Removing the Noise from Chaos Plus Noise

Steven P. Lalley
Department of Statistics
University of Chicago

November 5, 2000

Abstract

The problem of extracting a "signal" x_n generated by a dynamical system from a times series $y_n = x_n + e_n$, where e_n is an observational error, is considered. It is shown that consistent signal extraction is *impossible* when the errors are distributed according to a density with unbounded support and the underlying dynamical system admits *homoclinic pairs*. It is also shown that consistent signal extraction is *possible* when the errors are uniformly bounded by a suitable constant and when the underlying dynamical system has the "weak orbit separation property". Simple algorithms for signal recovery are described in the latter case.

1 Introduction

Is it possible to consistently recover a "signal" $\{x_n\}_{n \in \mathbb{Z}}$ generated by a chaotic dynamical system from a time series of the form

$$y_n = x_n + e_n \tag{1}$$

where e_n is observational noise? This is the *noise removal*, or *signal separation* problem, and has been discussed in a number of papers, including [4, 5, 8, 2, 1]. Various sophisticated methods for noise removal have been proposed, nearly all requiring a degree of smoothness in the underlying dynamical system, and some requiring rather detailed *a priori* knowledge of the dynamics. The issue of convergence seems not to have been broached, until now. The purpose of this paper is to state some general results concerning the theoretical possibility of *consistent filtering*, and to propose some fairly simple general-purpose filters for use in high signal/noise ratio problems.

In many circumstances, *scalar* measurements will be made on a dynamical system at equally spaced times to produce the series x_n, which is then observed with error. We shall assume here, however, that x_n is the actual *state vector* of the system at time n. This assumption is probably harmless, in view of the "Embedding Theorem" [9]. Moreover, we shall assume throughout that the *noise e_n* consists of i.i.d. mean zero random vectors that are independent of the state vectors x_n. Although we shall make only weak assumptions about the dynamics, we shall limit our attention to dynamical systems with compact invariant sets. Compactness is essential in Theorems 2 and 3 below.

Definition 1. *A* dynamical system *is a homeomorphism* $F : \Lambda \to \Lambda$ *of a compact subset* Λ *of a Euclidean space* $\mathbb{R}^d$*. For any point* $x \in \Lambda$*, the* orbit *of* x *is the doubly infinite sequence* $\{x_n = F^n(x)\}_{n \in \mathbb{Z}}$*, where* F^n *denotes the* $n-$*fold composition of* F*.*

Definition 2. *A* filter $\hat{x}$ *is a collection of functions*

$$\hat{x}_n(y_0, y_1, y_2, \ldots, y_m) = \hat{x}_n^{(m)}(y_0, y_1, y_2, \ldots, y_m).$$

A filter $\hat{x}$ *is* **weakly consistent** *if for* every *orbit* $\{x_n\}_{n \in \mathbb{Z}}$ *and every* $\varepsilon > 0$*,*

$$\frac{1}{m} \sum_{n=1}^{m} |\hat{x}_n - x_n| \xrightarrow{P} 0. \tag{2}$$

Informally, a filter is weakly consistent if, for large m, *most* of the fitted values $\hat{x}_n$ are close to the corresponding state vectors x_n. Other notions of consistency are undoubtedly worthy of consideration. The stipulation that the convergence in (2) hold for *every* orbit may be seen as overly restrictive – perhaps in some circumstances one would be happy with filters which achieve (2) only for "most" orbits. See section 6 below for further discussion of this point. On the other hand, in certain situations one might regard the requirement in (2) that only "most" points on the orbit be well approximated as too weak. For a stronger notion of consistency, see [6], Theorem 2.

