ELSEVIER

Signal Processing 53 (1996) 133–148

SIGNAL PROCESSING

Application of methods based on higher-order statistics for chaotic time series analysis

Olivier Michel*, Patrick Flandrin

Ecole Normale Supérieure de Lyon, Laboratoire de Physique (URA 1325 CNRS), 46 allée d'Italie, 69364 Lyon Cedex 07, France

Received 1 March 1995; revised 15 April 1996

Abstract

The aim of this paper is to illustrate some applications of HOS within the context of chaotic time series analysis. After reviewing briefly some of the most popular methods and approaches used in chaotic signal analysis, we show how HOS may lead to some significant improvement. First, an HOS expansion of the mutual information is shown to provide an easy way to estimate the reconstruction delay that must be used in the embedding reconstruction method. Then, a fourth-order extension of the local intrinsic dimension analysis (LID) is proposed. The ability of this HOS extension to separate between chaotic and stochastic behaviour is illustrated by examples on simulated data and experimental time series.

Zusammenfassung

Ziel dieses Beitrags ist die Illustration einiger HOS-Anwendungen im Zusammenhang mit der Analyse chaotischer Zeitreihen. Nach einem kurzen Rückblick auf die bekanntesten Methoden und Ansätze, die man zur Analyse chaotischer Signale verwendet, zeigen wir, wie HOS zu bedeutenden Verbesserungen führen können. Zunächst wird gezeigt, daß eine HOS-Entwicklung der gegenseitigen Information einen einfachen Weg zur Schätzung der Rekonstruktionsverzögerung bietet, die man in der eingebetteten Rekonstruktionsmethode anwenden muß. Dann wird eine Erweiterung vierter Ordnung für die Analyse lokaler intrinsischer Dimension (LID) vorgeschlagen. Die Fähigkeit der HOS-Erweiterung zur Trennung zwischen chaotischem und stochastischem Verhalten wird durch ein Beispiel mittels simulierter Daten und experimenteller Zeitreihen illustriert.

Résumé

Le but de cet article est de présenter quelques applications des SOS dans le contexte de l'analyse des séries temporelles chaotiques. Après une brève révision de quelques-unes des méthodes et approches les plus populaires utilisées en analyse des signaux chaotiques, nous montrons comment les SOS peuvent conduire à des améliorations significatives. Tout d'abord, on montre qu'une expansion en SOS de l'information mutuelle fournit un moyen aisé pour estimer le délai de reconstruction qui doit être utilisé dans la méthode de reconstruction par immersion. On propose également une extension au quatrième ordre de l'analyse de la dimension intrinsèque locale (DIL). La capacité de cette extension SOS à discriminer entre comportements chaotique et stochastique est illustrée par des exemples à partir de données simulées et de séries temporelles expérimentales.

*Corresponding author.

PII S0165-1684(96)00082-5

1. Introduction

When dealing with signals that exhibit irregular behavior, the most widely accepted approach consists of modelling it as the realization of some stochastic process. Until recently, this approach was the only one available. In such models, irregularity is implicitly associated with randomness within a generation mechanism. In this respect, one of the most common paradigms of signal processing consists in describing an irregular signal as the output of a linear system driven by a purely stochastic process, usually white Gaussian noise.

However, although such an approach may prove extremely useful in numerous engineering problems, it is nowadays well-recognized that, in some cases, irregularity may also stem from some purely deterministic (i.e., non-random) non-linear systems, exhibiting a very high sensitivity to initial conditions. This situation is referred to as *chaos*. Chaos offers therefore a new paradigm and it allows one to describe the corresponding signals from a completely new perspective, thus requiring the development of specific analysis tools. (General references concerning chaos and its analysis are, e.g., [1, 4, 13, 29, 32, 35]). Most of the studies aim at *measuring* chaos that may be embedded in noise, but rely on the assumption that the underlying process *is* actually chaotic. In these studies, *fractional dimension* plays a key role. However, it has been recently shown that a fractional dimension is not a feature of chaotic signal *only*. As an example, AR(1) or $1/f_\alpha$ processes [30, 4] may exhibit a low-dimensional and non-integer fractal dimension in the reconstructed phase space, although they are purely stochastic processes.

Because of the nature of chaos (non-linear systems, non-Gaussian statistics, ...), higher-order techniques are likely to play an important role in algorithms aimed at chaotic signals. It is the purpose of this paper to provide a brief introduction to chaotic signal analysis and its main algorithms (some classical, some new), emphasizing the usefulness and the relevance of higher-order concepts in this context. The detection problem (does the system present any chaotic behavior?) rather than the estimation problem (what are the characteristics of the observed chaotic system?) is emphasized.

In the next section, basic definitions and properties of chaos are briefly given. The whole study presented in this paper will be based on a reconstruction of the phase space of chaotic attractors, performed by using the celebrated time-delay embedding method. In Section 3, this latter embedding method is briefly reviewed and the role of HOS in the experimental determination of the embedding parameters is examined. A new insight into the interpretation of local intrinsic dimensionality (LID) is provided by the use of independent component analysis (ICA) [9]. This method, based on fourth-order cumulants, is shown to bring a partial answer to the deterministic versus stochastic separation problem, one of its most important features being that ICA relates directly to the underlying dynamical system, whereas the Grassberger–Proccacia Algorithm, for instance, is a measure of geometrical properties of the attractor.

2. A brief review of chaos

2.1. Definitions

An observed signal $x(t)$ is usually considered to partially represent information about a stochastic process (a realization). Here we will rather view it as partial information related to the deterministic evolution of a dynamical system. By definition, a *dynamical system* is characterized by a *state* $\boldsymbol{X} \in \mathbf{R}^n$, whose time evolution is governed by a *vector field* $f: \mathbf{R}^n \to \mathbf{R}^n$ according to the differential equation

$$\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} = f(\boldsymbol{X}). \tag{1}$$

The number n of coordinates in the state vector $\boldsymbol{X}$ characterizes the number of *degrees of freedom* involved in the system and the whole space to which $\boldsymbol{X}$ is allowed to belong is called the *phase space*. Notice that this latter quantity n defines what we will refer to as the *dimensionality* or the number of degrees of freedom of the system, as there is no overall accepted definition for it.

If f does not depend upon time (which will be assumed in the rest of the paper), the system is said

to be *autonomous*. Any solution of an autonomous system expressed by (1) can be written $\boldsymbol{X}(t) = \varphi_t(\boldsymbol{X}_0)$, where $\boldsymbol{X}_0$ stands for some initial conition. This solution is referred to as a *trajectory* in phase space, the mapping φ_t (such that $\varphi_t \circ \varphi_s = \varphi_{t+s}$ and $\varphi_0(\boldsymbol{X}) = \boldsymbol{X}$) being called the *flow*.

A solution of Eq. (1) is characterized asymptotically (i.e., as $t \to \infty$) by a *steady-state* behavior which must be bounded if it is to make sense. Accordingly, the trajectory tends to remain bounded within a subset of the phase space (called the *attractor*), the nature of which heavily depends on f. In the case of *linear* f's the only (bounded) attracting sets are *points* or *cycles*, which restricts the set of possible steady-state trajectories to be fixed points or quasi-periodic orbits (countable sum of periodic solutions). Moreover, the steady-state behavior is unique and is attained whatever the initial condition is.

The situation is quite different and much richer if one considers the case of non-linear dynamics. Different steady-state behaviors can be observed, depending on the initial conditions, and the solutions themselves are not restricted to quasi-periodicity: chaotic motion is one among the many possibilities.

As there is no unique and well-accepted definition of chaos, a signal will be considered here as chaotic if

- it results from a (non-linear) autonomous deterministic system, and
- the behaviour of this system is highly dependent on initial conditions in the sense that trajectories initiated from neighboring points in phase space diverge exponentially as functions of time.

Furthermore, it seems important to emphasize here that in the case of high values of n, most experimental signals may be modelled equally well by stochastic processes. The limit is rather empirical between these approaches. Therefore, we will restrict ourselves to situations which only involve a *low number of degrees of freedom*, i.e. a low dimensionality n.

