Dynamics of PDE, Vol.3, No.4, 259-279, 2006

Denoising Deterministic Time Series

Steven P. Lalley and A. B. Nobel

Communicated by Y. Charles Li, received May 5, 2006.

Abstract. This paper is concerned with the problem of recovering a finite, deterministic time series from observations that are corrupted by additive, independent noise. A distinctive feature of this problem is that the available data exhibit long-range dependence and, as a consequence, existing statistical theory and methods are not readily applicable. This paper gives an analysis of the denoising problem that extends recent work of Lalley, but begins from first principles. Both positive and negative results are established. The positive results show that denoising is possible under somewhat restrictive conditions on the additive noise. The negative results show that, under more general conditions on the noise, no procedure can recover the underlying deterministic series.

1. Introduction

Recent interest in chaos has drawn the attention of statisticians to deterministic phenomena that exhibit random behavior. While there is no universally accepted definition of chaos, phenomena termed "chaotic" have generally been studied in the context of dynamical systems, which provide mathematical models of physical systems that evolve deterministically in time. (Good introductions to dynamical systems and chaos for non-specialists can be found in the texts of Devaney [**15**] and Alligood *et al.* [**5**].) In what follows we will consider a standard model for dynamical systems, in which the relevant states of the system form a compact subset Λ of $\mathbb{R}^d$. The time evolution of the system is described by an invertible map $F : \Lambda \to \Lambda$. If at time i the system is in state $x \in \Lambda$, then at time $i+1$ it is in state Fx, and at time $i-1$ it is in state $F^{-1}x$. That descriptions of this sort are, in a precise sense, generic follows from Takens's embedding theorem [**34, 4, 33**]. We do not assume that F (or F^{-1}) is continuous. Starting from an initial state $x \in \Lambda$ at time zero, the complete time evolution of the system is described by the bi-infinite trajectory

state	...	$F^{-2}x$	$F^{-1}x$	x	Fx	F^2x	...
time	...	-2	-1	0	1	2	...

1991 *Mathematics Subject Classification.* 37, 34, 62, 82.
Key words and phrases. Denoising, additive noise, deterministic series.
Nobel's work was supported in part by NSF Grant DMS-9971964.

Here F^i is the i-fold composition of F with itself and F^{-i} is the i-fold composition of F^{-1}. This model is deterministic: from exact knowledge of the state of the system at any point in time, one may reconstruct all the past and future states of the system by repeated application of F and F^{-1}. To simplify notation in what follows, let $x_i = F^i x$, $i \in \mathbb{Z}$, so that the initial state x of the system is denoted by x_0.

To date, most statistical analyses of dynamical systems have been carried out in the context of dynamical noise models. In a dynamical noise model, the available observations are assumed to be generated according to a nonlinear autoregressive scheme of the form $x_{i+1} = Fx_i + \eta_i$, where $\{\eta_i\}$ are independent, mean zero random vectors. In this model, random noise is "folded" into the dynamics at each step, and the resulting sequence of states x_i is not purely deterministic. In the presence of dynamical noise, the observed states form a discrete time, continuous state Markov Chain, and estimating interesting features of the dynamics (e.g. the map F) can often be accomplished in part by an appeal to traditional time series techniques. Representative work can be found in references [**37, 19, 11, 12, 23, 31, 26, 35**]. An alternative approach to the map estimation problem is described in [**30**].

Of interest here is the so-called *observational noise model*, in which the available data are observations (or measurements) of an underlying deterministic system that are corrupted by additive noise. In this model our observations take the form $y_i = x_i + \varepsilon_i$, where $\{\varepsilon_i\}$ are independent, mean zero random vectors. In contrast with the dynamical noise model, the noise does not interact with the dynamics: the deterministic character of the system, and its long range dependence, are preserved beneath the noise. Due in part to this dependence, estimation in the observational noise model has not been broadly addressed by statisticians, though the model captures important features of many experimental situations. Here we are interested in the problem of how to recover the underlying time series $\{x_i\}$ from the observations $\{y_i\}$.

Denoising problem: Reconstruct the successive states $x_0, \ldots, x_n$ of the deterministic system from observations of the form

$$y_i \;=\; x_i + \varepsilon_i \;=\; F^i x_0 + \varepsilon_i \quad 0 \le i \le n \tag{1}$$

where $\varepsilon_0, \ldots, \varepsilon_n \in \mathbb{R}^d$ are independent random vectors with mean zero.

Several versions of the denoising problem, and associated methods, have previously been considered by a number of authors, including Kostelich and Yorke [**23**], Davies [**14**], Sauer [**32**], Kostelich and Schreiber [**22**]. The methods and results described here are motivated by recent work of Lalley [**24, 25**]. MacEachern and Berliner [**27**] studied the problem of distinguishing trajectories in the observational noise model when the noise distribution comes from a suitable exponential family and established the asymptotic normality of relevant likelihood ratios.

Though some features of denoising can be found in more traditional statistical problems such as errors in variables regression, deconvolution, and measurement error modeling (c.f. [**10**]), other features distinguish it from these problems and require new methods of analysis. For example, in the denoising problem the covariates x_i are deterministically related (not i.i.d. or mixing), the noise ε_i is often bounded (not Gaussian), and the noise distribution is usually unknown.

In the denoising problem the underlying states of the observed deterministic system are of primary interest. Denoising methods can also provide useful preprocessing for other statistical analyses. In the absence of noise, and under appropriate regularity conditions, $x_0, x_1, \ldots$ can be used to estimate the map F [**7, 3, 30**], its invariant measure, entropy, and Lyapunov exponents [**18**], or the fractal dimension of its attractor (see [**13**]). When observational noise is present, consistent reconstructions $\hat{x}_0, \ldots, \hat{x}_n$ can sometimes act as surrogates for the unobserved states in estimation problems of this sort. The surveys [**17, 6, 20, 21**] give an account of statistical problems in the study of dynamical systems. Formal limits to statistical inference from dependent processes can be found in [**2, 1, 29**]. From the viewpoint of statistical practice and theory, it is interesting to ask whether estimation is still possible when noise removal is not, but we will not address such issues here.

2. Summary

The next section contains several preliminary definitions and results that will be used throughout the paper. Section 4 describes two denoising procedures. The consistency of these procedures is established in Theorems 1 and 2 under a boundedness assumption on the noise. It is shown in Section 5 that, in a variety of settings, consistent denoising is not possible when this assumption is significantly relaxed. Proofs of the positive (consistency) results are given in Section 6; proofs of the negative results are given in Section 7.

3. Preliminaries

Throughout this paper we assume that $F : \Lambda \to \Lambda$ is an invertible map of a compact set $\Lambda \subseteq \mathbb{R}^d$. Of primary interest are maps that possess an elementary form of sensitive dependence on initial conditions. Recall that F is said to be expansive if there exists $\Delta > 0$ such that for every pair of vectors $x, x' \in \Lambda$ with $x \neq x'$,

$$\sup_{s \in \mathbb{Z}} |F^s x - F^s x'| \; > \; \Delta.$$

The constant Δ is called a *separation threshold* for F. If F is expansive then, beginning from any two distinct initial states, the corresponding bi-infinite trajectories of F will, at some (possibly negative) time i be at least Δ apart. Note that the separation threshold Δ does not depend on x or x'.

Definition: Let F be an expansive map with separation threshold $\Delta > 0$. The separation time for $x \neq x'$ is

$$s(x, x') = \min\{|s| : |F^s x - F^s x'| > \Delta\}.$$

For each $\alpha > 0$ define the *separation horizon*

$$H(\alpha) \; = \; \sup\{\, s(x, x') \; : \; |x - x'| \geq \alpha \,\}.$$

Note that $\alpha \leq \alpha'$ implies $H(\alpha) \geq H(\alpha')$. If $H(\alpha) < \infty$ for every $\alpha > 0$, then then F will be said to have *finite separation horizon.*

Proposition 1. *If F has finite separation horizon then the inverse function*

$$H^{-1}(k) \; = \; \inf\{\alpha > 0 : H(\alpha) \leq k\} \tag{2}$$

tends monotonically to zero as $k \to \infty$.

Proof: The monotonicity of H^{-1} follows from that of H. If $H^{-1}(k) \geq \alpha_0 > 0$ for every k, then $H(\alpha) = +\infty$ for $\alpha < \alpha_0$.

If $F : \Lambda \to \Lambda$ is invertible and continuous, then F^{-1} is continuous and F is a homeomorphism (see, *e.g.* [**36**]). An elementary argument shows that every expansive homeomorphism has finite separation horizon.

