λ, is discussed in section 5, particularly in the context of gradient descent noise reduction and state estimation. It might be useful, however, to contrast this proposition with classical shadowing theorems, which generally involve only two or three key parameters corresponding to our δ, ϵ and λ; some classical theorems roughly correspond to $\beta = 0$ and no reference to γ. If $\beta = 0$ in our proposition, then a classical result similar to Chow and Palmer [5] is recovered, because although not immediately obvious the conditions involving γ can be ignored, see proof of proposition 4. In classical shadowing theorems the parameters δ, ϵ and λ are not independent, similarly, our δ, ϵ, λ, β and γ are not independent, indeed there is an additional complex relationship between λ, β and γ, see section 5.1.

3 Detailed results

In this section the main results are stated in detail, with proofs deferred to the next section. Both this and the next section can be safely skipped on a first reading.

It is common to assume a dynamical system is modelled by a diffeomorphism $f\colon \mathbb{U} \to \mathbb{U}$, on a metric space $\mathbb{U}$ with everywhere identical tangent spaces, $T_x\mathbb{U} = \mathbb{V}$ for all $x \in \mathbb{U}$. Here we will make the stronger assumption that $\mathbb{U}$ is a Banach space $\mathcal{B}$, in which case $\mathbb{V} = \mathbb{U} = \mathcal{B}$, however, we will continue to use the different labels $\mathbb{U}$ and $\mathbb{V}$, because it is advantageous to make their identity apparent. We will also assume that f is twice differentiable and will define constants

$$D = \sup_{x\in\mathbb{U}} \|df(x)\|_\infty \quad \text{and} \quad J = \sup_{x\in\mathbb{U}} \|d^2 f(x)\|_\infty.$$

Suppose there is a sequence of model states $x = (x_i)_{i=1}^n \in \mathbb{U}^n$. Define on $\mathbb{U}^n$ the (nonlinear) *mismatch* operator $M\colon \mathbb{U}^n \to \mathbb{U}^{n-1}$, by $M = S - F$, where $S(x) = (x_{i+1})_{i=1}^{n-1}$ and $F(x) = (f(x_i))_{i=1}^{n-1}$. Furthermore, for $x \in \mathbb{U}^n$ and $v = (v_i)_{i=1}^n \in \mathbb{V}^n$ define the linear operator $L_x\colon \mathbb{V}^n \to \mathbb{V}^{n-1}$, by $L_x = S - A_x$, where $A_x(v) = (df(x_i)v_i)_{i=1}^{n-1}$ and the shift operator S similarly applies to $v \in \mathbb{V}^n$.

For any linear operator A on $\mathbb{V}^n$ and $\beta > 0$, define a semi-norm

$$\|A\|_{a,b,\beta} = \sup_{\|v\|_a=1} \inf_{\|w\|_b=1} \|A(v + \beta w)\|_a,$$

where subscripts a and b denote possibly different norms on $\mathbb{V}^n$. Note that this is a norm if $\beta < 1/\sup\{\|w\|_a\colon \|w\|_b = 1\}$.

Proposition 2 *Given $x \in \mathbb{U}^n$, with $\|M(x)\|_a \le \delta$, if there exists a right inverse of L_x, with $K = \max\{1, \|L_x^{-1}\|_{a,b,\beta}\}$ and*

$$\delta \le \frac{1}{2K^2 J},$$

then there exists $y \in \mathbb{U}^n$, with $\|M(y)\|_b \le \beta\epsilon$ and $\|x - y\|_a \le \epsilon$, where

$$\epsilon = \frac{1 - \sqrt{1 - 2K^2 J\delta}}{KJ}.$$

By itself this proposition is a fairly limited generalisation of a result of Chow and Palmer [5], and the proof we provide follows their proof fairly closely. For the purposes of proposition 1 the important task now is estimating useful bounds on $\|L_x^{-1}\|_{a,b,\beta}$. This is the subject of the following propositions.

For $0 < \gamma < 1$ define *γ-mean-max norms* to be a combination of three norms on $\mathcal{B}^n$, labelled a, b, and c, such that if $z \in \mathcal{B}^n$ then $\|z\|_c = \frac{1}{n}\sum_{i=1}^n \|z_i\|$, $\|z\|_b = \sup_i \|z_i\|$ and $\|z\|_a = \max\{\|z\|_b, \gamma^{-1}\|z\|_c\}$. Note that the norm on $\mathcal{B}$ does not need to be specified.

The following three propositions provide statements of bounds that were used, or approximated, in the statement of proposition 1. For the purposes of these propositions, define,

$$K(\lambda) = K_{\beta,\gamma,\theta,n}(\lambda) = \frac{1}{2\sin\theta/2}\left(2\frac{1-\lambda^{(n\gamma+1)/2}}{1-\lambda} - 1 - \beta\frac{1-\lambda^n}{1-\lambda}\right)$$

Proposition 3 *If f is hyperbolic with uniform bound λ and angle bound θ, then given γ-mean-max norms there exists a right inverse of L_x with $\|L_x^{-1}\|_{a,b,\beta} \leq K(\lambda)$.*

Proposition 4 *If f is weakly hyperbolic on $(\lambda_0, 1)$ with angle bound θ, then for γ-mean-max norms with $n^2\gamma^2 - 1 < 2n(n-1)\beta$, $0 < \beta < 1$, there exists a right inverse of L_x such that $\|L_x^{-1}\|_{a,b,\beta} \leq K(\lambda^\star)$, where $\lambda^\star \in (\lambda_0, 1)$ is a root of*

$$\beta(n-1)\lambda^n - \beta n\lambda^{n-1} - (n\gamma-1)\lambda^{(n\gamma+1)/2} + (n\gamma+1)\lambda^{(n\gamma-1)/2} + \beta - 2 = 0.$$

Proposition 5 *If f is partially hyperbolic on (λ_0, λ_1), then the results of proposition 4 apply provided $\lambda^\star \in (\lambda_0, \lambda_1)$ and $K(\lambda_1) < 0$.*

It is clear that the above propositions are meaningless unless $\theta > 0$, however, this is assured by the upper bound D on the derivative of f as the following proposition implies.

Proposition 6 *Let A be an invertible matrix with $\|A\|_\infty \leq D$, and u and v vectors such that for some $0 < \lambda < 1$, $\|Au\| \leq \lambda\|u\|$ and $\|Av\| \geq \lambda^{-1}\|v\|$. If ϕ is the acute angle between u and v, then $\phi > 0$ and*

$$\cos\phi \leq 1 - \frac{\lambda^{-2} - \lambda^2}{2D^2}.$$

4 Proofs

This first proof is a generalisation of the method of Chow and Palmer [5].