2.2. Characterizing chaos

According to the above definition, chaotic systems undergo many interesting specific characteristics. Different characterizations of chaotic signals exist, each of them putting emphasis on some specific property.

In phase space, trajectories of a chaotic system with few degrees of freedom converge towards a limit set, an *attractor* which only fills a low-dimensional subset of phase space.[1] Because of the assumptions of (i) boundedness and (ii) non-periodicity, this attractor has necessarily a very peculiar and intricate structure with possibly *fractal* properties, a situation referred to as a *strange attractor*. The existence of a fractal (i.e., non-integer) dimension for an attractor can therefore be used as a hint for a possibly chaotic behavior, the dimension itself being a lower bound for the number of degrees of freedom governing the dynamics of the system. Let us remark that the fractal structure of the attractor associated with a given signal should not be confused with fractal properties pertaining to the signal itself: for instance, some recent studies [30, 39, 41] laid stress on the fact that, e.g., fractional Brownian processes may behave like chaotic processes when studied by using some fractal dimension estimators, thus leading to erroneous conclusions about the deterministic or stochastic nature of the process (see Section 4.1).

Another consequence of the lack of periodicity is the *broadband* nature of the spectrum for chaotic signals, which is therefore a necessary (but, of course, not sufficient) condition for the assessment of chaos.

One of the main features of chaos is the strong dependence of the trajectories on the initial conditions. This property drastically limits any possibility of *long-term prediction*: Sensitivity to initial condition leads to the fact that (initially close) trajectories diverge exponentially with respect to time. A measure of this divergence is provided by the *Lyapunov exponents*. Consider, at some initial time $t = 0$, an infinitesimal hyper-sphere of radius $\varepsilon(0)$ centered on a (e.g. randomly chosen) point of the

[1] It will be assumed throughout the paper that the systems under study are undergoing a steady-state behavior, in the sense that all the observations correspond to state vectors belonging to the attractor (the signal is observed after sufficiently large time after the transient).

attractor. After a short time t, this hyper-sphere will be deformed by the action of the flow into some hyper-ellipsoid whose principal axes $\varepsilon_i(t)$, $i=1,\dots,p$ characterize contracting (respectively dilating) directions if $\varepsilon_i(t)<\varepsilon(0)$ (respectively $\varepsilon_i(t)>\varepsilon(0)$). In this picture, Lyapunov exponents are defined as

$$\lambda_i = \lim_{t\to\infty}\lim_{\varepsilon(0)\to 0}\frac{1}{t}\log\frac{\varepsilon_i(t)}{\varepsilon(0)}.$$

Therefore, a necessary condition for a possible situation of chaos is the existence of at least one *positive* Lyapunov exponent.

2.3. Examples

We will use two experimental chaotic time series as test signals. Both are sampled records measured on a chaotic electronic circuit from Chua's family (see [32]). Both series were sampled from the same circuit, proposed in [41] and described in Fig. 1.

The behavior of this circuit is described by the following set of coupled differential equations:

$$C_2\frac{dV_2}{dt} = \frac{V_R}{R_1} - V_2\left(\frac{1}{R_1}+\frac{1}{R_p}\right) + I_L,$$

$$C_1\frac{dV_R}{dt} = -\frac{V_R}{R_1} + \frac{V_2}{R_1} - g(V_R),$$

$$L\frac{dI_L}{dt} = -V_2 - R_s I_L,$$

where $g(V) = m_0 V + 0.5(m_1 - m_0)(|V+b| - |V-b|)$. This system may exhibit some chaotic behavior, depending on the value taken by R_1. Different tuning of this adjustable resistor allows us to get either periodic signals or chaotic ones, as those referred to as Expl or Exp2 in the rest of the paper.

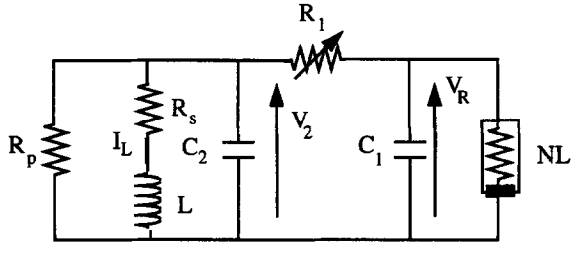

Fig. 1. Schematic representation of the double-scroll Chua circuit from which test signals are measured.

The voltage threshold b and the admittance values m_0 and m_1 are set by the negative impedance converter and the rectifiers used to construct the nonlinear device. For these experiments, we had $C_1 = 5.6$ nF, $C_2 = 47$ nF, $L = 7.5$ mH, $R_s = 3.3\ \text{k}\Omega$, $R_p = 33\ \text{k}\Omega$ and the adjustable resistor $R_{1\max} = 10\ \text{k}\Omega$, $b = 1.55$ V, $m_0 = 0.498\ \text{m}\Omega^{-1}$ and $m_1 = 0.802\ \text{m}\Omega^{-1}$. The experimental signals $V_2(t)$, referred to as *Expl* and *Exp2*, shown in Fig. 2 were obtained for different tuning of the resistor R_1. The initial conditions were set to identical values ($V_R = V_2 = 0$) for both experiments, though they may be considered as being random, due to the presence of (thermal and electromagnetic, as no shielding was used) noise. The corresponding time series were recorded at a sampling rate of 28.8 kHz, and with a 12 bit quantization. The records were performed once the system was "locked" on its attractor, thus insuring that no transient behaviour was present. A thorough discussion about the behavior of the circuit is to be found in [41].

Notice that Expl time series clearly exhibits some nonlinear characteristics, as its switching behavior, whereas Exp2 time series may easily be interpreted (at first glance) as to be a filtered stochastic process (with poles of the filter close to the unit circle). However, there exist some other representations and characteristics that allow to identify Exp2 as stemming from a deterministic system. These points are developed in the next sections.

3. Phase space reconstruction

As the characterization of chaotic signals considered so far are based on properties of state vectors within the phase space, at least n independent time series are to be measured and recorded on the experimental set, n standing for the expected dimension of the phase space. As it was already emphasized, n should be equivalently interpreted here as the number of degrees of freedom of the system under study. In most experimental situations, no sufficient information is available for

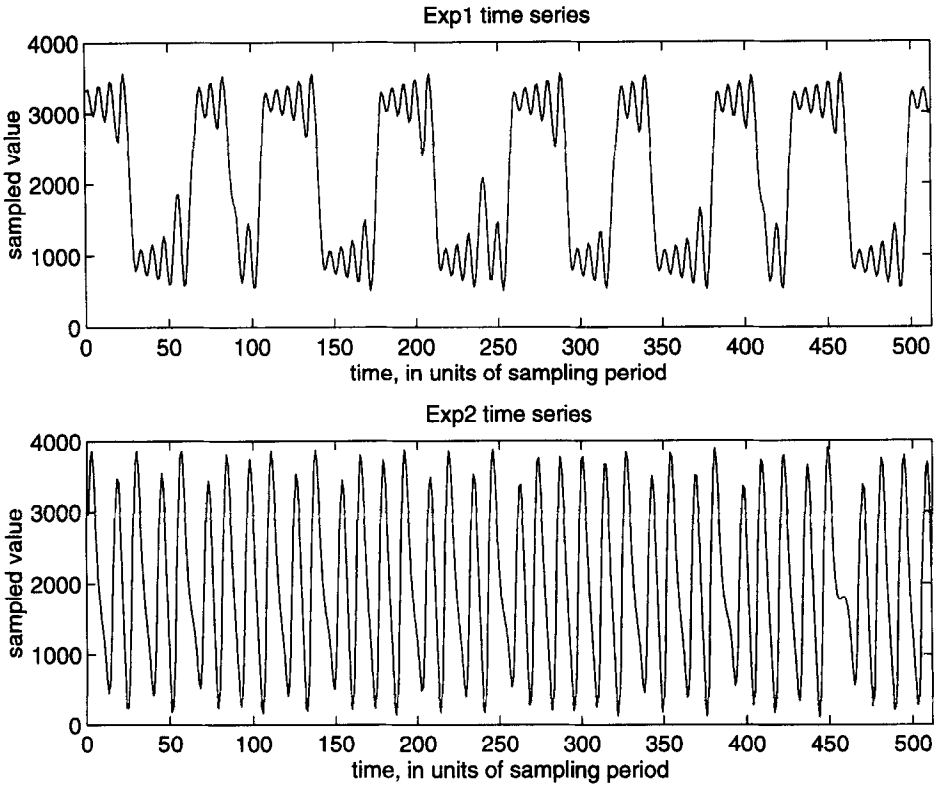

Fig. 2. Experimental time series, Exp1 (above) and Exp2 (below).

properly composing a state vector and studying its evolution. The most critical solution arises when only one measured time series can be recorded. In such a case, the problem is somewhat similar to the one faced within a stochastic framework, when ensemble quantities are to be inferred from the observation of only one realization.

