Phase-Space Reconstruction in Hamiltonian Systems
Through Multiple Time Series

Fatihcan M. Atay
Koç University
Mathematics Department
İstinye, İstanbul 80860, Turkey

Ersin Yurtsever
Koç University
Chemistry Department
İstinye, İstanbul 80860, Turkey

Preprint. Final version published in

Abstract

A method which combines the various time series originating from a single source in order to reconstruct the phase space dynamics is applied to a chaotic Hamiltonian system. It is shown that for a large energy range, the variation of the maximum Lyapunov exponents can be reproduced more accurately than the results obtained from a single time series, for both clean and noisy signals. Especially in cases where the maximum exponent is close to zero, using multiple series can give better results, providing a more reliable way of detecting chaos.

1 Introduction

One of the most important and widely used quantitative measures of complicated behavior of a dynamical system is the Lyapunov spectrum. The Lyapunov exponents quantify the rate of separation or approach of nearby trajectories, a positive exponent signaling exponential divergence, and hence, chaos. It is thus not surprising that much effort has gone into the accurate calculation of the exponents, especially in the last decade. When the differential equations which model the system are known, this calculation is more or less straightforward; however, in a large class of problems, all one has is a sequence of measurements with no knowledge of the governing equations. The analyst then faces the problem of extracting information about the dynamics of the system using a scalar time series

$$x(t_1), x(t_2), \ldots, x(t_N)$$

of measurements taken at (presumably equally spaced) time instants $t_k$. The classical linear tools such as the Fourier spectrum offers little help here because of the essential nonlinear nature of chaos. The starting point of a nonlinear analysis is often the embedding theorem [1, 2] which says that the essential phase space geometry may be reconstructed by employing the so-called delay coordinates to form $d$-dimensional vectors

$$z(t_k) = (x(t_k), x(t_{k+\tau}), \ldots, x(t_{k+(d-1)\tau})),$$
where τ is a positive integer called the delay. More precisely, when an infinite amount of noise-free data are available and if $d$, called the embedding dimension, is chosen large enough, then the collection of these vectors “embed” the underlying flow for basically any choice of $\tau$. Hence, the $d$-dimensional space of delay coordinates serves as a pseudo state-space which provides a natural setting to approximate the quantitative aspects of the dynamics.

Several algorithms have appeared in the literature to calculate the Lyapunov exponents using delay coordinates. One of the first was given by Wolf et al. [3]. Although several modifications and alterations followed it [4, 5, 6, 7, 8, 9], this algorithm remains among the most influential ones. All the various methods, however, have to contend with the following issues.

1. Parameter selection: In practice, the choice of the delay $\tau$ becomes important since one deals with finite data, contrary to the setting of the embedding theorem. Often, choosing the embedding dimension $d$ and the other parameters entering the particular algorithm also is a non-trivial matter. Although there are a couple of agreed-upon prescriptions for selecting the parameters, there is no proof that they lead to the “best” estimates for the Lyapunov exponents.

2. Noise: Again the use of the embedding theorem is made difficult by the fact that no real data set is perfectly clean from extraneous signals.

3. Data size: In general, the longer the time series, the more accurate are the calculations. However, in economic, financial, or historical studies the available data may be limited, say less than 2000 data points. Just how much data is needed for a reliable estimate of the Lyapunov exponents is hard to prescribe a priori.

4. Small exponents: Since a positive Lyapunov exponent is an indication of chaos, and the calculation of the exponents is not immune to error, a small positive value calculated for the maximum exponent is hard to interpret as evidence for chaos. Suppose that we find 0.05 as the value of the largest exponent; are we to conclude then that we are dealing with a chaotic signal?

