Convergence of the Least Squares Shadowing Method for Computing Derivative of Ergodic Averages

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1137/130917065">http://dx.doi.org/10.1137/130917065</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Society for Industrial and Applied Mathematics</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Sat Dec 22 05:48:07 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/88173">http://hdl.handle.net/1721.1/88173</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.</td>
</tr>
</tbody>
</table>
CONVERGENCE OF THE LEAST SQUARES SHADOWING METHOD FOR COMPUTING DERIVATIVE OF ERGODIC AVERAGES

QIQI WANG†

Abstract. For a parameterized hyperbolic system \( u_{i+1} = f(u_i, s) \), the derivative of an ergodic average \( \langle J \rangle = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} J(u_i, s) \) to the parameter \( s \) can be computed via the least squares shadowing method. This method solves a constrained least squares problem and computes an approximation to the desired derivative \( \frac{d\langle J \rangle}{ds} \) from the solution. This paper proves that as the size of the least squares problem approaches infinity, the computed approximation converges to the true derivative.

Key words. sensitivity analysis, linear response, least squares shadowing, hyperbolic attractor, chaos, statistical average, ergodicity

AMS subject classifications. 37M25, 93E24

DOI. 10.1137/130917065

1. Introduction. Consider a family of \( C^1 \) bijection maps \( f(u, s) : \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^m \) parameterized by \( s \in \mathbb{R} \). We are also given a \( C^1 \) function \( J(u, s) : \mathbb{R}^m \times \mathbb{R} \to \mathbb{R} \). We assume that the system is ergodic, i.e., the infinite time average

\[
\langle J \rangle = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} J(u_i, s), \quad \text{where } u_{i+1} = f(u_i, s), \ i = 1, \ldots,
\]

depends on \( s \) but does not depend on the initial state \( u_0 \). The least squares shadowing (LSS) method attempts to compute its derivative via the following theorem.

THEOREM LSS. Under ergodicity and hyperbolicity assumptions (details in section 6),

\[
\frac{d\langle J \rangle}{ds} = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} (DJ(u_i, s)) v_i^{(n)} + (\partial_s J(u_i, s)),
\]

where \( v_i^{(n)} \in \mathbb{R}^m, i = 1, \ldots, n \), is the solution to the constrained least squares problem

\[
\min \left\{ \frac{1}{2} \sum_{i=1}^{n} v_i^{(n)T} v_i^{(n)} \right\} \quad \text{subject to } v_i^{(n)} = (Df(u_i, s)) v_i^{(n)} + (\partial_s f(u_i, s)),
\]

\( i = 1, \ldots, n - 1 \). Here the linearized operators are defined as

\[
(DJ(u, s)) v := (D_v J)(u, s) := \lim_{\epsilon \to 0} \frac{J(u + \epsilon v, s) - J(u, s)}{\epsilon},
\]

\[
(Df(u, s)) v := (D_v f)(u, s) := \lim_{\epsilon \to 0} \frac{f(u + \epsilon v, s) - f(u, s)}{\epsilon},
\]

\[
(\partial_s J(u, s)) := \lim_{\epsilon \to 0} \frac{J(u, s + \epsilon) - J(u, s)}{\epsilon},
\]

\[
(\partial_s f(u, s)) := \lim_{\epsilon \to 0} \frac{f(u, s + \epsilon) - f(u, s)}{\epsilon};
\]
(D_J), (\partial_s J), (D_f), and (\partial_s f) are a $1 \times m$ matrix, a scalar, an $m \times m$ matrix, and an $m \times 1$ matrix, respectively, representing the partial derivatives.

Computation of the derivative $d(J)/ds$ represents a class of important problems in computational science and engineering. Many applications involve simulation of nonlinear dynamical systems that exhibit chaos. Examples include weather and climate, turbulent combustion, nuclear reactor physics, plasma dynamics in fusion, and multibody problems in molecular dynamics. The quantities that are to be predicted (the so-called quantities of interest) are often time averages or expected values $\langle J \rangle$. Derivatives of these quantities of interest with respect to parameters are required in applications including

- numerical optimization: The derivative of the objective function $\langle J \rangle$ with respect to the design, parameterized by $s$, is used by gradient-based algorithms to efficiently optimize in high dimensional design spaces;
- uncertainty quantification: The derivative of the quantities $\langle J \rangle$ with respect to the sources of uncertainties $s$ can be used to assess the error and uncertainty in the computed $\langle J \rangle$.

A scientific example is when the dynamical system is a climate model, and the ergodic average $\langle J \rangle$ is the long time averaged global mean temperature. Its derivative with respect to the amount of anthropogenic emissions would be a valuable quantity to study. An engineering example can be found in simulation of turbulent air flow over an aircraft, where the ergodic average $\langle J \rangle$ is the long time averaged drag. Its derivative with respect to shape parameters of the aircraft can help engineers increase the efficiency of their design. Although it is difficult to analyze theoretically whether these complex dynamical systems are ergodic, many of them have been observed to have ergodic quantities of interest, leading to the popular chaotic hypothesis [14, 8, 6, 9]. Efficient computation of the derivative of long time averaged quantities in these systems is an important and challenging problem.

Traditional transient sensitivity analysis methods fail to compute $d(J)/ds$ in chaotic systems. These methods focus on linearizing initial value problems to obtain the derivative of the quantities of interest. When the quantity of interest is a long time average in a chaotic system, the derivative of this average does not equal the long time average of the derivative. As a result, traditional adjoint methods fail, and the root of this failure is the ill-conditioning of initial value problems of chaotic systems [11].

The differentiability of $\langle J \rangle$ has been shown by Ruelle [15]. Ruelle also constructed a formula for the derivative. However, Ruelle’s formula is difficult to compute numerically [11, 7]. Abramov and Majda are successful in computing the derivative based on the fluctuation dissipation theorem [1]. However, for systems whose SRB measure [20] deviates strongly from Gaussian, fluctuation dissipation theorem based methods can be inaccurate. Recent work by Cooper and Haynes has alleviated this limitation by using a nonparametric method for estimating the stationary probability density function [5]. Several more recent methods have been developed for computing this derivative [17, 19, 3, 18]. In particular, the LSS method [18] is a method that computes the derivative of $\langle J \rangle$ efficiently by solving a constrained least squares problem. The primary advantage of this method is its simplicity. The least squares problem can be easily formulated and efficiently solved as a linear system. Compared to other methods, it is insensitive to the dimension of the dynamical system and requires no knowledge of the equilibrium probability distribution in the phase space.