Proof of proposition 2: Given $x \in \mathbb{U}^n$ and a right inverse of L_x with $K = \|L_x^{-1}\|_{a,b,\beta}$, we are required to find $z, w \in \mathbb{V}^n$ with $w = M(x+z)$, $\|z\|_a \leq \epsilon$ and $\|w\|_b \leq \beta\epsilon$.

The following sequence of equalities follow by straight forward substitutions and manipulations:

$$\begin{aligned}
M(x+z) &= w, \\
M(x+z) - A_x z &= w - A_x z, \\
Sx + Sz - F(x+z) - A_x z &= w - A_x z, \\
Sz - A_x z &= F(x+z) - Sx - A_x z + w, \\
L_x z &= (F(x+z) - Fx - A_x z) - (Sx - Fx) + w, \\
L_x z &= G_x(z) - M(x) + w,
\end{aligned}$$

where the last equality serves to define $G_x(z) = F(x+z) - Fx - A_x z$.

Now define $T_w(z) = L_x^{-1}(G_x(z) - M(x) + w)$. It follows from the definition of K that for each z there exists $w(z)$ such that

$$\|w(z)\|_b \leq \beta\|G_x(z) - M(x)\|_a \quad \text{and} \quad \|T_{w(z)}(z)\|_a \leq K\|G_x(z) - M(x)\|_a. \tag{1}$$

On the other hand, using the definition of J and δ

$$\|G_x(z) - M(x)\|_a \leq \|G_x(z)\|_a + \|M(x)\|_a \leq \frac{1}{2} J\|z\|_a^2 + \delta. \tag{2}$$

Let $\mathcal{Z}_\epsilon = \{z \in \mathbb{V}^n \colon \|z\|_a \leq \epsilon\}$ and $\mathcal{W}_{\beta\epsilon} = \{w \in \mathbb{V}^n \colon \|w\|_b \leq \beta\epsilon\}$. It follows from (1) and (2) that for every $z \in \mathcal{Z}_\epsilon$ there exists $w(z) \in \mathcal{W}_{\beta\epsilon}$, such that $T_{w(z)}(z) \in \mathcal{Z}_\epsilon$, provided ϵ can be chosen so that

$$\frac{1}{2} J\epsilon^2 + \delta \leq \epsilon \quad \text{and} \quad K \times \left(\frac{1}{2} J\epsilon^2 + \delta\right) \leq \epsilon. \tag{3}$$

If $K < 1$, then the first inequality implies the second, and if $K > 1$, then the second implies the first.

Consequently, if ϵ is chosen so that the appropriate inequality of (3) is solved as an equality, then $\widehat{T}(\mathcal{Z}_\epsilon) \subseteq \mathcal{Z}_\epsilon$, where $\widehat{T}(z) = T_{w(z)}(z)$. Since $\widehat{T}$ is continuous it follows from Brower's fixed point theorem that there exists $w^\star \in \mathcal{W}_{\beta\epsilon}$ and $z^\star \in \mathcal{Z}_\epsilon$ such that $T_{w^\star}(z^\star) = z^\star$. These are the required z and w. It is easily seen that to satisfy (3) the required ϵ is as stated in the proposition. $\square$

The proof of propositions 3, 4 and 5 require the following two results. Define $\mathcal{L}(\mathbb{V})$ to be the space of linear operators on $\mathbb{V}$.

Proposition 7 *Any right inverse of L_x can be written as $B^+C^+ + B^-C^-$ where C^+ and C^- are arbitrary $n \times n$ diagonal matrices over $\mathcal{L}(\mathbb{V})$ such that $C^+ + C^- = I$, and B^+ and B^- are the fixed $(n+1) \times n$ matrices over $\mathcal{L}(\mathbb{V})$ with*

$$B_{ij}^+ = \begin{cases} 0 & i \leq j \\ I & i = j+1 \\ A_{i-1} \dots A_{j+1} & i > j+1, \end{cases}, \quad B_{ij}^- = \begin{cases} 0 & i > j \\ -A_i^{-1} \dots A_j^{-1} & i \leq j, \end{cases}$$

where $A_i = df(x_i)$.

It might help to understand this result by writing out the case where $n = 3$, which gives,

$$L_x = \begin{pmatrix} -A_1 & I & 0 & 0 \\ 0 & -A_2 & I & 0 \\ 0 & 0 & -A_3 & I \end{pmatrix}$$

and

$$B^+ = \begin{pmatrix} 0 & 0 & 0 \\ I & 0 & 0 \\ A_2 & I & 0 \\ A_3A_2 & A_3 & I \end{pmatrix}, \quad B^- = \begin{pmatrix} -A_1^{-1} & -A_1^{-1}A_2^{-1} & -A_1^{-1}A_2^{-1}A_3^{-1} \\ 0 & -A_2^{-1} & -A_2^{-1}A_3^{-1} \\ 0 & 0 & -A_3^{-1} \\ 0 & 0 & 0 \end{pmatrix}.$$

Proof of proposition 7: First observe that B^+ and B^- are right inverses of L_x, and consequently, so is $B^+C^+ + B^-C^-$. Next observe that L_x is of full rank, because each $A_i = df(x_i)$ is invertible, because f is a diffeomorphism. By the rank nullity theorem the subspace of right inverses of L_x has dimension n in the space of $(n+1) \times n$ matrices. On the other hand, the subspace defined by $B^+C^+ + B^-C^-$ has dimension n, so it must include all the right inverses. $\square$

Proposition 8 *Let $\mathbb{U}^n$ be equipped with γ-mean-max norms. If f is partially hyperbolic and $\lambda \in (\lambda_0, \lambda_1)$, then there exists a right inverse of L_x, such that for all $v \in E_\lambda^* = E_\lambda^{(-1)} \oplus E_\lambda^{(+1)}$ with $\|v\|_a \leq 1$, there is a $w \in E_\lambda^*$, with $\|w\|_b \leq 1$, such that $\|L_x^{-1}(v + \beta w)\|_a \leq K(\lambda)$.*