The algorithms keep improving to deal with these problems when calculating the Lyapunov exponents from a time series. On the other hand, it is usually possible to measure more than one entity from the system under investigation, leading to several time series. Under the ideal conditions of the embedding theorem, each series is expected to yield the same Lyapunov spectrum. In practice, however, these conditions are not met; it may then be reasonable to expect that additional measurements are not redundant and provide additional information. This paper deals with the question of how to utilize the information obtained from various signals in order to get more accurate estimates of the average local Lyapunov exponents. We illustrate a method of processing several signals together improves. The method is a natural extension of the algorithm of Wolf et al., but to the authors’ knowledge has not appeared in the literature before. Possible advantages of the method are investigated regarding the four items in the above list. We demonstrate the ideas on a Hamiltonian system of
two nonlinearly coupled oscillators by comparing the maximum Lyapunov exponents calculated from the time series generated by the system to those calculated directly by using the equations of motion.

2 Combining multiple signals

Consider the case when several different quantities are available for measurement from the same system at the same instants of time:

\[
x_1(t_1), x_1(t_2), \ldots, x_1(t_N),
\]
\[
x_2(t_1), x_2(t_2), \ldots, x_2(t_N),
\]
\[
\vdots
\]
\[
x_n(t_1), x_n(t_2), \ldots, x_n(t_N).
\]

Since different finite, and possibly noisy, series are not necessarily redundant, the question of interest is how to combine the information content of each series to improve the estimation of the Lyapunov spectrum. Any scheme which claims to accomplish this should come up with a result which is better than each one of the constituent series. One sometimes finds references to methods which use a (weighted) average, or similar scalar combinations of the series. Clearly, such methods cannot give results better than those obtained from the single series giving the closest estimate. The approach we take is to utilize all the available information without modifying it, and is based on viewing and processing the data (3) as a vector time series.

Given the measurements (3), one can consider the sequence of arrays

\[
Z(t_k) = \begin{bmatrix}
  x_1(t_k) & x_1(t_{k+\tau}) & \ldots & x_1(t_{k+(d-1)\tau}) \\
  x_2(t_k) & x_2(t_{k+\tau}) & \ldots & x_2(t_{k+(d-1)\tau}) \\
  \vdots & \vdots & \ddots & \vdots \\
  x_n(t_k) & x_n(t_{k+\tau}) & \ldots & x_n(t_{k+(d-1)\tau})
\end{bmatrix},
\]

which evolve in the space \(\mathbb{R}^{n\times d}\), reflecting the topology of the flow. Note that since the dimension of the embedding space is \(nd\), the integer \(d\) can be smaller than the case of a single series. The algorithm of Wolf et al. can be extended in a straightforward fashion to calculate the Lyapunov exponents by following the evolution of the distances between points in this space. For instance, the euclidean distance between the two points \(Z(t_k)\) and \(Z(t_l)\) is given by the expression

\[
\sqrt{\sum_{i=1}^{n} \sum_{j=1}^{d} (z_{ij}(t_k) - z_{ij}(t_l))^2},
\]

where \(z_{ij}\) denotes the entries of the array (4). Other elements of the calculations are modified accordingly to work in this more general matrix–space. To calculate the largest exponent, for instance, one starts with a point of the form (4), searches the series to find another point which is closest to the first one in euclidean distance, and then follows the evolution of the two to estimate their rate of separation. When the
distance gets too large, a replacement step is attempted and another closest neighbor is searched. Continuing in this fashion, the whole time series is traversed while a running sum is updated, which gives an estimate of the desired quantity.

3 Model System and Computational Aspects

We illustrate the results on a Hamiltonian system. A Hamiltonian system is chosen for dual purposes: We intend to apply this analysis to signals coming from molecular systems (of quantal sources or systems with many particles). Also, from a technical point of view, the theorems on the Lyapunov spectra of Hamiltonian systems provide rigorous checks on computational aspects especially for the exact calculations of the local Lyapunov exponents [10]. The model system used in this work consists of two nonlinearly coupled anharmonic oscillators.