This paper provides a theoretical foundation for the least squares shadowing method by proving Theorem LSS for uniformly hyperbolic maps. Section 2 lays out
the basic assumptions, and introduces hyperbolicity for readers who are not familiar with this concept. Section 3 then proves a special version of the classic structural stability result, and defines the shadowing direction, a key concept used in our proof. Section 4 demonstrates that the derivative of $(J)$ can be computed through the shadowing direction. Section 5 then shows that the least squares shadowing method is an approximation of the shadowing direction. We consider this as a mathematically new and nontrivial result. Section 6 finally proves Theorem LSS by showing that the approximation of the shadowing direction makes a vanishing error in the computed derivative of $(J)$.  

2. Uniform hyperbolicity. In this section we consider a dynamical system governed by

\[(2.1) \quad u_{i+1} = f(u_i, s)\]

with a parameter $s \in \mathbb{R}$, where $u_i \in \mathbb{R}^m$ and $f : \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^m$ is $C^1$ and bijective in $u$. This paper studies perturbation of $s$ around a nominal value. Without loss of generality, we assume the nominal value of $s$ to be 0. We denote $f^{(0)}(u, s) \equiv u$ and $f^{(i+1)}(u, s) \equiv f^{(i)}(f(u, s), s)$ for all $i \in \mathbb{Z}$.

We assume that the map has a compact, global, uniformly hyperbolic attractor $\Lambda \subset \mathbb{R}^m$ at $s = 0$, satisfying:

1. for all $u_0 \in \mathbb{R}^m$, $\text{dist}(\Lambda, f^{(n)}(u_0, 0)) \xrightarrow{n \to \infty} 0$, where $\text{dist}$ is the Euclidean distance in $\mathbb{R}^m$;
2. there is a $C \in (0, \infty)$ and $\lambda \in (0, 1)$, such that for all $u \in \Lambda$, there is a splitting of $\mathbb{R}^m$ representing the space of perturbations around $u$;

\[(2.2) \quad \mathbb{R}^m = V^+(u) \oplus V^-(u),\]

where the subspaces are

- $V^+(u) := \{ v \in \mathbb{R}^m : \|(Df^{(i)}(u, 0)) v\| \leq C \lambda^{-i} \|v\|, \text{ for all } i < 0 \}$ is the unstable subspace at $u$, where $\| \cdot \|$ is the Euclidean norm in $\mathbb{R}^m$, and

$$
(Df^{(i)}(u, s)) v := \lim_{\epsilon \to 0} \frac{f^{(i)}(u + \epsilon v, s) - f^{(i)}(u, s)}{\epsilon} = (Df^{(i-1)}(f(u, s), s))(Df(u, s)) v;
$$

- $V^-(u) := \{ v \in \mathbb{R}^m : \|(Df^{(i)}(u, 0)) v\| \leq C \lambda^i \|v\|, \text{ for all } i > 0 \}$ is the stable subspace at $u$.

Both $V^+(u)$ and $V^-(u)$ are continuous with respect to $u$.

It can be shown that the subspaces $V^+(u)$ and $V^-(u)$ are invariant under the differential of the map $(Df)$, i.e., if $u' = f(u, 0)$ and $v' = (Df(u, 0)) v$, then [16]

\[(2.3) \quad v \in V^+(u) \iff v' \in V^+(u'), \quad v \in V^-(u) \iff v' \in V^-(u').\]

Uniformly hyperbolic chaotic dynamical systems are known as “ideal chaos.” Because of its relative simplicity, studies of hyperbolic chaos have generated enormous insight into the properties of chaotic dynamical systems [10]. Although most dynam-

---

1. A necessary condition for the applicability of our method is that the dynamical system settles down to an attractor after many iterations. The attractor can be a fixed point, a limit cycle, or a strange attractor. Empirically, this means that the system eventually reaches an equilibrium or quasi equilibrium.
ical systems encountered in science and engineering are not uniformly hyperbolic, many of them are classified as quasi hyperbolic. These systems, including the famous Lorenz system, have global properties similar to those of uniformly hyperbolic systems [4]. Results obtained on uniformly hyperbolic systems can often be generalized to quasi-hyperbolic ones. Scholars believe that very complex dynamical systems like turbulence behave like they are quasi hyperbolic [14, 8, 6, 9]. Although this paper focuses on proving the convergence of the least squares shadowing method for uniformly hyperbolic systems, it has been shown numerically that this method also works when the system is not uniformly hyperbolic [18].

3. Structural stability and the shadowing direction. The hyperbolic structure (2.2) ensures the structurally stability [13] of the attractor Λ under perturbation in s. Here we prove a specialized version of the structural stability result.

**Theorem 1.** If (2.2) holds and f is continuously differentiable, then for all sequences \{u^i, i ∈ Z\} ⊂ Λ satisfying \(u^i_{i+1} = f(u^i, 0)\), there is an \(M > 0\) such that for all \(|s| < M\), there is a unique sequence \(\{u^i_s, i ∈ Z\} ⊂ \mathbb{R}^m\) satisfying \(|u^i_s - u^i| < M\) and \(u^i_{i+1} = f(u^i_s, s)\) for all \(i ∈ Z\). Furthermore, \(u^i_s\) is i-uniformly continuously differentiable with respect to \(s\).

Note: i-uniformly continuous differentiability of \(u^i_s\) means for all \(s \in (-M, M)\) and \(\epsilon > 0\): \(\exists \delta : |s' - s| < \delta \Rightarrow \|\frac{du^i_s}{ds'}\| < \epsilon\) for all \(i\). Other than the i-uniformly continuous differentiability of \(u^i_s\), this theorem can be obtained directly from the shadowing lemma [12]. However, the uniformly continuous differentiability result requires a more in-depth proof. A more general version of this result has been proven by Ruelle [15].