LEMMA 1. *If $F : \Lambda \to \Lambda$ is an expansive homeomorphism, then F has finite separation horizon.*

Proof: Let $\Delta > 0$ be a separation threshold for F. If $H(\alpha) = +\infty$ for some $\alpha > 0$ then there exist pairs of states $(x_n, x'_n) \in \Lambda \times \Lambda$, $n \geq 1$, such that $|x_n - x'_n| \geq \alpha$ for each n and $s(x_n, x'_n) \to \infty$. As Λ is compact, there exist integers $n_1 < n_2 < \cdots$ and points $x, x' \in \Lambda$ such that $x_{n_k} \to x$ and $x'_{n_k} \to x'$. Clearly $|x - x'| \geq \alpha$. Moreover, as F is continuous and $s(x_{n_k}, x'_{n_k}) \to \infty$, for each $m \geq 1$,

$$\max_{|s| \leq m} |F^s x - F^s x'| \;=\; \lim_{k \to \infty} \max_{|s| \leq m} |F^s x_{n_k} - F^s x'_{n_k}| \;\leq\; \Delta.$$

It follows that $H(\alpha) \geq s(x, x') = \infty$, which is a contradiction.

3.1. Ergodic Transformations. Ergodic Transformation: Let μ be a probability measure on the Borel subsets of Λ. A map $F : \Lambda \to \Lambda$ is said to preserve μ if $\mu(F^{-1}B) = \mu(B)$ for each Borel set $B \subseteq \Lambda$. A μ-preserving map F is said to be *ergodic* if $F^{-1}B = B$ implies $\mu(B) \in \{0, 1\}$, i.e. every F-invariant set has μ-measure zero or one.

The ergodic theorem generalizes the ordinary law of large numbers and is an important tool in understanding the asymptotic behavior of dynamical systems. It states that the time average of a real-valued measurement along the trajectory of an ergodic map F will converge to the space average of that measurement.

THEOREM A (Ergodic Theorem). *If $F : \Lambda \to \Lambda$ is μ-preserving and ergodic, and $f : \Lambda \to \mathbb{R}$ is such that $\int |f| \, d\mu < \infty$, then $n^{-1} \sum_{i=0}^{n-1} f(F^i x) \;\to\; \int f \, d\mu$ with probability one and in mean.*

4. Consistent Denoising

In this section we describe two consistent denoising methods for deterministic time series, and provide a preliminary analysis of their theoretical performance.

4.1. Smoothing Algorithm D. We first describe a denoising method originally proposed by Lalley [**24**], called Smoothing Algorithm D. Let the available data be a sequence of vectors $y_0, \ldots, y_n$ defined as in (1), and let k be a positive integer less than $\log n$. For each $l = k, \ldots, n - k$ define the index set

$$A_n(l, k) \;=\; \{\, j \,:\, |y_{j+r} - y_{l+r}| \;\leq\; 3\Delta/5 \text{ for } |r| \leq k \,\}. \tag{3}$$

Note that $l \in A_n(l, k)$ so that $A_n(l, k)$ is always non-empty. For $l = k, \ldots, n - k$ define the denoising estimate

$$\hat{x}_{l,n} \;=\; \frac{1}{|A_n(l, k)|} \sum_{j \in A_n(l,k)} y_j \tag{4}$$

of x_l: set $\hat{x}_{l,n} = 0$ for other values of l. To see how the estimate is constructed, let $w(j, k) = (y_{j-k}, \ldots, y_{j+k})$ contain the observations in a window of length $2k + 1$

centered at y_j. The estimate $\hat{x}_{l,n}$ of x_l is obtained by averaging all those values y_j for which $w(j,k)$ is close, on a term by term basis, to $w(l,k)$.

THEOREM 1. *Let F be an expansive map with separation threshold $\Delta > 0$ and finite separation horizon. Suppose that $|\varepsilon_i| \le \Delta/5$ for each $i \ge 0$. If $k \to \infty$ and $k/\log n \to 0$ then*

$$\frac{1}{n-2k} \sum_{i=k}^{n-k} |\hat{x}_{i,n} - x_i| \;\to\; 0 \quad as \quad n \to \infty$$

with probability one for every initial vector $x \in \Lambda$.

The in-probablity consistency of Smoothing Algorithm D was first established in Theorem 1 of [**24**] under the condition that F is a C^2-diffeomorphism and Λ is a hyperbolic attractor (or the basin of attraction of such a set). A more general result for expansive homeomorphisms is stated in Theorem 2 of [**25**]. Here these conditions are replaced by the weaker assumption of finite separation horizon, and in-probability convergence is strengthened to convergence with probability one. The proof of Theorem 1 is given in Section 6.

4.2. Implementation. A naive implementation of smoothing algorithm D has running time $O(n^2)$, where n denotes the number of available observations. More efficient, approximate, versions of the algorithm with running time $O(n \log n)$ are investigated in [**25**]. In simulations, Algorithm D and its approximations have been used to successfully remove noise from trajectories of the logistic map, the Hénon attractor, and Smale's solenoid. Informal studies have illustrated the failure of the algorithm to remove uniform noise whose support is comparable to the diameter of the associated attractor. These simulations lend empirical support to Theorem 1 and the negative results discussed below.

4.3. Preliminary Analysis. Smoothing Algorithm D removes observation noise from the trajectory of an expansive map by judicious averaging. To understand why Theorem 1 holds, fix l between k and $n-k$. Together (1) and (4) imply that

$$|x_l - \hat{x}_{l,n}| \;\le\; \frac{\sum_{j \in A_n(l,k)} |x_l - x_j|}{|A_n(l,k)|} \;+\; \frac{|\sum_{j \in A_n(l,k)} \varepsilon_j|}{|A_n(l,k)|}. \tag{5}$$

The first term on the right hand side of (5) controls the bias of the estimate $\hat{x}_l$, and the second controls its stochastic variation. Regarding the bias, note that

$$\begin{aligned} j \in A_n(l,k) \;&\Rightarrow\; |y_{j+r} - y_{l+r}| \le 3\Delta/5 \;\text{ for }\; 1 \le |r| \le k \\ &\Rightarrow\; |x_{j+r} - x_{l+r}| \le \Delta \;\text{ for }\; 1 \le |r| \le k \\ &\Rightarrow\; k \le H(|x_l - x_j|) \;\Rightarrow\; |x_l - x_j| \le H^{-1}(k). \end{aligned}$$

Thus (5) implies that

$$|x_l - \hat{x}_{l,n}| \;\le\; H^{-1}(k) \;+\; \frac{|\sum_{j \in A_n(l,k)} \varepsilon_j|}{|A_n(l,k)|}. \tag{6}$$

This yields the following bound on the average denoising error:

$$\frac{1}{n-2k} \sum_{l=k}^{n-k} |x_l - \hat{x}_{l,n}| \;\le\; H^{-1}(k) \;+\; \frac{1}{n-2k} \sum_{l=k}^{n-k} \frac{|\sum_{j \in A_n(l,k)} \varepsilon_j|}{|A_n(l,k)|}. \tag{7}$$

The upper bound $H^{-1}(k)$ on the average bias depends on the map F and the window width k, but is independent of n and l. Moreover, $H^{-1}(k) \to 0$ by Proposition 1, as F has finite separation horizon and $k \to \infty$. Analysis of the stochastic variation is complicated by the fact that the ε_i are not independent when summed over the *random* index set $A_n(l,k)$. The details are given in the appendix (see in particular inequality (24) and Lemma 4).

The analysis above suggests a more adaptive version of Smoothing Algorithm D that offers improved performance under somewhat stronger conditions. Fix l for the moment and consider inequality (6). It can be seen that the window width k plays a role analogous to inverse bandwidth in kernel type estimators. Monotonicity of H^{-1} ensures that the bias of $\hat{x}_{l,n}$ decreases as k increases. On the other hand, as k increases, the index set $A_n(l,k)$ gets smaller, and the variability of the estimate will increase as one averages over fewer noise variables ε_j. One modification of Smoothing Algorithm D, analogous to local bandwidth selection, is to adaptively select a window width for each location l. This is considered in more detail below.

4.4. Denoising with a Variable Length Window. Here new denoising estimates $\tilde{x}_{l,n}$ are described. Let the index sets $A_n(l,k)$ be defined as in (3). The new estimates are based on windows whose widths are chosen adaptively to ensure that $|A_n(l,k)|$ is sufficiently large. For $l = \log n, \ldots, n - \log n$ define

$$k_{l,n} \;=\; \max\{1 \le k \le \log n \,:\, |A_n(l,k)| \ge n/\log n\}, \tag{8}$$

and set $k_{l,n} = 0$ if $|A_n(l,1)| < n/\log n$. For the same values of l, define denoising estimates

$$\tilde{x}_{l,n} \;=\; \frac{\sum_{j \in A_n(l,k_{l,n})} y_j}{|A_n(l,k_{l,n})|}. \tag{9}$$

Set $\tilde{x}_{l,n} = 0$ if $k_{l,n} = 0$. Strong consistency of the estimates $\tilde{x}_{l,n}$ requires that the trajectory under study exhibit a natural recurrence property.

Definition: A point $x \in \Lambda$ with trajectory $x_i = F^i x$ will be called strongly recurrent if there is a finite cover $\mathcal{O}$ of Λ such that (i) every $O \in \mathcal{O}$ has diameter less than $\Delta/5$, and (ii) for each $r \ge 1$ and each choice of sets $O_1, \ldots, O_r \in \mathcal{O}$ either

$$\sum_{i=0}^{\infty} I\{x_i \in O_1, \ldots, x_{i+r-1} \in O_r\} \;<\; \infty \tag{10}$$

or

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} I\{x_i \in O_1, \ldots, x_{i+r-1} \in O_r\} \;>\; 0. \tag{11}$$

Conditions (10) and (11) ensure that if the forward trajectory of F starting from x visits the product set $O_1 \times \cdots \times O_r$ infinitely often, then it does so a non-negligible fraction of the time.

