

Detecting Noise in Chaotic Signals through Principal Component Matrix Transformation

Ivan Michieli and Božidar Vojnović

Electronic Department, Ruđer Bošković Institute, Zagreb, Croatia

We study the reconstruction of continuous chaotic attractors from noisy time-series. A method of delays and principal component eigenbasis (defined by singular vectors) is used for state vectors reconstruction. We introduce a simple measure of trajectory vectors directional distribution for chosen principal component subspace, based on nonlinear transformation of principal component matrix. The value of such defined measure is dependent on the amount of noise in the data. For isotropically distributed noise (or close to isotropic), that allows us to set up window width boundaries for acceptable attractor reconstruction as a function of noise content in the data.

Keywords: time series, chaotic signals, strange attractor, principal component analysis (PCA), singular value decomposition (SVD), noise.

1. Introduction

In signal processing applications one usually analyzes data from experimental measurements in the form of a finite scalar time series or “chaotic like” signals, contaminated with the noise of various origins. To uncover underlying dynamical structure of the data, several methods have been proposed. Packard et al. [18] suggest that dynamically equivalent phase portrait could be reconstructed from time derivatives of the single variable time series. The method commonly used today is known as the “method of delays” and is based on Takens’s embedding theorem [22]. Takens showed that for n -dimensional dynamical system it is sufficient to use $d_E \geq 2n + 1$ (“E” denotes embedding) data from single variable measurements $x(t)$, as a coordinate of a respective trajectory

point in d_E dimensional space, to progressively reconstruct dynamically equivalent phase portrait of original trajectory. The scalar time series $x(t)$ is presumed to be the result of smooth measurement function on the n -dimensional manifold that contains original trajectory.

The dynamically equivalent phase portrait means that such defined map is an embedding i.e. that it preserves geometrical invariants such as the fractal dimension of the attractor, Lyapunov exponents of a trajectory, eigenvalues of a fixed point etc. Nevertheless, it is quite possible that the dimension less than $2n + 1$ is enough to reconstruct attractor [19].

The respective state “delay” vector in d_E -dimensional embedding space is given as:

$$\mathbf{x}_i^T = [x(t_i), x(t_i + \tau), x(t_i + 2\tau), \dots, x(t_i + (d_E - 1)\tau)],$$

where τ is the “lag time” defined as some integer multiple of sampling time ($\tau = n t_s$; $t_s = t_{i+1} - t_i$). The product $(d_E - 1)\tau$ is customarily known as “window width” denoted as τ_w .

For infinite noise free data sets the choice of reconstruction parameters such as lag time or window width is almost arbitrary, (providing that d_E is sufficiently high).

For limited noisy data sets using window widths near one of the characteristic periods of the system, local collapse of the attractor can occur. Likewise, in the presence of noise, excessive spreading of the attractor, using large τ_w , additionally unclear the attractor structure due to insufficient separation of opposite flow trajectory

paths. One of the major problems in attractor reconstruction from noisy data also occurs for small window widths. Although in ideal cases with infinite resolution noise free data, theoretically there are no limitations on smallness of delay time [9, 22], for noisy data, employing the small window widths is totally unsuitable. Such reconstruction forced the trajectory to occupy narrow m -dimensional ellipsoid along the main diagonal of the embedding space and the geometry of the attractor in all other directions could become completely obscured.

Accordingly, in practice, the choice of reconstruction parameters becomes very important and it is necessary to perform some optimization analysis for their selection. Various methods for choosing adequate delay time τ (and consequently τ_w) have been proposed. Schuster [20] fixed the delay time at the first zero of autocorrelation function (AFC) to ensure linear independence of the coordinates. Fraser and Swinney [8] used the first minimum of “mutual information” which generally measures dependence of two variables. Broomhead and King [5] proposed singular value decomposition – SVD, (also known as principal component analysis – PCA) of the trajectory matrix to obtain satisfactory orthonormal coordinate projection basis.

Recently a variety of different algorithms have been proposed by Liebert et al. [16], Kember and Fowler [11], Kantz and Olbrich [10], Libton and Dabke [17] and Zoltowski [23].

Most of aforementioned methods are based on maximization of some measure of expansion of the attractor in the phase space. Such methods are generally not directly related to the amount of noise in the data. Moreover, they are relatively insensitive to it (with the exception of the method introduced in [5]). Because of that, propositions of that kind are to some extent incomplete.

More thorough account on reconstruction of the chaotic attractors in the presence of noise was reported by Casdagli et al. [7] and Gibson et al. [9].

In the present paper we limited our study to continuous low-dimensional chaotic processes i.e. flows (discrete chaotic maps are not considered) contaminated with isotropic-like noise,

that is, the distribution of the noise vectors is assumed to be close to isotropic. Such distribution is characteristic of several types of noise such as white Gaussian noise, quantization noise (if it isn't too crude), noise that originates from round-off errors in computing process or from limiting resolution of experimental apparatus, etc.

We establish a simple, noise-dependent measure of anisotropy for selected principal component subspace that allows us to set up a criterion for choosing adequate window width boundaries for satisfactory reconstruction of the chaotic attractor. The details of the analysis are presented in Section 2.

2. Noise-Related State Vectors Distribution

2.1. Review of Principal Component Coordinates

Principal component analysis – PCA, also known as singular value decomposition – SVD (or as Karhunen-Loève decomposition), is the method that defines specific orthogonal projection basis.

If the trajectory points of the chaotic attractor are thought of as a collection of N point masses distributed in m -dimensional space, then principal components are projections of associated state vectors onto the principal axes of inertia of such mass distribution.