Given an observed one-dimensional time series $x(t)$, considered as only one component of an unknown n-dimensional state vector, the problem is therefore to *reconstruct* an approximate phase space, with the requirement that it be topologically equivalent to the true one.

3.1. Method of delays

A solution to the phase space reconstruction problem was proposed by Whitney [40] in 1936. In 1981 Takens proposed a theorem that extends Whitney's ideas, and provides strong mathematical support for it. This theorem states the following: given one (noiseless) infinitely long observation $x(t)$ and any arbitrary (non-zero) delay τ, the collection

$$x(t), x(t+\tau), \ldots, x(t+(p-1)\tau)$$

defines a reconstructed attractor which is guaranteed to be equivalent (up to some unknown diffeomorphism[2]) to the actual one, provided that $p \geqslant 2D_2 + 1$, where D_2 stands for the correlation dimension of the attractor (see Section 4.1 for

[2] Let $f: \boldsymbol{X} \to f(\boldsymbol{X})$ a function and f^{-1} its inverse, both bijective and continuous. Then f is a *homeomorphism*. Furthermore, if f is differentiable, it is called *diffeomorphism*. Note that continuity here has to be considered in the sense of a topological continuity: $\forall \boldsymbol{X}$, $\exists V(\boldsymbol{X})$, such that if $\boldsymbol{X}' \in V(\boldsymbol{X})$, then $f(\boldsymbol{X}') \in f(V(\boldsymbol{X}))$.

definition of the correlation dimension). Geometrical properties of the reconstructed attractor are therefore identical with the geometrical properties of the real attractor, which makes of this *method of delays* a tool of considerable interest.

When only N data points of a sampled time signal x_n are available, the technique consists of building the vectors

$$\boldsymbol{X}_i = (x_i, x_{i+d}, \ldots, x_{i+(p-1)d}), \quad i = 1, 2, \ldots, N_{pd}, \quad (2)$$

where $N_{pd} = N - (p-1)d$. d stands for elementary delay and p is the *embedding* dimension chosen (as previously) such that $p \geqslant 2D_2 + 1$.

The embedding dimension p is a priori a free parameter which, given a signal, can be used as a variable when looking for chaotic dynamics. Indeed, if we suppose that a dimension estimation of the (reconstructed) attractor is performed (e.g. estimating D_2 by using the Grassberger–Procaccia algorithm, see Section 4.1) with increasing embedding dimensions p's, it is expected that the estimated dimension will vary significantly until the effective dimension D_2 of the attractor is attained. Therefore, chaotic dynamics are expected to lead to a saturation of the estimated dimension (at a low, and generally non-integer, value), as a function of the embedding dimension p. The same behavior is evidently expected (parameter estimation exhibiting a saturation effect with increasing p's) when it comes to the estimation of the number of degrees of freedom of the system by using rank-based approaches, as discussed in Sections 4.2 and 4.3. Alternatively, stochastic signals (with many degrees of freedom) should explore as many directions within phase space as possible, thus presenting no saturation effect for increasing embedding dimensions.

3.2. The choice of the reconstruction delay

Takens's theorem states that the choice of the delay is theoretically of no importance in the case of noiseless observations of infinite length. However, this may become an important issue from a practical point of view: for delays that are too small, all the coordinates of the reconstruction remain strongly correlated and the estimated dimension tends towards 1; conversely, for too large delays, the coordinates are almost independent so that the dimension is generally close to the embedding dimension p, with no significant relationship with the number of degrees of freedom involved in the dynamics.

A "good" delay should therefore correspond to the smallest value for which the different coordinates of the reconstructed state vector are almost "unrelated" in a sense which has to be made precise. The most popular approach makes use of the first zero of the estimated cross-correlation function between the observation and its delayed version

$$\hat{C}_x(\tau) = \frac{1}{(N-\tau)(N-\tau-1)} \sum_{k=\tau+1}^{N} \tilde{x}_k \tilde{x}_{k-\tau}, \quad (3)$$

where $[\tilde{x}_k]_{k=1,N}$ stands for the centered sampled version of $x(t)$. This is clearly a gross indicator which only concerns second-order independence. A more global criterion, relying on the concept of *general* (not just second-order) independence, has therefore been proposed in terms of *mutual information* (see [31]) between the time series and a delayed version of it. The optimum delay is chosen so as to coincide with the first minimum of the mutual information function[3] [37, 17]. Efficient algorithms have been developed towards this end, thus permitting an improved phase space reconstruction, but at the expense of huge computation times (in the general case) [17, 18].

This problem of optimum selection of a delay is one instance in which techniques based on higher-order statistics may prove useful: they allow some improvement as compared to second-order based criteria, while remaining of a reduced complexity, as compared to "all-order" methods.

The general framework for *independent component analysis* of vectors can be described as follows [10, 9]. Let p_x and p_{x_i} be the probability

[3] Though the processes considered here are purely deterministic, the collection of all processes related to different initial conditions can be embedded within a stochastic framework [13]. This point of view will be underlying most of the concepts and ideas handled throughout this paper, with appropriate ergodic estimates in the case where only one time series (one "realization") is available.

density functions (pdf) of a set of vectors $\boldsymbol{x} = (x_1, \ldots, x_i, \ldots, x_p)^T$ and of its components x_i, respectively. If all the components are statistically independent,

$$p_x(\boldsymbol{u}) = \prod_{i=1}^{p} p_{x_i}(u_i). \tag{4}$$

The purpose of independent component analysis is therefore to search for a linear transformation that minimizes the statistical dependence between the vector components. To this end, a useful distance measure is the *Kullback divergence* which, in this case, is given by

$$d\left(p_x, \prod_{i=1}^{p} p_{x_i}\right) = \int p_x(\boldsymbol{u}) \log \frac{p_x(\boldsymbol{u})}{\prod_{i=1}^{p} p_{x_i}(u_i)} \, \mathrm{d}\boldsymbol{u}. \tag{5}$$

Eq. (5) also turns out to be the average *mutual information*

$$I_m\left(p_x, \prod_{i=1}^{p} p_{x_i}\right) = S(p_x) - S(p_{x|\prod_{i=1}^{p} p_{x_i}}), \tag{6}$$

where $S(p_x)$ stands for the Shannon differential entropy of $\boldsymbol{x}$, and $S(p_{x|y})$ denotes the entropy associated with the conditional pdf $p_{x|y}$. Let $J(p_x)$ be the negentropy of the distribution p_x, defined by

$$J(p_x) = S(\phi_x) - S(p_x), \tag{7}$$

where ϕ_x is the Gaussian probability density function having the same mean and variance as p_x. By substituting (7) in (6), it may proved that

$$I_m\left(p_x, \prod_{i=1}^{p} p_{x_i}\right) = -S(\phi_x) + \sum_{i=1}^{p} S(\phi_{x_i}) + J(p_x) - \sum_{i=1}^{p} J(p_{x_i}). \tag{8}$$

Remarks

- The negentropy of the distribution p_x may also be shown to be equal to the Kullback divergence between ϕ_x and p_x; this latter property, together with the positiveness property of the Kullback divergence, shows that the Gaussian distribution (with a given mean and variance) is maximum entropy distribution (among the set of distribution having regular covariance).
- Let $\bar{M}$ be a $p \times p$ non-singular matrix. $\boldsymbol{Y} = \bar{M}\boldsymbol{X}$ is a new random variable obtained from $\boldsymbol{X}$ through linear transformation. Then, one gets $S(p_y) = S(p_x) + \log(\det(\bar{M}^{-1}))$. As the definition of the negentropy of $\boldsymbol{y}$ implies that ϕ_y has the same variance as p_y, the same transform $\bar{M}$ has to be applied on both ϕ_x and p_x, from which it is seen that the negentropy is invariant with respect to changes of coordinates.
- Introducing a Gaussian probability density function as a reference distribution is natural here, as higher-order cumulants of a distribution measure the departure of that distribution from the Gaussian.