Recall that F is said to preserve a probability measure μ on the Borel subsets of Λ if $\mu(F^{-1}B) = \mu(B)$ for each Borel set $B \subseteq \Lambda$, and that μ-preserving map F is said to be ergodic if $F^{-1}B = B$ implies $\mu(B) \in \{0,1\}$, i.e. every F-invariant set has μ-measure zero or one. Strongly recurrent points are the norm in measure preserving systems.

PROPOSITION 2. *If F preserves a measure ν on Λ and is ergodic then ν-almost every $x \in \Lambda$ is strongly recurrent.*

Proof: Let $\mathcal{O}$ be any finite open cover of Λ by sets having diameter less than $\Delta/5$. Fix sets $O_1, \ldots, O_r \in \mathcal{O}$. Note that $x_i \in O_1, \ldots, x_{i+r-1} \in O_r$ if and only if $x_i = F^i x \in O'$ where $O' = \cap_{j=1}^r F^{-j+1} O_j$. If $\nu(O') > 0$, the ergodic theorem ensures that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} I\{x_i \in O_1, \ldots, x_{i+r-1} \in O_r\} \;=\; \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} I\{F^i x \in O'\} \;=\; \nu(O') > 0$$

with ν-probability one. On the other hand, if $\nu(O') = 0$ then $\sum_{i=1}^{\infty} \nu(F^{-i} O') = 0$ and consequently $\nu\{F^i x \in O' \text{ infinitely often}\} = 0$ by the first Borel Cantelli lemma.

THEOREM 2. *Let F be an expansive map with separation threshold $\Delta > 0$ and finite separation horizon. If $|\varepsilon_i| \le \Delta/5$ for each $i \ge 0$, then for every strongly recurrent initial vector $x \in X$,*

$$\max\{\, |\tilde{x}_{l,n} - x_l| \,:\, \log n \le l \le n - \log n \,\} \;\to\; 0$$

with probability one as n tends to infinity.

Performance bounds of this sort for Smoothing Algorithm D are established in [**24**] under the stronger assumption that F is a C^2-diffeomorphism and that Λ is an Axiom A basic set.

5. Negative Results

One distinctive (and restrictive) feature of Theorems 1 and 2 is the assumption that the noise ε_i is bounded in absolute value by a fraction of the separation threshold Δ. In light of the popularity and widespread study of Gaussian noise, it is natural to ask if denoising is possible when the ε_i are normally distributed, perhaps under some constraints on the component-wise variances. Surprisingly, the answer is often "no". Lalley [**24**] shows that for many smooth dynamical systems no scheme can successfully remove Gaussian noise, even in the weak sense of Theorem 1. In this section we extend and generalize this result. Our proof covers the Gaussian case, generalizations of the Gaussian case to noise distributions supported on all of $\mathbb{R}^d$ (stated in [**25**]), and the case of noise distributions with bounded support.

Suppose, as in the previous section, that $\{x_i = F^i x : i \in \mathbb{Z}\}$ is the trajectory of a fixed initial vector $x \in \Lambda$, and that observations of x_i are subject to additive noise, and can be modeled as random vectors

$$y_i \;=\; x_i + \varepsilon_i \quad i \in \mathbb{Z} \tag{12}$$

where $\ldots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \ldots \in \mathbb{R}^d$ are independent, mean-zero random vectors having a common distribution η on $\mathbb{R}^d$. We assume in what follows that the ε_i are defined on a common underlying probability space $(\Omega, \mathcal{F}, P)$. Of interest here are several related problems, which may be informally expressed as follows.

> **Problem 1:** Identify the initial state $x \in \Lambda$ from observation of the infinite sequence $\{y_i : i \in \mathbb{Z}\}$.
>
> **Problem 2:** Consistently identify the initial state $x \in \Lambda$ from observations $y_{-n}, \ldots, y_n$, in the limit as $n \to \infty$.

Problem 3: Estimate the states $x_1, \ldots, x_n \in \Lambda$ from observation of $y_1, \ldots, y_n$.

It is evident that Problem 1 is easier than Problem 2, as in the former we have access to all the available data at the outset. It is also clear that an answer to Problem 2 might be used, in conjunction with shifts of the observations, to answer Problem 3. Problem 3 is just the denoising problem considered in the previous section.

It is shown in Theorem 3 below that for distinguished states x and noise distributions η, neither Problem 1 nor Problem 2 has a solution. This negative result is then used to establish Theorem 4, which states that, for suitable dynamical maps F and noise distributions η, consistent denoising is impossible.

5.1. Distributional Assumptons. The negative results in Theorems 3 and 4 require that the distribution η of ε_i's be smooth and has sufficiently large support. Here we give a precise statement of these conditions.

Suppose first that η is absolutely continuous, having a density f with respect to d-dimensional Lebesgue measure λ. Recall that if A is a Borel subset of $\mathbb{R}^d$, $u \in \mathbb{R}^d$ is any vector and $r > 0$, then

$$A + u \;=\; \{v + u : v \in A\} \quad \text{and} \quad A^r \;=\; \{u \,:\, |u - v| < r \text{ for some } v \in A\}$$

are also Borel subsets of $\mathbb{R}^d$. For $v \in \mathbb{R}^d$ and $r > 0$, let $B(v, r) = \{u : |u - v| < r\}$ be the Euclidean ball of radius r centered at v. Let $S = \{v : f(v) > 0\}$ be the support of the density f of η. Let $\overline{S}$ and S^o denote the closure and interior of S, respectively, and let $\partial S = \overline{S} \setminus S^o$ be its boundary. Finally, let $\rho = \max\{|u - v| : u, v \in \Lambda\}$ be the diameter of Λ. Note that ρ is finite as Λ is compact. We make the following assumptions concerning η:

$$\limsup_{|z| \to 0} \frac{1}{|z|} \int_{S \cap (S - z)} \left| \log \frac{f(w + z)}{f(w)} \right| f(w)\, dw \;<\; \infty, \tag{13}$$

$$\limsup_{r \searrow 0} \frac{1}{r}\, \eta((\partial S)^r) \;<\; \infty, \text{ and} \tag{14}$$

$$S \supseteq\; B(0, 3\rho/2). \tag{15}$$

Assumption (13) states that $\log f$ is Lipschitz continuous on the average. Assumption (14) says that the measure of those points within distance r of ∂S decreases at least linearly with r. Assumption (15) states that S contains a sphere whose radius is significantly larger than the diameter ρ of Λ. It is enough that assumptions (14) and (15) hold for *some* version f of $d\eta/d\lambda$. Note that (14) and (15) are trivially satisfied if $S = \mathbb{R}^d$.

Example 1: If η is multivariate Gaussian and has a covariance matrix of full rank, then assumptions (14) and (15) are immediate, and one may readily verify that assumption (13) holds.

Example 2: Suppose that η has a density f with compact support S satisfying (15), and suppose further that f is Lipschitz continuous on S. Then f is bounded

away from zero and infinity on S and one may verify that (13) holds. Satisfaction of (14) requires, in addition, that the boundary of S be regular. To quantify this, let $N(\partial S, r)$ denote the least number of Euclidean balls of radius $r > 0$ needed to cover ∂S. If $N(\partial S, r) \leq c\,(1/r)^{d-1}$ for some $c < \infty$ and each $0 < r < r_0$, then

$$\eta((\partial S)^r) \;\leq\; c' \sup_{x \in S} |f(x)|\, N(\partial S, r) \cdot r^d \;\leq\; c'c \sup_{x \in S} |f(x)| \cdot r$$

for a suitable normalizing constant c', and (14) follows. The bound $N(\partial S, r) \leq c\,(1/r)^{d-1}$ implies, in particular, that the box counting dimension of ∂S is $d - 1$. Assumption (14) is satisfied, for example, if η is the uniform distribution on $B(0, 3\rho/2)$, or the uniform distribution on a cube of side length $3\rho/2$ centered at the origin.

5.2. Homoclinic Pairs. Let x and x' be distinct initial states in Λ, with corresponding trajectories $\{x_i = F^i x : i \in \mathbb{Z}\}$ and $\{x'_i = F^i x' : i \in \mathbb{Z}\}$. Suppose that we wish to distinguish x and x' on the basis of their trajectories. In the absence of noise, and with knowledge of F, this is always possible: from observation of any x_i one can recover x, and from observation of any x'_j one can recover x'. However, when observation noise is present, this simple inversion process is no longer applicable. Recall that $y_i = x_i + \varepsilon_i$, $i \in \mathbb{Z}$, are noisy observations of the trajectory of x. Let

$$y'_i \;=\; x'_i + \varepsilon_i \quad i \in \mathbb{Z} \tag{16}$$

be observations of the trajectory of x', corrupted by the same additive noise sequence. Define $\mathcal{X}$ to be the set of all bi-infinite sequences $\mathbf{v} = \ldots, v_{-1}, v_0, v_1, \ldots$ with $v_i \in \mathbb{R}^d$, and let $\mathcal{S}$ be the product sigma field for $\mathcal{X}$ generated by the finite dimensional Borel cylinder sets. For fixed x, x' the sequences

$$\mathbf{y} \;=\; (\ldots, y_{-1}, y_0, y_1, \ldots) \quad \text{and} \quad \mathbf{y}' \;=\; (\ldots, y'_{-1}, y'_0, y'_1, \ldots)$$

are random elements of $(\mathcal{X}, \mathcal{S})$, defined on the underlying probability space $(\Omega, \mathcal{F}, P)$. Consider the following special case of Problem 1 above.