2   Homoclinic Pairs

A common and important dynamical feature of many chaotic systems is the occurrence (or even abundance) of *homoclinic pairs*. Two distinct points x, x' are said to be *homoclinic* if their orbits $\{x_n\}_{n \in \mathbb{Z}}$ and $\{x'_n\}_{n \in \mathbb{Z}}$ satisfy

$$\lim_{n \to \pm\infty} |x_n - x'_n| = 0. \tag{3}$$

In smooth systems, it is commonly (but not always) the case that if the convergence (3) occurs then it is exponentially fast. Say that two homoclinic points x, x' are *strongly homoclinic* if their orbits satisfy

$$\sum_{n=-\infty}^{\infty} |x_n - x'_n| < \infty. \tag{4}$$

In uniformly hyperbolic systems, *most* points are members of strongly homoclinic pairs: If the stable and unstable manifolds through x intersect at x', then (x, x') is a strongly homoclinic pair. In systems admitting "symbolic dynamics" (that is, systems conjugate [or nearly conjugate] to subshifts of finite type), *all* points will be members of homoclinic pairs. This class includes all mixing *Axiom A* diffeomorphisms — see [6].

The occurrence of strongly homoclinic pairs is a fundamental obstruction to the existence of consistent filters. For any error density ϕ on $\mathbb{R}^d$, say that ϕ is in the class Φ if it it strictly positive, has mean zero, and satisfies

$$\limsup_{y \to 0} \frac{1}{|y|} \int \left| \log \frac{\phi(x+y)}{\phi(x)} \right| \phi(x) \, dx < \infty. \tag{5}$$

Note that all mean zero Gaussian densities are of class Φ.

Theorem 1. *If the noise density is of class Φ and the dynamical system admits strongly homoclinic pairs, then there is no sequence of (measurable) functions $\xi_n(y_{-n}, y_{-n+1}, \ldots, y_n)$ such that, for all orbits $x_n = F^n(x)$,*

$$\xi_n(x_{-n} + e_{-n}, x_{-n+1} + e_{-n+1}, \ldots, x_n + e_n) \xrightarrow{P} x \tag{6}$$

Proof Sketch. The argument is essentially the same as in the Axiom A case — see [6]. Let x, x' be a strongly homoclinic pair, with orbits $\{x_n\}_{n\in\mathbb{Z}}$ and $\{x'_n\}_{n\in\mathbb{Z}}$, respectively. Define probability measures Q, Q' on the sequence space $(\mathbb{R}^d)^{\mathbb{Z}}$ to be the distributions of the doubly infinite sequence $\{y_n\}_{n\in Z}$ when y_n is defined by

$$y_n = x_n + e_n \qquad (Q), \tag{7}$$
$$y_n = x'_n + e_n \qquad (Q'), \tag{8}$$

with the random vectors e_n i.i.d. from a density ϕ in class Φ. Then the measures Q, Q' are mutually absolutely continuous, because (4) and the assumption that $\phi \in \Phi$ guarantees the almost sure convergence of the infinite product

$$\frac{dQ}{dQ'} = \prod_{n=-\infty}^{\infty} \frac{\phi(y_n - x_n)}{\phi(y_n - x'_n)} \tag{9}$$

to a strictly positive limit. But if Q and Q' are mutually a.c., then there can be no sequence of functions $\xi_m(y_{-m}, \ldots, y_m)$ such that as $m \to \infty$,

$$\xi_m(y_{-m}, \ldots, y_m) \xrightarrow{Q} x \qquad \text{and} \qquad \xi_m(y_{-m}, \ldots, y_m) \xrightarrow{Q'} x'. \tag{10}$$

Although Theorem 1 does not by itself preclude the existence of weakly consistent filters, it indicates that consistent orbit identification is impossible when there are homoclinic pairs. Moreover, if homoclinic pairs are sufficiently common, then there may not be weakly consistent filters:

Corollary 1. *Supose that there exist an ergodic, $F-$invariant probability measure μ on Λ and a probability measure ν on $\Lambda \times \Lambda$ such that if the $\Lambda \times \Lambda-$ valued random vector (X, X') has distribution ν, then*
    (a) the marginal distributions of X and X' are both μ;
    (b) X and X' are either equal or strongly homoclinic, with probability one; and
    (c) with positive probability, X and X' are strongly homoclinic.
If the noise density is of class Φ, then there is no weakly consistent filter.