Making use of the fact that $J(p_x)$ is invariant with respect to orthogonal changes of coordinates and that

$$S(\phi_x) = \tfrac{1}{2}(p + p \log 2\pi + \log \det \boldsymbol{V}), \tag{9}$$

where $\boldsymbol{V}$ is the covariance matrix of $\boldsymbol{x}$, the mutual information (8) can be expressed as

$$I_m\left(p_x, \prod_{i=1}^{p} p_{x_i}\right) = J(p_{\tilde{x}}) - \sum_{i=1}^{p} J(p_{\tilde{x}_i}) - S(\phi_x) + \sum_{i=1}^{p} S(\phi_{x_i}), \tag{10}$$

where $\tilde{\boldsymbol{x}}$ is the standardized vector obtained from a Choleski factorization of $\boldsymbol{x}$.

It then becomes possible to approximate the mutual information function via a fourth-order expansion of $J(p_x)$ based on the *Edgeworth expansion* of probability density function [21, pp. 145–150]. For standarized data, and using the Einstein summation convention,[4] this expansion reads [9]

$$J(p_{\tilde{x}}) = \tfrac{1}{12} K^{ijk} K_{ijk} + \tfrac{1}{48} K^{ijkl} K_{ijkl} + \tfrac{3}{48} K^{ijk} K_{ijn} K_{kqr} K^{qrn} + \tfrac{1}{12} K^{ijk} K_{imn} K^{m}_{jr} K^{nr}_{k} - \tfrac{1}{8} K^{ijk} K_{ilm} K^{lm}_{jk} + \mathrm{O}(p^{-2}). \tag{11}$$

The $K_{ij\ldots q}$'s stand for the cumulants of the standarized variables $\tilde{x}_i, \tilde{x}_j, \ldots, \tilde{x}_q$ and p is the number of independent variables in $\boldsymbol{x}$. The convergence of

[4] Presence of the same index in superscript and subscript implies a summation over this index, e.g. $K^{il}K_{im} = \sum_i K_{im}K_{il}$.

such an expansion is related to the fact that cumulants of order i of a sum of p independent variables behave asymptotically as $p^{(2-i)/2}$ (see [21, 9] and references therein).

Fig. 3 shows a comparison between an estimate of the mutual information function (MIF) between the time series and its delayed versions based on Fraser's algorithm [18] and a fourth-order approximation of the MIF using the Edgeworth expansion. The signals under study consists of 2048 data points from Exp1 and Exp2 (cf. the preceeding sections) experimental sampled time series $[x_k]_{k=1,N}$. A set of two-dimensional vectors $(x_k, x_{k+\tau})^T$ is formed, and the preceeding equations are used with $(i, j, k, l) \in 1, 2$. All the cumulants involved in the Edgeworth expansion have been estimated using k-statistics [21, pp. 259–260] providing unbiased estimations.

The agreement between the curves (Fraser's estimation of MIF, and the fourth-order expansion of MIF) is good, especially when it comes to the location of local extrema, which are the quantity of major interest. Further advantages offered by the approach based on the fourth-order expansion are the following: (i) it allows one to compute the variance of the estimated mutual information, (ii) it does not need any empirical threshold (as needed to perform a direct estimation of the pdf p_x by using Fraser's algorithm [17]), (iii) it has an improved efficiency for larger p's (cf. Eq. (11)) and (iv) it requires generally less data points than Fraser's algorithm in the case of signals exhibiting complex trajectories in the phase space.[5]

One should finally note that the global approach reported here generalizes recent studies and supports the claim that, from an empirical point of view, the "embedding window" $(p-1)d$ may be chosen as the characteristic time for which some suitably chosen cumulants (up to order four) are simultaneously close to zero [3]. In fact, when most cumulants simultaneously vanish, so does the MIF. However, some alternative methods based on the "stability" of the reconstructed phase space when p is modified [22] may lead to a different choice for d. We have also noticed that this general approach based on minimum mutual information may fail to give the optimal delay, especially in the case where the time series contains strong spectral lines. This latter point has been briefly discussed in [25].

4. Dimensionality

As was previously stressed, a chaotic system is associated with a low-dimensional attractor whose complex geometry makes it a *fractal* object. Therefore, the estimation of a fractal[6] dimension of an attractor (true or reconstructed) has been considered a clue for giving evidence of chaotic behavior within a given system.

Different definitions of dimensions were proposed. The simplest one (*capacity* dimension) consists in covering the attractor by the minimum number $N(\varepsilon)$ of hypercubes of size ε, and evaluating the quantity

$$D_c = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log(1/\varepsilon)}. \tag{12}$$

A refinement of this definition takes into account the probability P_i with which each of the $N(\varepsilon)$ hypercubes is visited (i.e. the probability that a state vector "falls" within this hyper-sphere defined on the attractor), leading to the *Renyi's (generalized) information dimensions*

$$D_q = \frac{1}{q-1} \lim_{\varepsilon \to 0} \frac{\log(\sum_{i=1}^{N(\varepsilon)} P_i^q)}{\log(1/\varepsilon)}. \tag{13}$$

Though the estimation of any of these quantities is an important issue, it must be emphasized that, even in the case of a reliable estimation, the significance of a result based on purely geometrical

[5] In the general case, it remains difficult to accurately evaluate and compare the complexity of both approaches, as the number of operations involved in Fraser's algorithm depends heavily upon the structure of the attractor: if the structure of the attractor is homogeneous, the algorithm requires very few operations whereas for very intricate structures the convergence is very slow. On the opposite, the fourth-order method requires a constant number of operations, whatever the attractor is.

[6] Many different dimensions for characterizing fractals have been proposed. The interested reader should refer to, e.g., [6] or [27] for interesting discussion.