Equivalently, principal component coordinates coincide with an orthonormal basis that is generated by process of maximizing the root mean square (rms) projection of trajectory state vectors onto a set of m orthonormal vectors that span m -dimensional embedding space. The coordinate system is centered at the center of mass of the trajectory. The principal component matrix Y is given by matrix equation

$$Y = XS,$$

or equivalently;

$$\begin{bmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_N^T \end{bmatrix} = N^{-\frac{1}{2}} \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix} \begin{bmatrix} \cdot & \cdot & \cdot \\ \mathbf{s}_1 & \cdot & \mathbf{s}_m \\ \cdot & \cdot & \cdot \end{bmatrix}, \quad (1)$$

where \mathbf{X} is normalized $N \times m$ trajectory matrix with zero mean value and \mathbf{S} is $m \times m$ orthogonal rotation matrix with columns composed of orthonormal vectors that coincide with principal axes of inertia, ordered by the magnitude of corresponding moment of inertia. The matrix \mathbf{S} can be easily calculated by standard methods from linear algebra [15] i.e. by diagonalization of covariance matrix defined as $\mathbf{\Xi} = \mathbf{X}^T \mathbf{X}$.

Since $\mathbf{\Xi}$ is real symmetric $m \times m$ matrix we can write: $\mathbf{\Sigma}^2 = \mathbf{S}^{-1} \mathbf{\Xi} \mathbf{S}$, where \mathbf{S} is $m \times m$ orthogonal and $\mathbf{\Sigma}^2$ is $m \times m$ diagonal matrix with eigenvalues of σ_i^2 of $\mathbf{\Xi}$ on the diagonal ($i = 1, 2, \dots, m$, and $\sigma_1^2 \geq \sigma_2^2 \geq \dots \geq \sigma_m^2$).

The columns of \mathbf{S} are normalized eigenvectors (singular vectors) of $\mathbf{\Xi}$ with respective eigenvalues σ_i^2 and consequently each $N\sigma_i^2$ is respective moment of inertia. The set $\{\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2\}$ is known as a singular spectrum and σ_i are respective singular values.

The benefits of projecting delay coordinates on principal component basis, especially for noisy data, have been extensively studied [5, 9, 12, 13]. Here, we just wish to emphasize that by using high delay dimension d_E and projecting the trajectory onto first k principal component eigenvectors generally acts like noise filter. Furthermore, for noise vectors that are distributed isotropically, first k subset of principal components is the optimal choice for coordinate transformation, since those have maximum variance of the data and, therefore, maximum signal to noise ratio (SNR). This is because the projection of isotropic noise vectors on any direction has the same variance.

Gibson et. al. [9] showed that, for clean data and small window width, the eigenvalues σ_i^2 decreased exponentially with i and each σ_i^2 exhibits $(\tau_w)^{2i}$ power-law scaling with τ_w . That is the reason why, for small τ_w , the attractor projection on any p -dimensional subspace of d_E , spanned by respective p principal component eigenvectors, is stretched along the line defined by the respective largest eigenvalue. Furthermore, for small τ_w value interval, the structure of the attractor remains nearly the same (except for the scaling factor). When τ_w nears some characteristic recurrence time of the system, the structure begins to complicate and loses its original ‘‘simplicity’’. So, in the absence of noise,

using small τ_w , with appropriate linear scaling of higher principal components, adequate reconstruction can be realized.

The presence of noise violates such favorable conditions and forces us to increase τ_w up until the last of the important eigenvalues (σ_k^2) is sufficiently extracted from the noise floor [5, 9]. However, following such approach, when increasing τ_w by increasing ‘‘lag time’’ with fixed d_E , one can end up with highly complex attractor structure and, at the same time, with overall worse SNR, since higher unimportant eigenvalues could be increased too much at the expense of the first ones. This could happen because the trace of a matrix is invariant under similarity transformations and it is bounded with the total variance of the data in d_E -dimensional space, i.e. the following expression is valid:

$$\text{Tr } \mathbf{X}^T \mathbf{X} = \text{Tr } \mathbf{Y}^T \mathbf{Y},$$

or equivalently

$$N \sum_{i=1}^{d_E} \sigma_i^2 = d_E \langle x^2 \rangle. \quad (2)$$

Appropriate balance between these opposing effects strongly depends on the type of the attractor and on the further use of the reconstruction. So, to find acceptable unique rule for reconstruction, is an intricate task.

2.2. State Vectors Nonlinear Transformation

2.2.1. State Vectors Submatrix Definition

As explained in the previous chapter, for small τ_w , the attractor projection on any p -dimensional principal component subspace of d_E , looks like a thin ellipsoid with highly anisotropic state vectors distribution. As we increase τ_w , the ellipsoid becomes thicker and thicker. Consequently, anisotropy of the projection is progressively decreased to some point when complex oscillatory behavior starts.

If isotropic noise, with variance of the same magnitude as the largest projection eigenvalue, is added, the ellipsoid would be changed to a p -dimensional ball with almost ideal isotropic vector distribution. Deterministic character of chosen projection would be totally lost.

For some chosen d_E and τ_w , the effect of noise would be most dominant in higher subspaces where trajectory vectors' projections are not yet extracted from the noise floor. If we assume that the first j -dimensional projection is sufficient for embedding, our primary goal is to filter the noise from that primary j -subspace. That constitutes the standard noise filtering in global PCA approach i.e. we embed the attractor in relatively high d_E -dimensional space and then project it on the first j th principal components. Among them, the most critical is trajectory vectors' projection on the last j th eigenvector where the noise is more dominant than in the first ones. We can analyze that noise effect by considering the trajectory vectors directional distribution when projected on the higher p -dimensional subspace (where $p = d_E - j + 1$) that is fronted by that j th eigenvector.