The proof is essentially the same as in the case of an Axiom A diffeomorphism — see [6].

3   Sensitive Dependence on Initial Conditions

Definition 3. *The dynamical system $F : \Lambda \to \Lambda$ has **sensitive dependence on initial conditions** if there exists a constant $\Delta > 0$, called a **separation threshold**, such that for any two distinct points $x, x' \in \Lambda$, there exists $n \in \mathbb{Z}$ such that*

$$|F^n(x) - F^n(x')| > \Delta. \tag{11}$$

Dynamical systems with sensitive dependence on initial conditions often have homoclinic pairs: for instance, topologically mixing Axiom A diffeomorphisms have both sensitive dependence and homoclinic pairs. For systems with sensitive dependence on initial conditions, consistent noise removal is possible if the noise level is sufficiently low. Consistent filters are easily described and implemented.

Smoothing Algorithm D: The algorithm takes as input a finite sequence $\{y_n\}_{0 \leq n \leq m}$ and produces as output a sequence $\{\hat{x}_n\}_{0 \leq n \leq m}$ of the same length that will approximate the unobservable signal $\{x_n\}_{0 \leq n \leq m}$. Let κ_m be an increasing sequence of integers such that

$$\lim_{m \to \infty} \kappa_m = \infty \quad \text{and} \quad \lim_{m \to \infty} \frac{\kappa_m}{\log m} = 0; \tag{12}$$

e.g., $\kappa_m = \log m / \log \log m$. For each integer $1 \leq n \leq m$, define A_n to be the set of indices $\nu \in \{0, 1, \ldots, m\}$ such that

$$\max_{|j| \leq \kappa_m} |y_{\nu+j} - y_{n+j}| < 3\delta, \tag{13}$$

with the convention that $|y_j - y_i| = \infty$ if either of i or j is not in the range $[0, m]$. Observe that $n \in A_n$, so A_n is nonempty; and for $n \leq \kappa_m$ or $n \geq m - \kappa_m$, the set A_n is the singleton $\{n\}$. In rough terms, A_n consists of the indices of those points in the time series whose orbits "shadow" the orbit of x_n for κ_m time units. Now define

$$\hat{x}_n = \frac{1}{|A_n|} \sum_{\nu \in A_n} y_\nu. \tag{14}$$

Theorem 2. *Suppose that the dynamical system $f : \Lambda \to \Lambda$ has sensitive dependence on initial conditions, with separation threshold Δ. If the errors e_n have mean zero and are uniformly bounded in absolute value by δ, where $\delta < \Delta/5$, then Smoothing Algorithm D is weakly consistent.*

This is a generalization of Theorem 1 of [6], which applies only to smooth, uniformly hyperbolic systems, where sensitive dependence on initial conditions can be "quantified". Theorem 2 requires no smoothness of the underlying dynamical system at all. Furthermore, the hypotheses may be relaxed in several ways: (A) It is not necessary that the errors e_n be identically distributed. If, for instance, the distribution of e_n is allowed to depend on the state vector x_n, then weak consistency of Smoothing Algorithm D will still hold *provided* that the errors are conditionally independent, given the orbit $\{x_n\}_{n \in \mathbb{Z}}$, that they are uniformly bounded by δ, and that $E(e_n \,|\, \{x_n\}_{n \in \mathbb{Z}}) = 0$. (B) It is not necessary even that the errors be mutually independent. If $\{e_n\}$ is a mean zero, stationary sequence satisfying suitable mixing requirements, then the conclusion of Theorem 2 remains valid.