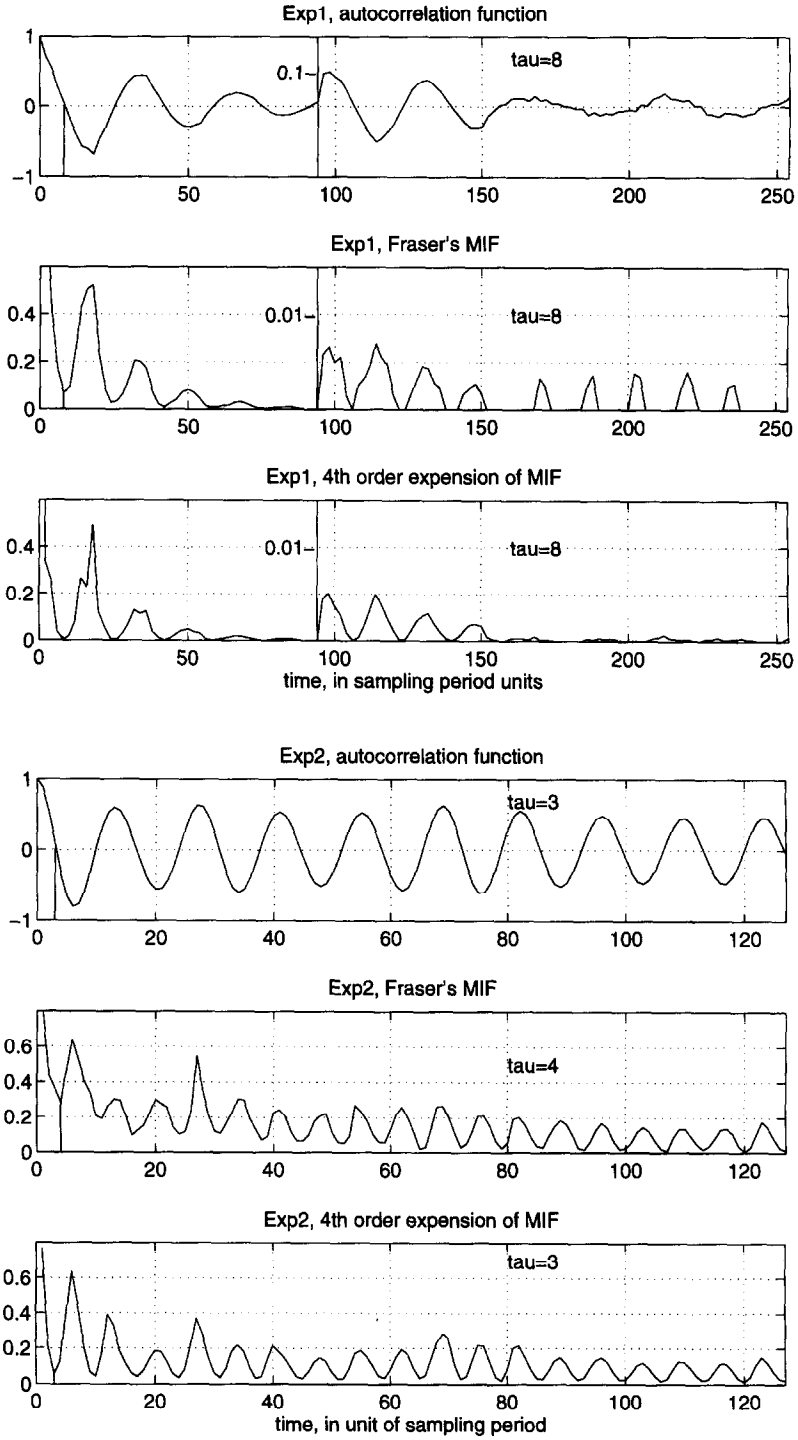

Fig. 3. Estimation of mutual information between experimental time series and delayed versions. The time delay is expressed in unit of sampling period. The upper curve on both plots shows the autocorrelation function, as a basis for comparison with the mutual information (MIF). The curves for the analysis of Exp1 have been rescaled for delays greater than 95 units, for sake of readability. The time series corresponding to Exp1 was here down-sampled by 2.

properties of the attractor should be questioned. It is, for example, not clear whether any dynamical information, such as the number of degrees of freedom governing the system, can be gained from such a static perspective. This problem is another instance in which higher-order statistics may prove useful, as we will illustrate later.

4.1. Correlation dimension

The most widely used fractal dimension is the *correlation dimension* (which is identical with Reny's D_2). Estimation of D_2 is usually performed by using the *Grassberger–Procaccia Algorithm* (GPA) [19]. This algorithm measures for each p, the number of pairs (X_i, X_j) whose distance is less than a given radius r: the following *correlation integral* is computed:

$$C_N(r;p) = \frac{1}{M(N_{pd}-1)} \sum_{i=1}^{M} \sum_{j=1, j\neq 1}^{N_{pd}} U(r - \|X_i - X_j\|), \quad (14)$$

where M stands for the number of test points X_i selected at random on the attractor and U is the unit step function. This approach is highly effective because of the fundamental property of $C_N(r;p)$

$$\lim_{N\to\infty} C_N(r;p) \sim r^{D_2}, \quad r \to 0, \quad p \geqslant 2D_2 + 1, \quad (15)$$

according to which the *correlation dimension D* can be estimated from a slope measurement in a log–log plot of $C_N(r;p)$.

In Fig. 4 we present some results obtained from GPA, applied to either chaotic signals (the

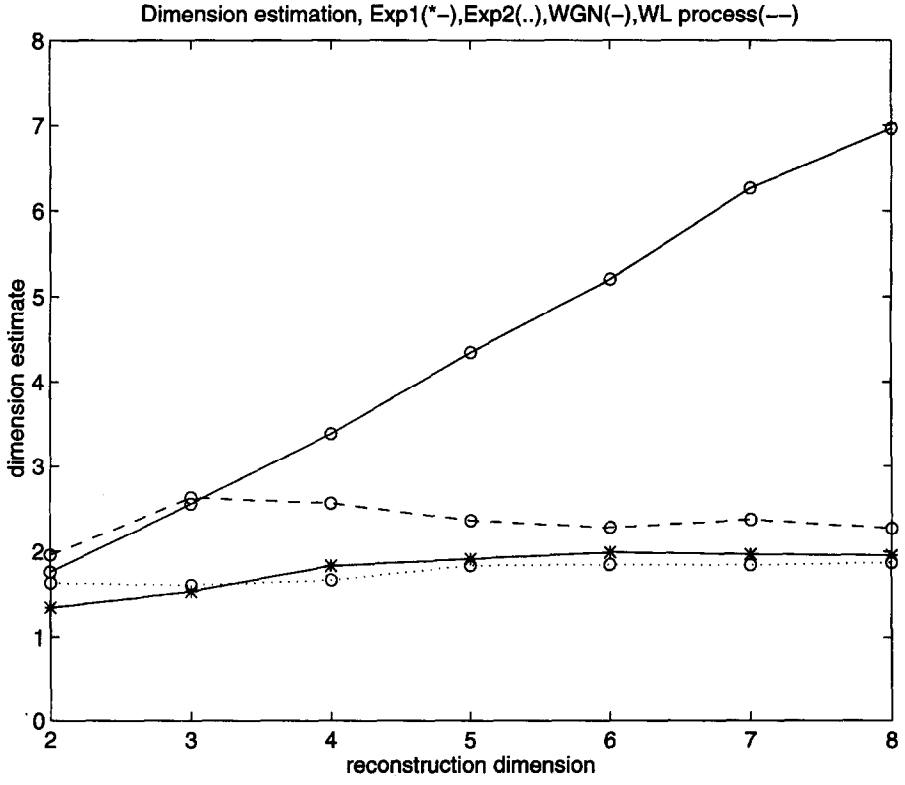

Fig. 4. Results of correlation dimension (D_2) estimation performed by GPA for both stochastic signals and chaotic time series. All estimations were performed from a 32 K point time series, and reconstruction delays were $d=1$, $d=64$ for WGN and WL signals, respectively, and $d=4$ for experimental series Exp1 and Exp2.

experimental time series depicted in Section 2.3 and observed in a "steady-state" regime, so as to guarantee that the observations lie on the attractor (limit set)) or a discrete-time white-noise signal and a Wiener–Lévy process. These results were partially to be expected in the sense that the estimated dimension saturates in the first case, when also the embedding dimension is increased, while in the case of white Gaussian noise it keeps on growing with p. However, a misleading behavior can be observed also in the case of some purely stochastic processes with a "$1/f$" spectrum, as illustrated here for the Wiener–Lévy process.

$$x_n = x_{n-1} + \varepsilon_n, \tag{16}$$

for which a thorough theoretical [30, 39, 41] and experimental [24] analysis justifies a saturation at the value $D_2 = 2$. This is to be explained as follows: although "$1/f$" stochastic processes are unrelated to any chaotic dynamics, they are fractal *signals*, thus leading to fractal reconstructed attractors. This counterexample gives evidence that, in the general case, GPA is much more an *estimation* algorithm (given chaotic dynamics, what is the dimension of the attractor?) than a *detection* one (is there any chaotic dynamics?).

Another drastic limitation of GPA concerns its prohibitive computational load and its reduced effectiveness for small data sets. Roughly speaking, a reliable estimation of a dimension D_2 requires approximately 10^{D_2} data points, which makes the method inapplicable as soon as D_2 exceeds some units [36].

4.2. Rank dimensions

The above limitations of GPA have motivated the search for improvement [2] and alternative methods. Another dimension estimation has been proposed, the simplicity of which made it very attractive: the motivation for the local intrinsic dimensionality (LID) [16, 34] (see also [5]) is to extract some information from local matrices associated with a phase space trajectory, and to reduce the problem of dimension estimation to that of rank determination. The algorithm proceeds as follows: (i) given an embedding dimension p, a number M of test points is selected at random on the attractor; (ii) for each of these points $\boldsymbol{X}_i$, the q-nearest neighbors $\boldsymbol{X}_{i(q)}$ are retained and organized in a $(p \times q)$ matrix

$$\bar{Y}(i) = (\boldsymbol{X}_{i(1)} - \boldsymbol{X}_i, \boldsymbol{X}_{i(2)} - \boldsymbol{X}_i, \ldots, \boldsymbol{X}_{i(q)} - \boldsymbol{X}_i), \tag{17}$$

whose rank is estimated by counting the number of significant singular or eigenvalues, and (iii) the LID is then obtained as the average of these local ranks over the M chosen points:

$$\hat{D} = \frac{1}{M} \sum_{i=1}^{M} \operatorname{rank}(\bar{Y}(i)\, \bar{Y}^{\mathrm{T}}(i)). \tag{18}$$

In this respect, LID may be interpreted as being related to principal component analysis (PCA) of the set of difference vectors in the neighbourhood of a given center.