> **Question 1:** Is there a measurable function $\phi : \mathcal{X} \to \mathbb{R}^d$ such that $\phi(\mathbf{y}) = x$ and $\phi(\mathbf{y}') = x'$ with probability one?

Intuitively, it will be more difficult to identify x and x' in the presence of noise if their trajectories stay close to each other across time. The notion of a strongly homoclinic pair is one way of making this precise.

Definition: A pair (x, x') of distinct states in Λ is said to be *strongly homoclinic* for F if their bi-infinite trajectories are such that

$$\sum_{i \in \mathbb{Z}} |F^i x - F^i x'| \;<\; \infty \tag{17}$$

As noted in [**24**], homoclinic pairs exist and are common in many smooth dynamical systems. It is worth noting that the existence of a separation threshold does not preclude the existence of homoclinic pairs, as the separation of $F^i x$ and $F^i x'$ need only occur for *one* value of i. Theorem 7 below shows that the answer to Question 1 is "no" when x and x' are strongly homoclinic. The proof is given in Section 7.

THEOREM 3. *Suppose that the distribution η of ε_i satisfies conditions (13)–(15). If x and x' are strongly homoclinic, then for every measurable function $\phi : \mathcal{X} \to \mathbb{R}^d$,*

$$E[\,|\phi(\mathbf{y}) - x\,| \,+\, |\phi(\mathbf{y}') - x'\,|\,] \;>\; 0.$$

Remark: Among the functions ϕ included in the theorem are those that incorporate knowledge of the dynamical map and the two possible initial states. Thus even with knowledge of $\{x, x'\}$ and F, and even with access to the entire noisy trajectory, one cannot recover the initial state of the system with certainty.

5.3. Negative Results for Denoising. Suppose now that $F : \Lambda \to \Lambda$ preserves a Borel measure μ on Λ and is ergodic. Let $X \sim \mu$ be independent of $\{\varepsilon_i\}$ and define

$$X_i \;=\; F^i X, \quad Y_i \;=\; X_i + \varepsilon_i \qquad i \in \mathbb{Z} \tag{18}$$

where the ε_i are i.i.d. with distribution η. Then $\{(X_i, Y_i) : i \in \mathbb{Z}\}$ is a stationary ergodic process taking values in $\mathbb{R}^d \times \mathbb{R}^d$. Our principal negative result applies to dynamical systems that admit a homoclinic coupling.

Definition: A μ preserving transformation $F : \Lambda \to \Lambda$ admits a *homoclinic coupling* if on some probability space one may define random vectors X and X' such that

(1) X and X' take values in Λ
(2) X and X' have distribution μ
(3) (X, X') is strongly homoclinic for F with positive probability.

For systems admitting a homoclinic coupling, strongly homoclinic pairs are relatively common. When a homoclinic coupling exists we may ensure, by means of a standard product construction, that the pair (X, X') is defined on the same probability space as, and is independent of, the noise variables ε_i. It is shown in [**24**] that many common models of smooth dynamical systems, for example uniformly hyperbolic (and Axiom A) C^2-diffeomorphisms, admit homoclinic couplings.

Definition: A denoising procedure is a collection of measurable maps $\psi_{n,i} : (\mathbb{R}^d)^n \to \mathbb{R}^d$, with $n \geq 1$, and $i = 1, \ldots, n$. The procedure $\{\psi_{n,i}\}$ is weakly consistent for a process $\{(X_i, Y_i)\}$ if

$$E\left[\frac{1}{n}\sum_{i=1}^{n} |\psi_{n,i}(Y_1, \ldots, Y_n) - X_i|\right] \;\to\; 0$$

as n tends to infinity.

THEOREM 4. *Suppose that $F : \Lambda \to \Lambda$ is a μ-preserving ergodic transformation that admits a homoclinic coupling (X, X'). If the distribution η of ε_i satisfies conditions (13) - (15) then no denoising procedure is weakly consistent for the process $\{(X_i, Y_i)\}$ defined in (18).*

Proof: Assume, without loss of generality, that X is the first component of a homoclinic coupling (X, X') for F. Let $X_i' = F^i X'$ and $Y_i' = X_i' + \varepsilon_i$ for $i \in \mathbb{Z}$. Fix

a denoising scheme $\{\psi_{n,i}\}$ and assume by way of contradiction that

$$E\left[\frac{1}{n}\sum_{i=1}^{n}|\psi_{n,i}(Y_1,\ldots,Y_n)-X_i|\right] \;\to\; 0. \tag{19}$$

The joint distribution of $\{(X_i,Y_i)\}$ is the same as that of $\{(X_i',Y_i')\}$ and, therefore (19) implies that

$$E\left[\frac{1}{n}\sum_{i=1}^{n}|\psi_{n,i}(Y_1',\ldots,Y_n')-X_i'|\right] \;\to\; 0. \tag{20}$$

For each $n \geq 1$ define

$$\phi_n(v_{-n},\ldots,v_n) \;=\; \frac{1}{n}\sum_{i=1}^{n}\psi_{n,i}(v_{1-i},\ldots,v_{n-i}).$$

The stationarity of $\{(X_i,Y_i)\}$ implies that

$$\begin{aligned} E|\,\phi_n(Y_{-n},\ldots,Y_n)\,-\,X\,| &= E\left|\frac{1}{n}\sum_{i=1}^{n}\psi_{n,i}(Y_{1-i},\ldots,Y_{n-i})\,-\,X\right| \\ &\leq \frac{1}{n}\sum_{i=1}^{n}E|\,\psi_{n,i}(Y_{1-i},\ldots,Y_{n-i})\,-\,X\,| \\ &= \frac{1}{n}\sum_{i=1}^{n}E|\,\psi_{n,i}(Y_1,\ldots,Y_n)\,-\,X_i\,|, \end{aligned}$$

which tends to zero by (19). An analogous argument using (20) shows that

$$E|\,\phi_n(Y_{-n}',\ldots,Y_n')\,-\,X'\,| \;\to\; 0.$$

If H is the event that (X,X') is strongly homoclinic for F then, letting $v_i^j = v_i,\ldots,v_j$,

$$\begin{aligned} 0 &= \lim_{n\to\infty}E\left[|\phi_n(Y_{-n}^n)\,-\,X|\,+\,|\phi_n(Y_{-n}^{'n})\,-\,X'|\right] \\ &\geq \liminf_{n\to\infty}E\left[(\,|\phi_n(Y_{-n}^n)\,-\,X|\,+\,|\phi_n(Y_{-n}^{'n})\,-\,X'|\,)\cdot I_H\right]. \end{aligned}$$

It follows from Theorem 3 and the assumption that $P(H) > 0$ that the last term above is positive. As this leads to an evident contradiction, (19) cannot hold, and the proof is complete.

5.4. Some Refinements. The proof of Theorem 4 shows that the values of $X_1, X_2, \ldots$ are not estimable, even if one is given access to the *entire sequence* $\ldots, Y_{-1}, Y_0, Y_1, \ldots$ generated by X and the noise. In particular, there is no function $\psi : \mathcal{X} \to \mathbb{R}^d$ such that

$$E\left[\frac{1}{n}\sum_{i=1}^{n}|\,\psi(\ldots,Y_{i-1},Y_i,Y_{i+1},\ldots)\,-\,X_i\,|\right] \;\to\; 0$$

as n tends to infinity.

Another question that arises is how Theorem 4 bears on the problem of denoising a trajectory arising from a fixed (non-random) initial vector $x \in \Lambda$. It follows

immediately that if $x_i = F^i x$ and $y_i = x_i + \varepsilon_i$, then there is no denoising procedure such that

$$E\left[\frac{1}{n}\sum_{i=1}^{n} |\,\psi_{n,i}(y_1,\ldots,y_n) \,-\, x_i\,|\right] \;\to\; 0$$

for μ-almost every initial state $x \in \Lambda$. For denoising procedures satisfying a natural fading-memory property, this conclusion may be strengthened. Let us say that a procedure $\{\psi_{n,i}\}$ has fading memory if, with Y_i defined as in (18), for each $k \geq 1$,

$$\lim_{n\to\infty} E\left[\frac{1}{n-k}\sum_{i=k+1}^{n} |\,\psi_{n-k,i-k}(Y_{k+1},\ldots,Y_n) \,-\, \psi_{n,i}(Y_1,\ldots,Y_n)\,|\right] \;=\; 0$$

Averaging methods such as Smoothing Algorithm D posess the fading memory property. Under the conditions of Theorem 4, it can be shown that if $\{\psi_{n,i}\}$ has fading memory, then

$$\limsup_{n\to\infty} E\left[\frac{1}{n}\sum_{i=1}^{n} |\,\psi_{n,i}(y_1,\ldots,y_n) \,-\, x_i\,|\right] \;>\; 0$$

for μ-almost every initial state $x \in \Lambda$. Thus successful denoising is not possible starting from almost any initial state.