Explanation of Theorem 2. A complete proof will be given in [7]; here we shall give only a brief indication of the argument. Observe that the average (14) may be rewritten as

$$\hat{x}_n = \frac{1}{|A_n|} \sum_{\nu \in A_n} x_\nu + \frac{1}{|A_n|} \sum_{\nu \in A_n} e_\nu. \tag{15}$$

Thus, to establish weak consistency, it suffices to show that for most of the indices $n \in [1, m]$, (a) the cardinality of A_n is large, and (b) if $\nu \in A_n$ then $|x_n - x_\nu|$ is

small. Property (b) will guarantee that the first average in (15) is close to x_n, while property (a), together with the law of large numbers, will imply (with some work!) that with high probability the second average is near zero.

Property (b) follows easily from sensitive dependence on initial conditions. This implies that if x, x' are any two points whose orbits remain within distance 5δ for all times $-\kappa < n < \kappa$, then $|x - x'|$ must be small, provided κ is large. Since the errors e_n are of magnitude less than δ, if $\nu \in A_n$ then, by the triangle inequality, the orbits of x_n and x_ν must remain within distance 5δ for all $j \in [-\kappa_m, \kappa_m]$. Thus, $\nu \in A_n$ implies that $|x_n - x_\nu|$ is small.

Property (a) follows from the assumption that $\kappa_m = o(\log m)$. Let H be a finite subset of Λ that is $\delta-$dense in Λ, and denote by H^* the set of all $H-$valued sequences of length $2\kappa_m + 1$. For every $F-$orbit segment of length $2\kappa_m + 1$, there is at least one sequence in H^* that $\delta-$shadows it. Since $\kappa_m = o(\log m)$, the cardinality of H^* is $o(m)$. Thus, by the pigeonhole principle, for most indices $n \in [1, m]$ there will be *many* indices ν such that

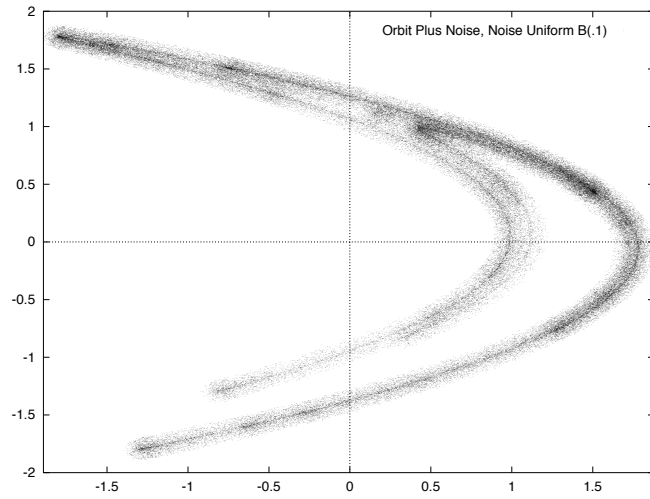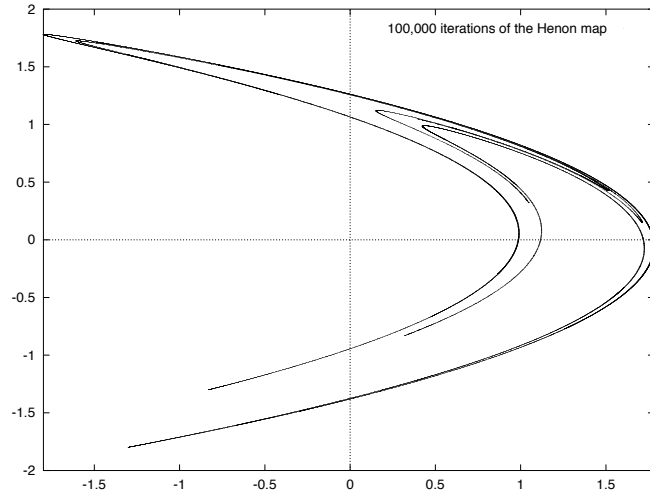$$\max_{|j| \le \kappa_m} |x_{n+j} - x_{\nu+j}| < 2\delta. \tag{16}$$