\[ H = \frac{1}{2}(p_x^2 + p_y^2) + \frac{1}{2}x^2 - 0.05x^3 + 0.00140625x^4 + 0.72y^2 - 0.0864y^3 + 0.002916y^4 + 0.1x^2y^2, \]  

with position coordinates \( x \) and \( y \) and momenta \( p_x \) and \( p_y \). The parameters of the potential are chosen such that oscillators along both modes are bound, they are highly anharmonic and possess only single minima (the presence of the double minima can seriously complicate the dynamics). The inflection points of both modes are close in energy. This system has been studied by one of us and we have shown that above \( E = 2 \) (in generalized units) the positive Lyapunov exponents are detected and after \( E = 8 \), almost all the trajectories are chaotic, those which have zero Lyapunov exponents are possibly the trajectories trapped in the valleys of the potential minima [11]. The full quantum mechanical treatment displays a strong nodal structure pointing out to a great deal of regularity; however, the quantal-classical mixed mode description of the system shows a strong energy dependence of the chaotic character [12, 13].

For the calculations of the actual exponents, a random set of initial conditions are chosen such that energy changes from \( E = 2 \) to 27 in relatively equal increments. The Hamilton’s equations of motion are integrated by a constant time step Runge-Kutta integrator of the fourth order. The evolution of trajectories and the four Lyapunov exponents are obtained from calculations of 200,000 time steps with Gram-Schmidt orthonormalizations at every 10th step. The plot of maximum exponents with respect to the energy follows the almost linear trend observed in our previous calculations of 2000 trajectories.

To calculate the exponents using the algorithm of Wolf et al., each of the four coordinates \( x, y, p_x, p_y \) of the trajectories were used to generate a time series of 20,000 points, although typically much shorter series suffice. All time series calculations are done with replacement of nearest neighbor at every 5 steps.

Presently, there is no automatic method that can take in a time series and return the Lyapunov spectrum reliably and without operator intervention; rather, one needs to do some preliminary analysis of the time series to determine the parameters that go into the algorithm. The situation is further complicated by the fact that our model system is conservative and hence does not have an attractor. The trajectories are
confined to surfaces of constant energy determined by the initial conditions. Thus, at every energy level the trajectories give rise to time series which are potentially of different character; so, the parameters need to be determined separately for each one. We apply the same parameter selection scheme for all the time series considered. The embedding dimension $d$ is taken to be 8. To choose $\tau$, we use the time it takes for the autocorrelation function of (1) to drop to one half of its initial value. Similar methods have been shown to yield good results; for instance, in [7] the delay is taken to be the time when the autocorrelation function is reduced to $1 - 1/e$ of its initial value, which results in a slightly smaller $\tau$ compared to our choice. See also [14] and the review article [15] for other approaches. In addition to such rules of thumb, the convergence of the calculated exponent as the time series is traversed, as well as comparison with the true value, are taken into account to zoom into the correct parameter range. We nevertheless emphasize that our aim is not to determine the optimum parameters, but to compare the exponents calculated from single versus multiple series under the same parameter selection scheme.

4 Results

Figure 1 shows a plot of the largest Lyapunov exponent calculated using the individual time series generated by the components $x$, $y$, $p_x$, and $p_y$ of the Hamiltonian system (5) at various values of the energy $E$ between 2 and 27. Also plotted are the values of the true exponent as calculated from the equations of motion. It is seen that the variation of the largest Lyapunov exponent is estimated reasonably well by each time series. However, the individual exponents have large errors especially at high energies. The results can possibly be improved by further fine-tuning of the parameters. Still, not only doing this separately at each energy level is burdensome, it is also difficult to realize the extent of the error committed when one deals with a “black box” system, where the actual exponents are unknown. Hence, a more robust method is indeed desirable.

We test the algorithm of Section 2 on the Hamiltonian system (5). As a preliminary check, we use all the four components together. Since these form the coordinates of the actual phase space of the system, we take $n = 4$ and $d = 1$; hence, the question of choosing the optimum time delay or the embedding dimension does not arise. Thus, by focusing on the essential mechanism of the algorithm instead of the parameter selection process, this method can be used as a benchmark against which different algorithms are compared. For our model system, the largest Lyapunov exponent calculated using all four components at various energy levels are compared to those obtained from individual time series. Table 1 lists the average of the errors committed over the energy range considered. Clearly, using the four series together is on the average superior to using single series. Perhaps more striking are the maximum errors. Calculations using a single series can err by as much as 75%, especially at low energy levels where the true exponent is close to zero, whereas the errors in multiple series calculations rarely exceed 10%. This robustness can be an important advantage of using multiple series.