Proof of proposition 8: Using proposition 7 write $L_x^{-1} = B^+C^+ + B^-C^-$ and set $C^+ = P^{(+1)} + \alpha_+ P^{(0)}$ and $C^- = P^{(-1)} + \alpha_+ P^{(0)}$, where $\alpha_+, \alpha_- \in \mathbb{R}$, $\alpha_+ + \alpha_- = 1$, and $P^{(\kappa)}$ is a projection onto $E_\lambda^{(\kappa)}$ where $P^{(+1)}$ projects perpendicularly onto E_λ^* then projects onto $E_\lambda^{(+1)}$ parallel to $E_\lambda^{(-1)}$, $P^{(-1)}$ projects similarly, and $P^{(0)}$ is a projection parallel to E_λ^*. (See figure 1.) Note that these C^+, C^-, are diagonal matrices over $\mathcal{L}(\mathbb{V})$ and $C^+ + C^- = I$. From the partial hyperbolicity of f it follows for $v \in E_\lambda^*$ that

$$\begin{aligned}\|L_x^{-1}v\|_a &= \|B^+C^+v + B^-C^-v\|_a \\ &\leq \frac{1}{2\sin\theta/2}\max\left\{\sup_i \sum_j \lambda^{|i-j-1|}\|v_j\|, \quad \frac{1}{\gamma}\sum_{i,j}\lambda^{|i-j-1|}\|v_j\|\right\}.\end{aligned}$$

The factor involving θ appears because this is the maximum expansion of the length $\|v_j\|$ as a consequence of the parallel projections. (See figure 1.) The constraint $\|v\|_a \leq 1$ defines a convex set and so on this subset of E_λ^* the supremum of $\frac{1}{\gamma}\sum_{i,j}\lambda^{|i-j-1|}\|v_j\|$ occurs when $\|v_j\| = 1$ for the terms with largest coefficient $\lambda^{|i-j-1|}$, subject to $\sum_j \|v_j\| = n\gamma$. This implies that for $v \in E_\lambda^*$

$$\sup_{\|v\|_a=1} \|L_x^{-1}v\|_a \leq \frac{1}{2\sin\theta/2}\left(2\frac{1-\lambda^{(n\gamma+1)/2}}{1-\lambda} - 1\right),$$

that is, choose the $\lfloor n\gamma \rfloor$ components v_i with largest coefficients to have $\|v_i\| = 1$. (In the above we bump up the power to $(n\gamma+1)/2$ to allow that $n\gamma$ is generally not integer.) On the other hand, if $\|w\|_b = 1$ and $w \in E_\lambda^*$, then there exists a w with

$$\sup_{\|w\|_b=1} \|L_x^{-1}w\|_a \geq \frac{1}{2\sin\theta/2}\left(\frac{1-\lambda^n}{1-\lambda}\right),$$

that is, let every component w_i have $\|w_i\| = 1$. The statement in the theorem now follows. □

Proof of proposition 3: If f is hyperbolic with uniform bound λ, then the result follows trivially from proposition 8, because $E_\lambda^* = E_\lambda^{(-1)} \oplus E_\lambda^{(+1)} = \mathbb{U}^n$, so the stated bound is a bound on $\|L_x^{-1}\|_{a,b,\beta}$. □

Proof of proposition 4: Define $F(\theta) = 1/(2\sin\theta/2)$. Observe $K(\lambda)$ and its derivative $K'(\lambda)$ have the properties that,

$$\begin{aligned}&K(0) = (1-\beta)F(\theta), \quad \lim_{\lambda\to1} K(\lambda) = n(\gamma-\beta)F(\theta),\\ &K'(0) = (2-\beta)F(\theta), \quad \lim_{\lambda\to1} K'(\lambda) = (\tfrac{1}{4}(n^2\gamma^2-1) - \tfrac{1}{2}n(n-1)\beta)F(\theta).\end{aligned}$$

The limit results are obtained by employing L'Hopital's rule. If $n^2\gamma^2 - 1 < 2n(n-1)\beta$, then $K(\lambda)$ is increasing at $\lambda = 0$ and decreasing at $\lambda = 1$, which implies $K(\lambda)$ has a global extremum for some $0 < \lambda^\star < 1$. This occurs where $K'(\lambda^\star) = 0$, that is, $\lambda^\star$ is the appropriate solution of

$$\beta(n-1)\lambda^n - \beta n\lambda^{n-1} - (n\gamma-1)\lambda^{(n\gamma+1)/2} + (n\gamma+1)\lambda^{(n\gamma-1)/2} + \beta - 2 = 0.$$

Furthermore, since $(1-\lambda^m)/(1-\lambda)$ is increasing for $\lambda \in [0,1)$, and the rate of increase increases with m, it is easily shown that if $n^2\gamma^2 - 1 \geq 2n(n-1)\beta$, then $K(\lambda)$ is increasing for $\lambda \in [0,1)$ and hence $K(\lambda) < n(\gamma-\beta)F(\theta)$.

Proof of proposition 5: If $K(\lambda_1) < 0$, then $K(\lambda) < 0$ for all $\lambda \in (\lambda_1, 1)$. If $K(\lambda) < 0$, then it follows from the argument of proposition 8 that the effect of any $v \in E_\lambda^* \ominus E_{\lambda_1}^*$ can be completely nullified by some $w \in \mathbb{V}^n$ with $\|w\|_b \leq \beta$. Hence, only $v \in E_{\lambda_1}^*$ need be considered. □

Proof of proposition 1: The first part of the proposition follows from proposition 2 using γ-mean-max norms. The approximate value of K for hyperbolic f follows from simplification of proposition 3. For the weakly hyperbolic f observe that from proposition 4 that the equation

$$\beta(n-1)\lambda^n - \beta n\lambda^{n-1} - (n\gamma - 1)\lambda^{(n\gamma+1)/2} + (n\gamma+1)\lambda^{(n\gamma-1)/2} + \beta - 2 = 0$$

to a first approximation, that is, when $\gamma \ll 1$, has a solution $\lambda^\star$ with

$$\lambda^\star \geq \left(\frac{2-\beta}{\beta(n-1)}\right)^{\frac{1}{n}}.$$

The case of partially hyperbolic f follows from simplification of proposition 5. □

Proof of proposition 6: Let $v = u + w$. We have the following inequalities

$$\begin{aligned} \lambda^{-2}\|v\|^2 \leq \|Av\|^2 &\leq \|A(u+w)\|^2 \\ &\leq \|Au\|^2 + \|Aw\|^2 \\ &\leq \lambda^2\|u\|^2 + D^2\|w\|^2 \end{aligned}$$

