

Short communication

Using information to generate derivative coordinates from noisy time series

B.P. Mann^{a,*}, F.A. Khasawneh^a, R. Fales^b^a Department of Mechanical Engineering & Material Science, Duke University, Durham, NC 27708, USA^b Department of Mechanical & Aerospace Engineering, University of Missouri, Columbia, MO 65211, USA

ARTICLE INFO

Article history:

Received 25 August 2010

Received in revised form 16 November 2010

Accepted 21 November 2010

Available online 3 December 2010

Keywords:

Information theoretic

Signal derivatives

Derivative coordinates

ABSTRACT

This paper describes an approach for recovering a signal, along with the derivatives of the signal, from a noisy time series. To mimic an experimental setting, noise was superimposed onto a deterministic time series. Data smoothing was then used to successfully recover the derivative coordinates; however, the appropriate level of data smoothing must be determined. To investigate the level of smoothing, an information theoretic is applied to show a loss of information occurs for increased levels of noise; conversely, we have shown data smoothing can recover information by removing noise. An approximate criterion is then developed to balance the notion of information recovery through data smoothing with the observation that nearly negligible information changes occur for a sufficiently smoothed time series.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

It is rarely practical to measure each state variable in an experimental setting. To circumvent this issue, methods exist for reconstructing a pseudo-state space from a small number of measured observables [1–3]. A primary benefit of the reconstruction is that it produces a pseudo-state space with dynamics equivalent to those of the original state space. As a consequence of the equivalence, an attractor in the reconstructed state space has the same invariants, such as Lyapunov exponents and dimension, as the original attractor [4].

While delayed embedding is the predominant choice for attractor reconstruction [3,5–8], the use of derivative coordinates is appealing, given their obvious physical meaning. For instance, many physical systems can be described by a set of equations with states that are related through their derivatives. While the numerical derivatives of a signal can be used for noise-free data, the presence of noise renders the signal derivatives to be poor approximates – owing to noise amplification in the signal derivatives.

The inherent goal of signal analysis is to extract useful information about a system from the observed data. Consider a continuous and deterministic system that has produced the scalar time series $q(t)$. To mimic realistic data from an experiment, noise is superimposed onto the noise-free data as follows: $\eta(t) = q(t) + \sigma\alpha(t)$, where $\alpha(t)$ is a time series with normally distributed random noise, a zero mean, and a standard deviation equal to σ . The underlying goal of this investigation is to extract $q(t)$ and the derivatives of $q(t)$ from the noisy time series $\eta(t)$.

This work investigates smoothing $\eta(t)$ to recover $q(t)$ and its derivatives. While the process of data smoothing is well known, the question of how much smoothing yields accurate derivative coordinates is unclear. To answer this question, we explored the use of an information theoretic, known as the average mutual information, to develop a criterion for the appropriate amount of data smoothing.

* Corresponding author. Tel./fax: +1 919 660 5214.

E-mail address: brian.mann@duke.edu (B.P. Mann).

The work of this paper is organized as follows. The next section describes the data smoothing technique and average mutual information tools used in our analyses. These discussions are followed by a series of example results that apply the average mutual information to investigate the influence of noise and an approximate criterion for the level of data smoothing.

2. Example implementation

The investigations that follow use synthetic data generated from a Duffing oscillator,

$$q'' + \mu q' + \omega^2 q + \beta q^3 = \Gamma \cos \Omega t, \quad (1)$$

where a prime denotes a derivative with respect to time. Numerical simulation was used to generate a chaotic time series for $q(t)$, $q'(t)$, and $q''(t)$ while using the following parameters $\mu = 0.2$, $\omega = 1$, $\beta = 1$, $\Gamma = 27$, and $\Omega = 1.33$; however, in an effort to mimic the realistic challenges of an experiment, we have assumed only the noisy time series is observable $\eta(t) = q(t) + \sigma \alpha(t)$. The remainder of this section describes the application of data smoothing for noise removal and the use of the average mutual information to determine the appropriate smoothing level to recover the derivatives of $q(t)$.