To model such general directional properties of attractor' projection vectors in the presence of noise, we introduce principal component submatrix associated with j th eigenvector as follows:

$$\mathbf{R}^{(j,m)} = \begin{bmatrix} \mathbf{r}_1^T \\ \cdot \\ \cdot \\ \mathbf{r}_N^T \end{bmatrix} = \mathbf{X} \begin{bmatrix} \cdot & \cdot & \cdot \\ s_j & s_i & s_m \\ \cdot & \cdot & \cdot \end{bmatrix}, \quad (3)$$

where \mathbf{r}_i is normalized (normalization factor is $N^{-1/2}$) projection of \mathbf{x}_i on p -subspace, s_i is the i th eigenvector, m denotes dimension of embedding space ($m = d_E$) and j indicates associated first eigenvector of p -subspace. For $j = 1$, submatrix $\mathbf{R}^{(j,m)}$ is identical to original principal component matrix i.e.: $\mathbf{R}^{(j,m)} = \mathbf{Y}$, and for $j = 2, 3, \dots, m$, resulting submatrix $\mathbf{R}^{(j,m)}$ represents projection of the trajectory vectors on a subspace spanned by the last $p = m - j + 1$ principal component eigenvectors.

2.2.2. Characterization of the State Vectors Directional Properties

The next step is characterization of attractor' directional properties through some quantitative measure associated with selected p -subspace that will clearly indicate the presence of noise in the chosen subspace and that will change gradually with attractor spreading, that is, as a function of window width τ_w .

If the p -subspace is dominated by isotropic noise (small τ_w), the power distribution among associated eigenvalues would be balanced since the noise variance is the same in any direction, that is, the following relation would be satisfied: $\sigma_j^2 \approx \sigma_{j+1}^2 \approx \dots \approx \sigma_m^2$.

Accordingly, it is tempting to use some function of σ_j^2 and its dependence on τ_w as a measure of noise impact on trajectory vector' directional distribution in p -subspace.

The role of higher non-prime singular values σ_i was also reported by others [3] in the context of the detection of determinism in a time series. The natural choice would be to use $\hat{\sigma}_j^2 = \sigma_j^2 / \sum_{i=j}^m \sigma_i^2$, since such normalized eigenvalue reflects, on average, the extraction of trajectory vectors' projections on j th principal axes from the noise floor and, at the same time, its value varies from $\approx 1/p$ (note that $p = m - j + 1$) for noise-dominated to ≈ 1 , for highly anisotropic and noise-free vectors' distribution in p -subspace (small τ_w and clean data). Nevertheless, such measure is generally too crude for quantitative assessment of noise impact on trajectory vectors' directional distribution, which is apparent from the following analysis:

The sole contribution of a particular k th trajectory vector to j th eigenvalue is as follows:

$$\begin{aligned} \Delta_k \sigma_j^2 &= N^{-1} (\mathbf{x}_k \bullet \mathbf{s}_j)^2 = (\mathbf{r}_k \bullet \mathbf{s}_j)^2 \\ &= (\|\mathbf{r}_k\| \cos \beta_k)^2 \\ &= (\|\bar{\mathbf{r}}_k + \mathbf{n}_k\|^2 \cos^2 \beta_k), \end{aligned} \quad (4)$$

where \mathbf{r}_k and $\bar{\mathbf{r}}_k$ are noisy and clean data vectors' projection on p -subspace, \mathbf{n}_k is the noise vector and β_k is the angle between \mathbf{r}_k and \mathbf{s}_j .

It is obvious from relation (4) that, for large trajectory vectors projections on p -subspace ($\|\bar{\mathbf{r}}_k\| \gg \|\mathbf{n}_k\|$), the noise doesn't significantly change their average contribution to σ_j^2 . For small data vectors' projections when $\|\bar{\mathbf{r}}_k\| \approx \|\mathbf{n}_k\|$, although noise alters their direction severely, their average impact on σ_j^2 is minor, because it is limited by their lengths. As the result, as soon as the majority of the largest data vectors projections \mathbf{r}_i float up from the noise floor, the dependence of σ_j^2 on noise becomes constrained by attractors projection-dominant vectors. This

could happen too soon i.e. for some value of τ_w , when the noise still dominates over attractors' internal structure.

In order to find the measure that is more susceptible to noise-alteration of data vectors directional distribution in p -subspace, we refer to equivalent formulation of PCA as a process that defines m orthonormal vectors by minimizing the sum of the squares of orthogonal distances of trajectory state vectors to the respective eigenvectors and thus is equivalent to what is termed orthogonal regression [15]. In a view of that, the j th eigenvector s_j is defined by minimizing the sum

$$L_j = \sum_{i=1}^N \|w_i[r_i - (r_i \cdot s_j)s_j]\|^2 \quad (5)$$

under constraint $\|s_j\| = 1$, where w_i are arbitrary weights (in standard PCA $w_i = 1$) and r_i are normalized projections of respective x_i onto p -subspace or, equivalently

$$r_i = N^{-1/2} \left[x_i - \sum_{k=1}^{j-1} (x_k \cdot s_k) s_k \right]. \quad (6)$$

Since previous analysis showed that the noise impact on state vectors directional distribution in p -subspace (and, consequently, their average contribution to σ_j^2) is inversely proportional to its p -projected length, we introduce weights $w_i = \frac{1}{\|r_i\|}$ into equation (5) and we

get: $L_j = \sum_{i=1}^N \left\| \left[\frac{r_i}{\|r_i\|} - \left(\frac{r_i}{\|r_i\|} \cdot s_j^w \right) s_j^w \right] \right\|^2$,

with the notation $c_i = \frac{r_i}{\|r_i\|}$, we finally may write

$$L_j = \sum_{i=1}^N \| [c_i - (c_i \cdot s_j^w) s_j^w] \|^2, \quad (7)$$

where s_j^w denotes j th weighted eigenvector. Associated eigenvalues are accordingly defined as:

$$\lambda_j^2 = N^{-1} \sum_{i=1}^N (c_i \cdot s_j^w)^2. \quad (8)$$

Equations (7) and (8) define weighted orthogonal regression or weighted PCA approach.