On the other hand the cosine rule states

$$\|w\|^2 = \|u\|^2 + \|v\|^2 - 2\|u\|\|v\|\cos\phi.$$

Eliminating $\|w\|$ obtains

$$\lambda^{-2}\|v\|^2 - \lambda^2\|u\|^2 \leq D^2(\|u\|^2 + \|v\|^2 - 2\|u\|\|v\|\cos\phi). \tag{4}$$

If $0 \leq \phi < \pi/2$, that is, the angle ϕ between u and v is acute and not a right angle, then

$$\lambda^{-2}\|v\|^2 - \lambda^2\|u\|^2 < D^2(\|u\|^2 + \|v\|^2). \tag{5}$$

Since we are only concerned about the angle ϕ between u and v, we can scale both vectors so that $\|u\| = \|v\| = 1$. With this substitution the inequality stated in the proposition follows immediately from (4). If $\phi = 0$ and $\|u\| = \|v\| = 1$, then (4) implies that $\lambda^{-2} - \lambda^2 \leq 0$, which contradicts $0 < \lambda < 1$, hence $\phi > 0$. Alternatively, making the substitution $\|u\| = \|v\| = 1$ into (5) implies that $\lambda^{-2} - \lambda^2 < 2D^2$, which when combined with the fact that $0 < \lambda < 1$ in turn implies the right hand side of the stated inequality is strictly positive and less than one. □

5 Applications and Discussion

One method of finding shadowing trajectories and pseudo-orbits is gradient descent of indeterminism (GDI). Suppose one is given an initial sequence of states $\theta = (\theta_i)_{i=0}^n$ and the aim is to find a pseudo-orbit, $x = (x_i)_{i=0}^n$ of a model f that is close to θ having smaller mismatches $x_i - f(x_{i-1})$. Define the *indeterminism* of x as

$$I(x) = \frac{1}{n}\sum_{i=1}^{n} \|x_i - f(x_{i-1})\|^2. \tag{6}$$

Defined in this way, indeterminism is the mean square size of the mismatches of a pseudo-orbit. (Note that $I(x) = 0$ if and only if x is a trajectory.) One can obtain from θ a family of pseudo-orbits $x(\xi)$ with progressively smaller mismatches by solving the differential equation

$$\frac{dx}{d\xi} = -\nabla I(x), \quad x(0) = \theta. \tag{7}$$

One can obtain an iterative algorithm for reducing indeterminism by integrating this differential equation by Euler's method. For example, using the Euclidean norm, choose a step size Δ, then iterate from $x = \theta$,

$$x_i \mapsto x_i - \frac{2\Delta}{n} \begin{cases} -A(x_i)(x_{i+1} - f(x_i)), & i = 0, \\ (x_i - f(x_{i-1})) - A(x_i)(x_{i+1} - f(x_i)), & 0 < i < n, \\ x_i - f(x_{i-1}), & i = n, \end{cases} \tag{8}$$

where $A(x)$ is the adjoint $df(x)^T$. This is a remarkably successful algorithm for obtaining pseudo-orbits with smaller mismatches. Typically, $2\Delta/n = 0.1$ is suitable. For some models the adjoint is not readily available, however, GDI algorithm is quite robust and still works with approximate adjoints, even as bad an approximation as a constant times the identity matrix [18]. The point being that to decrease indeterminism one does not necessary have to take the direction of steepest descent.

GDI has been applied to many systems, but we are most interested in applications to high dimensional systems, such as weather forecasting models. The author, for example, has applied GDI to a simple quasi-geostrophic model of the atmosphere [18] (dimension ≈ 1500) and an operational weather forecasting model at reduced resolution [15] (dimension $\approx 7.5 \times 10^5$). In both of these applications the model certainly has Lyapunov exponents near zero, and possibly local Lyapunov exponents that fluctuate about zero [8, 21, 22, 31]. Three important issues that arise in the application of GDI are:

1. Concern, often raised as a criticism, that GDI has no means to prevent wandering far from the initial sequence of states θ.

2. Since GDI will never converge to a trajectory, when should the gradient descent be stopped, that is, when has sufficient convergence been achieved?

3. How does model error influence the convergence of GDI? Can the mismatches of a pseudo-orbit give an indication of the nature of model error?

In the following we use proposition 1 to provide some guidance on these, and other, issues. Although we speak only of GDI it is likely that much of what we say applies to other algorithms for finding shadowing pseudo-orbits.

5.1 Numerical computations of λ and K

As stated previously, the parameters β, γ and λ are not independent. Figure 2(a) shows numerical computations of a bound on λ for various β and γ. This bound has been computed according to proposition 4, which is more precise than the bound stated in proposition 1.

Figure 2(b) shows numerical computations of a bound $K/F(\theta)$, that is, the K of proposition 1 without the factor $F(\theta) = 1/(2 \sin \theta/2)$. The numerical bound has been computed according to proposition 1 using the bound computed for λ. In the most optimistic situation one has $\theta = \pi/2$, so the value of K is a factor of $\sqrt{2}$ less than the value of $K/F(\theta)$ shown in figure 2. As $\lambda \to 1$ it is likely that $\theta \to 0$, which will inflate K (significantly) more than the value of $K/F(\theta)$ shown in figure 2.

The implications of figure 2 will be discussed further in the following.

5.2 Convergence of gradient descent algorithms

Proposition 1 and the computed bounds on λ and K shown in figure 2 can be used to appreciate the convergence properties of GDI. In the terms of proposition 1, GDI corresponds to finding successive y that reduce the value of β, with a trajectory obtained when $\beta = 0$. (Note the convergence has a non-trivial quadratic effect: the bound on the mismatch of y is $\beta\epsilon$, but ϵ depends on K, which depends on β.) It is also possible that γ varies as iteration proceeds.

Other things being equal, the rate of convergence of GDI is governed by the value of λ [28]; the smaller λ, the faster the convergence, and if $\lambda \to 1$, then convergence essentially stops. From figure 2(a) it can be seen that as β decreases, λ increases, and furthermore, unless γ is small, convergence will be slow, and will grind to a halt long before a trajectory is obtained. It should be stressed that figure 2(a) represents a worse case analysis, because convergence will depend on the nature of the initial mismatches of x and perhaps other factors discussed later.

5.3 Wandering

The parameter ϵ in proposition 1 is an explicit bound on how far y can be expected to wander from x. From proposition 1, if δ is small, then $\epsilon \approx K\delta$. Now K can be greater than one[2]. From proposition 1 when β is sufficiently close to one, then one could have $K < 1$. Of course, if $\beta = 1$, then $\beta\epsilon = \delta$ and y has the same size mismatches as x, which is of no value.