Whereas GPA is theoretically well-founded, LID is more difficult to justify.[7] In practice, LID tends to overestimate D_2, though it remains generally in good agreement with the estimation issued by GPA. However, LID enjoys the property of a very easy implementation at a low computational cost. This may also be somewhat confusing, as an estimate of the minimum number of coordinates necessary to reconstruct the phase space was expected from the explanations given in the preceding footnote, rather than an estimate of the correlation dimension. Actually, this latter dimension depicts a geometrical characteristic of the fractal attractor (see also next sections).

Results obtained from LID estimations are shown in Fig. 5. The rank of local correlation matrices was estimated by thresholding their sorted eigen-spectrum. The threshold was fixed arbitrarily so as to retain $t = 80\%$ of the total energy, spread over the most significant eigenvalues. The values shown on the plot are the average values of rank estimations performed from a set of m local correlation matrices. These latter were estimated over

[7] The main justification [16] for the LID algorithm stems from the equivalence between the number of independent parameters which are necessary in either a first order Taylor expansion of the non-linear dynamics at a given point $\boldsymbol{X}_i$ or a local Karhunen–Loève expansion of $\bar{Y}(i)$. This is however not trivial, and the interested reader is referred to [25] for a more detailed discussion.

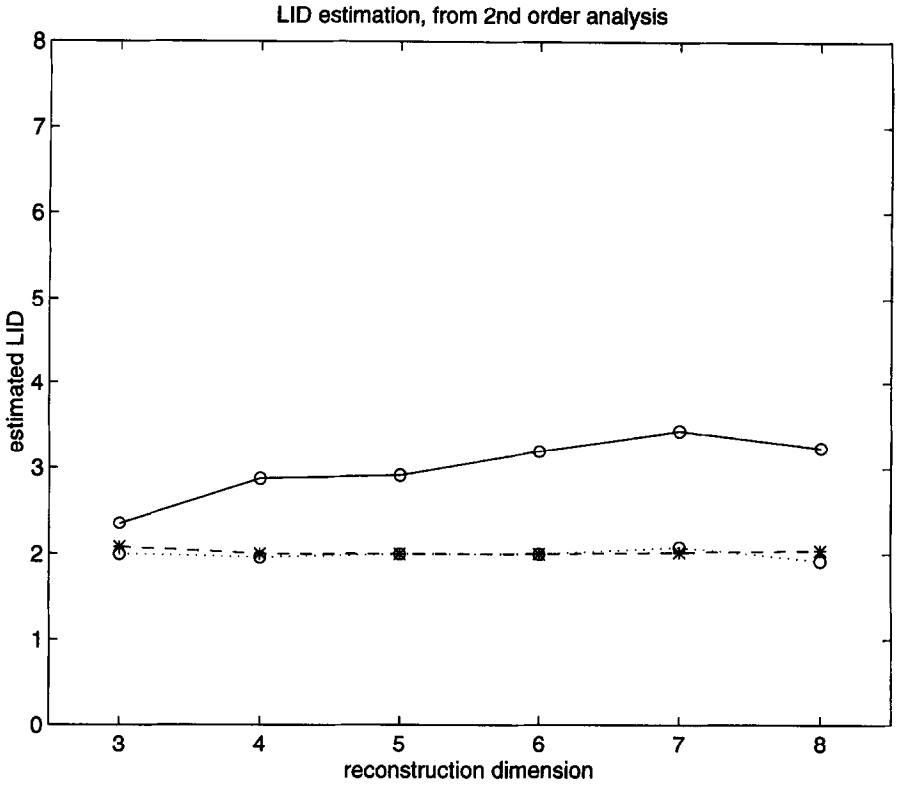

Fig. 5. Results of the local intrinsic dimensionality algorithm on chaotic Exp1 (dashed) and Exp2 (dotted) signals and Wiener–Lévy stochastic process (solid).

neighbourhoods that are randomly spread over the reconstructed phase space. In the example shown, 64 K points were considered for the phase space reconstruction; $m = 50$ different neighborhood containing $q = 80$ neighbors each are used for the LID estimation. Apart from the gain in terms of computational load, it can be checked that LID suffers from deficiencies similar to GPA when one wants to use it as a discriminator "chaos versus noise".

4.3. HOS extensions

The sensitivity of both GPA and LID towards additive noise leads to severe degradation of their performance, even at reasonable signal-to-noise ratios. Essentially motivated by an SNR improvement of the LID algorithm when the data points are corrupted by some additive Gaussian noise, a "higher-order version" of LID (referred to as HOLID) has recently been proposed [33]. It consists of introducing a fourth-order cumulant matrix $\bar{M}$:

$$\bar{M}_{ij} = \mathscr{E}\{x_i^3 x_j\} - 3\mathscr{E}\{x_i^2\}\mathscr{E}\{x_i x_j\}, \tag{19}$$

where $\mathscr{E}$ stands for the expectation operator and x_i for a component of the embedding vectors. Estimation is then performed on a neighborhood of q local data points, rank being deduced from the SVD of $\bar{M}$. As it was designed in order to get some improved rank estimation in the presence of Gaussian noise, HOLID leads basically to the same results as LID, though it has an improved behavior with respect to SNR.

It was previously stressed that LID is basically an eigenmethod, i.e. an algorithm that only tracks *uncorrelated* (i.e. *linearly* independent) components

in the embedding, while forcing the orthogonality of the decomposition vectors. This amounts to saying that the coordinates of (17) are decomposed according to

$$\boldsymbol{X}_{i(k)} - \boldsymbol{X}_i = \sum_{j=1}^{p} \mu_{jk} \boldsymbol{V}_j(\boldsymbol{X}_i), \tag{20}$$

where the μ_{jk} are uncorrelated and the $\boldsymbol{V}_j$ orthogonal (Karhunen–Loève expansion). Coming back to the interpretation of degrees of freedom (that is the minimum number of coordinates that must be considered to get some reliable representation of the system) in terms of unrelated coordinates of the state vector, it is more appealing to look for an *independent* component analysis, without imposing any orthogonality condition between the corresponding vectors, i.e. to only impose the requirement that the μ_{jk} be independent until fourth-order.

A global fourth-order matrix, generalizing (19), can therefore be constructed, containing all possible cumulants from a set of p-dimensional vectors. Forcing all of the cross-cumulants to be zero, one can form the basis of a new higher-order version of LID (referred to as LID4), which has been proposed and applied in [24]. However, this approach led to some difficulties in its interpretation, as it dealt with fourth-order cumulants, homogeneous to $\mathscr{E}\{x^4\}$ whereas any energy oriented threshold deals with the signal variance homogeneous to $\mathscr{E}\{x^2\}$.

Another appealing approach consists in considering the problem in terms of source separation, in the spirit of [9]. This corresponds again to a decomposition of the form of Eq. (20) under a constraint of statistical independence up to fourth-order. This leads however to a new distribution of the *energy* of the local observations, in the sense that all vectors in the phase space are expressed in new basis in such a way that all coordinates are again uncorrelated, thus allowing to consider the total energy of the process as being equal to the sum of the energy of each component. This point is illustrated in the next paragraphs. This kind of linear separation of independent components has been recently addressed in [8, 9]. We briefly outline the principle of separation algorithms in the following. Exhaustive justification and theoretical developments, together with a discussion of different issues concerning the source separation problem (for the first time referred to as independent component analysis (ICA) by Jutten and Hérault, see [20]), are discussed in several papers, e.g. [8–11, 20]. A detailed presentation of the algorithm within the context of chaotic signal analysis is presented in [25].