6. Proof of Theorems 1 and 2

6.1. McDiarmid's Inequality. McDiarmid's inequality is a special case of what is known as the concentration of measure phenomena. The basic idea is the following. If $f(x_1,\ldots,x_n)$ be a function that does not depend too strongly on the value of any single argument, and if $X_1,\ldots,X_n$ are independent random variables, then $f(X_1,\ldots,X_n)$ will be close to $Ef(X_1,\ldots,X_n)$ with high probability. Put another way, the distribution of $f(X_1^n)$ will be *concentrated* around its mean. For a proof and discussion of the following result, see [**28, 16**].

THEOREM B (McDiarmid). *Let $X_1,\ldots,X_n$ be independent random variables taking values in a set $A \subseteq \mathbb{R}$ and let $f : A^n \to \mathbb{R}$. For $i = 1,\ldots,n$ define*

$$v_i \;=\; \sup |f(x_1^n) - f(x_1^{i-1}, x_i', x_{i+1}^n)|, \tag{21}$$

where the supremum is over all numbers $x_1,\ldots,x_n,x_i' \in A$. Then for every $t > 0$

$$P\{|f(X_1^n) - Ef(X_1^n)| > t\} \;\leq\; 2\,\exp\left\{\frac{-2t^2}{\sum_{i=1}^n v_i^2}\right\}. \tag{22}$$

6.2. Analysis of Stochastic Variability. Here we derive exponential inequalities for the final term in (6), which governs the stochastic variability of the estimate $\hat{x}_{l,n}$. Define $U_n(l,k) = \sum_{j\in A_n(l,k)} \varepsilon_j$.

LEMMA 2. *If $H(\Delta/5) \leq k < n/2$ and $k \leq l \leq n-k$ then*

$$U_n(l,k) \;=\; \sum_{j=k}^{n-k} \varepsilon_j\, I\{|x_l - x_j| \leq \Delta/5\} \prod_{1\leq|s|\leq k} I\{|y_{l+s} - y_{j+s}| \leq 3\Delta/5\}$$

Proof: Note that $U_n(l,k)$ can be written in the form

$$U_n(l,k) \;=\; \sum_{j=k}^{n-k} \varepsilon_j \prod_{|s|\leq k} I\{|y_{l+s} - y_{j+s}| \leq 3\Delta/5\}$$

Fix j and define the quantities

$$W_0 \;=\; \prod_{|s|\le k} I\{|y_{l+s} - y_{j+s}| \le 3\Delta/5\}$$

and

$$W_1 \;=\; I\{|x_l - x_j| \le \Delta/5\} \prod_{1\le|s|\le k} I\{|y_{l+s} - y_{j+s}| \le 3\Delta/5\}.$$

It suffices to show that $W_0 = W_1$. If $|x_l - x_j| \le \Delta/5$ then $|y_l - y_j| \le 3\Delta/5$ and the desired equality is immediate. Suppose then that $|x_l - x_j| > \Delta/5$, in which case $W_1 = 0$. If in addition $W_0 = 1$, then $|x_{l+s} - x_{j+s}| \le \Delta$ for $|s| \le k$, which implies that $|x_l - x_j| \le H^{-1}(k) \le \Delta/5$. As this is a contradiction, W_0 must be zero, and the lemma is established.

LEMMA 3. *Let $L = \Delta/5$ be an upper bound on $|\varepsilon_i|$. Fix $n \ge 1$ and integers l, k satisfying the conditions of Lemma 2. Then for every $t > 0$,*

$$P\{|U_n(l,k)| > t\} \;\le\; 2\exp\left\{\frac{-2t^2}{nL^2(2k+1)^2} + \frac{4t}{nL(2k+1)}\right\},$$

and in particular

$$P\{|U_n(l,k)| > t\} \;\le\; 2\exp\left\{\frac{-t^2}{2nL^2(2k+1)^2}\right\}$$

for $t \ge 2L(2k+1)$.

Proof: Define $\tilde{U}$ by excluding indices $j = l-k, \ldots, l+k$ from the sum defining $U_n(l,k)$, more precisely

$$\tilde{U} \;=\; \left(\sum_{j=k}^{l-k-1} + \sum_{j=l+k+1}^{n-k}\right) \varepsilon_j\, I\{|x_l - x_j| \le \Delta/5\} \prod_{1\le|s|\le k} I\{|y_{l+s} - y_{j+s}| \le 3\Delta/5\},$$

with the understanding that the first sum is zero if $l \le 2k$, and the second sum is zero if $l \ge n - 2k$. Then $|U_n(l,k) - \tilde{U}| \le (2k+1)L$, and as ε_j is independent of the other products in the j'th summand, $E\tilde{U} = 0$. Suppose for the moment that the values of $\varepsilon_{l-k}, \ldots, \varepsilon_{l+k}$ have been fixed. In this case $y_{l-k}, \ldots, y_{l+k}$ are fixed, and $\tilde{U}$ is a function of $n - (2k+1)$ independent random variables $\Theta = \{\varepsilon_j : j = 1, \ldots, l-k-1, l+k+1, \ldots, n\}$. Let f be such that $\tilde{U} = f(\Theta)$. Changing any $\varepsilon_j \in \Theta$ will change y_j, and can affect at most $2k+1$ terms in the sum defining $\tilde{U}$; thus the coefficient v_j defined in (21) is at most $(2k+1)L$. As $E(\tilde{U} \,|\, \varepsilon_{l-k}^{l+k}) = E\tilde{U} = 0$, McDiarmid's inequality implies that

$$\begin{aligned} P(|\tilde{U}| > t \mid \varepsilon_{l-k}^{l+k}) \;&\le\; 2\exp\left\{\frac{-2t^2}{\left(\sum_{j=k}^{l-k-1} + \sum_{j=l+k+1}^{n-k}\right)((2k+1)L)^2}\right\} \\ &\le\; 2\exp\left\{\frac{-2t^2}{n\,(2k+1)^2\,L^2}\right\}. \end{aligned}$$

Taking expectations, the same inequality holds for $P\{|\tilde{U}| > t\}$. The first of the stated inequalities follows from the fact that $|U_n(l,k) - \tilde{U}| \le (2k+1)L$, and the second follows from the first by a straightforward calculation.

Definition: Let $V_n(l,k) = |A_n(l,k)|^{-1} \sum_{j \in A_n(l,k)} \varepsilon_j$ be the stochastic term appearing in (6). For each $m \geq 1$ and $1 \leq k \leq n/2$ define

$$L_n(m,k) = \{l : |A_n(l,k)| \geq m \text{ and } k \leq l \leq n-k\}$$

to be the set of indices l for which at least m length-k matches are found.

As an immediate corollary of Lemma 3 we may derive bounds on the probability that one of the terms $V_n(l,k)$ with $|A_n(j,k)| \geq m$ exceeds a given constant $\delta > 0$. In particular, treating a maximum over the empty set as zero, we find that

$$\begin{aligned} P\left\{\max_{l \in L_n(m,k)} |V_n(l,k)| > \delta\right\} &= P\left\{\max_{l \in L_n(m,k)} \frac{|U_n(l,k)|}{|A_n(l,k)|} > \delta\right\} \\ &\leq P\left\{\max_{l \in L_n(m,k)} |U_n(l,k)| > \delta m\right\} \\ &\leq n \cdot \max_l P\{|U_n(l,k)| > \delta m\} \\ &\leq 2n \exp\left\{\frac{-2\delta^2 m^2}{nL^2(2k+1)^2} + \frac{4\delta m}{nL(2k+1)}\right\} \qquad (23) \\ &\leq 2n \exp\left\{\frac{-\delta^2 m^2}{2nL^2(2k+1)^2}\right\} \text{ if } \delta m \geq 2L(2k+1). \qquad (24) \end{aligned}$$

Inequality (24) is used below, in conjunction with the Borel Cantelli Lemma, to establish the almost sure consistency of the estimates $\hat{x}_{l,n}$ and $\tilde{x}_{l,n}$. Neither result makes full use of the inequality, which shows, for example, that for each $\alpha \in (0, 1/2)$,

$$n^\alpha \cdot \max_{l \in L_n(m,k)} \frac{|\sum_{j \in A_n(l,k)} \varepsilon_j|}{|A_n(l,k)|} \to 0 \qquad (25)$$

with probability one, provided that $k = O(\log n)$ and $m \geq n^\beta$ with $\beta \in (\alpha + 1/2, 1)$. The next lemma appears in ([**24**]); we include the proof for completeness.