From the values of K implied by figure 2(b) it is seen that there can be significant movement away from x as GDI proceeds, that is as β decreases, especially when γ is large. This assertion could be a little misleading, because, as the preceding discussions point out, for large γ convergence grinds to halt before it wanders this far. Even so, there is still the potential for significant wandering when γ is small.

Once again it should be stressed that figure 2(b) represents a worse case analysis, however, this upper bound should be more of a concern than it was for convergence issues, because there is more potential for wandering when the model is imperfect. This should be taken as a warning not to force convergence of GDI with imperfect models, or at least to watch for wandering.

5.4 Effects of model error on gradient descent

Model error and observational errors tend to have different characteristics: observational errors tend to be random, whereas model errors tend to be systematic. Even when model errors are effectively random, there is typically a strong systematic component too, because model error often results from the model having an incorrect functional form or incorrect parameter values. If model errors have a systematic component, then model trajectories tend to drift away from true trajectories [25]. The effects of systematic model error on GDI can be understood in terms of the parameter γ. (See also the following section for a discussion of how the interpretation of γ may need to be adjusted for spatially extended systems, like weather models.)

First consider the nature of γ given a perfect model. If an initial $x = \theta$ were influenced by observational errors, say independent Gaussian errors with mean zero and variance σ^2, then γ should be significantly less than one, because for most x_i the error will be smaller than σ, but some may be larger than 2σ. As GDI proceeds it typically reduces mismatches in proportion to their magnitude, because it follows the gradient of the mismatches. Thus, in a situation of a perfect model and random state errors one should anticipate that as GDI proceeds the value of γ is initially significantly less than one and will remain so. From figure 2 one can also expect GDI to obtain good convergence toward a trajectory of the model.

Consider now an imperfect model and a trajectory x of the true system that is perfectly observed. This x will be a pseudo-orbit of the imperfect model f. Systematic model error (drift) will be revealed in x through the mismatches $x_i - f(x_{i-1})$ being in a similar direction, and of similar magnitude, for consecutive states. Even if there is a random component, the

[2]Chow and Palmer made a direct numerical estimate of $K \approx 113277$ for the Henon map [5], but they were considering the situation where $\beta = 0$.

consecutive mismatches will be highly correlated. If this is the case, then γ will be close to one, because all the mismatches are of similar size. (In contrast with random errors as discussed above for which γ should be significantly less than one.) According to proposition 1 and figure 2, if γ is close to one, then GDI cannot be expected to obtain pseudo-orbits with much smaller mismatches, that is, progress of GDI will be slow, small, or non-existent, and convergence to a trajectory is unlikely.

If the model is imperfect and the initial θ also includes observational errors that are large relative to model error, then γ can be significantly less than one. In this case GDI can be expected to at first reduce the random observational error component of the mismatches, but as the random errors are reduced to a size similar to the systematic model error, the value of γ will increase toward one, and GDI will slow to a halt well short of obtaining a trajectory. When this happens the residual mismatches may provide useful information about the nature of the systematic model error.

On the other hand, model error need not imply γ is close to one. It could happen that model error is larger in some states than others. Then the model may be fairly good most of the time and have occasional bad periods; the model error is localised in time. In this case γ can be small, and proposition 1 implies good shadowing pseudo-orbits can be found.

5.5 Spatially extended systems

For spatially extended systems, like weather and climate models, the discussion of the preceding section concerning the role of γ may need to be modified. The problem that arises is γ typically always appears to be close to one, regardless of the nature of the errors. For example, in the weather model discussed later, γ is typically around 0.9, but GDI still achieves significant reduction of mismatches. This would appear to contradict the indications of proposition 1 and figure 2, but really it requires a re-evaluation of the role of γ. First we consider two ways that γ can be close to one, and GDI will still be effective. We then discuss how γ can be re-interpreted in these cases.

In spatially extended systems observational errors are typically independent on long length scales. Suppose, for example, each component of $\theta_i \in \mathbb{R}^d$ has an independent normally distributed error with mean zero and variance σ^2. The square of the magnitude of the error will have a χ^2-distribution with d degrees of freedom. If d is large, then the magnitude of the error will have a very small variance. Consequently, γ will be approximately one.

Similarly, in spatially extended systems model errors can be localized in space and move around in time. As a consequence, if the localized errors are of a similar size, then, in a typical norm, γ will be close to one.

We can re-interpret γ as follows. The value of γ is defined by the norm on state space $\mathcal{B}$. In proposition 1 the underlying norm on $\mathcal{B}$ is not specified, which means this norm can be any convenient or pertinent norm. If the system is spatially extended one could choose a norm that is spatially localised, emphasizing differences in locales of interest. For example, using a norm that emphasized differences over North America, or Japan, or over the tropical Pacific Ocean. With a localised normed, observational and model errors are only large when they directly effect the locality. Viewed in this way, it can be understood why GDI is still effective with localised errors. Whether there is any advantage in using a localised norm, rather than a global norm, in practice needs further investigation.

In situations like this an indication of the effective value of γ can be determined by reducing γ in proportion to the size of the region where errors are localised relative to the size of the whole domain.

5.6 Application to a weather forecasting model

The author has applied GDI to an operation weather forecasting model using real atmospheric observations [15]. The model used was the Navy Operational Global Atmospheric Prediction System (NOGAPS) [2]. Various experiments involving GDI have been performed using operational observations from March and October 2003. The results discussed here used a T79L30 resolution model with a 6 hour forecast period. The initial pseudo-orbit x was 7 days of analyses (29 states) obtained by three dimensional variational assimilation at the model's resolution of atmospheric observations using the NAVDAS implementation [6]. GDI was implemented as equation (8), with $2\Delta/n = 0.1$, and $A(x)$ a dry atmosphere approximation of the adjoint [15].

Figure 3 shows how the indeterminism typically varies as the GDI algorithm is iterated. Note how there is an initial rapid decrease in the first 5 to 10 steps, followed slower decrease. Normalising to have $\delta = 1$, it was found that $\epsilon \approx 1.44$ and $\beta \approx 0.43$ after about 10 iterations of the algorithm used, and $\beta \approx 0.27$ after 50 iterations, with a very slow rate of convergence by this stage. Using a global spectral norm a direct computation gives $\gamma \approx 0.9$, but this is misleading because of the highly localised spatio-temporal nature of the initial mismatches. Viewing zonal averages of the mismatches suggests significant localization and an effective γ of around 0.3 to 0.6.