In agreement with (7) and (8), our referent component submatrix $\mathbf{R}^{(j,m)}$ is transformed to:

$$\mathbf{R}^{(j,m)} \implies \mathbf{C}^{(j,m)} = N^{-1/2} \begin{bmatrix} c_1^T \\ \vdots \\ c_N^T \end{bmatrix},$$

where $c_i = \frac{r_i}{\|r_i\|}$. (9)

Generally, $\mathbf{C}^{(j,m)}$ represents transformed p -projection of attractor trajectory. Geometrically, it means that trajectory points are placed on the unit p -dimensional sphere. The directions of projected state vectors are preserved in this mapping, but their lengths are changed to unity. The contribution of, in this manner, transformed individual p -projected vectors c_i to the first principal component eigenvalue of the new trajectory submatrix is highly dependent on the amount of noise in the data. Severe noise-alteration of small data vectors directions is now properly taken into account by respective weights. Consequently, the above approach more uniformly traces overall data vectors extraction from the noise floor.

So, we proceed with minimization of the sum in equation (7) using standard methods of linear algebra, that is, we diagonalize new covariance real symmetric $p \times p$ matrix $\Theta^{(j,m)}$ defined as:

$$\Theta^{(j,m)} = (\mathbf{C}^{(j,m)})^T \mathbf{C}^{(j,m)}. \quad (10)$$

After diagonalization we get:

$(\Lambda^{(j,m)})^2 = T^{-1} \Theta^{(j,m)} T$, where T is $p \times p$ orthogonal and $(\Lambda^{(j,m)})^2$ is $p \times p$ diagonal matrix with eigenvalues λ_i^2 of $\Theta^{(j,m)}$ on the diagonal ($i = 1, 2, \dots, p$, and $\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_p^2$).

Eigenvalues λ_i^2 have some obvious properties: If the p -dimensional subspace is dominated by isotropic noise, all eigenvalues are approximately the same due to isotropic distribution of noise variance i.e., $\lambda_i^2 \approx 1/p$ since $\sum_j \lambda_i^2 = 1$.

On the contrary, for clean data and small window widths, the relations amongst λ_i^2 are qualitatively the same as amongst last p respective σ_i^2 (exponential law), since the trajectory points are nested along the main principal axes of the

p -subspace, i.e., in such situation, λ_1^2 is close to unity.

Therefore, we choose the largest eigenvalue λ_1^2 of associated real symmetric $p \times p$ matrix $\Theta^{(j,m)}$, to be our quantitative measure of attractor's directional properties for selected p -subspace that will reveal the presence of noise in respective subspace and that will change gradually as a function of window width τ_w . We term this quantity *anisotropy index* of selected p -subspace, and denote it $\mathbf{I}_{j,p}$. Such measure strongly depends on the amount of noise in the data and partially reflects the behavior of the last important singular value σ_j of trajectory matrix. It eventually changes as we increase window width, due to attractor spreading throughout the p -subspace. It turns out (and will be demonstrated in the next chapter) that such functional dependence could be used for setting adequate window width, that is related to chosen total embedding dimension, for suitable reconstruction.

Finally, we select subspace spanned by last $p = m - j + 1$ eigenvectors to be our referent subspace for calculation that will result in setting up the criterion for adequate reconstruction in the presence of noise. By definition, this determines submatrix $\mathbf{C}^{(j,m)}$.

Since the fractal dimension of the attractor and the content of noise in the data are generally not known, in a real experiment selection of embedding parameters, such as d_E and p , is established through iterative procedure comprising the analysis of system's dynamical invariants [1, 2]. Nevertheless, in PCA approach it is a common practice to choose high d_E for filtering purposes [5, 12], and some smaller j -dimensional subspace spanned by first j eigenvectors ($j < d_E$), that presumptively assures embedding, as a starting point of iteration process. A more thorough account on defining minimal embedding j -subspace is given at the end of next chapter.

3. Common Examples

We apply the above-described concepts to three well-known low dimensional attractors in chaotic regime i.e. Lorenz, Rössler and Duffing, as they were described by Broomhead and King

[5], Fraser and Swinney [8] and Buzug et al. [6] respectively. Calculations of anisotropy index $\mathbf{I}_{j,p}$ were performed with various number of data points, different sample times and embedding dimensions. Several types and different amounts of noise were added to the data sets.

To allow qualitative comparison of the results for different dimensions of p -subspaces (different embedding dimensions) we introduce normalization of $\mathbf{I}_{j,p}$ as follows:

$$(\mathbf{I}_{j,p})_{\text{norm}} = \check{\mathbf{I}}_{j,p} = (\mathbf{I}_{j,p} - 1/p)/(1 - 1/p).$$

Normalized in such way, $\check{\mathbf{I}}_{j,p}$ ranges theoretically from zero for isotropic, to unity for ideally anisotropic state vector space distribution. The calculations were performed for $m = d_E = 5$ up to $m = 35$ and with the addition of flat (uniform) noise with zero mean, ranging from zero to about 40% noise content. The same was repeated with simulated equivalent quantization error of approximately equal signal to noise ratio (data were quantized in the range from 3 to 8 bits of precision). To examine the stability of $\check{\mathbf{I}}_{j,p}$, we use various number of data points. Specifically, 2 000, 5 000 and 20 000 points of $\mathbf{x}(t_i)$ were used for calculations.

Since we are dealing with known low dimensional attractors, we use 3-dimensional subspace ($j = 3$) that assures embedding [5, 6, 8], so we choose $p = d_E - 3 + 1$, in all three cases.

3.1. Results and Discussion

We used Mathematica v. 4.0 program for solving differential equation systems. We adopted resulting time series of 20 000 points of $\mathbf{x}(t_i)$ that describes the attractors with original time resolution of 0.005, 0.0314 and 0.01 for Lorenz, Rössler and Duffing equations respectively. The typical result of $\check{\mathbf{I}}_{3,13}$ ($d_E = 15$) calculations for Lorenz system, using 5 000 data points and 4 different levels of noise, is presented in Fig. 1.