LEMMA 4. *If $k = o(\log n)$ then for every $\epsilon > 0$,*

$$\frac{1}{n} \sum_{j=0}^{n} I\{|A_n(j,k)| \leq n^{1-\epsilon}\} \to 0 \text{ as } n \to \infty.$$

Proof: As Λ is compact, there exists a finite set set $S \subseteq \Lambda$, such that

$$\max_{u \in \Lambda} \min_{v \in S} |u - v| \leq \frac{\triangle}{10}.$$

Let S^{2k+1} be the collection of sequences $\underline{s} = (s_{-k}, \cdots, s_k)$ with $s_i \in S$. For each $x \in \Lambda$ there is some $\underline{s} \in S^{2k+1}$, such that $\max_{|i| \leq k} |s_i - F^i x| < \triangle/10$. Thus if we define

$$J_n(\underline{s}) = \left\{j : 0 \leq j \leq n \text{ and } \max_{|i| \leq k} |s_i - F^{i+j} x| \leq \frac{\triangle}{10}\right\} \quad \underline{s} \in S^{2k+1}$$

then each integer $j = k, \ldots, n-k$ is in contained in at least one set $J_n(\underline{s})$. Moreover, if $j_1, j_2 \in J_n(\underline{s})$ then

$$\max_{|i| \leq k} |x_{j_1+i} - x_{j_2+i}| < \frac{\triangle}{5} \quad \text{and} \quad \max_{|i| \leq k} |y_{j_1+i} - y_{j_2+i}| < \frac{3\triangle}{5},$$

and therefore $j_1 \in A_n(j_2,k)$ and $j_2 \in A_n(j_1,k)$. It follows from this last observation that $|A_n(j,k)| \le N$ and $j \in J_n(\underline{s})$ imply $|J_n(\underline{s})| \le N$. Fix $0 < \epsilon < 1$. As $k = o(\log n)$, $|S^{2k+1}| = |S|^{2k+1} = o(n^{\epsilon/2})$. Let $\sum_{\underline{s}}$ denote the sum over S^{2k+1}. When n sufficiently large,

$$\begin{aligned}
\sum_{j=0}^{n} I\{|A_n(j,k)| \le n^{1-\epsilon}\} &\le \sum_{j=0}^{n}\sum_{\underline{s}} I\{|A_n(j,k)| \le n^{1-\epsilon}\} I\{j \in J_n(\underline{s})\} \\
&\le \sum_{\underline{s}}\sum_{j=0}^{n} I\{|J_n(\underline{s})| \le n^{1-\epsilon}\} I\{j \in J_n(\underline{s})\} \\
&= \sum_{\underline{s}} |J_n(\underline{s})|\, I\{|J_n(\underline{s})| \le n^{1-\epsilon}\} \\
&\le \sum_{\underline{s}} |J_n(\underline{s})| I\left\{|J_n(\underline{v})| \le n^{1-\epsilon/2}\cdot |S|^{-2k-1}\right\} \\
&\le n^{1-\frac{\epsilon}{2}}.
\end{aligned}$$

As the last term above is $o(n)$ the result follows.

Proof of Theorem 1: Fix $\beta \in (1/2,1)$. The stochastic term in inequality (7) can be bounded as follows:

$$\begin{aligned}
\frac{1}{n-2k}\sum_{i=k}^{n-k} \frac{|\sum_{j\in A_n(l,k)} \varepsilon_j|}{|A_n(l,k)|} &\le \max_{l\in L_n(n^\beta,k)} \frac{|\sum_{j\in A_n(l,k)} \varepsilon_j|}{|A_n(l,k)|} \\
+\frac{\Delta}{5(n-2k)}\sum_{i=k}^{n-k} I\{|A_n(i,k)| \le n^\beta\}&
\end{aligned}$$

Inequality (24) ensures that the first term on the right hand side tends to zero with probability one. The second term tends to zero by Lemma 4.

PROPOSITION 3. *Let the window widths k_l be defined as in (8). If x is strongly recurrent then* $\min\{k_{j,n} : \log n \le j \le n - \log n\} \to \infty$ *as* $n \to \infty$.

Proof: Let $\mathcal{O}$ be a given finite cover of Λ by sets having diameter less than $\Delta/5$. Fix $K \ge 1$ and define γ to be the set of all Cartesian products $O_{-k} \times \cdots \times O_k$ with $1 \le k \le K$ and such that each $O_i \in \mathcal{O}$. Let C_k denote any product of $2k+1$ sets from $\mathcal{O}$. As x is assumed to be strongly recurrent, $\gamma = \gamma_0 \cup \gamma_1$ where

$$\begin{aligned}
\gamma_0 &= \bigcup_{k=1}^{K}\left\{C_k : \sum_{i=k}^{\infty} I\{x_{i-k}^{i+k} \in C_k\} < \infty\right\} \quad \gamma_1 \\
&= \bigcup_{k=1}^{K}\left\{C_k : \liminf_n \frac{1}{n}\sum_{i=k}^{n} I\{x_{i-k}^{i+k} \in C_k\} > 0\right\}
\end{aligned}$$

As γ_0 is finite, there exists an integer $N < \infty$ such that $x_{j-k}^{j+k} \in C_k \in \gamma_1$ for every $k \le K$ and every $j \ge \log N$. Moreover, if x_{i-k}^{i+k} and x_{j-k}^{j+k} lie in the same set $C_k \in \gamma_1$ and $\log n \le i, j \le n - \log n$ then it is clear that $i \in A_n(j,k)$. Thus when $n \ge N$,

$$|A_n(j,k)| \ge \min_{C_k\in\gamma_1} \sum_{i=k}^{n-k} I\{x_{i-k}^{i+k} \in C_k\} \quad \text{for} \quad k \le K \text{ and } j = \log n, \ldots, n - \log n.$$

The definition of γ_1 ensures that $|A_n(j,k)| \geq n/\log n$ for n sufficiently large and k, j as above. Therefore $\liminf_n \min_j k_{j,n} \geq K$ and the result follows as K was arbitrary.

Proof of Theorem 2: Let $\kappa = \log n$ and $m = n/\log n$. It follows from inequality (6) and the definition of $k_{l,n}$ that

$$\begin{aligned}\max_{\kappa \leq l \leq n-\kappa} |x_l - \hat{x}_{l,n}| &\leq \max_{\kappa \leq l \leq n-\kappa} H^{-1}(k_{l,n}) + \max_{\kappa \leq l \leq n-\kappa} \frac{|\sum_{j \in A_n(l,k_{l,n})} \varepsilon_j|}{|A_n(l,k_{l,n})|} \\ &\leq H^{-1}\left(\min_{\kappa \leq l \leq n-\kappa} k_{l,n}\right) + \max_{1 \leq k \leq \kappa} \max_{l \in L_n(m,k)} \frac{|\sum_{j \in A_n(l,k)} \varepsilon_j|}{|A_n(l,k)|}\end{aligned}$$

If x is strongly recurrent then the first term on the right hand side tends to zero by an application of Proposition 1 and Proposition 3. Inequality (24) and a standard Borel-Cantelli argument show that the second term tends to zero with probability one.

7. Proof of Theorem 3

Throughout this section (x, x') is a fixed strongly homoclinic pair for F. Define $x_i = F^i x$, $x'_i = F^i x'$, $y_i = x_i + \varepsilon_i$ and $y'_i = x'_i + \varepsilon_i$ as above. As (x, x') is strongly homoclinic,

$$\sum_{i \in \mathbb{Z}} |x_i - x'_i| < \infty. \tag{26}$$

LEMMA 5. *If conditions (14) and (15) hold, then there exist sets $A_i^* \subseteq \mathbb{R}^d$, $i \in \mathbb{Z}$, such that*

a. *$A_i^* \subseteq (S + x_i) \cap (S + x'_i)$ for each i, and*
b. *$P\{y_i \in A_i^* \text{ and } y'_i \in A_i^* \text{ for all } i \in \mathbb{Z}\} > 0$.*

Proof: For each $i \in \mathbb{Z}$ define $A_i = (S + x_i) \cap (S + x'_i)$. Note that

$$\begin{aligned}P\{y_i \notin A_i \text{ or } y'_i \notin A_i\} &\leq P\{y_i \notin A_i\} + P\{y'_i \notin A_i\} \\ &= P\{\varepsilon_i \notin (S + (x'_i - x_i))\} + P\{\varepsilon_i \notin (S + (x_i - x'_i))\} \\ &= \eta(\, S \setminus (S + (x'_i - x_i))\,) + \eta(\, S \setminus (S + (x_i - x'_i))\,) \\ &\leq 2\,\eta(\,(\partial S)^{|x_i - x'_i|}\,)\end{aligned}$$

Assumption (14) implies that $\eta(\,(\partial S)^{|x_i - x'_i|}\,) \leq c\,|x_i - x'_i|$ for some constant $c < \infty$, and it then follows from (26) that

$$\sum_{i \in \mathbb{Z}} P\{y_i \notin A_i \text{ or } y'_i \notin A_i\} < \infty.$$

By an application of the Borel Cantelli Lemma, there exists an integer N such that

$$P\{y_i \in A_i \text{ and } y'_i \in A_i \text{ for all } |i| > N\} \geq 1/2. \tag{27}$$

Define $A_i^* = A_i$ for $|i| > N$. Clearly (a) holds for each $|i| > N$.