2.1. Noise removal with smoothing

Cubic splines are often applied to empirical data to estimate interim points or data points that lie between two measurements. The basic idea is to fit the data with a piecewise polynomial,

$$s(t) = b_{i0} + b_{i1}(t - t_i) + b_{i2}(t - t_i)^2 + b_{i3}(t - t_i)^3, \quad (2)$$

where the polynomial coefficients b_{i0} , b_{i1} , b_{i2} , and b_{i3} have subscripts that denote their validity between two neighboring data points - the time interval from t_i to t_{i+1} . Therefore, cubic splines provide a piecewise fit to the data while simultaneously giving a functional relationship for the derivatives of the data, i.e. after differentiation of Eq. (2). Furthermore, cubic splines invoke continuity in the signal and its first two derivatives at the intersection of the neighboring time steps [9].

Smoothing splines, which differ from the typical cubic spline fitting operation, provide a refinement to the idea using a polynomial to fit empirical data. The basic difference lies in the introduction of a smoothing parameter which reduces noise amplification in the signal derivatives by balancing a fit between the measured data $\eta(t)$ and the smoothness in the second derivative [10]. In the results that follow, cubic smoothing splines have been implemented; this approach obtains the polynomial coefficients for Eq. (2) by minimizing

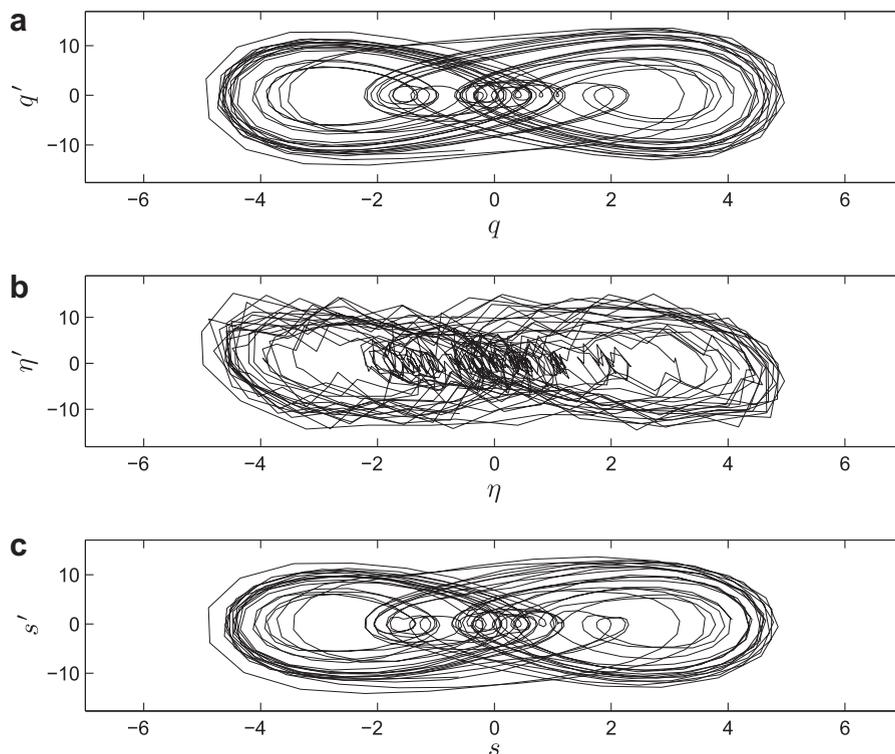


Fig. 1. Two dimensional state space with graphs showing: (a) the actual q and q' ; (b) the noisy time series η and the numerical derivative η' when $\sigma = 0.1$; and (c) the smoothed version s of the noisy time series and the derivative of the smoothed signal s' obtained when $\gamma = 99.95 \times 10^{-2}$.