Working with the computed values of ϵ, β and γ, it is seen that the observed convergence behaviour is consistent with the bounds indicated by figure 2(a), that is, the slowing down after an initial rapid convergence is consist with approaching the λ bound for the computed β and γ values.

The ϵ value is quite small given the worse case bounds of figure 2(b); even with $\gamma = 0.2$ we could have expected ϵ up to 2. This is encouraging. Even if we inflate ϵ by $1/0.3$, as a crude assumption that mismatches *only* occur in 30% of the spatial variables, then this would give $\epsilon \approx 4.8$, which is still consistent with figure 2(b).

We conclude that the convergence properties and wandering of the gradient descent algorithm we applied to an operational weather model are consistent with the theory developed, and therefore indicates the results obtained with the weather model are both reasonable and useful.

6 Conclusions and further work

The introduction of the idea of *shadowing pseudo-orbits* has provided a useful tool to understand and analyse gradient descent algorithms for finding shadowing trajectories. The main theorem shows that provided a non-hyperbolic system is sufficiently like a hyperbolic system, then shadowing pseudo-orbits exist and some of their properties can be explicitly stated. Since gradient descent algorithms iteratively improve pseudo-orbits, the main theorem can be used to assess the convergence in non-hyperbolic systems. It can be expected that the algorithm's convergence will slow to an almost halt before it convergences to a trajectory. (Although partial convergence may be useful, because, for example, it may improve state estimates for forecasting or give better indications of model error.)

The definitions of *hyperbolic* and *partially hyperbolic* were in terms of a single growth rate parameter λ. This choice was made to simplify the proofs. It is common to define hyperbolicity in terms of two parameters, one for unstable modes, and another for stable modes. This can be done for partial hyperbolicity too. By doing so one should be able to obtain tighter bounds than those stated in propositions 1, 3 and 4.

Clearly, further work needs to be done on application to spatially extended systems, because the results as currently stated are not immediately applicable, some adjustment was necessary.

References

[1] D.V. Anosov. Geodesic flows and closed Riemannian manifolds with negative curvature. *Proc. Steklov Inst. Math.*, 90, 1967.

[2] N. E. Baker, R. Daley, R. Gelaro, J. Goerss, T. Hogan, R. Langland, R. Pauley, M. Rennick, C. Reynolds, G. Rohaly, T. Rosmond, and S. Swadley. The navy operational global atmospheric prediction system: A brief history of past, present, and future developments. Technical report, Naval Research Laboratory, Monterey, California, 1998.

[3] R. Bowen. ω-limit sets for axiom A diffeomorphisms. *J. Diff. Eqns.*, 18:333–339, 1975.

[4] J. Brocker, U. Parlitz, and M. Ogorzalek. Nonlinear noise reduction. *Proceedings of the IEEE*, 90(5):898–918, 2002.

[5] S.E. Chow and K.J. Palmer. On the numerical computation of orbits of dynamical systems: The higher dimensional case. *Journal of complexity*, 8:398–423, 1992.

[6] R. Daley and E. Barker. NAVDAS: Formulation and diagnostics. *Monthly Weather Review*, 129:869–883, 2001.

[7] M. Davies. Noise reduction by gradient descent. *International Journal of Bifurcation and Chaos*, 3(1):113–118, 1992.

[8] S. Dawson, C. Grebogi, T. Sauer, and J.A. Yorke. Obstructions to shadowing when a lyapunov exponent fluctuates about zero. *Phys. Rev. Lett.*, 73:1927, 1994.

[9] P. Diamond, P. Kloeden, V. Kozyakin, , and A. Pokrovskii. Semihyperbolic mappings. *Journal of Nonlinear Science*, 5:419–431, 1995.

[10] P. Diamond, P. Kloeden, V. Kozyakin, and A. Pokrovskii. Robustness of observed behaviour of semi-hyperbolic dynamical systems. *Avtomatika i Telemekhanika*, 11, 1994.

[11] I. Gilmour. *Nonlinear model evaluation: iota-shadowing, probabilistic prediction and weather forecasting*. PhD thesis, Mathematical Institute, Oxford University, 1998.

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

[13] S. M. Hammel, J. A. Yorke, and C. Grebogi. Numerical orbits of chaotic processes represent true orbits. *Bulletin (New Series) of the American Mathematical Society*, 19, 1987.

[14] K. Judd. Nonlinear state estimation, indistinguishable states and the extended Kalman filter. *Physica D*, 183:273–281, 2003.

[15] K. Judd, C.A. Reynolds, and T. Rosmond. Toward shadowing in operational weather prediction. Technical Report NRL/MR/7530-04-18, Naval Research Laboratory, Monterey, CA, 93943-5502 USA., 2004.

[16] K. Judd and L.A. Smith. Indistinguishable states I : perfect model scenario. *Physica D*, 151:125–141, 2001.

[17] K. Judd and L.A. Smith. Indistinguishable states II : imperfect model scenarios. *Physica D*, 196:224–242, 2004.

[18] K. Judd, L.A. Smith, and A. Weisheimer. Gradient free descent: shadowing and state estimation with limited derivative information. *Physica D*, 190:153–166, 2004.

[19] A Katok and B. Hasselblatt. *Introduction to the modern theory of dynamical systems*, volume 54 of *Encylopedia of Mathemtics and its Applications*. Cambridge Univeristy Press, 1995.

[20] P.E. Kloeden, J. Ombach, and A.V. Pokrovksii. Continuous and inverse shadowing. *J. Functional Differential Eqns.*, 6:135–151, 1999.

[21] E. Kostelich, I. Kan, C. Grebogi, E. Ott, and J.A. Yorke. Unstable dimension variability: A source of nonhyperbolicity in chaotic systems. *Physica D*, 109:81, 1997.

[22] Y.C. Lai, C. Grebogi, and J. Kruths. Modeling of deterministic chaotic systems. *Physical Review E*, 59(3):2907–2910, 1999.

[23] A. Livsic. Certain properties of the homology of Y-systems. *Mat. Zametki*, 10:555–564, 1971.

[24] V. Nitica and M Pollicott. Transitivity of euclidean extensions of anosov diffeomorphisms. *Ergodic Theory and Dynamical Systems*, 200?

[25] D. Orrell, L.A. Smith, J. Barkmeijer, and T. Palmer. Model error in weather forecasting. *Nonlinear Processes in Geophysics*, 8:357–371, 2001.