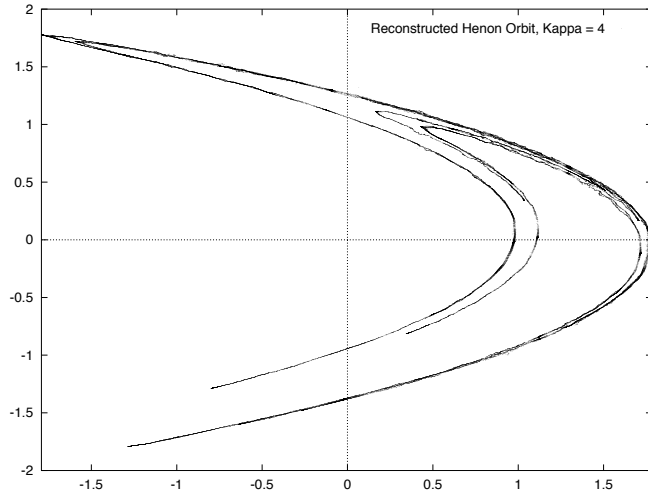All such indices ν must be included in A_n.

Note that this is not a complete proof, because the sets A_n are random, not fixed, and so the use of the Law of Large Numbers is problematic.

4   Example: The Henon Mapping

Smoothing Algorithm D is easily implemented, and simple variations of the algorithm can be made to run in $O(m \log m)$ steps. In practical terms this means that, for simple low-dimensional systems, with $m = 10^5$, the procedure can be run in "real time" (e.g., 10 to 20 seconds on a 200 MHz Power Macintosh). This implies that experimentation with the parameters κ_m and δ may be done in real time. In simple examples, choosing δ to be one-fifth to one-tenth the apparent diameter of the attractor has been effective; and, for $m \approx 10^5$, choosing κ_m so that most bins A_n have 20 to 50 points has provided the best results.

The following sequence of figures illustrate the results of using the filter for a noise-corrupted orbit of length 10^5 generated by the Henon mapping. The figures show (a) 10^5 points on the orbit of a randomly chosen point in the basin of attraction of the attractor Λ; (b) the orbit corrupted by noise; and (c) the reconstructed orbit. The author is indebted to JASON STOVER for coding the algorithm. Similar figures for a noise-corrupted orbit of Smale's solenoid mapping may be found on the author's web page `http://galton.uchicago.edu/~lalley`.

100,000 iterations of the Henon map

Orbit Plus Noise, Noise Uniform B(.1)

Reconstructed Henon Orbit, Kappa = 4

5   The Weak Orbit Separation Property

Although sensitive dependence on initial conditions is sometimes taken to be a necessary condition for chaos (see, e.g., [3], section 1.8), there are important systems for which sensitive dependence does not hold, which nonetheless share many of the dynamical features of chaotic systems. Noteworthy among these are the time-1 mappings induced by smooth hyperbolic flows. If $\phi_t : \Lambda \to \Lambda$ is a smooth flow, and if $F = \phi_1$, then sensitive dependence cannot hold, for an obvious and trivial reason: if two points x, x' are on the same flow line (that is, if $x' = \phi_s(x)$ for some $s \neq 0$) then their orbits (under F) remain on the same flow line, at (roughly) the same distance, forever. However, if the flow ϕ_t is hyperbolic (see [10] for the definition) then the orbits of all neighboring points *not on a common flow line* will eventually separate. Such systems satisfy a *weak orbit separation property*, defined as follows.

Definition 4. *For any pair of points $x, x' \in \Lambda$ and any $\Delta > 0$, define*

$$\tau_+^\Delta(x, x') = \min\{n \geq 0 \,:\, |F^n(x) - F^n(x')| > \Delta\}, \tag{17}$$

$$\tau_-^\Delta(x, x') = \max\{n \leq 0 \,:\, |F^n(x) - F^n(x')| > \Delta\}, \tag{18}$$