It remains to select sets A_i^* for $|i| \leq N$. To this end, let $v^* \in \mathbb{R}^d$ be any vector such that for some $\delta > 0$

$$\sup_{v \in \Lambda} |v - v^*| \leq \rho - \delta$$

and define $A_i^* = B(v^*, (\rho+\delta)/2)$ for $|i| \le N$. Then for each $v \in \Lambda$,

$$\sup_{u \in (A_i^* - v)} |u| \;\le\; \frac{\rho+\delta}{2} \;+\; |v^* - v| \;<\; \frac{3}{2}\,\rho,$$

which implies that $(A_i^* - v) \subseteq B(0, 3\rho/2) \subseteq S$. Thus (a) holds for $|i| \le N$. Moreover, for each such i,

$$\begin{aligned} P\{y_i \in A_i^* \text{ and } y_i' \in A_i^*\} &= P\{\varepsilon_i \in (A_i^* - x_i) \cap (A_i^* - x_i')\} \\ &= \eta((A_i^* - x_i) \cap (A_i^* - x_i')). \end{aligned}$$

The inequality $|\,|v^* - x_i| - |v^* - x_i'|\,| \le |x_i - x_i'| \le \rho$ implies that $(A_i^* - u_i) \cap (A_i^* - v_i)$ has positive Lebesgue measure. As the intersection is also contained in S, the last probability above is greater than zero. Conclusion (b) of the lemma follows from this observation and (27), as the ε_i's are independent.

Let Q and Q' be probability measures on $(\mathcal{X}, \mathcal{S})$ equal to the respective probability distributions of the random elements $\mathbf{y}$ and $\mathbf{y}'$. Using the sets A_i^* from Lemma 5, define the Cartesian product

$$\Gamma \;=\; \prod_{i \in \mathbb{Z}} A_i^* \;\in\; \mathcal{S}. \tag{28}$$

It follows from part (b) of Lemma 5 that $Q(\Gamma), Q'(\Gamma) > 0$.

LEMMA 6. *The measures Q and Q' are mutually absolutely continuous on Γ: for each $B \in \mathcal{S}$ contained in Γ, $Q(B) = 0$ if and only if $Q'(B) = 0$.*

Proof: Let $\mathcal{S}_n \subseteq \mathcal{S}$ denote the sigma field generated by the coordinate functions $\pi_i(\mathbf{x}) = x_i$, with $|i| \le n$. Let Q_n and Q_n' be the restrictions of Q and Q' to $\mathcal{S}_n$, respectively. Then clearly

$$dQ_n(\mathbf{v}) \;=\; \prod_{i=-n}^{n} f(v_i - x_i)\, dv_{-n} \cdots dv_n \;\text{ and }\; dQ_n'(\mathbf{v}) \;=\; \prod_{i=-n}^{n} f(v_i - x_i')\, dv_{-n} \cdots dv_n.$$

Furthermore, Lemma 5 ensures that Q_n and Q_n' are mutually absolutely continuous on Γ, with derivative

$$\frac{dQ_n}{dQ_n'}(\mathbf{v}) \;=\; \prod_{i=-n}^{n} \frac{f(v_i - x_i)}{f(v_i - x_i')} \qquad \mathbf{v} \in \Gamma.$$

For each $n \ge 1$ let $\Gamma_n = \{\mathbf{v} : v_i \in A_i^* \text{ for } |i| \le n\}$, and define the $\mathcal{S}_n$-measurable function

$$R_n(\mathbf{v}) \;=\; \frac{dQ_n}{dQ_n'}(\mathbf{v}) \cdot I\{\mathbf{v} \in \Gamma_n\}.$$

Suppose that $B \in \mathcal{S}_n$. Then clearly $B \cap \Gamma_{n+1} \in \mathcal{S}_{n+1}$ and $B \cap \Gamma_n \in \mathcal{S}_n$, and therefore

$$\begin{aligned} \int_B R_{n+1}\, dQ' &= \int_{B \cap \Gamma_{n+1}} \frac{dQ_{n+1}}{dQ_{n+1}'}\, dQ' \;=\; Q_{n+1}(B \cap \Gamma_{n+1}) \\ &= Q(B \cap \Gamma_{n+1}) \;\le\; Q(B \cap \Gamma_n) \;=\; \int_B R_n\, dQ'. \end{aligned}$$

Thus $(R_n, \mathcal{S}_n)$ is a non-negative super-martingale. By the martingale convergence theorem, R_n converges with Q'-probability one to a non-negative random variable R^*.

We now wish to establish the following relation, which will imply that $Q' << Q$ on Γ (see the argument below):

$$Q'\{\mathbf{v} \in \Gamma : R^*(\mathbf{v}) = 0\} = 0. \tag{29}$$

By condition (13) there exists numbers $\delta_0 > 0$ and $c < \infty$ such that

$$\int_{S \cap (S-z)} \left| \log \frac{f(w+z)}{f(w)} \right| f(w)\, dw \;\le\; c\,|z| \tag{30}$$

whenever $|z| \le \delta_0$. By (26) there is an integer m such that $|u_i - v_i| \le \delta_0$ for $|i| \ge m$. As $R_m(\mathbf{v}) > 0$ for each $\mathbf{v} \in \Gamma$, the equality (29) will follow from

$$\int_\Gamma \left| \log \frac{R^*}{R_m} \right| dQ' \;<\; \infty. \tag{31}$$

To establish (31), note that by Fatou's lemma

$$\begin{aligned}
\int_\Gamma \left| \log \frac{R^*}{R_m} \right| dQ' &= \int_\Gamma \liminf_{n\to\infty} \left| \log \frac{R_n}{R_m} \right| dQ' \;\le\; \liminf_{n\to\infty} \int_\Gamma \left| \log \frac{R_n}{R_m} \right| dQ' \\
&\le \liminf_{n\to\infty} \int_{\Gamma_n} \left| \log \frac{R_n}{R_m} \right| dQ' \;=\; \liminf_{n\to\infty} \int_{\Gamma_n} \left| \log \frac{R_n}{R_m} \right| dQ'_n.
\end{aligned}$$

Moreover, for each $n > m$,

$$\begin{aligned}
\int_{\Gamma_n} \left| \log \frac{R_n}{R_m} \right| dQ'_n &= \int_{\Gamma_n} \left| \sum_{m \le |i| \le n} \log \frac{f(v_i - x_i)}{f(v_i - x'_i)} \right| \prod_{j=-n}^{n} f(v_j - x'_j)\, dv_{-n} \cdots dv_n \\
&\le \sum_{m \le |i| \le n} \int_{A_i^*} \left| \log \frac{f(v_i - x_i)}{f(v_i - x'_i)} \right| f(v_i - x'_i)\, dv_i.
\end{aligned}$$

By an elementary change of variables, our choice of m and the inequality (30) imply that

$$\int_{A_i^*} \left| \log \frac{f(v_i - x_i)}{f(v_i - x'_i)} \right| f(v_i - x'_i)\, dv_i \;\le\; c\,|x_i - x'_i|.$$

Combining the results of the last three displays, it follows that

$$\int_\Gamma \left| \log \frac{R^*}{R_m} \right| dQ' \;\le\; \sum_{i=-\infty}^{\infty} |x_i - x'_i|.$$

The sum is finite by (26), which establishes (31) and the relation (29).

Suppose now that $B \in \mathcal{S}$ is such that $B \subseteq \Gamma$ and $Q'(B) > 0$. For $n \ge 1$ define events $B_n = \{\mathbf{v} : \exists \mathbf{v}' \in B \text{ s.t. } v_i = v'_i \text{ for } |i| \le n\} \supseteq B$. By another application of Fatou's Lemma,

$$\begin{aligned}
Q(B) &= \lim_{n\to\infty} Q(B_n) \;=\; \liminf_{n\to\infty} Q_n(B_n) \;=\; \liminf_{n\to\infty} \int \frac{dQ_n}{dQ'_n} I_{B_n}\, dQ' \\
&\ge \int \liminf_{n\to\infty} \frac{dQ_n}{dQ'_n} I_{B_n}\, dQ' \;\ge\; \int_B R^*\, dQ'.
\end{aligned}$$

The last inequality above follows from the definition of R^* and the fact that $B_n \supseteq B$. As $Q'(B) > 0$, the relation (29) implies that the last integral and $Q(B)$ are positive.

Thus $Q' \ll Q$ on Γ. An identical argument, exchanging the roles of Q and Q', shows that $Q \ll Q'$ on Γ as well.

LEMMA 7. *If (x, x') is homoclinic and conditions (13)–(15) hold, then for every measurable function $\phi : \mathcal{X} \to \mathbb{R}^d$,*

$$E[\,|\phi(\mathbf{y}) - x\,| \,+\, |\phi(\mathbf{y}') - x'\,|\,] \;\geq\; |x - x'| \int_\Gamma \min\left[\frac{dQ'}{dQ}, 1\right] dQ \;>\; 0 \tag{32}$$

where $\Gamma \subseteq \mathcal{X}$ is defined as in (28).

