Geometric noise reduction for multivariate time series

M. Eugenia Mera and Manuel Morán

Departamento de Análisis Económico I

Universidad Complutense, 28223 Madrid. Spain

e-mail: mmoranca@ccee.ucm.es

Abstract

We propose a noise reduction algorithm, based on a maximum likelihood criterion, for chaotic multivariate time series corrupted by observational noise. We also propose a noise reduction measure based on the mean distance of the points of the cleaned time series to the attractor. We give evidence of the convergence of the empirical measure of the cleaned time series to the underlying invariant measure, which means the possibility of recovering the long run behavior of the

1

true dynamics.

The noise reduction problem arises when a sequence of states (time series) of a system governed by a deterministic law is recorded using a measurement process subject to error (measurement noise). Many algorithms have been proposed in order to minimize the loss, due to noise, of information about the behavior of the system.

In this paper we consider noise reduction for multivariate time series. This case is relevant to laboratory experiments or real world processes in which the state variables of a multivariate dynamical system can be measured through time. We propose a noise reduction algorithm based, as are many existing algorithms, on best local linear fits for the unknown smooth dynamics. However, while existing algorithms use a least squares approach, ours is based on the statistical theory of measurement error models (regression models wherein, as happens in noise reduction problems, both dependent and independent variables are measured with error).

The other main point of our approach is that it aims to recover the long run statistical regularity of the underlying dynamics, rather than to separate the noise and the true signal, as do the existing algorithms. Our results seem to indicate the existence

of intrinsic bounds for the exact pointwise prediction of the true dynamics although the exact prediction of the long run behavior of the underlying dynamics could be possible, at least in the low noise limit.

We give empirical evidence of the efficiency of our algorithm in the cleaning of Hénon and Lorenz dynamics corrupted by noise of low and high amplitudes, and for time series ranging from 500 to 50000 data points. In the case of short time series, in terms of distance to the attractor, up to a 80% noise reduction is achieved. In the case of larger data sets, greater reductions (up to a 95%) are possible. This allows us to recover fine details of the geometric structure of the attractor. We prove in the Appendix that our algorithm together with some of the most widely used algorithms can be understood in a common framework: all of them are based on orthogonal projections, with respect to some metric, onto optimal linear subspaces.

1 Introduction

Many problems related to noise reduction can be described by the equations

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k, \xi_k),$$

$$\mathbf{X}_k = g(\mathbf{x}_k) + \mathbf{e}_k,$$

where $\mathbf{x}_k \in \mathbb{R}^d$ is the unobservable state vector of a system at period k; f is a smooth deterministic dynamical law; $g : \mathbb{R}^d \to \mathbb{R}^p$ is a smooth vectorial function or *observable;* ξ_k, $\mathbf{e}_k$ are multivariate random variables; and $\mathbf{X}_k$, $k = 1, 2, ..., N$, is the available data set or time series. In this scheme ξ_k is referred to as the *dynamical noise*, and $\mathbf{e}_k$ as the *measurement noise.*

There is a rich literature on noise reduction algorithms (see for instance the reviews of Grassberger et al[1], Kostelich and Schreiber[2] and Davies[3]). Most such studies address the case in which there exists only measurement noise and the time series is a scalar one.[1],[4],[5],[6],[7] These algorithms reconstruct the scalar signal as an m-dimensional time series $\mathbf{X}_k^m = (X_k, X_{k+1}, ..., X_{k+m-1})$, $k = 1, 2, ..., N-m+1$. Takens's theorem guarantees[8] that, if $m \geq 2d + 1$, in absence of noise, and for generic observable g, the reconstructed time series $\mathbf{X}_k^m$, $k = 1, 2, ..., N - m + 1$, provides a diffeomorphic image of the chaotic attractor associated with f. Then the guess for the clean

data underlying the noisy time series is the projection $\widetilde{\mathbf{X}_k^m}$ of each $\mathbf{X}_k^m$ onto the d-dimensional linear subspace which best fits, in a least squares sense, the data in a small neighborhood of $\mathbf{X}_k^m$. The algorithms then iterate this procedure, starting from the projections $\widetilde{\mathbf{X}_k^m}, k = 1, 2, ..., N - m + 1$.

A further specialization of the general problem which has also been studied is the case in which the evolution law f is known (or at least there is known a clean time series)[9],[10],[11] and one must obtain the most likely trajectory $\mathbf{x}_k$, $k = 1, 2, ..., N$, from the data set.

The dynamical noise problem is relevant if no deterministic evolution law can fully explain the time evolution of the observed system. The interested reader can find information on this issue and on the related shadowing theorems in Refs.[12],[13],[14]

In this paper we restrict our attention to the noise reduction problem in the case in which there does not exist dynamical noise, f is unknown, g is the identity map, $\mathbf{x}_k$ and $\mathbf{X}_k$ are multivariate vectors, and the measurement noise $\mathbf{e}_k$ is an $i.i.d.$ stochastic process, i.e.

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k),$$

$$\mathbf{X}_k = \mathbf{x}_k + \mathbf{e}_k, \;\; k = 1, ..., N.$$

There exist only two algorithms addressed to[15] or adaptable to[4],[16]

multivariate time series. These algorithms extend in a natural way the method described above, i.e, they compute best linear local estimates, in a least squares sense, of the empirical one-time map[4],[16] or they use orthogonal local projections of data points embedded in higher dimensional spaces[15] (total least squares method).

The two main features of our algorithm are the following:

1) We use unbiased local linear estimates of the dynamics (see equation (1)). It is known that, in linear models wherein the independent variables are also measured with error, the least squares criterion used in the algorithms discussed above does not give unbiased estimators, the bias being larger for increasing variances or seriously correlated errors.

The theory that treats the models wherein both dependent and independent variables are measured with error is that of measurement error models[17]. We incorporate this theory in the design of our algorithm, taking advantage of the structure of the covariance matrix of the errors in order to obtain a new time series having almost the same statistical and geometric properties as the true dynamics, even in the case of high noise amplitudes or when the variances of the components of the error are different and/or correlated. This may be applied in experimental settings in order to exploit

the information available a priori on the specific degree of uncertainty of each state coordinate, as would happen if each were measured by a different device. It also may exploit the information on the correlation between errors of different coordinates as would be the case if, for instance, one of these coordinates were the rate of change of another. In Section 3 we empirically test the efficiency of the total least squares method against that of measurement error models.

2) The objective of our algorithm is a geometric noise reduction.

We say that the cleaned time series has the same geometric properties as the true dynamics (geometric cleaning) when it converges, in the Hausdorff metric, to the true time series (or to the attractor). We present empirical evidence (see Section 3) of the weak convergence of the empirical measure of the cleaned time series (the probability measure that gives the same weight to each point of the time series) to the underlying invariant measure. In this sense both clean and cleaned time series have the same statistical properties; they generate the same empirical measure and therefore they assign the same probability to each ball of the phase space. Notice that in chaotic dynamics high accuracy can be attained in the determination of the probability of a set of states although there exists an intrinsic limit to the prediction of the

exact state of the system. Thus our goal is to obtain a new time series close to the true time series in a geometric and statistical sense rather than in the sense of pointwise convergence between the two time series.

Our starting point is the algorithm of Kostelich and Yorke[4]. These authors have already pointed out the unsuitability of the least squares method for computing the best linear maps, and they suggested the use of the theory of measurement error models. Jaeger and Kantz[18] have also proposed the use of the theory of measurement error models in the problem of estimation of the dynamics underlying a noisy time series. This is an area closely related to noise reduction, in which the bad behavior of the least squares estimator and the development of techniques based on maximum likelihood criteria are well established[19],[20],[21],[22].

The paper is organized as follows. In Section 2 we describe our algorithm and a new measure of noise reduction, which is intended to measure the geometric and statistical proximity of two time series. We also recall from the literature the standard measures of noise reduction based on pointwise and dynamic proximity. Section 3 is devoted to the results obtained by applying our algorithm to noisy multivariate time series from the Hénon map and the Lorenz dynamics. In the Appendix we show that our method

9

can be regarded as a local linear projection onto an optimal subspace, with respect to a special metric that incorporates the information contained in the error covariance matrix. We also show that the various local projection methods used in the literature can be understood as local projections with respect to suitable metrics.

2 The algorithm and noise reduction measures

We now define more precisely the elements of the problem. Our hypothesis is that the observed multivariate time series $\{\mathbf{X}_i, i = 1, ..., N\} \subset \mathbb{R}^d$ is obtained by adding a noise component $\mathbf{e}_i \in \mathbb{R}^d$ to a deterministic signal $\mathbf{x}_i$, that is, $\mathbf{X}_i = \mathbf{x}_i + \mathbf{e}_i$, where $\mathbf{x}_{i+1} = f(\mathbf{x}_i)$, $1 \leq i \leq N - 1$, and f is a smooth evolution law. We assume that the $\mathbf{e}_i$ are independently Gaussian distributed with null mean.