In this approach, we propose to represent the set of points in the neighborhood $\boldsymbol{X}_i$ through the following new expansion around $\boldsymbol{X}_i$:

$$\boldsymbol{X}_{i(k)} - \boldsymbol{X}_i = \sum_{j=1}^{p} z_{i(k),j} \boldsymbol{U}_{i,j}, \tag{21}$$

where the projection coordinates $z_{i(k),j}$ on $\boldsymbol{U}_{i,j}$ are independent of the $z_{i(k),l}$ but the set of vectors $\boldsymbol{U}_{i,j}$ are no longer required to be orthogonal to each other. The difference between the KLE and the expansion (21) resides in these latter properties that both the projections coordinates and the basis vectors are required to achieve. This analysis leads one to express the correlation matrix of the 'points' within the neighborhood as follows:

$$\bar{C} = \frac{1}{q-1} \bar{Y}(i) \bar{Y}^T(i) = \bar{W} \bar{D}^2 \bar{W}^T, \tag{22}$$

where the vectors $\boldsymbol{U}_{i,k}$ in (21) are given by the kth row of $\bar{W}$, and $\bar{D}$ is a diagonal matrix. $\bar{W}$ is constructed so that the expression of the vectors in the neighborhood under consideration are such that their coordinates with respect to this new basis minimize their cross-mutual information (at least the fourth-order expansion of it) [9, 25]. Furthermore, $\bar{W}$ has normalized columns, so that one has

$$\begin{aligned} \mathrm{Trace}(\bar{C}) &= \mathrm{Trace}(\bar{W} \bar{D}^2 \bar{W}^T) \\ &= \mathrm{Trace}(\bar{W}^T \bar{W} \bar{D}^2) \\ &= \mathrm{Trace}(\bar{D}^2) \end{aligned} \tag{23}$$

as the Trace of a product of matrices is invariant by permutation, and the diagonal of the product $\bar{W}^T \bar{W}$ is identically equal to one. In this case, representations given by either LID or ICA-LID retain the same energy, namely the energy associated with the set of vectors in the neighborhood of interest. For the LID algorithm, the energy is spread over the entire eigenspectrum of the local correlation matrix, whereas the energy is given by

the Trace of the diagonal matrix $\bar{D}^2$ in the case of the ICA-LID approach.

An illustration of the effectiveness of this approach for improving the discrimination chaos versus noise, is given in Fig. 6. The rank estimation was performed by thresholding the sorted diagonal values of $\bar{D}^2$, so as to retain a given percentage of the total energy. The parameters for these estimations were the same as those used for the LID estimation in the previous paragraph (namely, $m=50$, $N=64$ K points, $q=80$ and $t=80\%$). Notice that both chaotic time series exhibit a saturation of their ICA-LID estimation when p is increased, and alternatively that ICA-LID keeps on growing with p for the stochastic Wiener–Lévy process. Thus, ICA-LID approach turns out to be much more efficient than second-order-based methods when it comes to separate between chaotic and stochastic processes. However, this was to be expected as we mentioned that the main motivation for introducing ICA approach was to estimate the number of necessary coordinates for mimicking the dynamical system, whereas GPA and LID are known to measure geometrical features of the attractor.

As for the second-order LID algorithm, a weakness of this higher-order method is the need to estimate the effective rank of higher-order matrices. This is generally done from empirical thresholding methods. Nevertheless, first investigations conducted up to now suggest that, in comparable circumstances, fourth-order algorithms tend to provide a result more related to the number of degrees of freedom involved in the dynamics than to the geometrical dimension of the attractor (see [14, 25]).

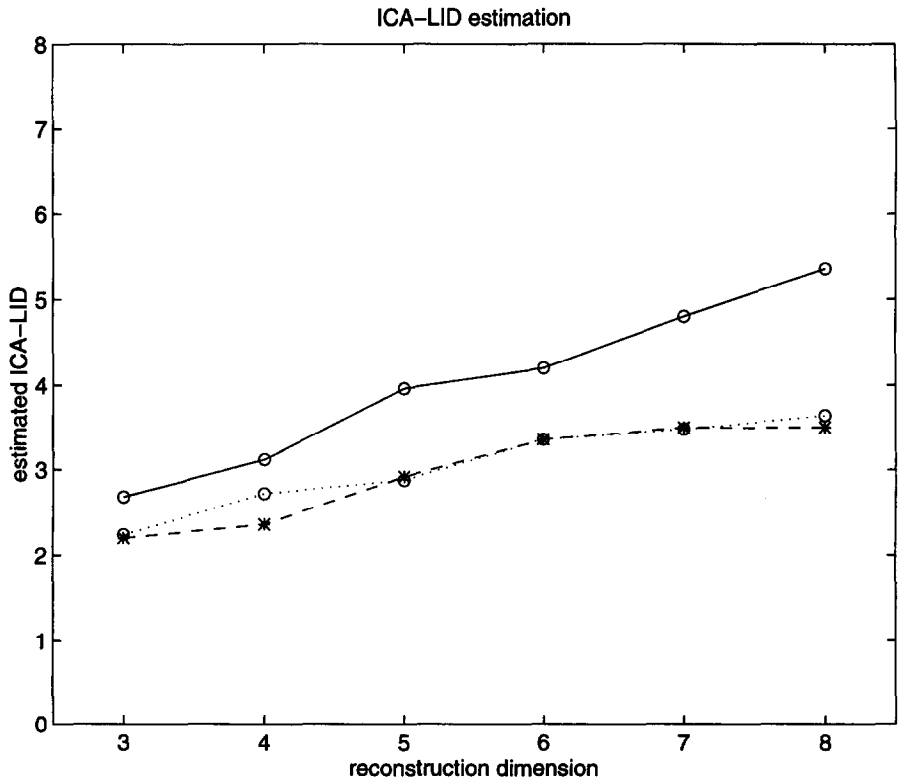

Fig. 6. Results of the ICA-LID algorithm computed on chaotic Exp1 (dashed) and Exp2 (dotted) signals and Wiener–Lévy stochastic process (solid).

4.4. LID or ICA-LID?

There is no simple relation between the number of coordinates of a chaotic system and the fractal dimension of its attractor. This latter is a purely geometrical characteristic, that is in some way a static property of the whole attractor, stemming from the simultaneous effects of positive Lyapunov exponents (to stretch the initial hyper-sphere) and non-linearities (to fold it on itself). The number of coordinates necessary to represent the system is closely related to the dynamics, which appears explicitly in the time-delayed structure of the coordinates that are expanded around a given position.

KLE imposes a double orthogonality on the expansion, which means that it only allows one to track the only principal directions into which all the points in the neighborhood are to be mapped. Thus, it gives an information about how the "cloud" of points in the neighborhood "covers" locally different orthogonal directions in the phase space, thus leading to some local geometrical characteristic. It may easily be seen that two independent modes associated to close trajectories will not be revealed through KLE analysis. Thus, this method will only yield one important eigenvalue (proportional to the sum of the directions), the second eigenvalue falling generally within the estimation variance of the eigenspectrum. The power of ICA-LID resides in its ability to separate independent components, even in the case where they are related to directions that are very close to each other, as the new basis is not required to be orthonormal.

5. Conclusion

The purpose of this paper was to show that some more insight can be gained in the characterization of chaotic signals by using higher-order based techniques. For instance, although it has been stressed that the problem of *detecting* chaos (or, in other words, of testing for determinism) cannot be restricted to an *estimation* problem based on chaos-oriented algorithms, fourth-order estimation algorithms have been shown to outperform second-order (classical) ones within a detection (stochastic versus deterministic) context.

However, with the few examples presented here, we are far from exhausting the potential usefulness of higher-order techniques within the context of chaotic signal processing. We mention for example that only *low-dimensional* chaos has been considered, although increasing the number of degrees of freedom also leads to important problems. Fully developed turbulence is a physical example of such a situation, and its characterization relies heavily on higher-order concepts such as the so-called "structure functions" [26]. (By definition, a structure function measures the higher-order moment of the increment of a velocity field, and its scaling behavior (in the small-scale limit) directly provides informations about deviations from Gaussian fluctuations.)