Proof: Lemma 6 shows that $Q' \ll Q$ on Γ. Let $(dQ'/dQ)(\mathbf{v})$ be the associated derivative, which is defined for each $\mathbf{v} \in \Gamma$. The expectation above can be written equivalently as

$$\begin{aligned}
\int |\phi - x|\, dQ \,+\, \int |\phi - x'|\, dQ' \;&\geq\; \int_\Gamma |\phi - x|\, dQ \,+\, \int_\Gamma |\phi - x'|\, dQ' \\
&=\; \int_\Gamma \left[|\phi - x| \,+\, |\phi - x'| \,\frac{dQ'}{dQ} \right] dQ \\
&\geq\; |x - x'| \int_\Gamma \min\left[\frac{dQ'}{dQ}, 1\right] dQ.
\end{aligned}$$

As $(dQ'/dQ)(\mathbf{v})$ is positive for Q-almost every $\mathbf{v} \in \Gamma$, and $Q(\Gamma) > 0$, the last integral is positive.

The lower bound in Lemma 7 bears further discussion. Suppose for the moment that the distribution η of the noise satisfies (13) and is supported on all of $\mathbb{R}^d$, which is the case if the ε_i are Gaussian. Then we may take $A_i^* = \mathbb{R}^d$ for each i, so that $\Gamma = \mathcal{X}$. In this case, further evaluation leads to a simplification of the integral in (32):

$$\begin{aligned}
\int \min\left[\frac{dQ'}{dQ}, 1\right] dQ \;&=\; Q\left\{\frac{dQ'}{dQ} \geq 1\right\} \,+\, Q'\left\{\frac{dQ'}{dQ} < 1\right\} \\
&=\; 1 \,-\, Q\left\{\frac{dQ'}{dQ} < 1\right\} \,+\, Q'\left\{\frac{dQ'}{dQ} < 1\right\} \\
&=\; 1 \,-\, ||Q - Q'||.
\end{aligned}$$

Here $||Q - Q'|| \;=\; \sup_{B \in \mathcal{S}} |Q(B) - Q'(B)|$ is the total variation distance between Q and Q'. As Q and Q' are mutually absolutely continuous, $||Q - Q'|| < 1$ and we see again that the lower bound in Lemma 7 is positive. When $\Gamma \neq \mathcal{X}$ one may derive a similar, but more complicated, expression for the integral in (32).

Although no scheme can reliably distinguish between the elements of a strongly homoclinic pair (x, x') from noisy observations of their trajectories, we may say that a scheme ϕ is optimal for this pair if it achieves the lower bound above. One may readily check that the maximum likelihood scheme

$$\phi(\mathbf{v}) \;=\; \begin{cases} x & \text{if } \frac{dQ'}{dQ}(\mathbf{v}) \leq 1 \\ x' & \text{otherwise} \end{cases}$$

is optimal in this sense.

References

[1] Adams, T.M., Families of ergodic processes without consistent density or regression estimates. Preprint, 1997.

[2] Adams, T.M. and Nobel, A.B., On density estimation from ergodic processes, *Ann. Probab.*, vol.26 , pp.794–804, 1998.

[3] Adams, T.M. and Nobel, A.B., Finitary reconstruction of a measure preserving transformation, *Israel Journal of Mathematics*, vol.126, pp.309-326, 2001.

[4] Aeyels, D., Generic observability of differential systems, *SIAM J. Control and Optimization*, vol.19, pp.595–603, 1981

[5] Alligood, K., Sauer,T., and Yorke, J.A., *Chaos: An Introduction to Dynamical Systems*, Springer, New York, 1996.

[6] Berliner, L.M., Statistics, probability, and chaos, *Stat. Sci.*, vol.7, pp.69-122 (with discussion), 1992.

[7] Bosq, D. and Guégan, D., Nonparametric estimation of the chaotic function and the invariant measure of a dynamical system. *Stat. and Prob. Let.*, vol.25, pp.201-212, 1995.

[8] Bowen, R., Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms. *Springer Lecture Notes in Math. 470*, 1975.

[9] Bowen, R., On Axiom A Diffeomorphisms, *Regional Conference Series in Math.*, 35, American Math. Soc., Providence, RI, 1978.

[10] Carroll, R.J., Ruppert, D. and Stefanski, L.A., *Measurement Error in Nonlinear Models*, Chapman and Hall, London, 1995

[11] Casdagli, M., Nonlinear prediction of chaotic time series, *Physica D*, vol.35, pp.335-356, 1989.

[12] Casdagli, M., Chaos and deterministic *versus* stochastic non-linear modeling, *J. R. Stat. Soc. B*, vol.54, pp.303–328, 1992.

[13] Cutler, C.D., A review of the theory and estimation of fractal dimension, In *Dimension estimation and models*, pp.1-107, Nonlinear Time Ser. Chaos I, World Sci. Publishing, 1993.

[14] Davies, M., Noise reduction schemes for chaotic time series, *Physica D*, vol.79, pp.174-192, 1992.

[15] Devaney, R.L., *An Introduction to Chaotic Dynamical Systems*, Second Edition, Addison-Wesley, 1989.

[16] Devroye, L., Györfi, L., and Lugosi, G., *A Probabilistic Theory of Pattern Recognition*, Springer Verlag, New York, 1996.

[17] Eckmann, J.-P., and Ruelle, D., Ergodic theory of chaos and strange attractors, *Rev. Mod. Phys.*, vol.57, pp.617-656, 1985.

[18] Eckmann, J.-P., Kamphorst, S.O., Ruelle, D. and Ciliberto, S., Lyapunov exponents from time series, *Phys. Rev. A*, vol.34, pp.4971-4979, 1986.

[19] Farmer, J.D. and Sidorowich, J.J., Predicting chaotic time series, *Phys. Rev. Let.*, vol.59 pp.845-848, 1987.

[20] Isham, V., "Statistical aspects of chaos: a review", in *Networks and Chaos – Statistical and Probabilistic Aspects*, O.E. Barndorff-Nielsen, J.L. Jensen, and W.S. Kendall editors, Chapman and Hall, London, 1993.

[21] Jensen, J.L., "Chaotic dynamical systems with a view towards statistics: a review" in *Networks and Chaos – Statistical and Probabilistic Aspects*, O.E. Barndorff-Nielsen, J.L. Jensen, and W.S. Kendall editors, Chapman and Hall, London, 1993.

[22] Kostelich, E.J. and Schreiber, T., Noise reduction in chaotic time-series data: a survey of common methods, *Phys. Rev. E*, vol.48, pp.1752-1763, 1993.

[23] Kostelich, E.J. and Yorke, J.A., Noise reduction: finding the simplest dynamical system consistent with the data, *Physica D*, vol.41, pp.183-196, 1990.

[24] Lalley, S.P., Beneath the noise, chaos, *Ann. Stat.*, vol.27, pp.461-479, 1999.

[25] Lalley, S.P., Removing the noise from chaos plus noise, *Nonlinear Dynamics and Statistics*, A.I. Mees editor, Birkhauser, pp.233-244, 2001.

[26] Lu, Z.-Q., and Smith, R. L., Estimating local Lyapunov exponents, *Fields Institute Communications*, vol.11, pp.135-151, 1997.

[27] MacEachern, S.N. and Berliner, L.M., Asymptotic inference for dynamical systems observed with error, *J. Stat. Plan. and Inf.*, vol.46, pp.277-292, 1995.

[28] McDiarmid, C., On the method of bounded differences, In *Surveys in Combinatorics*, pp.148-188, Cambridge University Press, Cambridge, UK, 1989.

[29] Nobel, A.B., Limits to classification and regression estimation from ergodic processes, *Annals of Statistics*, vol.27, pp.262-273, 1999.
[30] Nobel. A.B., "Consistent estimation of a dynamical map", in *Nonlinear Dynamics and Statistics*, edited by A.I. Mees, Birkhauser, Boston, 2001.
[31] Nychka, D., Ellner, S., Gallant, A.R., and McCaffrey, D., Finding chaos in noisy systems, *J. R. Stat. Soc. B*, vol.54, pp.399-426, 1992.
[32] Sauer, T. (1992) A noise reduction method for signals from nonlinear systems. *Physica D*, 58:193-201.
[33] Sauer, T., Yorke, J.A., and Casdagli, M. (1991). Embedology. *J. Stat. Phys.*, 65:579–616.
[34] Takens, F. (1980) Detecting strange attractors in turbulence In *Dyanmical Systems and Turbulence, Warwick 1980*, D.A. Rand and L.-S. Young, editors. Springer Lecture Notes in Mathematics 898.
[35] Tong, H. (1990). *Non-linear Time Series: a Dynamical System Approach*. Oxford University Press.
[36] Walters, P. (1981). *An Introduction to Ergodic Theory*. Springer, New York.
[37] Wolf, A., Swift, J.B., Swinney, H.L., and Vastano, J.A. (1985). Determining Lyapunov exponents from a time series. *Physica D*, 16:285–315.

Department of Statistics, University of Chicago, Chicago, IL 60637
E-mail address: lalley@galton.uchicago.edu

Department of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260
E-mail address: nobel@email.unc.edu