To prove the theorem, we denote \(u = \{u_i, i ∈ Z\}\). The norm

\[
\|u\|_B = \sup_{i ∈ Z} \|u_i\|
\]

defines a Banach space \(B\) of uniformly bounded sequences in \(\mathbb{R}^m\). Define the map \(F : B × \mathbb{R} → B\) as \(F(u, s) = \{u_i - f(u_{i-1}, s), i ∈ Z\}\). We use the implicit function theorem to complete the proof, which requires \(F\) to be differentiable and its derivative to be nonsingular at \(u^0\).

**Lemma 2.** Under the conditions of Theorem 1, \(F\) has a Fréchet derivative at all \(u ∈ B\):

\[
(DF(u, s))v = \{v_i - (Df(u_{i-1}, s))v_{i-1}\}, \quad \text{where } v = \{v_i\}.
\]

**Proof.** Because \(\|u\|_B = \sup_{i} \|u_i\| < \infty\), we can find \(C > 2\|u_i\|\) for all \(i\). Because \(f ∈ C^1\), its derivative \((Df)\) is uniformly continuous in the compact set \(\{u : \|u\| ≤ C\}\). For \(\|v\|_B < C/2\), we apply the mean value theorem to obtain

\[
\frac{f(u_i + v_i, s) - f(u_i, s)}{\|v\|_B} - \frac{(Df(u_i, s))v_i}{\|v\|_B} = \frac{(Df(u_i + \xi v_i, s)) - (Df(u_i, s))}{\|v\|_B} v_i,
\]

where \(0 ≤ \xi ≤ 1\). Because \(\|u_i + \xi v_i\| ≤ \|u_i\| + \|v_i\| < C\) for all \(i\), uniform continuity of \((Df)\) implies that for all \(\epsilon > 0\), \(\exists \delta\) such that for all \(\sup \|v_i\| < \delta\),

\[
\left\|\frac{(Df(u_i + \xi v_i, s)) - (Df(u_i, s))}{\|v\|_B} v_i\right\| ≤ \left\|Df(u_i + \xi v_i, s) - (Df(u_i, s))\right\| < \epsilon
\]
for all $i$. Therefore,

$$
\frac{F(u + v, s) - F(u, s)}{\|v\|_B} = \left\{ \frac{v_i - f(u_{i-1} + v_{i-1}, s) - f(u_{i-1}, s)}{\|v\|_B} \right\} \rightarrow \left\{ v_i - (Df(u_{i-1}, s)) v_{i-1} \right\} / \|v\|_B
$$

in the $\mathcal{B}$ norm. Now we only need to show that the linear map $\{v_i\} \rightarrow \{v_i - (Df(u_{i-1}, s)) v_{i-1}\}$ is bounded. This is because $(Df)$ is continuous, thus it is uniformly bounded in the compact set $\{u : \|u\| \leq C\}$. Denote the bound in this compact set as $\|Df\| < A$, then $\|\{v_i - (Df(u_{i-1}, s)) v_{i-1}\}\|_B \leq (1 + A) \|\{v_i\}\|_B$.

**Lemma 3.** Under conditions of Theorem 1, the Fréchet derivative of $F$ at $u^0$ and $s = 0$ is a bijection.

**Proof.** The Fréchet derivative of $F$ at $u^0$ and $s = 0$ is

$$(DF(u^0, 0)) v = \{v_i - (Df(u^0_{i-1}, 0)) v_{i-1}\}.$$ 

We only need to show that for every $r = \{r_i\} \in \mathcal{B}$, there exists a unique $v = \{v_i\} \in \mathcal{B}$ such that $v_i - (Df(u^0_{i-1}, 0)) v_{i-1} = r_i$ for all $i$.

Because of (2.2), we can first split $r_i = r_i^+ + r_i^-$, where $r_i^+ \in V^+(u^0_i)$ and $r_i^- \in V^-(u^0_i)$. Because $V^+(u)$ and $V^-(u)$ are continuous to $u$ and $\Lambda$ is compact,

$$\inf_{u \in \Lambda, r^+ \in V^+(u)} \frac{\|r^+ + r^-\|}{\max(\|r^+\|, \|r^-\|)} = \beta > 0.$$ 

(This is because if $\beta = 0$, then by the continuity of $V^+(u), V^-(u)$, and the compactness of $\{u, r^+, r^- \in \Lambda \times \mathbb{R}^m \times \mathbb{R}^m : \max(\|r^+\|, \|r^-\|) = 1\}$, there must be a $u \in \Lambda, r^+ \in V^+(u), r^- \in V^-(u)$ such that $\max(\|r^+\|, \|r^-\|) = 1$ and $r^+ + r^- = 0$, which contradicts the hyperbolicity assumption (2.2).) Therefore,

$$\max(\|r^+_i\|, \|r^-_i\|) \leq \frac{\|r_i\|}{\beta} \leq \frac{\|r\|_B}{\beta} \text{ for all } i.$$ 

Now let

$$v_i = \sum_{j=0}^{\infty} (Df^{(j)}(u^0_{i-j}, 0)) r^-_{i-j} - \sum_{j=1}^{\infty} (Df^{(-j)}(u^0_{i+j}, 0)) r^+_{i+j}.$$ 

By combining

$$\sum_{j=0}^{\infty} (Df(u^0_{i-1}))(Df^{(j)}(u^0_{i-j-1})) r^-_{i-j-1} = \sum_{j=0}^{\infty} (Df^{(j+1)}(u^0_{i-j-1})) r^-_{i-j-1} = \sum_{j=1}^{\infty} (Df^{(j)}(u^0_{i-j})) r^-_{i-j}$$