Our algorithm starts by taking a neighborhood $\mathcal{U}_i$ for each point $\mathbf{X}_i$ and it uses the theory of measurement error models to obtain the matrix $\mathbf{A}_i$ such that

$$\mathbf{X}_{j+1} - \langle \mathbf{X}_{i+1} \rangle_{\mathcal{U}_i} \approx \left(\mathbf{X}_j - \langle \mathbf{X}_i \rangle_{\mathcal{U}_i} \right) \mathbf{A}_i, \ \mathbf{X}_j \in \mathcal{U}_i \tag{1}$$

where $\langle \mathbf{X}_i \rangle_{\mathcal{U}_i}$ denotes the center of mass of the points of the neighborhood $\mathcal{U}_i$ and $\langle \mathbf{X}_{i+1} \rangle_{\mathcal{U}_i}$ is the center of mass of their images. This theory (i) guarantees that $\mathbf{A}_i$ is the unbiased maximum likelihood estimate for the matrix of the underlying linear model, and (ii) gives, for each $\mathbf{X}_j \in \mathcal{U}_i$, the maximum likelihood estimates $\widehat{\mathbf{x}}_{j+1}$ and $\widehat{\mathbf{x}}_j$ for $\mathbf{x}_{j+1}$ and $\mathbf{x}_j$ respectively.

2.1 The algorithm

Let Σ be the empirical covariance matrix of the error terms of the $2d$-dimensional time series $\{(\mathbf{X}_{i+1}, \mathbf{X}_i)\}_{i=1,\dots,N-1}$. We assume first that Σ is known. If it were unknown we would take $\Sigma = \sigma^2 \mathbf{I}$, for arbitrary σ, at the first iteration.

The steps of the algorithm are the following:

1) For each point $\mathbf{X}_i$ of the noisy time series, construct a neighborhood $\mathcal{U}_i$ consisting of the NV points of the noisy time series closest to $\mathbf{X}_i$.

2) Compute the center of mass,

$$\langle \mathbf{X}_i \rangle_{\mathcal{U}_i} := \frac{1}{NV} \sum_{j:\mathbf{X}_j \in \mathcal{U}_i} \mathbf{X}_j,$$

of the points in $\mathcal{U}_i$ and the center of mass,

$$\langle \mathbf{X}_{i+1} \rangle_{\mathcal{U}_i} := \frac{1}{NV} \sum_{j:\mathbf{X}_j \in \mathcal{U}_i} \mathbf{X}_{j+1},$$

11

of their images.

3) Compute the $2d \times 2d$ matrix

$$M_{\mathbf{ZZ}} := \frac{1}{NV} \sum_{j:\mathbf{X}_j \in \mathcal{U}_i} \mathbf{Z}_j^t \mathbf{Z}_j$$

where $\mathbf{Z}_j \in \mathbb{R}^{2d}$ is defined as

$$\mathbf{Z}_j := (\mathbf{X}_{j+1} - \langle \mathbf{X}_{i+1} \rangle_{\mathcal{U}_i}, \mathbf{X}_j - \langle \mathbf{X}_i \rangle_{\mathcal{U}_i})$$

4) Compute the eigenvalues and an orthonormal basis of eigenvectors of the matrix $M_{\mathbf{ZZ}}$ in the metric of Σ.

A set of vectors $\{\mathbf{w}_1, ..., \mathbf{w}_{2d}\} \subset \mathbb{R}^{2d}$ is called *orthonormal basis of eigenvectors of $M_{\mathbf{ZZ}}$ in the metric of Σ* if they satisfy i) there exist real numbers $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{2d}$ such that $M_{\mathbf{ZZ}}\mathbf{w}_i = \lambda_i \Sigma \mathbf{w}_i, 1 \leq i \leq 2d$, and ii) $\mathbf{w}_i^t \Sigma \mathbf{w}_j = \delta_{ij}$, for all i, j. The scalars λ_i are called eigenvalues of $M_{\mathbf{ZZ}}$ in the metric of Σ. Such basis of eigenvectors is given by the columns of the matrix $Q\Lambda^{-1/2}H$, where the columns of Q are a basis of orthonormal eigenvectors of Σ (in the Euclidean metric), Λ is a diagonal matrix of eigenvalues of Σ, and the columns of H are an orthonormal basis of eigenvectors of the matrix $\Lambda^{-1/2}Q^t M_{\mathbf{ZZ}} Q\Lambda^{-1/2}$ (see details in Ref.[17], page 391).

5) The estimate for $\mathbf{z}_j := (\mathbf{x}_{j+1} - \langle \mathbf{x}_{i+1} \rangle_{\mathcal{U}_i}, \mathbf{x}_j - \langle \mathbf{x}_i \rangle_{\mathcal{U}_i})$ is

$$\widehat{\mathbf{z}}_j = (\mathrm{I} - \Sigma BB^t)\mathbf{Z}_j^t \tag{2}$$

12

where B is the matrix which has as columns the eigenvectors $\{\mathbf{w}_{d+1}, ..., \mathbf{w}_{2d}\}$ of $M_{\mathbf{ZZ}}$, in the metric of Σ, corresponding to the d smallest eigenvalues. In Theorem 4.1.1 of Ref.[17] it is proved that $\widehat{\mathbf{z}}_j$ is an unbiased maximum likelihood estimate of $\mathbf{z}_j$ under the hypotheses of linearity of the model and of Gaussian errors.

6) Take as estimate of $(\mathbf{x}_{i+1}, \mathbf{x}_i)$ the values

$$(\widehat{\mathbf{x}_{i+1}, \mathbf{x}_i}) := (\langle \mathbf{X}_{i+1} \rangle_{\mathcal{U}_i}, \langle \mathbf{X}_i \rangle_{\mathcal{U}_i}) + \widehat{\mathbf{z}}_i \tag{3}$$

7) For the first and the last point of the time series, we have a single estimate of the deterministic part. For each of the remaining points $\mathbf{X}_i$, $i \neq 1, N$, there are the two estimates obtained from the neighborhoods $\mathcal{U}_{i-1}$ and $\mathcal{U}_i$ respectively. Since we have empirical evidence that the estimate corresponding to the neighborhood $\mathcal{U}_{i-1}$ works better we take it as the estimate of $\mathbf{x}_i$.

8) Take as the estimate of Σ for iteration $k+1$ the empirical covariance matrix of the estimated errors at iteration k. The estimated errors at iteration k are defined by $\widehat{\mathbf{u}}_i := (\widehat{\mathbf{x}}_{i+1}^k - \widehat{\mathbf{x}}_{i+1}^{k-1}, \widehat{\mathbf{x}}_i^k - \widehat{\mathbf{x}}_i^{k-1})$ where $\widehat{\mathbf{x}}_i^k$ is the estimate of the i-th point of the time series at the iteration k and $\widehat{\mathbf{x}}_i^0 = \mathbf{X}_i$.

9) With $\{\widehat{\mathbf{x}}_i\}$ $i = 1, ..., N$, as the new data set and Σ as in step 8, we iterate the procedure until the results cease to improve.

Remark 1 *If the error terms are uncorrelated and have the same known or unknown variance σ^2, that is $\Sigma = \sigma^2 I$, then the matrix B above has as columns $\{\sigma^{-1}\mathbf{v}_j, j = d+1, ..., 2d\}$ where $\mathbf{v}_j, j = 1, ..., 2d$, are an orthonormal basis, respect to the Euclidean metric, of eigenvectors of the matrix $M_{\mathbf{ZZ}}$. Then $\widehat{\mathbf{z}}_j$ above is the orthogonal projection of $\mathbf{Z}_j$ onto the d-dimensional linear subspace which minimizes the Euclidean distance to the data (see Remark below). Thus, at the first iteration, the core of our method essentially coincides with the algorithm proposed independently by Cawley and Hsu[5] and Sauer[23] for scalar time series. However, the algorithms differ at later iterations.*

Remark 2 *In the Appendix we prove that $\widehat{\mathbf{z}}_j$ in equation (2) is the projection of $\mathbf{Z}_j$ onto the best d-dimensional linear subspace T_d with respect to the metric induced by the inverse of the covariance matrix of the errors. That is, T_d is the subspace for which the minimum in T of*

$$\sum_{j: \mathbf{X}_j \in \mathcal{U}_i} (\mathbf{Z}_j - P_T \mathbf{Z}_j) \Sigma^{-1} (\mathbf{Z}_j - P_T \mathbf{Z}_j)^t$$

is attained, where $P_T \mathbf{Z}_j := \arg \min_{\mathbf{y} \in T} (\mathbf{Z}_j - \mathbf{y}) \Sigma^{-1} (\mathbf{Z}_j - \mathbf{y})^t$ is the orthogonal projection, with respect to the metric induced by Σ^{-1}, of $\mathbf{Z}_j$ onto T. Notice that this shows the relation between our method and those proposed by Grassberger

et al[1] and Hegger and Schreiber[15]. The difference stems from the choice of metric. These authors choose, instead of Σ, a matrix P which gives almost all the weight to the central coordinates and practically null weight to the remaining coordinates of $\mathbf{Z}_j$.