The dashed curves in Fig. 1. present $\check{\mathbf{I}}_{3,13}$ for 4- and 6-bit quantization noise respectively. Below 4-bit quantization, the discrepancy from isotropic distribution becomes too large and thus misleading. For 8-bit quantization noise,

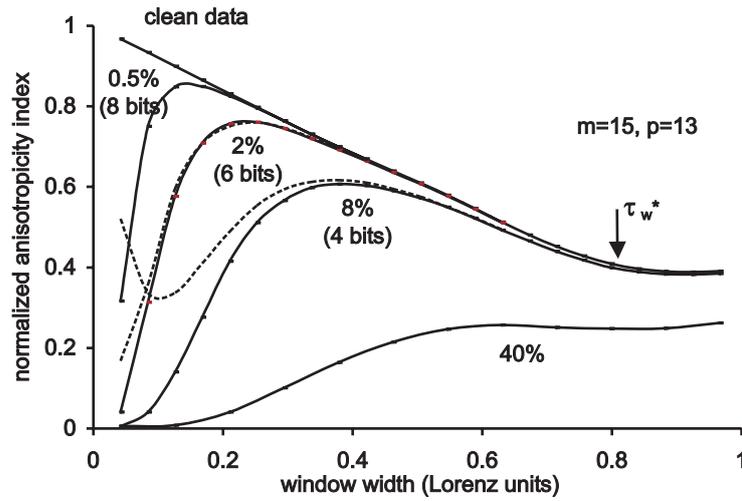


Fig. 1. Normalized anisotropy index $\check{\mathbf{I}}_{3,13}$ curves for the Lorenz attractor with different noise content ($\tau_s = 0.005$, $N = 5000$).

$\check{\mathbf{I}}_{3,13}$ curves are indistinguishable from respective isotropic noise curves (the content of isotropic noise was chosen to have approximately the same rms as quantization noise). Results for Rössler and Duffing systems were essentially the same, except for different window width scales.

In Fig. 1., the upper curve (almost straight line) with negative slope, representing $\check{\mathbf{I}}_{3,13}$ for clean data, is typical for deterministic (noise-free) behavior and it reflects gradual spreading of attractor projection in p -dimensional subspace due to increase of window width. When we add isotropic noise to the data, $\check{\mathbf{I}}_{3,13}$ starts at some lower level (defined by the amount of noise) and steeply increases until reaching its maximum at some τ_w . After that point, it begins to fall and approximate clean data behavior.

So, we clearly recognize two different regions of $\check{\mathbf{I}}_{3,13}$, one prior to the maximum that we indicated as “noise-dominated” region in p -subspace, and another after the maximum, which is comparable to clean data behavior. For smaller noise content and larger τ_w , the clean data curve and the noise curve become practically indistinguishable.

The term “noise-dominated” in presented context could be more precisely qualified as follows: Gradual increase of $\check{\mathbf{I}}_{j,p}$ traces the extraction of σ_j from the noise floor in p -subspace. At some point the extraction is slowed down due to the increase of deterministic component of

higher singular values which prevails over the noise vectors. Prior to that point there is a noise-dominated region in p -subspace, because directional properties presented by $\check{\mathbf{I}}_{j,p}$ are essentially governed by the superior value of the projection of noise vectors (constant noise floor). Soon after, deterministic directional properties, due to spreading of the attractor in higher dimension, start to dominate and $\check{\mathbf{I}}_{j,p}$ curves approximate clean data behavior. The above described qualitative analysis suggests that window width $(\tau_w)_{\max}$, for which the maximum of respective $\check{\mathbf{I}}_{j,p}$ curves occurs (the turning point), is approximately defined by following relation:

$$\sum_{i=j+1}^m \bar{\sigma}_i^2 \approx (p-1) \langle \eta^2 \rangle, \quad (11)$$

where $\bar{\sigma}_i^2$ denotes clean data eigenvalues, and $\langle \eta^2 \rangle$ is noise variance.

Since for isotropic noise process and infinite data series $\sigma_i^2 = \bar{\sigma}_i^2 + \langle \eta^2 \rangle$, the relation (11) can be rewritten as

$$\sum_{i=j+1}^m \sigma_i^2 \approx [(p-1) + (m-j)] \langle \eta^2 \rangle$$

and with $p = m - j + 1$ we finally get:

$$\sum_{i=j+1}^m \sigma_i^2 \approx 2(p-1) \langle \eta^2 \rangle. \quad (12)$$

Thus, the maximum of $\check{\mathbf{I}}_{j,p}$ curves defines boundary, after which extraction of the last important eigenvalue from the noise floor is no more dominant for directional properties of attractor projection on p -subspace. Therefore, it is reasonable to select that window width as the lower boundary for suitable reconstruction of the attractor, judged by its p -subspace noise-related directional properties. Furthermore, relation (12) could be used to estimate the noise content in the data, assuming that the noise process is close to isotropic. We checked it for all three attractors and for all 4 different levels of noise. The agreement was within 3% of the percentage noise content.

Repeated calculations of $\check{\mathbf{I}}_{3,13}$ using 2 000 and 20 000 points produced the same results i.e. the maximums of $\check{\mathbf{I}}_{3,13}$ curves occurred at the same window width. However, the $\check{\mathbf{I}}_{3,13}$ values were somewhat different due to the variation in average noise effect when a smaller set was used.

That shows stability of $\check{\mathbf{I}}_{j,p}$ as a measure of attractor spreading, and, consequently, enables us to perform calculations with smaller sets of trajectory points.