and

$$\sum_{j=1}^{\infty} (Df(u^0_{i-1}))(Df^{(-j)}(u^0_{i+j-1})) r^+_{i+j-1} = \sum_{j=1}^{\infty} (Df^{(-j+1)}(u^0_{i+j-1})) r^+_{i+j-1} = \sum_{j=0}^{\infty} (Df^{(-j)}(u^0_{i+j})) r^-_{i+j},$$
we can obtain that
\[ v_i - (Df(u_{i-1}^0))v_{i-1} = v_i - \sum_{j=1}^{\infty} (Df^{(j)}(u_{i-j}^0))r_{i-j}^- + \sum_{j=0}^{\infty} (Df^{(-j)}(u_{i+j}^0))r_{i+j}^+ , \]
\[ = r_i^- + r_i^+ = r_i . \]
Thus it can be verified that \( v_i - (Df(u_{i-1}^0))v_{i-1} = r_i , \) and by the definition of \( V^+(u) \) and \( V^-(u) , \)
\[ ||v_i|| \leq \sum_{j=0}^{\infty} \left| (Df^{(j)}(u_0^0))r_{i-j}^- \right| + \sum_{j=1}^{\infty} \left| (Df^{(-j)}(u_0^0))r_{i+j}^+ \right| \]
\[ \leq \sum_{j=0}^{\infty} C\lambda^j ||r_{i-j}^-|| + \sum_{j=1}^{\infty} C\lambda^j ||r_{i+j}^+|| \leq \frac{2C}{1 - \lambda^r/\lambda^\beta} . \]
Therefore, \( v_i \) is uniformly bounded for all \( i \). Thus \( v \in B \).

Because of linearity, uniqueness of \( v \) such that \( v_i - (Df(u_{i-1}^0))v_{i-1} = r_i \) only needs to be shown for \( r = 0 \). To show this, we split \( v_i = v_i^+ + v_i^- \), where \( v_i^+ \in V^+(u_i^0) \) and \( v_i^- \in V^-(u_i^0) \). Because the spaces \( V^+(u_i^0) \) and \( V^-(u_i^0) \) are invariant \((2.3)\),
\[ 0 = r_i = (v_i^+ - (Df(u_{i-1}^0))v_{i-1}^+)+(v_i^- - (Df(u_{i-1}^0))v_{i-1}^-) , \]
where the contents of the two parentheses are in \( V^+(u_i^0) \) and \( V^-(u_i^0) \), respectively. Because \( V^+(u_i^0) \cap V^-(u_i^0) = \{0\} \), both parentheses in the equation above must be 0 for all \( i \), and
\[ v_i^+ = (Df(u_{i-1}^0))v_{i-1}^+ = \cdots = (Df^{(-j)}(u_{i-j}^0))0 v_j^+ , \]
\[ v_i^- = (Df(u_{i-1}^0))v_{i-1}^- = \cdots = (Df^{(-j)}(u_{i-j}^0))0 v_j^- \quad \text{for all } i > j . \]
By the definition of \( V^+(u_i^0) \) and \( V^-(u_i^0) \), \( ||v_i^+|| \leq C\lambda^{i-j}||v_j^+||, \) \( ||v_i^-|| \leq C\lambda^{i-j}||v_j^-|| \). If \( v_j^+ \neq 0 \) for some \( j \), then
\[ \frac{||v_i||}{\beta} \geq ||v_i^+|| \geq \frac{\lambda^{j-i}}{C} ||v_j^+|| \quad \text{for all } i > j , \]
and \( \{v_i, i \in \mathbb{Z}\} \) is unbounded. Similarly, if \( v_i^- \neq 0 \) for some \( i \), then
\[ \frac{||v_j||}{\beta} \geq ||v_j^-|| \geq \frac{\lambda^{j-i}}{C} ||v_j^-|| \quad \text{for all } j < i , \]
and \( \{v_i, i \in \mathbb{Z}\} \) is unbounded. Therefore, for \( \{v_i\} \) to be bounded, we must have \( v_i = v_i^+ + v_i^- = 0 \) for all \( i \). This proves the uniqueness of \( v \) for \( r = 0 \).

Proof of Theorem 1. \( F(u_0^0, 0) = \{u_0^0 - f(u_{i-1}^0, 0)\} = 0 \). So \( u^0 \) is a zero point of \( F \) at \( s = 0 \). The combination of this and the two lemmas enables application of the implicit function theorem. Thus there exists \( M > 0 \) such that for all \( |s| < M \) there is a unique \( u^s = \{u_i^s\} \) satisfying \( ||u^s - u^0||_s < M \) and \( F(u^s, s) = 0 \). Furthermore, \( u^s \) is continuously differentiable with respect to \( s \), i.e., \( \frac{du^s}{ds} \in B \) is continuous with respect to \( s \) in the \( B \) norm. By the definition of derivatives (in \( B \) and in \( \mathbb{R}^m \), \( \frac{du^s}{ds} = \{ \frac{du_i^s}{ds} \} \)). Continuity of \( \frac{du^s}{ds} \) in \( B \) then implies that \( \frac{du_i^s}{ds} \) is \( i \)-uniformly continuous with respect to \( s \).

Theorem 1 states that for a series \( \{u_i^0\} \) satisfying the governing equation \((2.1)\) at \( s = 0 \), there is a series \( \{u_i^s\} \) satisfying the governing equation at nearby values of \( s \).
In addition, \( u^0_s \) shadows \( u^0 \), i.e., \( u^0_s \) is close to \( u^0 \) when \( s \) is close to 0. Also, \( \frac{du^0_s}{ds}|_{s=0} \) exists and is \( i \)-uniformly bounded.

**Definition 4.** The **shadowing direction** \( v^{(\infty)}_i \) is defined as the uniformly bounded series

\[
v^{(\infty)} := \left\{ v^{(\infty)}_i \right\} := \left\{ \frac{du^0_s}{ds} \right|_{s=0} \right\} = \frac{d\hat{u}^0}{ds},
\]

where \( u^0_s \) is defined by Theorem 1.

The shadowing direction is the direction in which the shadowing series \( u^0_s \) moves as \( s \) increases from 0. It provides a vehicle by which we prove Theorem LSS. We show that the derivative of the ergodic mean \( \langle J \rangle \) with respect to \( s \) can be obtained if the shadowing direction \( v^{(\infty)}_i \) was given (section 4). We then show that \( v^{(n)}_i \), the solution to the constrained least squares problem (1.3), sufficiently approximates the shadowing direction \( v^{(\infty)}_i \) when \( n \) is large (section 5). We finally show (in section 6) that the same derivative can be obtained from the least squares solution \( v^{(n)}_i \).