2.1.1 Further details of the algorithm

In this section we mention some further details of the algorithm that we have found empirically efficient.

1. **Increased statistical consistency in the estimations**

   Notice that by equations (2) and (3) we have, in addition to an estimate of $\mathbf{x}_{i+1}$, estimates of the remaining points belonging to the neighborhood $\mathcal{U}_i$. We store such estimates for the $Nprom$ points in $\mathcal{U}_i$ closest to the base point $\mathbf{X}_i$, and we use them in order to give more robust averaged estimates of each point in the cleaned time series.

2. **Effects of nonlinearity**

   We have obtained better results working with neighborhoods containing larger numbers of points, especially for time series with high noise amplitudes, but this procedure causes a little distortion due to nonlin-

ear effects. Such distortion can be reduced by replacing the center of mass of the neighborhoods $\mathcal{U}_i$ with the center of mass of a small number $NCT << NV$ of closest neighbors of $\mathbf{X}_i$. In all the results of the next section we choose $NCT = 25$ for large data sets and $NCT = 5$ for time series with $N \leq 1000$.

Sauer[23] was the first to point out that, due to the presence of small nonlinearities, locally linear and projective methods introduce errors which are systematic and make the mean of corrections nonzero. He proposed to force the corrections of all the points of the neighborhood to have zero mean in order to compensate for the nonlinearity effect. This procedure was also recommended by Grassberger et al[1] and Kostelich and Schreiber[2]. However, we have not found significant improvement using this device.

3. Badly conditioned linear fits

Linear fits may be badly conditioned when the variance of the independent variables is small in comparison to that of the noise. This may cause the algorithm to make anomalously large corrections; a problem that some authors[1],[4],[5] solve through controls that anchor the pre-

dictions to the data. We have observed badly conditioned linear fits only occasionally, when we have worked with very long time series and we have taken a small number of points in the neighborhoods. This problem becomes more relevant for low noise amplitudes, because for high amplitudes we always take large neighborhoods in order to see the local geometry of the deterministic part of the time series. We solve the problem by taking in each neighborhood a minimum number of points depending on the length, N, of the time series ($1\%N$, for instance), instead of introducing controls for avoiding anomalous predictions as do other authors.

2.2 Noise reduction measures

In this section we introduce some measures of noise reduction. Some of them require knowledge of the deterministic signal or the functional expression of the dynamics, and some others estimate how much noise is taken out by using only the noisy time series. The first two measures (see 1 and 2 below) are commonly used in noise reduction literature [1],[3],[4],[5],[6],[15] and they aim to measure the pointwise proximity of the true and the adjusted time series or the dynamic self consistency of the adjusted time series. We use them for

17

testing purposes. The goal of our algorithm is to reduce the noise reduction measure 3 below, which estimates the geometric and statistical proximity. This measure of noise reduction always improves as the number of iterations increases (we stop the process when this improvement is not relevant). In the case of other noise reduction measures the optimal reduction is attained during the early iterations of the algorithm and in the following iterations becomes poorer. In the results in Section 3 we give the values and the iteration at which these measures of reduction give the best results.

Recall that the noisy time series is denoted by $\{\mathbf{X}_i : i = 1, ..., N\}$, the clean time series is $\{\mathbf{x}_i : i = 1, ..., N\}$, the cleaned time series at the iteration k is $\{\widehat{\mathbf{x}}_i^k : i = 1, ..., N\}$, and the evolution law is denoted by f.

1. **Measure of the pointwise closeness**

   The *rms* distance between the time series $\mathbf{x}$ and $\widehat{\mathbf{x}}^k$ is defined as

   $$E(\mathbf{x}, \widehat{\mathbf{x}}^k) := \left(\frac{1}{N} \sum_{i=1}^{N} \left\| \widehat{\mathbf{x}}_i^k - \mathbf{x}_i \right\|^2 \right)^{1/2}.$$

   If $E(\mathbf{x}, \widehat{\mathbf{x}}^k) < E(\mathbf{x}, \mathbf{X})$ then the noise has been reduced. The parameter that gives the percentage of reduction is

   $$R = 100 \left(1 - \frac{E(\mathbf{x}, \widehat{\mathbf{x}}^k)}{E(\mathbf{x}, \mathbf{X})} \right).$$

18

2. Measure of dynamic closeness

We define two measures, E_{dyn} and $\widehat{E}_{dyn}$, of dynamic proximity: The definition of the first one assumes that f is known, whereas in the definition of the second one f is supposed unknown and is replaced by local estimates $\widehat{f}_i$ at $\mathbf{x}_i$. E_{dyn} and $\widehat{E}_{dyn}$ measure the deviation from the deterministic behavior of a time series and are defined as

$$E_{dyn}(\widehat{\mathbf{x}}^k) := \left(\frac{1}{N-1} \sum_{i=1}^{N-1} \left\| \widehat{\mathbf{x}}_{i+1}^k - f(\widehat{\mathbf{x}}_i^k) \right\|^2 \right)^{1/2} \quad \text{and}$$

$$\widehat{E}_{dyn}(\widehat{\mathbf{x}}^k) := \left(\frac{1}{N-1} \sum_{i=1}^{N-1} \left\| \widehat{\mathbf{x}}_{i+1}^k - \widehat{f}_i(\widehat{\mathbf{x}}_i^k) \right\|^2 \right)^{1/2} .$$

If $E_{dyn}(\widehat{\mathbf{x}}^k) < E_{dyn}(\mathbf{X})$ (or $\widehat{E}_{dyn}(\widehat{\mathbf{x}}^k) < \widehat{E}_{dyn}(\mathbf{X})$) then the adjusted time series is more consistent with the dynamics than the original noisy time series and the noise has been reduced. The parameters which give the percentage of reduction are

$$R_{dyn} := 100 \left(1 - \frac{E_{dyn}(\widehat{\mathbf{x}}^k)}{E_{dyn}(\mathbf{X})} \right) \quad \text{and} \quad \widehat{R}_{dyn} := 100 \left(1 - \frac{\widehat{E}_{dyn}(\widehat{\mathbf{x}}^k)}{\widehat{E}_{dyn}(\mathbf{X})} \right) .$$

3. Measures of geometric and statistical closeness

We have empirically observed that the parameter R defined above can be small for a time series having geometric and statistical properties very similar to the original deterministic time series. Think for example

of a situation in which the adjusted time series is very close to the attractor (we have global convergence) without there having pointwise convergence to the original deterministic time series (see for example the experiment in $Figs.1$ below). The main measure of noise reduction we propose is based on the Hausdorff distance between time series. We define the mean distance between the cleaned time series $\{\widehat{\mathbf{x}}_i^k\}$ and the deterministic time series $\{\mathbf{x}_i\}$ as

$$d_m(\widehat{\mathbf{x}}^k, \mathbf{x}) := \frac{1}{N} \sum_{j=1}^{N} \min_{i=1,\dots,N} \left\| \widehat{\mathbf{x}}_j^k - \mathbf{x}_i \right\|.$$

Notice that, for long data sets, $d_m(\widehat{\mathbf{x}}^k, \mathbf{x})$ provides the mean distance of $\{\widehat{\mathbf{x}}_i^k\}$ to the attractor. The noise reduction $\langle R_h \rangle$ is then

$$\langle R_h \rangle := 100 \left(1 - \frac{d_m(\widehat{\mathbf{x}}^k, \mathbf{x})}{d_m(\mathbf{X}, \mathbf{x})} \right).$$

We also consider the Hausdorff distance between time series

$$d_h(\widehat{\mathbf{x}}^k, \mathbf{x}) := \max_{j=1,\dots,N} \min_{i=1,\dots,N} \left\| \widehat{\mathbf{x}}_j^k - \mathbf{x}_i \right\|,$$

and the corresponding noise reduction measure

$$R_h := 100 \left(1 - \frac{d(\widehat{\mathbf{x}}^k, \mathbf{x})}{d(\mathbf{X}, \mathbf{x})} \right).$$

By the definition of d_h, poor performance of the algorithm at even a single point may cause global inefficiency in the noise reduction level

based on R_h. Conversely, a high level of noise reduction in an R_h sense gives strong evidence of the convergence of the support of the empirical measure associated to the cleaned time series to that of the underlying invariant measure. These measures of noise reduction require the use of some technique of fast neighbor search such as the box-assisted method[24] in order to make the algorithm more efficient. We also use such a technique in the construction of the neighborhoods.

3 Simulation results

In this section we show the results of our algorithm for time series generated by the Hénon and the Lorenz dynamics corrupted by noise. The Hénon map is given by the equations

$$x(k+1) = 1 - ax(k)^2 + y(k)$$

$$y(k+1) = bx(k)$$

and we use the parameter values $a = 1.4$ and $b = 0.3$. The Lorenz dynamic is defined by

$$x' = \sigma(y - x)$$

$$y' = x(R - z) - y$$

$$z' = xy - bz$$

and the parameter values we use are $\sigma = 16$, $R = 45.92$ and $b = 4$. The equations were integrated using a fourth order Runge-Kutta algorithm with an integration step of $\Delta t = 0.001$, and the data were recorded with a sampling time $\tau = 0.03$.