Furthermore, we have chosen here not to present any application of higher-order spectrum estimation for chaotic signal analysis. One may find good illustrations of the power and usefulness of higher-order spectra within this context in Refs. [12, 23]. Some examples of bispectral analysis of chaotic time series may also be found in [25], and a discussion on estimation issues together with an extended bibliography is given in [28].

From another point of view, it is clear that, beyond analysis, challenging problems of *prediction* are offered by chaotic signals and that non-linear models are to be explored further from such a perspective [7].[8]

References

[1] H.D.I. Abarbanel, "Chaotic signals and physical systems", *IEEE Internat. Conf. Acoust. Speech Signal Process., ICASSP-92*, San Francisco, 1992, pp. IV.113–IV.116.

[2] A.M. Albano, J. Muench and C. Schwartz, "Singular value decomposition and the Grassberger–Procaccia algorithm", *Phys. Rev. A*, Vol. 38, No. 6, 1988, pp. 3017–3026.

[3] A.M. Albano, A. Passamante and M.E. Farrell, "Using higher-order correlations to define an embedding window", *Physica D*, Vol. 54, 1991, pp. 85–97.

[8] Prediction within the context of chaotic signal processing is limited to short-term prediction, as the existence of positive Lyapunov exponents theoretically precludes any possibility of performing long-range prediction from finite precision measurements.

[4] P. Bergé, Y. Pomeau and C. Vidal, *Order within Chaos*, Wiley/Interscience, New York, 1984.
[5] D.S. Broomhead and G.P. King, "Extracting qualitative dynamics from experimental data", *Physica D*, Vol. 20, 1986, pp. 217–236.
[6] A.B. Cambel, *Applied Chaos Theory, A Paradigm for Complexity*, Academic Press, New York, 1993.
[7] M. Casdagli, "Nonlinear prediction of chaotic time series", *Physica D*, Vol. 35, 1989, pp. 335–356.
[8] P. Comon, "Analyse en composantes indépendantes et identification aveugle", *Traitement du Signal*, Vol. 7, No. 5, special issue "Non-Gaussien, Non-Lineaire", 1990, pp. 435–450.
[9] P. Comon, "Independent component analysis", *Internat. Signal Proc. Workshop on Higher-Order Statistics*, Chamrousse (France), 1991, pp. 111–120.
[10] J.F. Cardoso and P. Comon, "Tensor-based independent component analysis", *EUSIPCO-90*, Barcelona (Spain), Elsevier, Amsterdam, 1990, pp. 673–676.
[11] P. Duvaut, "Non-linear filtering in signal processing", *Internat. Signal Proc. Workshop on Higher Order Statistics*, Chamrousse (France), 1991, pp. 41–50.
[12] S. Elgar and V. Chandran, "Higher order spectral analysis to detect non linear interactions in measured time series and an application to Chua's circuit", *Internat. J. Bifurcation Chaos*, Vol. 3, No. 1, 1993, pp. 19–34.
[13] J.P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors", *Rev. Mod. Phys.*, Vol. 57, No. 3, 1985, pp. 617–656.
[14] P. Flandrin and O. Michel, "Chaotic signal analysis and higher order statistics", *EUSIPCO-92*, Brussels (Belgium), Elsevier, Amsterdam, 1992, pp. 179–182.
[15] P. Flandrin and O. Michel, "Higher order in chaos", *IEEE-SP Workshop on Higher-Order Statistics*, South Lake Tahoe, CA, 1993.
[16] K. Fukunaga and D. R. Olsen, "An algorithm for finding intrinsic dimensionality of data", *IEEE Trans. Comput.*, Vol. C-20, No. 2, 1971, pp. 176–183.
[17] A. Fraser and H.L. Swinney, "Independent coordinates for strange attractors from mutual information", *Phys. Rev. A*, Vol. 33, No. 2, 1986, pp. 1134–1140.
[18] A. Fraser, "Information and entropy in strange attractors", *IEEE Trans. Inform. Theory*, Vol. IT-35, No. 2, 1980, pp. 245–262.
[19] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors", *Physica D*, Vol. 9, 1983, pp. 189–208.
[20] C. Jutten and J. Herault, "Blind seperation of sources, Part I", *Signal Processing*, Vol. 24, No. 1, July 1991, pp. 1–10.
[21] M.G. Kendall, *The Advanced Theory of Statistics* – Vol. I, C. Griffin and Co. Ltd, London, 1952.
[22] W. Liebert and H.G. Schuster, "Proper choice of the time delay for the analysis of chaotic time series", *Phys. Lett. A*, Vol. 142, Nos. 2–3, 1989, pp. 107–111.
[23] L.D. Lutes and D.C.K. Chen, "Trispectrum for the response of nonlinear oscillator", *Internat. J. Non Linear Mech.*, Vol. 26, No. 6, 1991, pp. 893–909.
[24] O. Michel and P. Flandrin, "An investigation of chaos-oriented dimensionality algorithms applied to AR(1) processes", *IEEE Internat. Conf. Acoust. Speech Signal Process., ICASSP-92*, San Francisco, 1992.
[25] O. Michel and P. Flandrin, in: C.T. Leondes, ed., *Higher Order Statistics for Chaotic Signal Analysis*, Academic Press Theme Volumes on DSP Techniques and Applications, Vol. 75, 1996, pp. 105–154.
[26] J.F. Muzy, E. Bacry and A. Arnéodo, "Multifractal formalism for fractal signals: The structure-function approach versus the wavelet-transform modulus-maxima method", *Phys. Rev. E*, Vol. 47, No. 2, 1993.
[27] J.S. Nicolis, *Chaos and Information Processing, An Heuristic Outline*, World Scientific, Singapore, 1991.
[28] C.L. Nikias and M.R. Raghuveer, "Bispectrum estimation, A digital signal processing framework", *Proc. IEEE*, Vol. 75, No. 7, 1987, pp. 869–891.
[29] A.V. Oppenheim, G.W. Wornell, S.H. Isabelle and K.M. Cuomo, "Signal processing in the context of chaotic signals", *IEEE Internat. Conf. Acoust. Speech Signal Process., ICASSP-92*, San Franscisco, 1992, pp. IV.117–IV.120.
[30] A.R. Osborne and A. Provenzale, "Finite correlation dimension for stochastic systems with power-law spectra", *Physica D*, Vol. 35, 1989, pp. 357–381.
[31] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill International Editions, New York, 1984, pp. 527–535.
[32] T.S. Parker and L.O. Chua, *Practical Numerical Algorithms for Chaotic Systems*, Springer, New York, 1989.
[33] A. Passamante and M.E. Farrell, "Characterizing attractors using local intrinsic dimension via higher-order statistics", *Phys. Rev. A*, Vol. 43, No. 10, 1991, pp. 5268–5274.
[34] A. Passamante, T. Hediger and M. Gollub, "Fractal dimension and local intrinsic dimension", *Phys. Rev. A*, Vol. 39, No. 7, 1989, pp. 3640–3645.
[35] D. Ruelle, *Chaotic Evolution and Strange Attractors*, Cambridge Univ. Press, Cambridge, 1989.
[36] D. Ruelle, "Deterministic chaos: The science and the fiction", *Proc. Roy. Soc. London*, Vol. A427, 1990, pp. 241–248.
[37] R. Shaw, "Strange attractors, chaotic behavior, and information flow", *Z. Naturforschung*, Vol. 36A, No. 1, 1981, pp. 80–112.
[38] F. Takens, "Detecting strange attractors in turbulence", *Lecture Notes in Mathematics*, Vol. 898, 1981, pp. 366–381.
[39] J. Theiler, "Some comments on the correlation dimension of $1/f^{\alpha}$ noise", *Phys. Lett. A*, Vol. 155, Nos. 8–9, 1991, 480–493.
[40] H. Whitney, "Differentiable manifolds", *Ann. Math.*, Vol. 37, No. 3, 1936, pp. 645–680.
[41] R.C.L. Wolff, "A note on the behaviour of the correlation integral in the presence of a time series", *Biometrika*, Vol. 77, No. 4, 1990, pp. 689–697.