*with the convention that $\tau_+^\Delta = \infty$ and/or $\tau_-^\Delta = -\infty$ if there are no such integers n. The dynamical system $F : \Lambda \to \Lambda$ has the **weak orbit separation property** if there exist constants $\Delta > 0$ (a* separation threshold*) and $\alpha > 0$ such that for any two distinct points $x, x' \in \Lambda$, the inequality*

$$|F^n(x) - F^n(x')| > \alpha|x - x'| \tag{19}$$

holds for all integers n satisfying

$$0 \leq n \leq \tau_+^\Delta(x, x') \qquad if \qquad \tau_+^\Delta(x, x') < \infty; \qquad (20)$$

$$0 \geq n \geq \tau_-^\Delta(x, x') \qquad if \qquad \tau_-^\Delta(x, x') > -\infty; \ and \qquad (21)$$

$$-\infty < n < \infty \qquad if \qquad \tau_+^\Delta(x, x') = -\tau_-^\Delta(x, x') = \infty. \qquad (22)$$

Perhaps the simplest nontrivial dynamical systems satisfying the weak orbit separation property are the rotations R_α of the unit circle. It is trivial to verify that the w.o.s.p. holds, because

$$|R_\alpha^n x - R_\alpha^n y| = |x - y| \qquad \forall n \in \mathbb{Z} \text{ and } \forall x, y. \qquad (23)$$

The weak orbit separation property holds not only for highly rigid, non-chaotic systems such as rotations, but also for highly chaotic systems, such as topologically mixing, Axiom A diffeomorphisms restricted to their nonwandering sets. This is not difficult to check. There are other examples that arise naturally, for instance, when an Axiom A system is weakly coupled with an almost periodic system. In particular, the product

$$S \times R : X \times \mathbb{T}^d \longrightarrow X \times \mathbb{T}^d$$

of an Axiom A diffeomorphism $S : X \to X$ restricted to its nonwandering set X with a system $R : \mathbb{T}^d \to \mathbb{T}^d$ that is bi-Lipschitz conjugate to an ergodic rotation of the $d-$torus $\mathbb{T}^d$ satisfies the weak orbit separation property. Finally, the most important dynamical systems satisfying the weak orbit separation property are the time-1 mappings of smooth flows with compact, hyperbolic invariant sets Λ.

Consistent noise removal is possible for dynamical systems satisfying the weak orbit separation property, provided the noise level is sufficiently low. A consistent filter is easily described, although it is not so easily implemented as Smoothing Algorithm D above.

Smoothing Algorithm W: The filter is defined by averaging, as in Smoothing Algorithm D, but the selection of indices over which to average is now done differently. Let κ_m be a sequence of integers satisfying the conditions (12), and, for each $1 \leq n \leq m$, let A_n again be the set of indices $\nu \in \{0, 1, \ldots, m\}$ for which inequality (13) is satisfied. Define B_n to be the subset of A_n consisting of those $\lceil |A_n| / \log |A_n| \rceil$ indices ν for which the residual sums of squares

$$SS(\nu, n; m) = \sum_{|j| \leq \kappa_m} |y_{n+j} - y_{\nu+j}|^2 \qquad (24)$$

are the smallest. Now define

$$\hat{x}_n = \frac{1}{|B_n|} \sum_{\nu \in B_n} y_\nu. \qquad (25)$$

Theorem 3. *Suppose that the dynamical system $F : \Lambda \to \Lambda$ satisfies the weak orbit separation property, with separation threshold Δ. If the errors e_n have mean zero and are uniformly bounded by $\delta < \Delta/5$, then Smoothing Algorithm W is weakly consistent.*

Explanation of Theorem 3. A complete proof is given in [7]; what follows is the skeleton of the argument. For dynamical systems that satisfy the weak orbit separation property, orbits of nearby points need not diverge to a fixed distance Δ, and

so neighboring points cannot be identified in the same simple manner as in the case of dynamical systems with sensitive dependence on initial conditions. In particular, it is no longer necessarily the case that $\nu \in A_n$ (that is, $|y_{n+j} - y_{\nu+j}| < 3\delta$ for all $|j| < \kappa_m$) will guarantee that $|x_n - x_\nu|$ is small, even for large m. However, the weak orbit separation property *does* imply that if two orbits fail to diverge in time κ_m (either forwards in time or backwards in time), and if κ_m is sufficiently large, then

$$|x_{n+j} - x_{\nu+j}| > \alpha|x_n - x_\nu| \tag{26}$$

either for all $0 \leq j \leq \kappa_m$, or for all $-\kappa_m \leq j \leq 0$, or for all $-\kappa_m \leq j \leq \kappa_m$.