The dependence of $(\tau_w)_{\max}$, associated with $(\check{\mathbf{I}}_{3,p})_{\max}$ on embedding dimension d_E , for all three attractors, was calculated and is presented in Figure 2. The data with 2% noise content were used in calculations. For small d_E , slight increase of $(\tau_w)_{\max}$ with d_E is obvious. The increase gradually slows down and for $d_E \geq 10$, the curves (almost constant) asymptotically approach some limiting $(\tau_w)_{\max}$ that is dependent on the noise content in the data (Fig. 1.).

In view of the condition (11), such asymptotic behavior is expected, because respective eigenvalues $\bar{\sigma}_i^2(d_E)$ are roughly proportional to d_E and $p = d_E - j + 1$.

Since in PCA applications we generally use high embedding dimensions for filtering the noise, we choose that approximate limiting window width ($\tau_w = (\tau_w)_{\max}$; $d_E \geq 10$) as the lower boundary for suitable reconstruction of the attractor. We conclude that proposed measure is effectively defined by the window width, regardless of the specified time delay and embedding dimension. The importance of the window width in the reconstruction process was reported by other authors also [2, 14].

Although $\check{\mathbf{I}}_{j,p}$ and $(\tau_w)_{\max}$ were defined in fixed dimensional embedding space, in practical PCA application we should generally increase d_E by fixed lag time τ (=sample time t_s) until we reach a sufficiently large window width for our reconstruction ($\tau_w = \tau(d_E - 1)$). Such approach leads to maximal filtering effect for the chosen window width.

Accumulation of $(\check{\mathbf{I}}_{3,13})_{\max}$ with higher noise content, near some limiting τ_w^* where the $\check{\mathbf{I}}_{3,13}$ curve slope for clean data starts to change, is obvious (indicated by arrow in the Fig. 1.). In our example, this occurs at $\tau_w^* \sim 0.8$. It exhibits similar asymptotic behavior with the increase of d_E as $(\tau_w)_{\max}$. It presents the upper boundary for reconstruction in fixed dimensional space, above which no significant improvement in SNR could be expected.

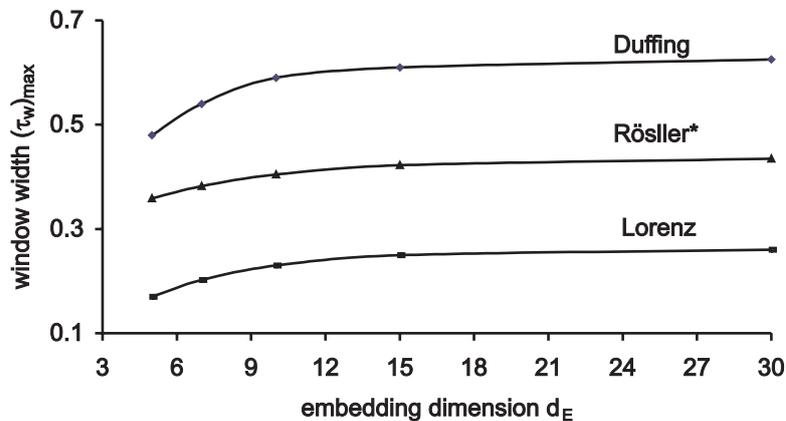


Fig. 2. Dependence of $(\tau_w)_{\max}$ on embedding dimension

*To accommodate the scaling in the figure, window width values for the Rössler system are divided by 2π .

What initial value of window width in the deterministic interval ($(\tau_w)_{\max} < \tau_w < \tau_w^*$) of p -subspace should one use for reconstruction? As a general rule, for the data with small noise content ($< 2\%$), we propose τ_w that lay well below τ_w^* where attractor' structure maintains its simple shape. So, we propose τ_w that satisfies simple condition:

$$\tau_w \approx ((\tau_w)_{\max} + \tau_w^*)/c, \quad (13)$$

where constant c complies with: $j - 1 < c \leq j$.

For the data with higher noise content the user should pick τ_w that is closer to $(\tau_w)_{\max}$ to avoid possibly complex attractor structure for too large τ_w , without significant increase of SNR. Accordingly, for such cases we propose more subtle relation:

$$\begin{aligned} \tau_w &\approx (\tau_w)_{\max} + (\tau_w^* - (\tau_w)_{\max})^2 / (c\tau_w^*), \\ \text{with } c &= j. \end{aligned} \quad (14)$$

In both (13) and (14) equations we define constant c proportional to the minimal embedding dimension of the attractor i.e., $c \sim j$. That is due to the fact that sizeable extraction of the higher eigenvalues from the noise floor occurs at higher τ_w [5, 9], that is, $(\tau_w)_{\max}$ is roughly proportional to j . Because of that, setting $c \sim j$ prevents us from using too large window widths when encountering higher j -dimensional attractors.

It is important to notice that for high noisy data the $\check{\mathbf{I}}_{j,p}$ curves deviate from their typical

shape, because the beginning of deterministic-dominated region of p -subspace enters into oscillatory area of clean data curves. For such cases the maximum of $\check{\mathbf{I}}_{j,p}$ curves is no more well defined and we cannot tell, with certainty, if the p -subspace deterministic region could be reached at all.

If the noise content exceeds a threshold that is dependent on data measurement sample rate, it is generally recommended to perform some pre-filtering of the data before PCA based reconstruction process. It is also dependent on the user application. When using reconstruction for dimension calculations that are highly sensitive to noise [14], pre-filtering is almost unavoidable. Generally, it is recommended for data with noise content over 10%. Nevertheless, even for the data with smaller noise content the use of PCA is effective only if sufficiently high sampling is available ($t_s \ll \tau_w^*$). To avoid destruction of dynamic invariants we recommend the use of SVD-based FIR (finite impulse response filters) linear filtering such as described in [4, 21]. For predictive modeling, besides filtering the data, it is more important to avoid complex attractor structure when too large window width is used [9], which is accomplished through equations (13) and (14).