**4. Ergodic mean derivative via the shadowing direction.** This section proves an easier version of Theorem LSS that replaces the solution to the constrained least squares problem \( v^{(n)}_i \), \( i = 1, \ldots, n \), by the shadowing direction \( v^{(\infty)}_i \).

**Theorem 5.** If (2.2) holds and \( f \) is continuously differentiable, for all continuously differentiable functions \( J(u, s) : \mathbb{R}^m \times \mathbb{R} \rightarrow \mathbb{R} \) whose infinite time average

\[
\langle J \rangle := \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} J(f^{(i)}(u_0, s), s)
\]

is independent of the initial state \( u_0 \in \mathbb{R}^m \), let \( \{v^{(\infty)}_i , i \in \mathbb{Z}\} \) be the sequence of shadowing direction in Definition 4; then

\[
\frac{d\langle J \rangle}{ds} \bigg|_{s=0} = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \left( (DJ(u^0_0, 0)) v^{(\infty)}_i + (\partial_s J(u^0_0, 0)) \right).
\]

**Proof.** This proof is essentially an exchange of limits through uniform convergence. Because \( \langle J \rangle \) in (4.1) is independent of \( u_0 \), we set \( u_0 = u^0 \) in Theorem 1 (thus \( f^{(i)}(u^0_0, s) = u^0_s \)) and obtain

\[
\frac{d\langle J \rangle}{ds} \bigg|_{s=0} = \lim_{s \to 0} \frac{\langle J \rangle|_{s=s} - \langle J \rangle|_{s=0}}{s} = \lim_{s \to 0} \frac{1}{n} \sum_{i=1}^{n} \frac{J(u^0_s, s) - J(u^0, 0)}{s}.
\]

Denote

\[
\gamma_i = \frac{dJ(u^0_s, s)}{ds} = (DJ(u^0_s, s)) \frac{du^0_s}{ds} + (\partial_s J(u^0_s, s))
\]

and using the mean value theorem, we obtain

\[
\frac{d\langle J \rangle}{ds} \bigg|_{s=0} = \lim_{s \to 0} \frac{1}{n} \sum_{i=1}^{n} \gamma_i \xi_i(s), \text{ where all } |\xi_i(s)| \leq |s|.
\]

Because \( J \) is continuously differentiable, we can choose a compact neighborhood of \( \Lambda \times \{0\} \subset \mathbb{R}^m \times \mathbb{R} \) in which both \( (DJ(u, s)) \) and \( (\partial_s J(u, s)) \) are uniformly continuous. When \( s \) is sufficiently small, this neighborhood of \( \Lambda \times \{0\} \) contains \( (u^*_s, s) \) for all \( i \) because \( u^*_0 \in \Lambda \) and \( u^*_s \) are \( i \)-uniformly continuously differentiable (from Theorem 1) and therefore are \( i \)-uniformly continuous. Also, \( \frac{du^*_s}{ds} \) are \( i \)-uniformly continuous. Therefore,
for all $\epsilon > 0$, there exists $M > 0$, such that for all $|\xi| < M$,

$$
||\gamma_i^\xi - \gamma_i^0|| < \epsilon \quad \text{for all } i.
$$

Therefore, for all $|s| < M$, $|\xi_i(s)| \leq |s| \leq M$ for all $i$; thus for all $n > 0$,

$$
\left| \frac{1}{n} \sum_{i=1}^{n} \gamma_i^\xi(s) - \frac{1}{n} \sum_{i=1}^{n} \gamma_i^0 \right| \leq \frac{1}{n} \sum_{i=1}^{n} \left| \gamma_i^\xi(s) - \gamma_i^0 \right| < \epsilon
$$

and thus,

$$
\lim_{n \to \infty} \left| \frac{1}{n} \sum_{i=1}^{n} \gamma_i^\xi(s) - \frac{1}{n} \sum_{i=1}^{n} \gamma_i^0 \right| \leq \epsilon.
$$

Therefore,

$$
d\langle J \rangle = \lim_{s \to 0} \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \gamma_i^\xi(s) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \gamma_i^0.
$$

This completes the proof via the definition of $\gamma_i^0$ and $v_i^{(\infty)}$. □

With Theorem 5, we are one step away from the main theorem (Theorem LSS)—the shadowing direction $v_i^{(\infty)}$ in Theorem 5 needs to be replaced by the solution $v_i^{(n)}$ to the least squares problems (1.3). The next section proves a bound of the distance between $v_i^{(\infty)}$ and $v_i^{(n)}$.

5. Computational approximation of shadowing direction. This section assumes all conditions of Theorem 1, and focuses on when $s = 0$. We denote $u_i^s$ by $u_i$ in this and the next section.

The main task of this section is providing a bound for

(5.1) \[ e_i^{(n)} = v_i^{(n)} - v_i^{(\infty)}, \quad i = 1, \ldots, n, \]

where $v_i^{(n)}$ is the solution to the least squares problem

(5.2) \[ \min \frac{1}{2} \sum_{i=1}^{n} v_i^{(n)}T v_i^{(n)} \text{ subject to } v_i^{(n+1)} = (Df(u_i, 0)) v_i^{(n)} + (\partial_s f(u_i, 0)), \quad i = 1, \ldots, n-1. \]

This is a mathematically new result in the following sense. The shadowing lemma guarantees the existence of a shadowing trajectory, but provides no clear way to numerically compute it or its derivative. This section suggests that the solution to the least squares problem (5.2) is a useful approximation to the derivative of the shadowing trajectory, and proves a bound of the approximation error. This bound will then enable us to show that the difference between $v_i^{(n)}$ and $v_i^{(\infty)}$ makes a vanishing difference in (4.2) as $n \to \infty$.

**Lemma 6.** $e_i^{(n)}$ as defined in (5.1) satisfies

(5.3) \[ e_{i+1}^{(n)} = (Df(u_i, 0)) e_i^{(n)}, \quad i = 1, \ldots, n-1. \]

In addition, their components in the stable and unstable directions, $e_i^{(n)+} \in V^+(u_i)$ and $e_i^{(n)-} \in V^-(u_i)$, where $e_i^{(n)+} + e_i^{(n)-} = e_i^{(n)}$, satisfies

(5.4) \[ e_{i+1}^{(n)+} = (Df(u_i, 0)) e_i^{(n)+}, \quad e_{i+1}^{(n)-} = (Df(u_i, 0)) e_i^{(n)-}, \quad i = 1, \ldots, n-1. \]
Proof. By definition, \( u_{i+1}^s = f(u_i^s, s) \) for all \( s \) in a neighborhood of 0. By taking the derivative with respect to \( s \) on both sides, we obtain

\[
v_{i+1}^{(\infty)} = (Df(u_i, 0))v_i^{(\infty)} + (\partial_s f(u_i, 0)).
\]

Subtracting this from the constraint in (5.2), we obtain (5.3).