Now consider $SS(n, \nu; m)$: since the random vectors e_n have mean zero, with high probability,

$$\begin{aligned} SS(n, \nu; m) &= \sum_{|j| \leq \kappa_m} |y_{n+j} - y_{\nu+j}|^2 \\ &\approx \sum_{|j| \leq \kappa_m} |x_{n+j} - x_{\nu+j}|^2 + \sum_{|j| \leq \kappa_m} |e_{n+j} - e_{\nu+j}|^2 \\ &\approx \sum_{|j| \leq \kappa_m} |x_{n+j} - x_{\nu+j}|^2 + 4\kappa_m E|e_0|^2. \end{aligned} \tag{27}$$

Thus, by (26), $SS(\nu, n; m)$ is, with high probability, considerably smaller for those indices ν such that $|x_n - x_\nu|$ is small. Selection of indices according to the values of $SS(\nu, n; m)$, will, therefore, tend to identify those ν such that x_ν is near x_n. Averaging over these indices will, with high probability, yield an estimate close to x_n, by an argument like that used in the proof of Theorem 2.

6 Concluding Remarks

(1) Much of the published work on the signal separation problem (and, indeed, most work on statistical inference for chaotic dynamical systems) makes no distinction between discrete-time systems and continuous-time systems. However, the results above suggest that there may, in fact, be a significant difference, at least for the signal separation problem. This is certainly the case for hyperbolic systems: discrete-time hyperbolic systems have the sensitive dependence property, but continuous-time systems do not — they satisfy only the weak orbit separation property.

(2) The case of *hyperbolic flows* deserves further attention. It may be shown that certain large classes of hyperbolic flows — including (a) mixing geodesic flows on compact, negatively curved manifolds, and (b) ergodic suspensions of hyperbolic toral automorphisms — admit homoclinic pairs. However, it may also be shown that for such flows homoclinic pairs are *rare*, in the sense that the set of points x that belong to such pairs has SRB-measure 0. Thus, it may be possible to construct weakly consistent filters, or filters which, although not weakly consistent in the sense of Definition 2, neverthless satisfy the consistency relation (2) for *almost every* orbit.

(3) Practical aspects of the signal separation problem have not been systematically studied. Various authors have investigated the efficacy of various filtering schemes for one or two low-dimensional systems, but no comparative studies have

been made of the relative merits of these schemes. Perhaps somewhere an enterprising graduate student will find this to be a worthwhile project.

References

[1] H. Abarbanel. *Analysis of observed chaotic data*. Springer-Verlag, 1996.

[2] M. Davies. Noise reduction schemes for chaotic time series. *Phys. D*, 79:174–192, 1992.

[3] R. Devaney. *An Introduction to Chaotic Dynamical Systems*. Benjamin and Cummings, 1986.

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

[5] E. Kostelich and J. Yorke. The simplest dynamical system consistent with the data. *Physica D*, 41:183–196, 1990.

[6] S. Lalley. Beneath the noise, chaos. *Annals of Statistics*, 27, 1999.

[7] S. Lalley. More noise, more chaos. 1999.

[8] T. Sauer. A noise reduction method for signals from nonlinear systems. *Physica D*, 58:193–201, 1992.

[9] T. Sauer, J. Yorke, and M. Casdagli. Embedology. *J. Statistical Phys.*, 65:579–616, 1991.

[10] S. Smale. Differentiable dynamical systems. *Bull. Amer. Math. Soc.*, 73:747–817, 1967.