We consider uncorrelated errors with equal and unequal variances. We assume that we do not know the covariance matrix of the errors, so it is estimated from the data. We give the results for a simplified version of our algorithm, that takes for all points of the time series and all iterations of the algorithm the same number, NV, of points in each neighborhood. This simplified version of our algorithm is very easy to use because it leaves as parameters to be chosen only NV and $Nprom$ (the number $Itera$ of iterations of the algorithm is the value for which the measure of geometric noise reduction ceases to improve significantly). Recall that the measures of reduction

are denoted by $R, R_{dyn}, \widehat{R}_{dyn}, R_h$ and $\langle R_h \rangle$, for the pointwise distance, the dynamical distance, the estimated dynamical distance, the Hausdorff distance, and the mean distance respectively. We denote by σ_x, σ_y, σ_{ε_x} and σ_{ε_y} the standard deviation of the components of the clean signal and of the errors. We denote by σ_S and σ_ε the standard deviation of the multivariate clean signal and error respectively.

In $Fig.$ 1 (a) we see a clean time series (x_i, y_i) of 10000 points generated by the Hénon map, and in (b) the noisy time series obtained by adding to the clean time series an uncorrelated Gaussian noise component with $\sigma_{\varepsilon_x} = 1\%\sigma_S$ and $\sigma_{\varepsilon_y} = 1\%\sigma_S$. In (c) we have plotted the clean time series together with the cleaned time series. It can be seen that the Hausdorff distance between the two time series is very small $(d_h = 0.0101)$. The numerical results for the degree of reduction are $R = 40\%$ (at iteration 3), $R_{dyn} = 51\%, \widehat{R}_{dyn} = 88\%,$ $R_h = 61\%$ and $\langle R_h \rangle = 84\%$.

The experiment above shows that the noise reduction is significantly bigger for $\langle R_h \rangle$ than for the standard noise reduction measure R. Moreover, $\langle R_h \rangle$ always decreases when the number of iterations increases, although R does not. Recall that if the covariance matrix of the errors is a multiple of the identity matrix, our method is equivalent to the total least squares method,

23

so the discrepancy between the values R and $\langle R_h \rangle$ is not due to the choice of the metric.

In $Fig.$ 2 we have plotted a noisy time series of 10000 points generated by the Hénon map, with a highly heteroskedastic Gaussian noise. We have plotted the cleaned time series that our algorithm gives as output together with the noisy time series. The levels of reduction are $R = 83\%$ (at iteration 4), $R_{dyn} = 80\%, \widehat{R}_{dyn} = 97\%, R_h = 95\%$ and $\langle R_h \rangle = 94\%$.

We have designed an algorithm based on the total least squares criterion (TLSA) in order to compare its efficiency with that of our algorithm based on the maximum likelihood criterion (MLA). TLSA differs from MLA in that it uses local orthogonal projections with respect to the Euclidean metric rather than with respect to the metric induced by the inverse of the covariance matrix of the errors. We show in $Table\ I$ the results of the two algorithms, when applied to time series generated by Hénon dynamics corrupted with homoskedastic noises. Since in the homoskedastic case the two methods are essentially equivalent, the levels of noise reduction, in a $\langle R_h \rangle$ sense, attained by the two algorithms are similar and high. The slightly better performance of our method is due to the fact that the errors resulting from the second iteration are not homoskedastic. Compare also the high values of noise re-

duction in a R_h sense attained by MLA to the smaller ones attained by $TLSA$.

In the heteroskedastic case (see $Table\ II$) there is a significant difference between the levels of noisy reduction attained by each algorithm. We present two experiments in which $\sigma_\varepsilon = 11.2\%\ \sigma_s$. The first experiment corresponds to a low level of heteroskedasticity and the second one to a higher level. In the first case each algorithm gives a large noise reduction, MLA about 12% larger than $TLSA$, and in the second case the noise reduction level is smaller, but now for MLA is about 25% larger than for $TLSA$. These experiments also show the high efficiency of our algorithm for short data sets.

We have plotted in $Fig.$ 3 the output of the MLA and $TLSA$ algorithms for a noisy time series of 500 data points generated by the same dynamics as that shown in $Fig.$ 2. The difference between the two methods for a short time series can be seen. The levels of reduction are $\langle R_h \rangle = 80\%$ and $R_h = 78\%$ for MLA and $\langle R_h \rangle = 63\%$ and $R_h = 40\%$ for $TLSA$. Also noteworthy is the high number of iterations, $Itera = 45$ in the experiment above, at which the algorithms cease to improve for short time series (see also $Table\ II$).

In $Table\ III$ we present some experiments for noise Lorenz time series. The results are similar to those obtained for the Hénon dynamics. Such re-

sults and those given for the Hénon dynamics in $Table\ II$ give an indication of the convergence of the output of the algorithm when the number of data points increases. Better results would require the use of adaptive neighborhoods (other authors that have proposed adaptive neighborhoods are Kern et al[28]).

We have verified, using the Jarque-Bera test of normality and the histograms of frequencies, that if the noise component is Gaussian, the distribution of the errors in the succeeding iterations of the algorithm are also Gaussian, so the theory can be applied again with unbiased maximum likelihood estimators. In the case of non-Gaussian errors the theory ensures strong consistency of the estimators. This fact is confirmed by our numerical results for noisy time series where the noise has uniform distribution. We have obtained a value $\langle R_h \rangle = 86\%$ for a 10000 data point time series generated by a Lorenz dynamics corrupted by a uniform noise with $\sigma_{\varepsilon_x} = 20\%\sigma_x, \sigma_{\varepsilon_y} = 10\%\sigma_y$ and $\sigma_{\varepsilon_z} = 20\%\sigma_y$.

We have contrasted the efficiency of our algorithm, comparing for the three time series (clean, cleaned and noisy) of the dynamics in $Fig.\ 2$:

(i) the Lyapunov exponents computed with the Eckmann and Ruelle algorithm[25],[26] and

(ii) the degrees of freedom[27] of the 2-embedded (and 4-dimensional) time series, i.e. we compare the degrees of freedom of the time series $(\mathbf{x}_i, \mathbf{x}_{i+1})$, $(\mathbf{X}_i, \mathbf{X}_{i+1})$ and $(\widehat{\mathbf{x}}_i, \widehat{\mathbf{x}}_{i+1}), i = 1, ..., N - 1$.

(i) The true Lyapunov exponents of the Hénon map are $\lambda_1 \sim 0.419$ and $\lambda_2 \sim -1.623$. Since the Eckmann and Ruelle algorithm uses local estimations of the tangent map at the points of the time series it cannot give the exact values of the Lyapunov exponents even if we compute them from the clean time series. The 95% confidence intervals for the values obtained by the algorithm for the true, cleaned and noisy time series, for different orbits and realizations of the error term, can be seen in $Table\ IV$. The parameters of the algorithm are kept fixed at the values that work correctly for the clean time series. The values obtained for the cleaned time series are reasonable estimates of the true values although we can not obtain good estimates of the negative exponent from the noisy time series.

(ii) The graphs in $Fig.$ 4 show the mean proportion of the total variability of the data points in balls of small radius, explained for each one of the principal components of the covariance matrix of such data points. The clean and cleaned time series have only two significantly non null singular values which is an indication of the 2-dimensionality of the true dynamics.

27

However, the noisy time series has the four singular values significantly non null. Observe furthermore that the clean and the cleaned time series have almost the same behavior at all scales with the only exception being at very small scales, which shows that more data points will be needed in order to capture the microstructure of the data and that an improvement in the noise reduction might be possible. This experiment, together with the high levels of noise reduction in an R_h sense, gives a strong indication that not only does the support of the empirical measure, μ_N, of the cleaned time series converges to the support of the invariant measure μ but also the measure μ_N itself converges weakly to the measure μ.

4 Conclusions

We propose an empirically efficient algorithm with a consistent theoretical basis. It allows us to obtain significant noise reductions for both low and high noise amplitudes and for short and long time series, giving as output a time series with geometric and statistical properties very similar to those of the clean time series. Further improvement can be made by introducing adaptive neighborhoods which exploit the geometry of the different parts of the time

series or at least by reducing the number of points of the neighborhoods as the number of iterations increase (and consequently the noise level decreases).

Our algorithm requires the empirical adjustment of only two parameters and it works efficiently without introducing controls for avoiding anomalous predictions as do the existing algorithms.

When compared to the previous work in the area, the results of our research seem to indicate that the uncertainty derived from the noisy corruption of chaotic dynamics puts an intrinsic limit on the short run predictability of the dynamics, whereas the long run behavior could be entirely recovered even for large noise amplitudes. The adequacy of the approach of noise reduction measurement based on the mean distance to the attractor $\langle R_h \rangle$, instead of on the pointwise distance R, is supported by empirical evidence showing a steady improvement of $\langle R_h \rangle$, in contrast to the worsening of R, as the number of iterations of the algorithm increases.