As an illustration of adequacy of relation (14) we present the PCA reconstruction of the Lorenz attractor ($\tau_s = 0.005$, $N = 5000$) using 8% noisy $\mathbf{x}(t)$ data in Figures 3 and 4.

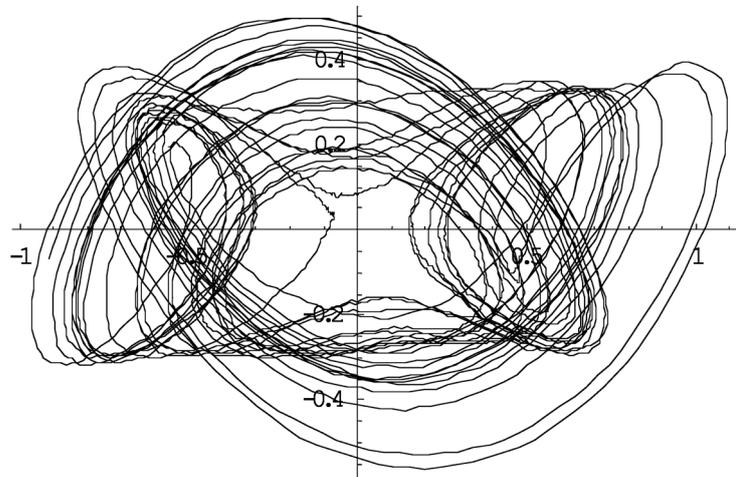


Fig. 3. Lorenz attractor embedded from 8% noisy $\mathbf{x}(t)$ data and projected on the plane spanned by the second and third eigenvectors using $\tau_w = 0.44$ ($N = 5000$, $d_E = 89$, $\tau_s = 0.005$).

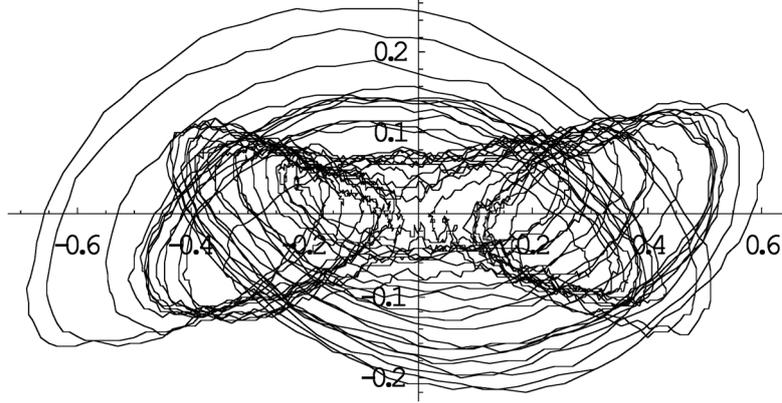


Fig. 4. Lorenz attractor embedded from 8% noisy $\mathbf{x}(t)$ data and projected on the plane spanned by the second and third eigenvectors using $\tau_w = 0.25$ ($N = 5\,000$, $d_E = 51$, $\tau_s = 0.005$).

	Lorenz			Rössler			Duffing		
τ_w	0.32			1.95			0.75		
d_E	3	5	7	3	5	7	3	5	7
τ_w [11]	0.26	0.3	0.32	2.1	2.5	2.6	0.42	0.5	0.53
τ_w [16]	—	—	—	1.75	—	—	—	—	—
τ_w [9]	0.32	—	—	—	—	—	—	—	—

Table 1. Comparison of the window width τ_w with the results by other authors.

Figure 3 presents projection of the Lorenz attractor in the plane spanned by the second and third eigenvector using the window width according to relation (14) with $c = j = 3$, while Figure 4 presents the same projection using inadequate window width (undersized d_E).

Comparison of the results for window width τ_w for clean data using relation (13) and $c = 2.5$ with available data from relevant literature is presented in Table 1.

Agreement of the results from various sources with our approach is very good. It reflects the fact that all of them are based on the consideration of some integral spreading features of the attractors.

Since for generally unknown chaotic process we have no information on intrinsic attractor dimension, calculation of some dynamic invariants such as correlation dimension can be used to find minimal j for reconstruction [1,2].

A good choice is to start with $\tau_w \approx ((\tau_w)_{\max} + \tau_w^*) / (j_{\min} - 1)$ for some assumed j_{\min} as the initial value for calculations. The minimal dimension for which the convergence of attractor correlation dimension is established defines appropriate minimal j value. The approximate values of $(\tau_w)_{\max}$ and τ_w^* could be determined using a smaller set of data with fixed $d_E \geq 10$. It is also possible to use false nearest neighbors (FNN) algorithms [1] for determination of final j_{\min} after filtering and after selection of proper τ_w through presented PCA approach.

4. Conclusion

We have presented the noise-related characterization of chaotic trajectories that provides users with quantitative measure for adequate reconstruction of chaotic attractors in the realm of PCA methodology. Introduced approach defines the lower boundary for “deterministic”

reconstruction, based on attractor' directional properties when projected on suitably chosen principal component subspace. The method is applicable for low-dimensional continuous chaotic dynamical systems (flows) contaminated with isotropic-like noise. Such noise is usually encountered in most experimental measurements of chaotic signals from various artificial (electric circuits, chemical reactions, fluid mechanics etc.) biological or environmental sources [3].

Proposed values of window width for "deterministic" reconstruction are closely related to the amount of noise in the data. Such relation balances two opposing effects in embedding process i.e. filtering out the noise through attractor spreading in high dimensional embedding space and avoiding ineffective excessive spreading when too complex attractor structure occurs. The calculations of introduced anisotropy index can be performed on relatively small data set (≈ 2000 data points) without violating the results.