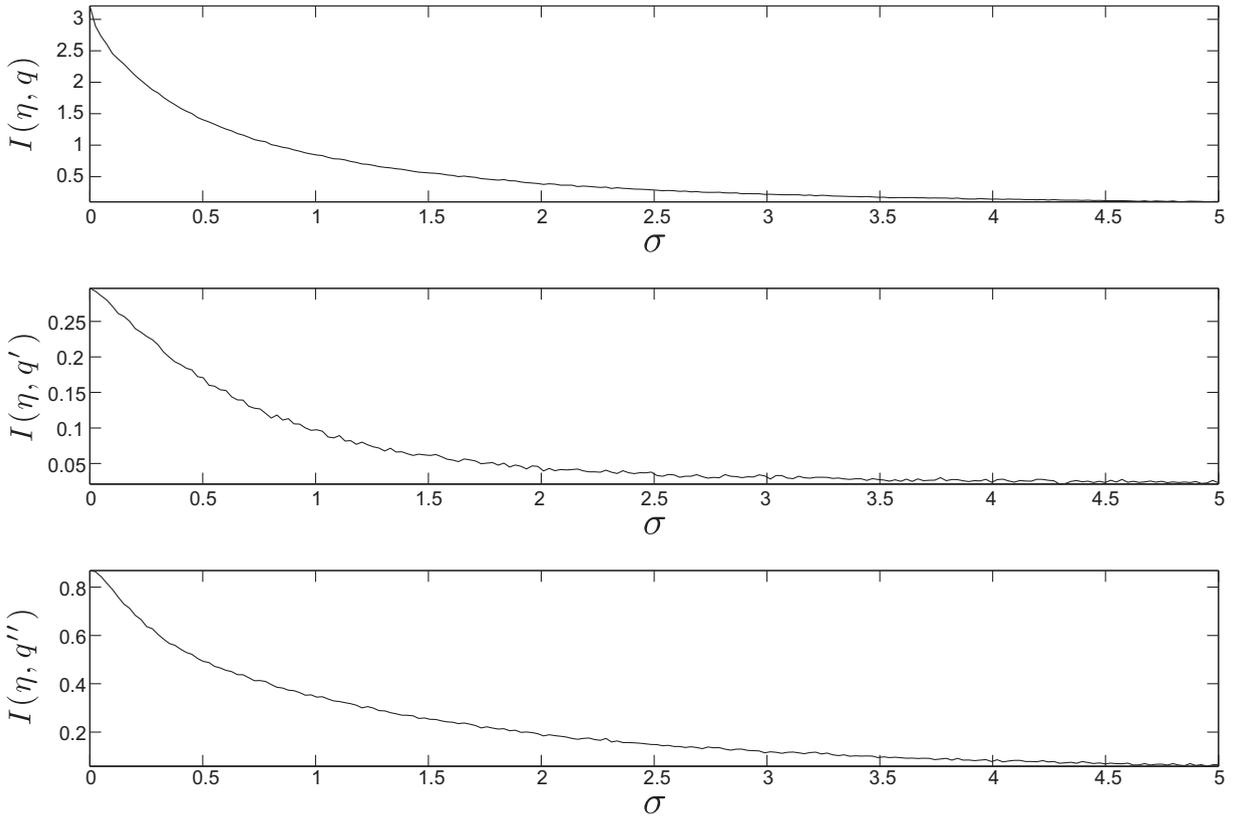


Fig. 2. Average mutual information for different levels of noise intensity. Graphs denote the mutual information between the observable signal η and the following signals: (a) q , (b) q' , (c) q'' .

$$\sum_{i=1}^N (1 - \gamma) |\eta_i - s_i|^2 + \gamma \int_{t_1}^{t_N} \left| \frac{d^2 s(t)}{dt^2} \right| dt, \tag{3}$$

where the notation $\eta_i = \eta(t_i)$ and $s_i = s(t_i)$ has been applied over N -data points. The smoothing parameter γ , provides a weight to balance the smoothness in the second derivative with a fit to the measured data.

The smoothing process can be illustrated with the following example. Consider the superposition of the scalar time series $q(t)$, generated from Eq. (1), with normally distributed random noise; this takes the form $\eta(t) = q(t) + \sigma\alpha(t)$. In the absence of noise, accurate derivative coordinates may be obtained from the numerical derivatives of the observed signal since $\eta(t) = q(t)$ and $\sigma = 0$ (see attractor in Fig. 1a). If normally distributed random noise, with a standard of deviation $\sigma = 0.1$, is added to the signal, Fig. 1b shows that the attractor generated by $\eta(t)$ and $\eta'(t)$ will significantly differ from the attractor given by the states $q(t)$ and $q'(t)$. While the numerical derivative, denoted by $\eta'(t)$, contains an amplified amount of noise, the noise in the higher-order numerical derivatives often overtakes the deterministic component of the signal. Here we note that the numerical derivative and derivative of the a cubic spline fit, one without smoothing, are identical and both yield the time series shown in Fig. 1b. The attractor of Fig. 1c shows the potential for smoothing the data since the time series for $s(t)$ and $s(t)'$ closely replicate the time series shown in Fig. 1a for $q(t)$ and $q(t)'$.