Further attempts, based on local linear fits, to find a reordering of the points of the output of our algorithm which reduces the pointwise distance $E(\mathbf{x}, \widehat{\mathbf{x}}^k)$ have proved useless. We think that this is due to the high degree of dynamical self-coherence indicated by the small residuals in $\widehat{E}_{dyn}(\widehat{\mathbf{x}}^k)$. This conjecture is supported by the accuracy in the estimation through the cleaned

29

time series of the highly sensitive negative Lyapunov exponent. This and the local microstructure of the empirical measure revealed by the test of degrees of freedom also indicates a convergence, in the weak sense, of the empirical measure μ_N of the cleaned time series to the natural invariant measure μ.

Acknowledgment

APPENDIX

In this appendix we prove the assertions of Remark 2.

Given an $n \times n$ symmetric positive definite matrix A we shall denote by δ_A the metric induced by the inner product with matrix A:

$$\delta_A(\mathbf{v}, \mathbf{w}) := \langle \mathbf{v} - \mathbf{w}, \mathbf{v} - \mathbf{w} \rangle_A^{1/2} := \left((\mathbf{v} - \mathbf{w}) A (\mathbf{v} - \mathbf{w})^t \right)^{1/2}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n.$$

Let $\mathcal{L}_p$ be the set of of p-dimensional linear subspaces of $\mathbb{R}^n$. The orthogonal projection of $\mathbf{v} \in \mathbb{R}^n$ onto $T \in \mathcal{L}_p$ in the metric δ_A is

$$\mathrm{P}_T \mathbf{v} := \arg \min_{\mathbf{w} \in T} \delta_A(\mathbf{v}, \mathbf{w}).$$

The δ_A−distance between $\mathbf{v} \in \mathbb{R}^n$ and $T \in \mathcal{L}_p$ is $\delta_A(\mathbf{v}, T) := \delta_A(\mathbf{v}, P_T \mathbf{v})$.

Let M be an $n \times n$ symmetric matrix. Recall that a system of vectors $\{\mathbf{w}_1, ..., \mathbf{w}_n\} \subset \mathbb{R}^n$ is called *orthonormal system of eigenvectors of M in the metric δ_A* if they satisfy: i) there exist real numbers λ_i such that $M\mathbf{w}_i = \lambda_i A\mathbf{w}_i$, $1 \leq i \leq n$, and ii) $\mathbf{w}_i^t A\mathbf{w}_j = \delta_{ij}$, for all i, j. The scalars λ_i are called eigenvalues of M in the metric δ_A.

The next theorem is an adaptation of one given by Cawley and Hsu [5] for the case of the Euclidean metric. The result, applied in the text to the covariance matrix, Σ, of the errors, also holds for an arbitrary $n \times n$ symmetric positive definite matrix A.

Theorem 3 *Let $\{\mathbf{Z}_k, k = 1, ..., NV\} \in \mathbb{R}^n$, and let A be an $n \times n$ symmetric positive definite matrix. Let $\sigma_1 \geq \cdots \geq \sigma_n$ be the eigenvalues of the $n \times n$ matrix $M_{\mathbf{ZZ}} := \frac{1}{NV} \Sigma_{k=1}^{NV} \mathbf{Z}_k^t \mathbf{Z}_k$ in the metric of A, let $\{\mathbf{w}_i, i = 1, ..., n\}$ be the corresponding orthonormal eigenvectors in the metric δ_A and, for $1 \leq p \leq n$, let B be the $n \times (n - p)$ matrix whose columns are the last $n - p$ eigenvectors $\{\mathbf{w}_{p+1}, ..., \mathbf{w}_n\}$. Then:*

i) the linear subspace $\mathcal{T}_p$ in $\mathcal{L}_p$ which minimizes the mean square $\delta_{A^{-1}}$-distance to the points $\{\mathbf{Z}_k, k = 1, ..., NV\}$ is $\mathcal{T}_p := span\{A\mathbf{w}_1^t, ..., A\mathbf{w}_p^t\}$.

ii) $\mathrm{P}_{\mathcal{T}_p}\mathbf{Z} := (\ I - ABB^t)\mathbf{Z}^t$, $\mathbf{Z} \in \mathbb{R}^n$.

Proof. By its definition, $\mathcal{T}_p = \arg\min_{T \in \mathcal{L}_p} \mathcal{E}(T)$, where

$$\mathcal{E}(T) := \frac{1}{NV} \sum_{k=1}^{NV} (\mathbf{Z}_k - P_T \mathbf{Z}_k) A^{-1} (\mathbf{Z}_k - P_T \mathbf{Z}_k)^t$$

Let $T \in \mathcal{L}_p$ and let $\{\mathbf{u}_1, ..., \mathbf{u}_n\}$ be an orthonormal basis of $\mathbb{R}^n$ with respect to the metric $\delta_{A^{-1}}$, such that $T = span\{\mathbf{u}_1, ..., \mathbf{u}_p\}$. From $\mathbf{Z}_k = \sum_{j=1}^n \alpha_j^k \mathbf{u}_j$ with $\alpha_j^k = \langle \mathbf{u}_j, \mathbf{Z}_k \rangle_{A^{-1}}$ there follows

$$\mathbf{Z}_k - P_T \mathbf{Z}_k = \sum_{j=p+1}^n \alpha_j^k \mathbf{u}_j, k = 1, ..., NV.$$

Therefore,

$$\mathcal{E}(T) = \frac{1}{NV} \sum_{k=1}^{NV} \sum_{j=p+1}^n \sum_{l=p+1}^n \alpha_j^k \alpha_l^k \mathbf{u}_j A^{-1} \mathbf{u}_l^t = \frac{1}{NV} \sum_{k=1}^{NV} \sum_{j=p+1}^n \left(\alpha_j^k \right)^2 =$$

$$\frac{1}{NV} \sum_{k=1}^{NV} \sum_{j=p+1}^n \mathbf{u}_j A^{-1} \mathbf{Z}_k^t \mathbf{Z}_k A^{-1} \mathbf{u}_j^t = \sum_{j=p+1}^n \mathbf{u}_j A^{-1} M_{\mathbf{ZZ}} A^{-1} \mathbf{u}_j^t. \qquad (4)$$

Thus, $\min_{T \in \mathcal{L}_p} \mathcal{E}(T) = \min_{\mathcal{U}} \sum_{j=p+1}^n \mathbf{u}_j A^{-1} M_{\mathbf{ZZ}} A^{-1} \mathbf{u}_j^t$ where $\mathcal{U}$ is the set of all subsets of vectors of $\mathbb{R}^n$ orthonormal with respect to the metric $\delta_{A^{-1}}$ and having cardinality $n - p$.

Let $Q : \mathbb{R}^n \to \mathbb{R}$ be the quadratic form with matrix $A^{-1} M_{\mathbf{ZZ}} A^{-1}$, and let $\mathbf{v} \in \mathbb{R}^n$ with $\mathbf{v} A^{-1} \mathbf{v}^t = 1$. Let $(v_1, ..., \nu_n)$ be the coordinates of $\mathbf{v} \in \mathbb{R}^n$ in the basis $\{A \mathbf{w}_1^t, A \mathbf{w}_2^t, ..., A \mathbf{w}_n^t\}$, which is an orthonormal basis with respect

to the metric $\delta_{A^{-1}}$. Then $1 = \mathbf{v}A^{-1}\mathbf{v}^t = \sum_{j=1}^n \nu_j^2$ and

$$Q(\mathbf{v}) = \left(\sum_{j=1}^n \nu_j \mathbf{w}_j A\right) A^{-1} M_{\mathbf{ZZ}} A^{-1} \left(\sum_{l=1}^n \nu_l \mathbf{w}_l A\right)^t = \sum_{j=1}^n \sum_{l=1}^n \nu_j \nu_l \sigma_l \mathbf{w}_j A \mathbf{w}_l^t =$$

$$\sum_{j=1}^n \sigma_j \nu_j^2 = \sigma_n (1 - \sum_{j=1}^{n-1} \nu_j^2) + \sum_{j=1}^{n-1} \sigma_j \nu_j^2 = \sigma_n + \sum_{j=1}^{n-1} (\sigma_j - \sigma_n) \nu_j^2 \geq \sigma_n = Q(A\mathbf{w}_n^t)$$