References

- [1] H.D.I. ABARBANEL, R. BROWN, J.J. SIDOROVICH AND L.S. TSIMRING, "The analysis of observed chaotic data in physical systems", *Rev. Mod. Phys.*, Vol. 65, No. 4, (1993), pp. 1331–1392.
- [2] A.M. ALBANO, J. MUENCH AND C. SCHWARTZ, "Singular-value decomposition and the Grassberger-Procaccia algorithm", *Physical Review A*, Vol. 38, No. 6., (1988), pp. 3017–3026.
- [3] J. BHATTACHARYA AND P.P. KANJILAL, "On the detection of determinism in a time series", *Physica D*, Vol. 132, (1999), pp. 100–110.
- [4] D.S. BROOMHEAD, J.P. HUKU AND M.R. MULDOON, "Linear Filters and Non-linear Systems", *J.R. Statistic Society*, Vol. 54, No. 2, (1992), pp. 373–382.
- [5] D.S. BROOMHEAD AND G.P. KING, "Extracting qualitative dynamics from experimental data", *Physica D*, Vol. 20, (1986), pp. 217–236.
- [6] TH. BUZUG AND G. PFISTER, "Optimal delay time and embedding dimension for delay- time coordinates by analysis of the global static behavior of strange attractors", *Physical Review A*, Vol. 45, No. 10, (1992), pp. 7073–7084.
- [7] M. CASDAGLI, S. EUBANK, J.D. FARMER AND J. GIBSON, "State space reconstruction in the presence of noise", *Physica D*, Vol. 51, (1991), pp. 52–98.
- [8] A.M. FRASER AND H.L. SWINNEY, "Independent coordinates for strange attractors from mutual information", *Physical Review A*, Vol. 33, No. 2, (1986), pp. 1134–1140.
- [9] J.F. GIBSON, J.D. FARMER, M. CASDAGLI AND S. EUBANK, "An analytic approach to practical state space reconstruction", *Physica D*, Vol. 57, (1992), pp. 1–30.
- [10] H. KANTZ AND E. OLBRICH, "Scalar observations from a class of high dimensional chaotic systems — limitations of the time delay", *Chaos*, Vol. 7, No. 3, (1997), pp. 423–429.
- [11] G. KEMBER AND A.C. FOWLER, "A correlation function for choosing time delays in phase portrait reconstructions", *Physics Letters A*, Vol. 179, (1993), pp. 72–80.
- [12] G.P. KING, R. JONES AND D.S. BROOMHEAD, "Phase portraits from a time series", *Nuclear Physics B (Proc. Suppl.)*, Vol. 2, (1987), pp. 379–390.
- [13] E.J. KOSTELICH, "Noise reduction in chaotic time-series data: A survey of common methods", *Physical Review E*, Vol. 48, No. 3, (1993), pp. 1752–1763.
- [14] D. KUGIUMTZIS, "State space reconstruction parameters in the analysis of chaotic time series — the role of the time window length", *Physica D*, Vol. 95, (1996), pp. 13–28.
- [15] D.C. LAY, *Linear Algebra and its Applications*, 2nd ed., Addison-Wesley, 1997.
- [16] W. LIEBERT, K. PAWELZIK AND H.G. SCHUSTER, "Optimal embeddings of chaotic attractors from topological considerations", *Europhysics Letters*, Vol. 14, No. 6, (1991), pp. 521–526.
- [17] J.M. LIPTON AND K.P. DABKE, "Reconstructing the state space of continuous time chaotic systems using power spectra", *Physics Letters A*, Vol. 210, No. 4–5, (1996), pp. 290–300.
- [18] N.H. PACKARD, J.P. CRUTCHFIELD, J.D. FARMER AND R.S. SHAW, "Geometry from a time series", *Phys. Review Letters*, Vol. 45, (1980), pp. 712–716.
- [19] T.S. PARKER AND L.O. CHUA, "Chaos: A tutorial for engineers", *Proc. IEEE*, Vol. 75, No 8, (1987), pp. 982–1008
- [20] H.G. SCHUSTER, *Deterministic Chaos*, 2nd ed., Physik Verlag, Weinheim, 1988.
- [21] K. SHIN, J.K. HAMMOND AND P.R. WHITE, "Iterative SVD Method for Noise Reduction of Low-dimensional Chaotic Time Series", *Mechanical Systems and Signal Processing*, Vol. 13, No. 1, (1999), pp. 115–124.
- [22] F. TAKENS, "Detecting strange attractors in turbulence", *Lecture Notes in Mathematics*, Vol. 898, eds. D.A. RAND AND L.S. YOUNG, Springer, Berlin, (1981), pp. 366–381.
- [23] M. ZOLTOWSKI, "Adaptive reconstruction of chaotic attractors out of their single trajectories", *Signal Processing*, Vol. 80, No. 6, (2000), pp. 1099–1113.

Received: April, 2002
Revised: March, 2003
Accepted: March, 2003

Contact address:

Ivan Michieli
Electronic Department
Ruđer Bošković Institute
Bijenička 54
10000 Zagreb, Croatia
Phone: +385 1 4561 023
e-mail: mic|hieli@faust.irb.hr

IVAN MICHIELI received his M.S. and Ph.D. degrees in nuclear engineering from the Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb. Since 1986 he has been with Ruđer Bošković Institute and since 1998 as scientific associate at the Electronics Department. His research interests are in the fields of stochastic and nonlinear data characterization and processing as well as in mathematical modeling of random processes from nuclear and other energy systems. He is the author and co-author of more than 30 scientific papers and technical reports. (Home page, <http://faust.irb.hr/~mic|hieli>)

BOŽIDAR VOJNOVIĆ graduated and got his M.S. and Ph.D. degrees in electronic engineering from the Faculty of Electrical Engineering, University of Zagreb. Since 1962 he has been with the Institute Ruđer Bošković, Zagreb, involved in many research projects as participant or leading scientist. His research interest includes fast pulse electronics, measurement and analysis of stochastic signals and processes as well as system reliability and effectiveness. He is a member of IEEE, IMEKO and MIDEM societies.