Next, a more practical case is considered where only two different entities are available for measurement. The two time series are used together (with $n = 2$ and
Figure 1: Variation of the largest Lyapunov exponent with the energy level. The solid line is calculated by integrating the equations of motion, hence represents the true value of the exponent. Other points are estimated from time series of single components using an embedding dimension of 8.

Table 1: Percent errors in the calculation of the largest Lyapunov exponent from various time series. For the single series the embedding dimension is 8.

<table>
<thead>
<tr>
<th>Time series</th>
<th>Average error</th>
<th>Maximum error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>19.9</td>
<td>64.7</td>
</tr>
<tr>
<td>$y$</td>
<td>18.3</td>
<td>67.0</td>
</tr>
<tr>
<td>$p_x$</td>
<td>18.1</td>
<td>70.6</td>
</tr>
<tr>
<td>$p_y$</td>
<td>13.3</td>
<td>76.6</td>
</tr>
<tr>
<td>All four</td>
<td>6.8</td>
<td>11.3</td>
</tr>
</tbody>
</table>
Figure 2: Comparison of the accuracy of the estimates for the largest Lyapunov exponent. The $x$ and $y$ components provide reasonable estimates separately, but processing their information content together in the form of a multiple time series gives better results. The embedding dimension is 8 for single series and 4 for the combination.

d = 4), and the results are compared to using them separately. Figure 2 shows the relative errors when using the series for the positions $x$ and $y$. Again, using multiple series almost always gives better results, and limits the relative error to roughly below 10%, whereas the errors for the single series vary widely as the energy is changed. In fact, we observed that the performance of the algorithm using these two series together is very similar to using all four. Using other pairs of series such as $x$ and $p_x$, or $y$ and $p_y$, or using different parameters like an embedding dimension of 4, gave qualitatively equivalent pictures.

The performance of multiple series is also very good for noisy signals, as shown in Figure 3. Here, each series is contaminated by 4% additive noise: The noisy series is obtained by adding $0.04G$ to each term, where $G$ is a Gaussian random variable with mean zero and variance equal to the variance of the individual series. The time series is a long one (20,000 points); still, at low energy–low exponent regime, errors as high as 150% are observed from single series. The multiple series calculations improve these results dramatically.

Finally, the calculations are carried on much shorter series, where the algorithms are more prone to numerical errors, with 4% additive noise and at low energy levels. At such energies, the maximum Lyapunov exponent is small, so estimating them accurately is important for the detection of chaos. Figure 4 shows that using two series together gives superior results than either one alone. It is also possible to see more clearly in this figure that the variation of the error for the multiple series
somewhat mimics the variations for the single series, clearly showing the effects of
the individual components, but nevertheless leading to smaller errors.

5 Conclusion

We have presented a modified method to calculate the Lyapunov exponents from
multiple time series originating from the same source. Although theoretically a single
time series is enough to reconstruct the topology of the phase space, it may not be a
trivial matter to estimate the exponents from practical data sets. When more than
one entity can be measured, their information content can be combined to advantage
by using the multiple series approach. In cases involving noisy and relatively short
data series, or when the maximum exponent is small, it is shown that the multiple
series provides superior accuracy compared to using only a single series. These are
clearly important considerations in studying “black box” systems, where the correct
ranges of the exponents and the noise level are not known a priori and it is difficult
to ascertain the judiciousness of the choice of the parameters. In such cases, using
all the measurements available from the system in a multiple time series can add
confidence to the results of the analysis.

Acknowledgments. We would like to thank Yigit Altintas for his help in cal-
culations with noisy data.
Figure 4: The largest exponent calculated from shorter series (5000 data points), under 4% additive noise, at low energies. Using multiple series greatly reduces the error committed.

References