By substituting \( e_i^{(n)} = e_i^{(n)+} + e_i^{(n)-} \) into (5.3), we obtain

\[
\left( e_{i+1}^{(n)+} - (Df(u_i, 0)) e_i^{(n)+} \right) + \left( e_{i+1}^{(n)-} - (Df(u_i, 0)) e_i^{(n)-} \right) = 0.
\]

Because the spaces \( V^+(u) \) and \( V^-(u) \) are invariant (2.3),

\[
(Df(u_i, 0)) e_i^{(n)\pm} \in V^\pm(u_{i+1}), \quad \text{and} \quad \left( e_{i+1}^{(n)\pm} - (Df(u_i, 0)) e_i^{(n)\pm} \right) \in V^\pm(u_{i+1}).
\]

Because they sum to 0, the contents of both parentheses must be in \( V^+(u_{i+1}) \cap V^-(u_{i+1}) = \{0\} \). This proves (5.4). \( \square \)

Lemma 6 indicates that for all \( \epsilon^+ \) and \( \epsilon^- \),

\[
v_i^{(n)} = v_i^{(n)} + \epsilon^+ e_i^{(n)+} + \epsilon^- e_i^{(n)-}
\]

satisfies the constraint in problem (5.2), i.e.,

\[
v_{i+1}^{(n)} = (Df(u_i, 0)) v_i^{(n)} + (\partial_s f(u_i, 0)), \quad i = 1, \ldots, n - 1.
\]

Because \( v_i^{(n)} \) is the solution to problem (5.2), it must be true that

\[
\sum_{i=1}^{n} v_i^{(n)} T v_i^{(n)} \leq \sum_{i=1}^{n} v_i^{(n)} T v_i^{(n)} \quad \text{for all} \quad \epsilon^+ \text{ and } \epsilon^-.
\]

By substituting the definition of \( v_i^{(n)} \) into (5.5), and using the first order optimality condition with respect to \( \epsilon^+ \) and \( \epsilon^- \) at \( \epsilon^+ = \epsilon^- = 0 \), we obtain

\[
\sum_{i=1}^{n} v_i^{(n)} T e_i^{(n)+} = \sum_{i=1}^{n} v_i^{(n)} T e_i^{(n)-} = 0.
\]

By substituting \( e_i^{(n)} = v_i^{(\infty)} + e_i^{(n)+} = v_i^{(\infty)} + e_i^{(n)+} + e_i^{(n)-} \) into (5.6), we obtain

\[
\sum_{i=1}^{n} (v_i^{(\infty)}) T e_i^{(n)+} + \sum_{i=1}^{n} (e_i^{(n)+}) T e_i^{(n)+} + \sum_{i=1}^{n} (e_i^{(n)-}) T e_i^{(n)+} = 0,
\]

\[
\sum_{i=1}^{n} (v_i^{(\infty)}) T e_i^{(n)-} + \sum_{i=1}^{n} (e_i^{(n)+}) T e_i^{(n)-} + \sum_{i=1}^{n} (e_i^{(n)-}) T e_i^{(n)-} = 0.
\]

To transform (5.7) into bounds on \( e_i^{(n)+} \) and \( e_i^{(n)-} \), we need the following lemma.

**Lemma 7.** The hyperbolic splitting of \( e_i^{(n)} \) as defined in (5.1) satisfies

\[
\|e_i^{(n)+}\| \leq C\lambda^{n-i}\|e_i^{(n)+}\|, \quad \|e_i^{(n)-}\| \leq C\lambda^i\|e_i^{(n)-}\|.
\]

**Proof.** This is a direct consequence of (5.4) and the definition of \( V^+ \) and \( V^- \) in (2.2). \( \square \)
By combining the first equality in (5.7) with Lemma 7 and using the Cauchy–Schwarz inequality, we obtain
\[
\|e_n^{(n)}\| \leq \sum_{i=1}^{n} (e_i^{(n)})^T e_i^{(n)} = -\sum_{i=1}^{n} (v_i^{(\infty)})^T e_i^{(n)} - \sum_{i=1}^{n} (e_i^{(n)})^T e_i^{(n)}
\]
\[
\leq \sum_{i=1}^{n} \|v_i^{(\infty)}\| \|e_i^{(n)}\| + \sum_{i=1}^{n} \|e_i^{(n)}\| \|e_i^{(n)}\|
\]
\[
\leq \sum_{i=1}^{n} C \lambda^{n-i} \|v_i^{(\infty)}\| \|e_i^{(n)}\| + \sum_{i=1}^{n} C^2 \lambda^n \|e_0^{(n)}\| \|e_i^{(n)}\|.
\]

Therefore,
\[
\|e_n^{(n)}\| \leq \frac{C}{1-\lambda} \|v^{(\infty)}\|_B + nC^2 \lambda^n \|e_0^{(n)}\|,
\]

where the \(B\) norm is as defined in section 3, and is finite by Theorem 1. Similarly, by combining the second equality in (5.7) with Lemma 7 and using the Cauchy–Schwarz inequality,
\[
\|e_0^{(n)}\| \leq \frac{C}{1-\lambda} \|v^{(\infty)}\|_B + nC^2 \lambda^n \|e_0^{(n)}\|.
\]

When \(n\) is sufficiently large such that \(nC^2 \lambda^n < \frac{1}{3}\), we can substitute both inequalities into each other and obtain
\[
\|e_n^{(n)}\| \leq \frac{2C}{1-\lambda} \|v^{(\infty)}\|_B , \quad \|e_0^{(n)}\| \leq \frac{2C}{1-\lambda} \|v^{(\infty)}\|_B.
\]

These inequalities lead to the following theorem that bounds the norm of \(e_i^{(n)}\), the difference between the least squares solution \(v_i^{(n)}\) and the shadowing direction \(v_i^{(\infty)}\).