Let $\mathbf{v} \in (A\mathbf{w}_n^t)^\perp$ with $\mathbf{v}A^{-1}\mathbf{v}^t = 1$ and let $(\nu_1, ..., \nu_{n-1}, 0)$ be the coordinates

of $\mathbf{v}$ in the basis $\{A\mathbf{w}_1^t, A\mathbf{w}_2^t, ..., A\mathbf{w}_n^t\}$. Then

$$Q(\mathbf{v}) = \sum_{j=1}^{n-1} \sigma_j \nu_j^2 = \sigma_{n-1} + \sum_{j=1}^{n-2} (\sigma_j - \sigma_{n-1}) \nu_j^2 \geq \sigma_{n-1} = Q(A\mathbf{w}_{n-1}^t),$$

and so on. Thus $\mathcal{E}(\mathrm{T}) = \sum_{j=p+1}^n Q(\mathbf{u}_j) \geq \sum_{j=p+1}^n \sigma_j = \sum_{j=p+1}^n Q(A\mathbf{w}_j^t) =$

$\mathcal{E}(\mathcal{T}_p)$ for any $\mathrm{T} \in \mathcal{L}_p$. This completes the proof of $i)$. Since $\{A\mathbf{w}_1^t, A\mathbf{w}_2^t, ..., A\mathbf{w}_n^t\}$

is an orthonormal basis of $\mathbb{R}^n$ in the metric $\delta_{A^{-1}}$, the projection of any vector

$\mathbf{Z} \in \mathbb{R}^n$ on $\mathcal{T}_p$ is $\mathrm{P}_{\mathcal{T}_p}\mathbf{Z} = \mathbf{Z} - \sum_{i=p+1}^n \alpha_i \mathbf{w}_i A$, where $\alpha_i := \langle \mathbf{Z}, A\mathbf{w}_i^t \rangle_{A^{-1}} = \mathbf{Z}\mathbf{w}_i^t$.

Then

$$\mathrm{P}_{\mathcal{T}_p}\mathbf{Z} = \mathbf{Z} - (\alpha_{p+1}, ..., \alpha_n)B^t A = \mathbf{Z} - (\mathbf{Z}\mathbf{w}_{p+1}^t, ..., \mathbf{Z}\mathbf{w}_n^t)B^t A = \mathbf{Z}(\mathrm{I} - BB^t A).$$

∎

In the next corollary we give an equivalent expression for $\mathrm{P}_{\mathcal{T}_p}\mathbf{Z}$, which

shows more clearly that the algorithms proposed in [1] and [15] are also

based on projecting onto optimal linear subspaces optimal with respect to a

particular metric $\delta_{A^{-1}}$. In Refs.[1],[15] the authors consider a diagonal matrix

A which gives practically null weight to the first and last components of the diagonal entries and weight one to the p central ones. Then they take $\widehat{\mathbf{z}}$ $= (\mathrm{I} - A^{1/2}DD^t A^{-1/2})\mathbf{Z}^t$ as the estimate of $\mathbf{z}$, where D is a matrix whose $n-p$ columns are the orthonormal eigenvectors of the matrix $A^{-1/2}M_{\mathbf{ZZ}}A^{-1/2}$ corresponding to the $n - p$ smallest eigenvalues. We show that $\widehat{\mathbf{z}} = \mathrm{P}_{\mathcal{T}_p}\mathbf{Z}$ where $\mathcal{T}_p$ is the linear subspace in $\mathcal{L}_p$ which minimizes the mean square $\delta_{A^{-1}}-$distance.

Let $A = Q\Lambda Q^t$ be the spectral decomposition of the symmetric positive definite matrix A, i.e. Λ is the diagonal matrix which has as entries the eigenvalues of A, Q is the matrix which has as columns the eigenvectors of A and $Q^t Q = \mathrm{I}$. We now show that the linear map $g(\mathbf{Z}) := \mathbf{Z}Q\Lambda^{-1/2}$ transforms the orthogonal projection in the metric $\delta_{A^{-1}}$ into an orthogonal projection with respect to the Euclidean metric.

Corollary 4 *Let $\sigma_1 \geq \cdots \geq \sigma_n$ be the eigenvalues of the $n \times n$ matrix $M_{g(\mathbf{Z})g(\mathbf{Z})} := \frac{1}{NV}\Sigma_{k=1}^{NV}g(\mathbf{Z}_k)^t g(\mathbf{Z}_k)$ and $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ the corresponding orthonormal eigenvectors in the Euclidean metric. Then, for $1 \leq p \leq n$, and for $\mathbf{Z} \in \mathbb{R}^n$,*

$$\mathcal{T}_p = span\{g^{-1}(\mathbf{v}_1), ..., g^{-1}(\mathbf{v}_p)\} \text{ and } \mathrm{P}_{\mathcal{T}_p}\mathbf{Z} := (\mathrm{I} - Q\Lambda^{1/2}DD^t\Lambda^{-1/2}Q^t)\mathbf{Z}^t$$

34

where D is the $n \times (n-p)$ matrix whose columns are the eigenvectors $\{\mathbf{v}_{p+1}, ..., \mathbf{v}_n\}$.

If A is a diagonal matrix then

$$M_{g(\mathbf{z})g(\mathbf{z})} = A^{-1/2} M_{\mathbf{ZZ}} A^{-1/2} \text{ and } \mathrm{P}_{\mathcal{T}_p} \mathbf{Z} := (\mathrm{I} - A^{1/2} DD^t A^{-1/2}) \mathbf{Z}^t.$$

Proof. Let $\mathrm{T} \in \mathcal{L}_p$ and let $\{\mathbf{u}_1, ..., \mathbf{u}_n\}$ be an orthonormal basis of $\mathbb{R}^n$ with respect to the metric $\delta_{A^{-1}}$ such that $\mathrm{T} = span\{\mathbf{u}_1, ..., \mathbf{u}_p\}$. Using the spectral decomposition of A^{-1} and (4) we get

$$\mathcal{E}(\mathrm{T}) = \sum_{j=p+1}^{n} \mathbf{u}_j A^{-1} M_{\mathbf{ZZ}} A^{-1} \mathbf{u}_j^t = \sum_{j=p+1}^{n} \mathbf{u}_j Q \Lambda^{-1} Q^t M_{\mathbf{ZZ}} Q \Lambda^{-1} Q^t \mathbf{u}_j^t =$$

$$\sum_{j=p+1}^{n} g(\mathbf{u}_j) \Lambda^{-1/2} Q^t M_{\mathbf{ZZ}} Q \Lambda^{-1/2} g(\mathbf{u}_j)^t = \sum_{j=p+1}^{n} g(\mathbf{u}_j) M_{g(\mathbf{z})g(\mathbf{z})} g(\mathbf{u}_j)^t.$$

Since $g(\mathbf{u}_i)g(\mathbf{u}_j)^t = \mathbf{u}_i Q \Lambda^{-1/2} \Lambda^{-1/2} Q^t \mathbf{u}_j = \delta_{ij}$ we have that the set $\{g(\mathbf{u}_j) : j = 1, ..., n\}$ is an orthonormal basis of $\mathbb{R}^n$ with respect to the usual inner product in $\mathbb{R}^n$. Thus, $\min_{\mathrm{T} \in \mathcal{L}_p} \mathcal{E}(\mathrm{T}) = \min_{\mathcal{V}} \sum_{j=p+1}^{n} \mathbf{v}_j \left(M_{g(\mathbf{z})g(\mathbf{z})} \right) \mathbf{v}_j^t$ where $\mathcal{V}$ is the set of all subsets of vectors of $\mathbb{R}^n$ orthonormal with respect to the Euclidean metric and having cardinality $n-p$. Using a similar argument to that given in the proof of Theorem 3 we can see that $\mathcal{E}(\mathrm{T})$ attains its minimum value when we take as $\mathbf{v}_j$ the eigenvectors $\mathbf{v}_{p+1}, ..., \mathbf{v}_n$ of $M_{g(\mathbf{z})g(\mathbf{z})}$ corresponding to the last $n-p$ eigenvalues and $\mathcal{T}_p := span\{g^{-1}(\mathbf{v}_1), ..., g^{-1}(\mathbf{v}_p))\}$. Since $\mathrm{P}_{\mathcal{T}_p} \mathbf{Z} = \mathbf{Z} - \sum_{i=p+1}^{n} \alpha_i g^{-1}(\mathbf{v}_i)$ with $\alpha_i := \langle \mathbf{Z}, g^{-1}(\mathbf{v}_i) \rangle_{A^{-1}} = \mathbf{Z} A^{-1} Q \Lambda^{1/2} \mathbf{v}_i^t$ we

get

$$P_{\mathcal{T}_p}\mathbf{Z} = \mathbf{Z} - (\alpha_{p+1}, ..., \alpha_n)D^t\Lambda^{1/2}Q = \mathbf{Z}(I - \mathbf{A}^{-1}Q\Lambda^{1/2}DD^t\Lambda^{1/2}Q^t) =$$

$$\mathbf{Z}(I - Q\Lambda^{-1/2}DD^t\Lambda^{1/2}Q^t).$$

The proof for the case of diagonal A is obvious using that $Q = I$ and $\Lambda = A$.

∎

References

[1] P. Grassberger, R. Hegger, H. Kantz, C. Schaffrath and T. Schreiber. "On noise reduction methods for chaotic data." Chaos **3**, 2, 127-141 (1993).