[26] S.Y. Pilyugin and O.B. Plamenevskaya. Shadowing is generic. *Topology and its Applications*, 97:253–266, 1999.

[27] M. Pollicott. *Lectures on Ergodic Theory and Pesin Theory on Compact Manifolds*, volume 180. Cambridge Unversity Press, Cambridge, 1993.

[28] D. Ridout and K. Judd. Convergence properties of gradient descent noise reduction. *Physica D*, 165:27–48, 2001.

[29] T. Sauer, C. Grebogi, and J. Yorke. How long do numerical chaotic solutions remain valid? *Physical Review Letters*, 79(1):59–62, 1997.

[30] L.A. Smith and I. Gilmour. Accountability and internal consistency in ensemble formation. In *Proceedings of a workshop on predicability*, volume CXXXIII, ECMWF, Shinfield Park, Reading RG2 9AX, October 1997.

[31] R.L. Viana, J.R.R. Barbosa, C. Grebogi, and A. Batista. Simulating a chaotic process. *Brazilian Journal of Physics*, 35(1):1–9, 2005.

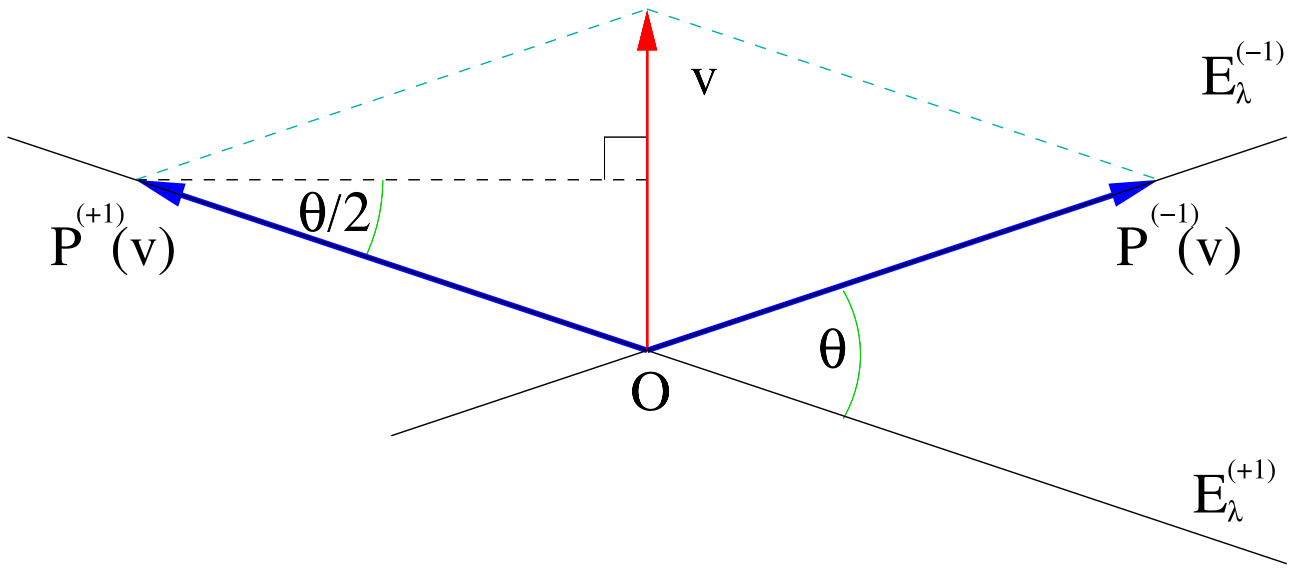
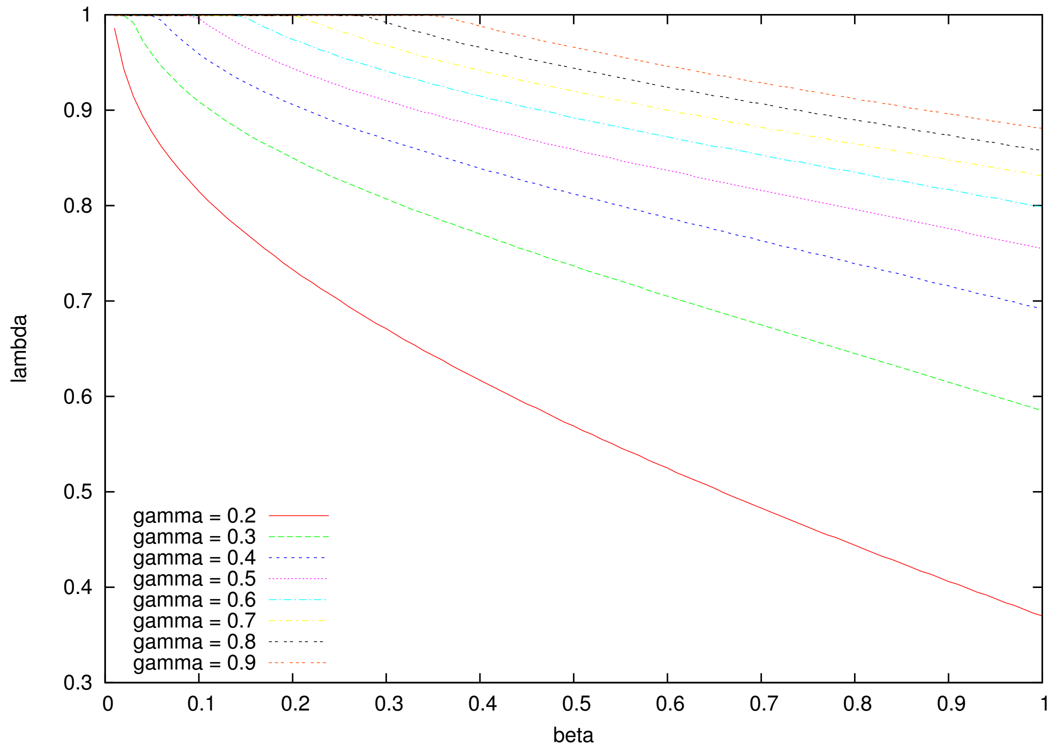


Figure 1: Parallel projections of v onto $E_\lambda^{(+1)}$ and $E_\lambda^{(-1)}$, illustrating how the expansion factor can be up to $1/2\sin(\theta/2)$.