**Theorem 8.** If \(n\) is sufficiently large such that \(3nC\lambda^n < 1\), then \(e_i^{(n)}\) as defined in (5.1) satisfies
\[
\|e_i^{(n)}\| < \frac{2C^2}{1-\lambda} \|v^{(\infty)}\|_B (\lambda^i + \lambda^{n-i}) , \quad i = 1, \ldots, n.
\]

**Proof.** From the hyperbolicity assumption (2.2) and Lemma 7,
\[
\|e_i^{(n)}\| \leq \|e_i^{(n)}\| + \|e_i^{(n)}\| \leq C \lambda^{n-i} \|e_i^{(n)}\| + C \lambda^i \|e_0^{(n)}\|.
\]

The theorem is then obtained by substituting (5.8) into \(\|e_n^{(n)}\|\) and \(\|e_0^{(n)}\|\) in the inequality above.

This theorem shows that \(v_i^{(n)}\) is a good approximation of the shadowing direction \(v_i^{(\infty)}\) when \(n\) is large and \(-\log \lambda \ll i \ll n + \log \lambda\). The next section shows that the approximation has a vanishing error in (1.2) as \(n \to \infty\). Combined with Theorem 5, we then prove a rigorous statement of Theorem LSS.

**6. Convergence of least squares shadowing.** This section uses the results of the previous sections to prove our main theorem.

**Theorem LSS.** For a \(C^1\) map \(f : \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^n\), assume \(f(\cdot, 0)\) is bijective and defines a compact global hyperbolic attractor \(\Lambda\). For a \(C^1\) map \(J : \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}\) whose infinite time average \(\langle J \rangle\) defined in (4.1) is independent of the initial state \(u_0 \in \mathbb{R}^m\). For a sequence \(\{u_i, i \in \mathbb{Z}\} \subset \Lambda\) satisfying \(u_{i+1} = f(u_i, 0)\), denote \(v_i^{(n)} \in \mathbb{R}^n\).
Theorem LSS is the following. Therefore,\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \left( (DJ(u_i, 0)) v_i^{(n)} + (\partial_s J(u_i, 0)) \right) = \frac{d(J)}{ds} \bigg|_{s=0} .
\]

Proof. Because $J$ is $C^1$ and $\Lambda$ is compact, $(DJ(u_i, 0))$ is uniformly bounded, i.e., there exists $A$ such that $\| (DJ(u_i, 0)) \| < A$ for all $i$. Let $e_{i}^{(n)}$ be defined as in (5.1), whose norm is bounded by Theorem 8; then for large enough $n$,
\[
\left| \frac{1}{n} \sum_{i=1}^{n} \left( (DJ(u_i, 0)) v_i^{(n)} + (\partial_s J(u_i, 0)) \right) - \frac{1}{n} \sum_{i=1}^{n} \left( (DJ(u_i^0, 0)) v_i^{(\infty)} + (\partial_s J(u_i^0, 0)) \right) \right|
\]
\[
\leq \frac{1}{n} \sum_{i=1}^{n} \left( (DJ(u_i, 0)) v_i^{(\infty)} \right) \| e_{i}^{(n)} \|
\]
\[
< \frac{1}{n} \sum_{i=1}^{n} \frac{2A C^2}{1-\lambda} \left( \lambda^i + \lambda^{n-i} \right) \| v^{(\infty)} \| \xrightarrow{n \to \infty} 0 .
\]
Therefore,
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \left( (DJ(u_i, 0)) v_i^{(n)} + (\partial_s J(u_i, 0)) \right)
\]
\[
= \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \left( (DJ(u_i, 0)) v_i^{(\infty)} + (\partial_s J(u_i, 0)) \right) = \frac{d(J)}{ds} \bigg|_{s=0}
\]
by Theorem 5. \(\square\)

7. The least squares shadowing algorithm. A practicable algorithm based on Theorem LSS is the following.

1. Choose large enough $n_0$ and $n$, and an arbitrary starting point $u_{-n_0} \in \mathbb{R}^m$.
2. Compute $u_{i+1} = f(u_i, s), i = -n_0, \ldots, 0, 1, \ldots, n$.
3. For large enough $n_0, u_0, u_1, \ldots, u_n$ are approximately on the global attractor $\Lambda$.
4. Solve the system of linear equations
\[
\begin{align*}
v_{i+1} & = (D f(u_i, s)) v_i + (\partial_s f(u_i, s)) , \quad i = 1, \ldots, n-1 , \\
w_{i+\frac{1}{2}} & = (D f(u_i, s))^T w_{i+\frac{3}{2}} + v_i , \quad i = 1, \ldots, n , \\
w_{\frac{3}{2}} & = w_{n+\frac{1}{2}} = 0 ,
\end{align*}
\]
which is the first order optimality condition of the constrained least squares problem (1.3), and gives its unique solution $v_i^{(n)}, v_i^{(s)}$. Note that a linear relation between $w_{i+\frac{1}{2}}, w_{i+\frac{3}{2}},$ and $w_{n+\frac{1}{2}}$ can be obtained by substituting the second equation into the first one. The resulting matrix system is block tridiagonal, where the block size is the dimension of the dynamical system $m$. A banded matrix solver can then be used to solve the system.

4. Compute the desired derivative by
\[
\frac{d(J)}{ds} = \frac{1}{n} \sum_{i=1}^{n} \left( (DJ(u_i, 0)) v_i^{(n)} + (\partial_s J(u_i, 0)) \right) .
\]
Most of the computation time in this algorithm is spent on solving the block-tridiagonal system in step 3. Because the $nm \times nm$ matrix has a bandwidth of
4m − 1, the computational cost of a banded solver (e.g., Lapack’s dgbsv routine [2]) is bounded by $O(nm^3)$. Here $n$ is the length of the trajectory, and $m$ is the dimension of the dynamical system. $O(nm^3)$ is the leading term in the number of operations of the algorithm presented in this paper.