[2] E. J. Kostelich and T. Schreiber. "Noise reduction in chaotic time-series data: A survey of common methods." Physical Review E **48**, 3, 1752-1763 (1993).

[3] M. Davies. "Noise reduction schemes for chaotic time series." Physica D **79**, 174-192 (1994).

[4] E. J. Kostelich and J. A. Yorke. "Noise reduction: Finding the simplest dynamical system consistent with the data." Physica D **41**, 183-196 (1990).

[5] R. Cawley and G-H. Hsu. "Local-geometric-projection method for noise reduction in chaotic maps and flows." Physical Review A **46**, 6, 3057-3082 (1992).

[6] T. Schreiber and P. Grassberger. "A simple noise-reduction method for real data.". Physics Letters A **160**, 411-418 (1991).

[7] A. Leontitsis, T. Bountis and J. Pagge. "An adaptive way for improving noise reduction using local geometric projection." Chaos **14**, 1, 106-110 (2004).

[8] F. Takens. "Dynamical Systems and turbulence." D. A. Rang and L. S. Young eds. Lecture Notes in Mathematic*s* **898**, 396 (1981).

[9] S. M. Hammel. "A noise reduction method for chaotic systems." Physics Letters A **148**, 421-428 (1990).

[10] J. D. Farmer and J. Sidorowich. "Optimal shadowing and noise reduction." Physica D **47**, 373-392 (1991).

[11] J. Bröcker and U. Parlitz. "Efficient noncausal noise reduction for deterministic time series." Chaos **11**, 2, 319-326 (2001).

[12] R. Bowen. "Markov partitions for Axiom A diffeomorphisms." American Journal of Mathematics **92**, 725-747 (1970).

[13] S.M. Hammel, J.A. Yorke and C. Gregobi. "Do numerical orbits of chaotic process represent true orbits?." Journal of Complexity **3**, 136-145 (1987).

[14] S. Y. Pilyugin. "Shadowing in Dynamical Systems." Lecture Notes in Mathematics 1706. Springer (1999).

[15] R. Hegger and T. Schreiber. "A noise reduction method for multivariate time series." Physics Letters A **170**, 305-310 (1992).

[16] E. J. Kostelich and J. A. Yorke. "Noise reduction in dynamical systems." Physical Review A **38**, 3, 1649-1652 (1988).

[17] W. A. Fuller. "Measurement error models." John Wiley & Sons (1987).

[18] L. Jaeger and H. Kantz. "Unbiased reconstruction of the dynamics underlying a noisy time series." Chaos **6**, 449-450 (1996).

[19] P.E. McSharry and L. A. Smith. "Better nonlinear models from noisy data: attractors with maximum likelihood." Physical Review Letters **83**, 21, 4285-4288 (1999).

[20] V.F. Pisarenko and D. Sornette. "Statistical methods of parameter estimation for deterministically chaotic time series." Physical Review E **69**, 036122 (2004).

[21] R. Meyer and N. Christensen. "Bayesian reconstruction of chaotic dynamical systems." Physical Review E **62**, 3, 3535-3542 (2000).

[22] K. Judd. "Chaotic-time-series reconstruction by Bayesian paradigm: Right results by wrong methods." Physical Review E **67**, 026212 (2003).

[23] T. Sauer. "A noise reduction method for signals from nonlinear systems." Physica D **58,** 193-201 (1992).

[24] P. Grassberger, T. Schreiber and C. Schaffrath. "Nonlinear time sequence analysis." International Journal of Bifurcation and Chaos **1**, 3, 521-547 (1991).

[25] J-P. Eckmann, S.O. Kamphorst, D. Ruelle. and S. Ciliberto. "Lyapunov exponents from time series." Physics Review A **34**, 6, 4971-4979 (1986).

[26] M.E. Mera and M. Morán. "Convergence of the Eckmann and Ruelle algorithm for the estimation of the Lyapunov exponents." Ergodic Theory and Dynamical Systems **20**, 531-546 (2000).

[27] M.E. Mera and M. Morán. "Degrees of Freedom of a Time Series." Journal of Statistical Physics **106**, 112, 125-145 (2002).

[28] A. Kern, W.-H. Steeb and R. Stoop. "Projective noise cleaning with dynamic neighborhood selection." International Journal of Modern Physics C **11**, 1, 125-146 (2000).

FIGURE CAPTIONS

FIG 1. Clean (a), noisy (b) and clean together with cleaned (c) time series from a Hénon dynamics corrupted by uncorrelated homoskedastic Gaussian noise $(\varepsilon_x, \varepsilon_y)$ with $\sigma_{\varepsilon_x} = 1\%\sigma_S$ and $\sigma_{\varepsilon_y} = 1\%\sigma_S$. The parameters of the algorithm are $N = 10000$, $Itera = 16$, $NV = 300$, and $Nprom = 25$.

FIG 2. Noisy and cleaned time series from a Hénon dynamics corrupted by uncorrelated and highly heteroskedastic noise with $\sigma_{\varepsilon_x} = 1\%\sigma_S$ and $\sigma_{\varepsilon_y} = 15\%\sigma_S$. The parameters of the algorithm are $N = 10000$, $Itera = 6$, $NV = 2500$, and $Nprom = 200$.

FIG 3. Noisy, cleaned (MLA), cleaned (TLSA) for a short time series (500 data points) from the noisy Hénon dynamics of $Fig.$ 2. The parameters for either algorithm are $NV = 25$, $NCT = Nprom = 5$ and $Itera = 45$.

FIG. 4 Average normalized singular values, $S_{r_n}(j)$, $j = 1,...,4$ of the covariance matrix of the data in balls of radius r for the 2-embedded time series (clean, cleaned and noisy) as a function of the normalized radius $r_n :=$ $\frac{r}{diam}$ of the balls where $diam$ is the diameter of the time series.

TABLE CAPTIONS

TABLE I. Noise reduction measures $< R_h >$ and R_h obtained by MLA and $TLSA$ for a time series of 10000 data points from Hénon dynamics corrupted by uncorrelated homoskedastic noises.

TABLE II. Noise reduction measure $< R_h >$ obtained by MLA and $TLSA$ for time series of various lengths from Hénon dynamics corrupted by uncorrelated heteroskedastic noises.

TABLE III. Noise reduction measure $< R_h >$ obtained by MLA for time series of various lengths from Lorenz dynamics corrupted by uncorrelated heteroskedastic noise.

TABLE IV. Estimation of the Lyapunov exponents obtained for the clean, cleaned and noisy time series of the experiment in Fig. 2.

TABLE I

HOMOSKEDASTIC CASE

$< R_h >$ **and** R_h **values**

$N = 10000,\ Nprom = 25$		$< R_h >$		R_h	
$(NV, ITERA)$	$\sigma_{\varepsilon_x},\ \sigma_{\varepsilon_y},\ \sigma_\varepsilon$	MLA	$TLSA$	MLA	$TLSA$
$(300, 16)$	$\sigma_{\varepsilon_x} = \sigma_{\varepsilon_y} = 1\%\sigma_S$ $\sigma_\varepsilon = 1.4\%\sigma_S$	**84**	**84**	**61**	**50**
$(900, 14)$	$\sigma_{\varepsilon_x} = \sigma_{\varepsilon_y} = 5\%\sigma_S$ $\sigma_\varepsilon = 7.1\%\sigma_S$	**90**	**82**	**62**	**42**
$(1200, 16)$	$\sigma_{\varepsilon_x} = \sigma_{\varepsilon_y} = 10\%\sigma_S$ $\sigma_\varepsilon = 14.1\%\sigma_S$	**91**	**84**	**68**	**50**

TABLE II

HETEROSKEDASTIC CASE

$< R_h >$ **values**

$\sigma_\varepsilon = 11.2\%\sigma_s$					
$\sigma_{\varepsilon_x} = 10\%\sigma_S = 11.5\%\sigma_x$ $\sigma_{\varepsilon_y} = 5\%\sigma_S = 17.4\%\sigma_y$			$\sigma_{\varepsilon_x} = 11\%\sigma_S = 12.6\%\sigma_x$ $\sigma_{\varepsilon_y} = 2\%\sigma_S = 7\%\sigma_y$		
$(N, ITERA)$ $(NV, NPROM)$	MLA	$TLSA$	$(N, ITERA)$ $(NV, NPROM)$	MLA	$TLSA$
$(10000, 13)$ $(800, 25)$	**87**	**73**	$(10000, 15)$ $(600, 25)$	**81**	**52**
$(5000, 18)$ $(400, 25)$	**84**	**68**	$(5000, 14)$ $(300, 25)$	**78**	**50**
$(1000, 23)$ $(80, 5)$	**78**	**68**	$(1000, 19)$ $(60, 5)$	**63**	**44**
$(500, 25)$ $(40, 5)$	**65**	**54**	$(500, 38)$ $(25, 5)$	**57**	**32**