(a)



(b)

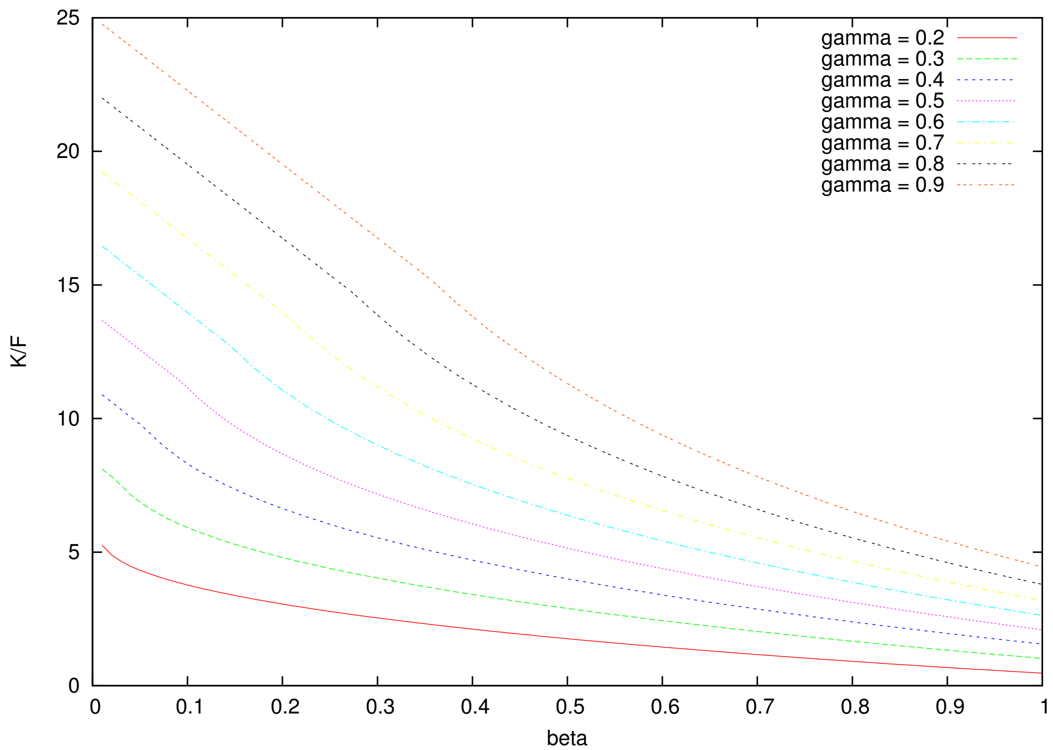


Figure 2: (a) The bound on λ computed according to proposition 4, (b) $K/F(\theta)$, that is, the bound on K stated in proposition 1 without the factor $F(\theta) = 1/(2\sin\theta/2)$ involving the angle bound θ.

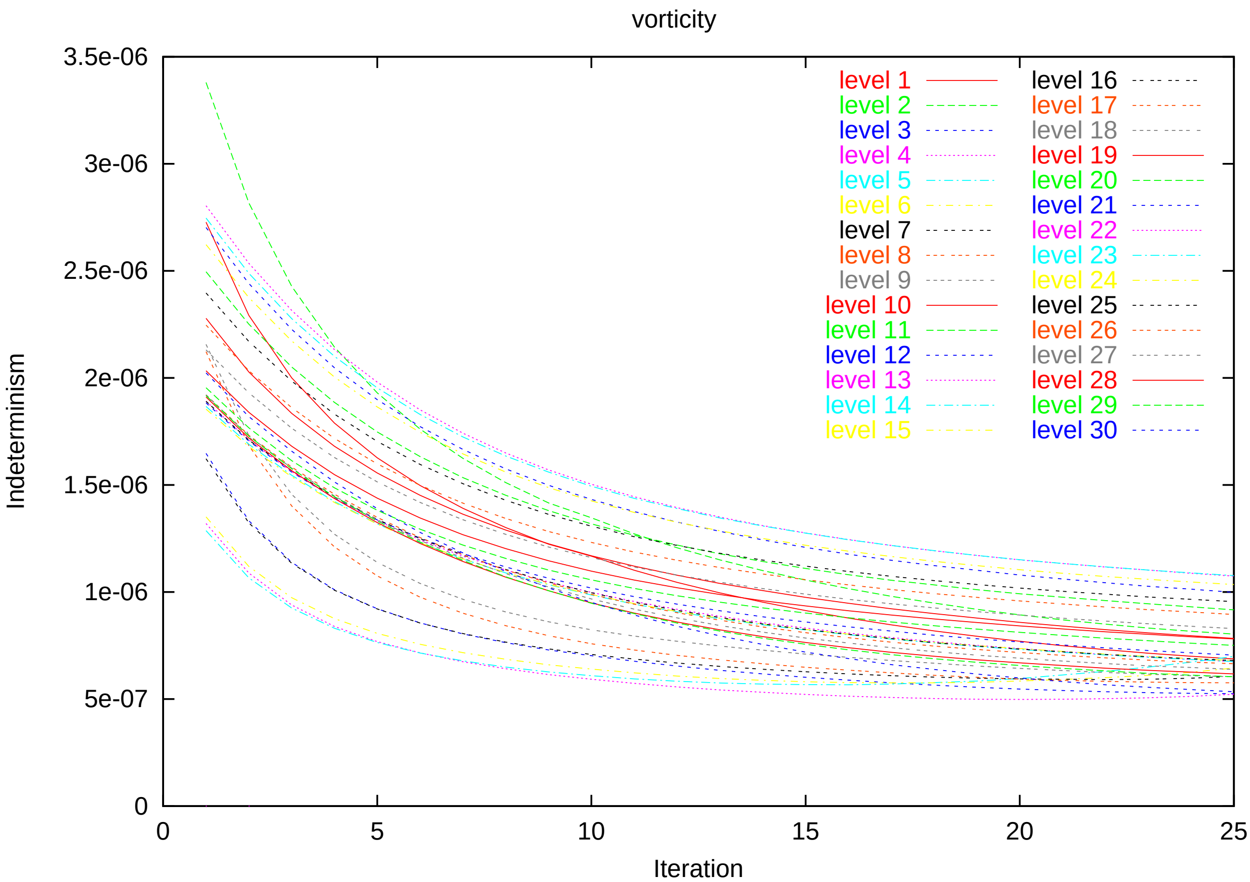


Figure 3: Convergence of indeterminism for the T79L30 NOGAPS model. This figure shows the indeterminism computed separately for each level of the vorticity field. The other prognostic variables (divergence, temperature, specific humidity and surface pressure) behave similarly.