Theorem LSS shows that the computed derivative is accurate for large $n$. The approximation error of (7.1) originates from two sources,

$$
\frac{d(J)}{ds} - \frac{1}{n} \sum_{i=1}^{n} ((DJ(u_i, 0)) v_i^{(n)} + (\partial_s J(u_i, 0))) = E_1 + E_2,
$$

where

$$
E_1 = \frac{d(J)}{ds} - \frac{1}{n} \sum_{i=1}^{n} ((DJ(u_i^0, 0)) v_i^{(\infty)} + (\partial_s J(u_i^0, 0))).
$$

Theorem 5 guarantees that $E_1 \xrightarrow{n \to \infty} 0$. This error represents the difference between an ergodic mean and an average over a finite trajectory. If the dynamical system is mixing, the central limit theorem implies that $E_1 \sim O(\sqrt{n})$. The other part of the error is

$$
E_2 = \frac{1}{n} \sum_{i=1}^{n} (DJ(u_i^0, 0)) (v_i^{(\infty)} - v_i^{(n)}).
$$

Theorem 8 guarantees that $E_2 \sim O(n^{-\frac{3}{2}})$. Because $E_1$ has a slower rate of decay, the rate of convergence of the algorithm presented in this paper is $O(n^{-\frac{3}{2}})$ for sufficiently large $n$.


The algorithm is tested on the Smale–Williams solenoid attractor. The map that defines this attractor in cylindrical coordinates is

$$
u_{n+1} = \begin{bmatrix}
    r_{n+1} \\
    \theta_{n+1} \\
    z_{n+1}
\end{bmatrix} = \begin{bmatrix}
    s + (r_n - s)/4 + (\cos \theta_n)/2 \\
    2\theta_n \\
    z_n/4 + (\sin \theta_n)/2
\end{bmatrix}.
$$

The map has a single parameter $s$, whose effect is qualitatively shown in Figure 8.1. We define the quantity of interest

$$
J(u) = \sqrt{r^2 + z^2},
$$

and focus on computing the derivative of the long time averaged quantity of interest $\langle J \rangle$ to the parameter $s$.

This particular map is chosen such that the shadowing direction has a rare analytic form. It is straightforward to verify that the constant sequence $v_i^{(\infty)} \equiv [r = 1, \theta = 2\pi]$. All the numerical results in this section are obtained by running revision fa82e4241ad3d3260332244a4a54c9500f6224 of this code hosted on github.

3Although the map is defined on cylindrical coordinates, the $L^2$ norm in $R^5$ is the Euclidean distance in Cartesian coordinates. In the numerical implementation of this map, the Cartesian coordinates of $u_n$ are transformed to cylindrical coordinates, then the map is applied to obtain $u_{n+1}$ in cylindrical coordinates before it is transformed back to Cartesian coordinates.
Fig. 8.1. Visualization of the Smale–Williams solenoid attractor defined by the map in (8.1). The left plots show the attractor at $s = 1$. The right plots show the attractor at $s = 1.4$.

Fig. 8.2. The $l^2$ norm of the least squares shadowing error $e_i^{(n)} = v_i^{(n)} - v_i^{(\infty)}$ for a trajectory of length $n = 100$ at $s = 2.0$.

$0, z = 0$] satisfies the tangent map $v_{i+1}^{(\infty)} = (Df(u_i, 0))v_i^{(\infty)} + (\partial_s f(u_i, 0))$ for any sequence $\{u_i\}$. This analytic form of the shadowing direction allows us to numerically evaluate the least squares shadowing error $e_i^{(n)}$ as defined in (5.1). Figure 8.2 shows that the error is order 1 at both the beginning and end of a trajectory, but decreases exponentially to numerical precision towards the middle portion of the trajectory. This trend is consistent with the error bound provided by Theorem 8.

The values of $d(J)/ds$ computed from the least squares shadowing algorithm are plotted in Figure 8.3 and compared against finite difference derivatives. The derivatives computed on trajectories of length $n = 100$ have significant error because $e_i^{(n)}$ is large on a significant portion of the trajectory. The derivatives computed with
Fig. 8.3. \( \frac{d\langle J \rangle}{ds} \) computed using the least squares shadowing algorithm. Red X's represent those computed with trajectories of length \( n = 100 \). Green dots represent those computed with \( n = 1000 \). Blue lines represent those computed with \( n = 10000 \). Each calculation is repeated several times at the same value of \( s \). The black bars represent the 3σ confidence interval of finite difference derivatives. Each finite difference derivative is computed by differencing the mean of 10000 trajectories at \( s + 0.05 \) and the mean of 10000 trajectories at \( s - 0.05 \). Each of these 20000 trajectories has length 10000.

\[
\begin{align*}
\frac{d\langle J \rangle}{ds} & \quad \text{\( n = 100 \) and 10000 appear to be at least as accurate as the finite difference values. It is worth noting that each finite difference calculation involves trajectories of total length 200000000, and takes orders of magnitude longer computation time than a least squares shadowing calculation with \( n = 10000 \).}
\end{align*}
\]

Figure 8.4 illustrates that the least squares shadowing algorithm converges at a rate of \( O(n^{-1}) \) at relatively small values of \( n \), then transitions to a rate of \( O(n^{-\frac{1}{2}}) \) at...
This behavior is consistent with the error analysis in section 7. For small \( n \), \( E_1 \) as in (7.3), which has a decay rate of \( O(n^{-1}) \), dominates. For larger \( n \), \( E_2 \) as in (7.4), which has a slower decay rate of \( O(n^{-\frac{3}{2}}) \), dominates. They lead to a two-stage convergence pattern as seen in Figure 8.4.

Acknowledgments. The author gratefully acknowledges David Moro and Dr. Si Li for helpful discussion on the proofs.

REFERENCES


\[ \frac{d(J)}{ds}\bigg|_{s=1} \text{ used in this convergence analysis is } 0.931450 \pm 0.000017 \text{ with } 99.7\% \text{ confidence. This value is obtained by averaging over 1100 least squares shadowing calculations, each of length 100000. The first 20 steps and the last 20 steps of each trajectory are removed from the averaging in order to remove the bias caused by } E_1. \text{ The value of } 20 \text{ is motivated by Figure 8.2.} \]