TABLE III

$< R_h >$ values for MLA

$\sigma_\varepsilon = 17\%\sigma_S$			
$\sigma_{e_x} = 20\%\sigma_x = 10.9\%\sigma_S$ $\quad$ $\sigma_{e_y} = 10\%\sigma_x = 6.2\%\sigma_S$ $\quad$ $\sigma_{\varepsilon_z} = 20\%\sigma_z = 11.4\%\sigma_S$			
	$N = 50000$	$N = 10000$	$N = 3000$	$N = 1000$
$< R_h >$	**90**	**85**	**75**	**60**
$(NV, ITERA, NPROM)$	$(4000, 11, 50)$	$(800, 10, 50)$	$(300, 6, 25)$	$(50, 20, 5)$

TABLE IV

LYAPUNOV EXPONENTS

	CLEAN	*CLEANED*	*NOISY*
$\widehat{\lambda}_1$	$\mathbf{0.421} \pm 0.029$	$\mathbf{0.405} \pm 0.031$	$\mathbf{0.479} \pm 0.042$
$\widehat{\lambda}_2$	$-\mathbf{1.605} \pm 0.032$	$-\mathbf{1.751} \pm 0.077$	$-\mathbf{0.948} \pm 0.108$

FIG. 1(a)
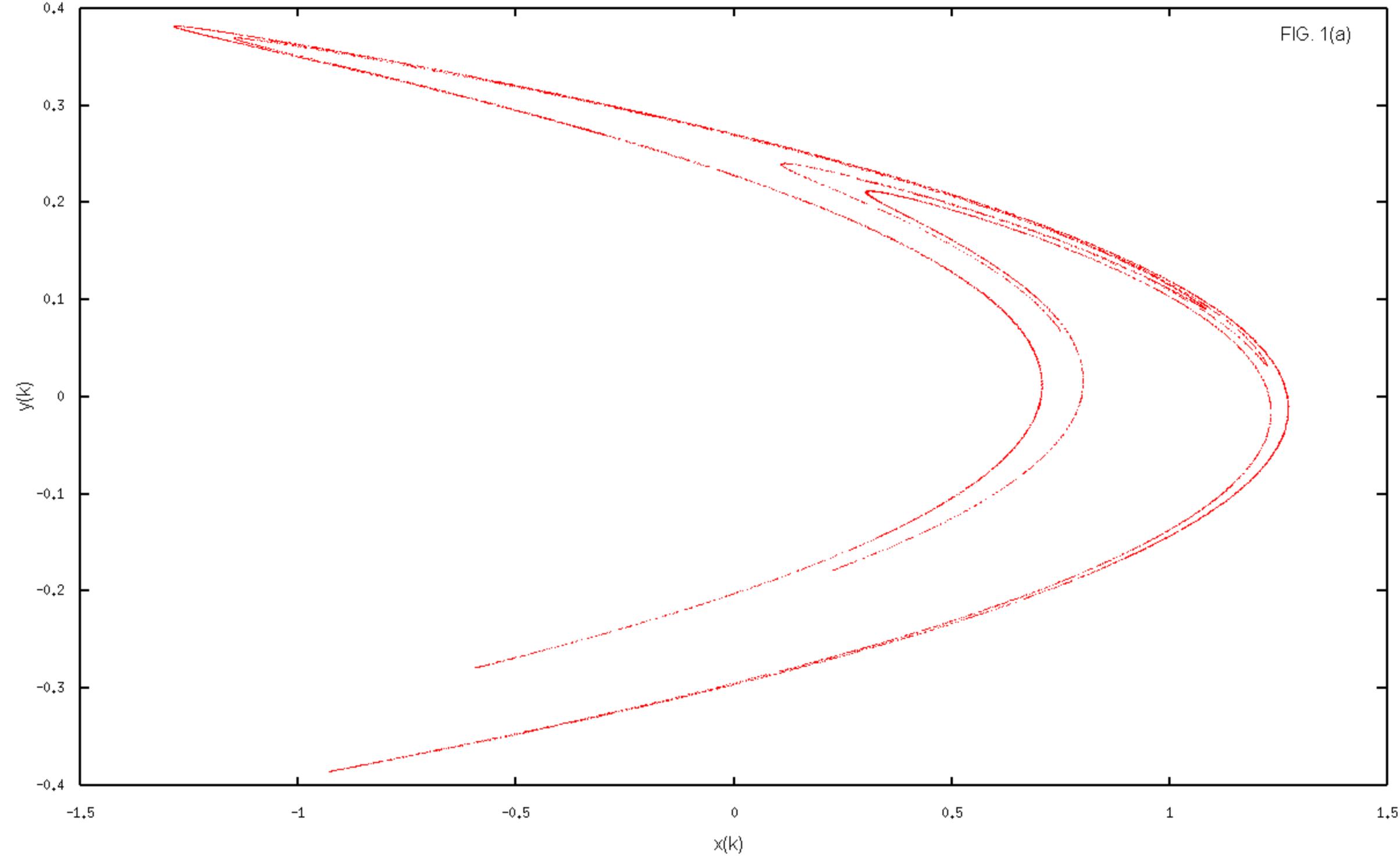
FIG. 1(b)
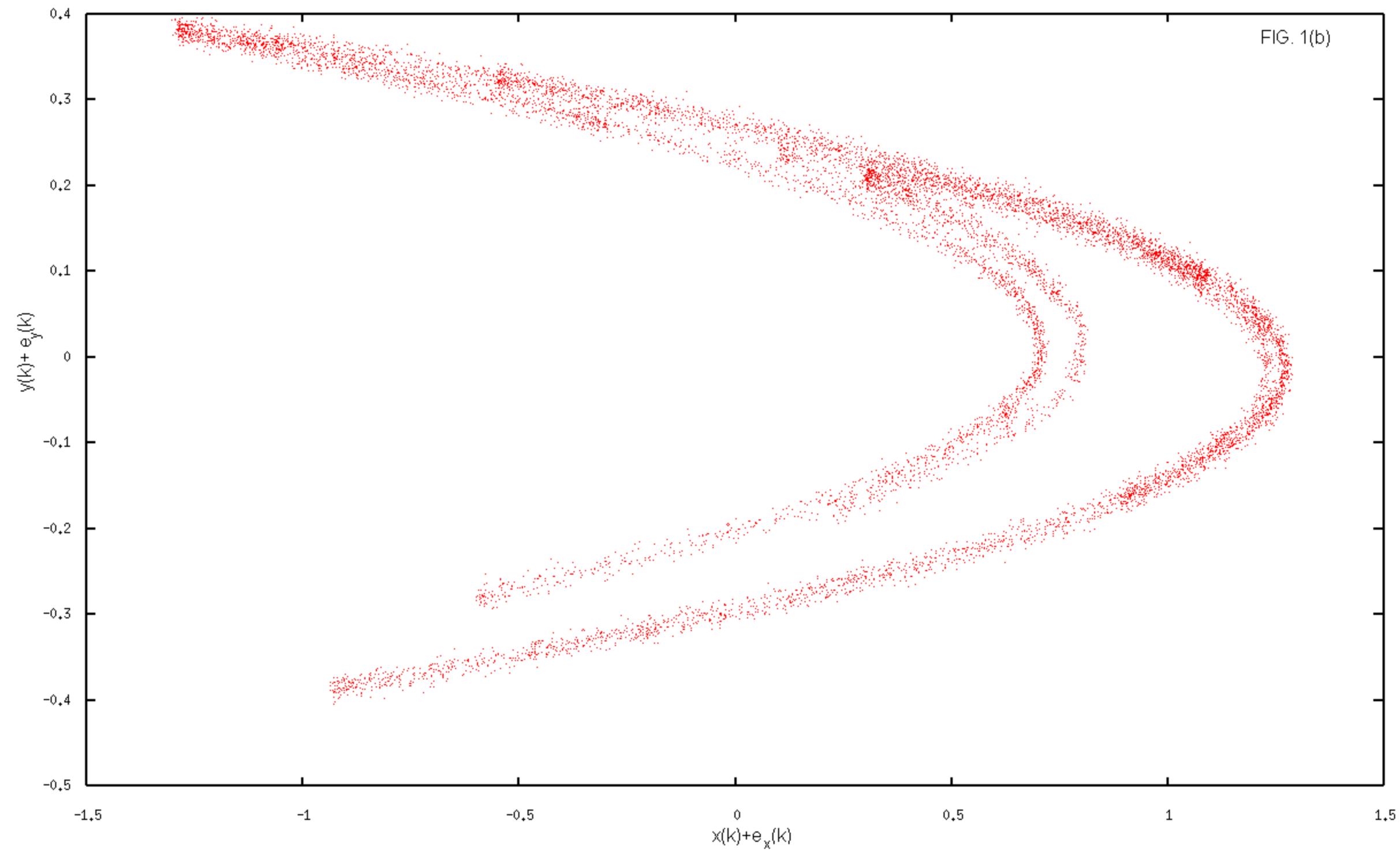
FIG. 1(C)

CLEAN

CLEANED

FIG. 2

FIG.3

FIG. 4