2.2. Average mutual information

The mutual information between the measurement $x_i = x(t_i)$, drawn from a set of measurements x , and a measurement $y_j = y(t_j)$, drawn from a set of measurements y , is the amount learned by the measurement of x_i about the measurement of y_j . The average amount of information learned by comparing the measurements of signal x to the measurements of the signal y is

$$I(x, y) = \sum_{x_i, y_j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}, \tag{4}$$

where \log was taken to be the natural log and $p(x_i, y_j)$ is the joint probability density that a measurement drawn from the signals x and y will result in the values of x_i and y_j [3,11]. The individual probability densities for the measurements x_i in

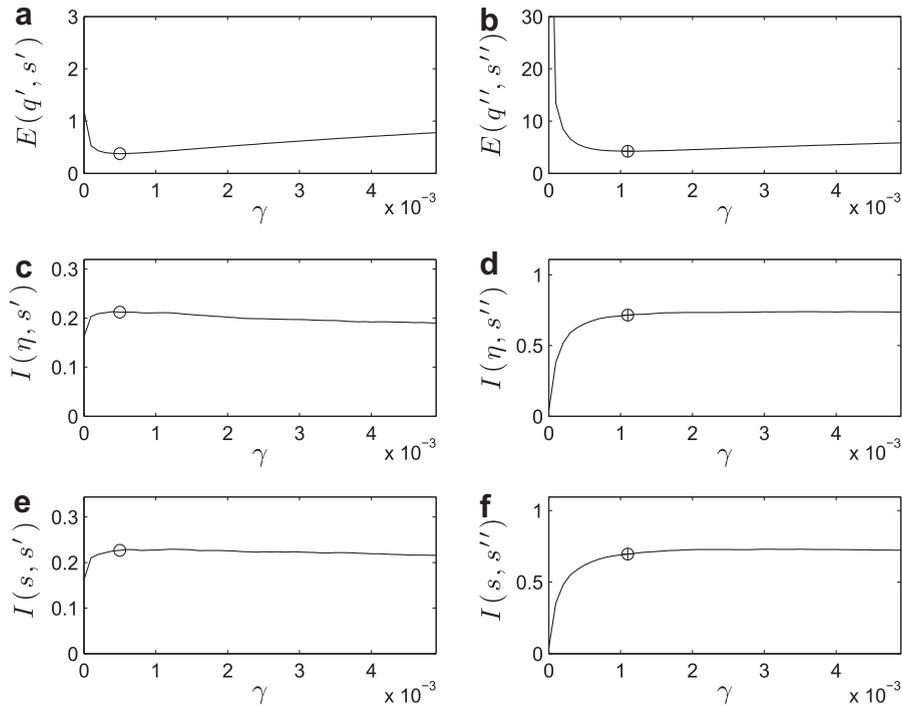


Fig. 3. Average mutual information and error results for a range of smoothing values and a fixed noise level $\sigma = 0.1$. Graphs (a) and (b) are error plots for the smoothed velocity and acceleration, respectively. Graphs (c)–(f) show average mutual information plots for the corresponding variables. The smoothing level that gives the smallest error is shown by a \circ on the plots with s' and by a \oplus on the plots with s'' .

x and y_i in y are given by $p(x_i)$ and $p(y_j)$, respectively. If the two signals are completely independent of each other, which indicates no information transfer, the joint probability factorizes to $p(x_i, y_j) = p(x_i)p(y_j)$ [12]. For computational purposes, it is convenient to rewrite Eq. (4) as

$$I(x, y) = \sum_{x_i, y_j} p(x_i, y_j) \log p(x_i, y_j) - \sum_{x_i} p(x_i) \log p(x_i) - \sum_{y_j} p(y_j) \log p(y_j). \tag{5}$$

Fig. 2 shows a series of plots comparing the mutual information between the observable signal η and the signals $q(t)$, $q'(t)$, and $q''(t)$. For each case, the shared information between the signals rapidly decreases for an increase in the noise intensity; conversely, these graphs also show that a reduction in noise, either from smoothing the data or through an alternative approach, increases the shared information between the signals.

Before investigating the usefulness of the shared information and noise level trends, the following error equation is introduced to ascertain the goodness of fit between any two signals

$$E(x, y) = \frac{1}{N} \sum_i^N |x_i - y_i|, \tag{6}$$

where $E(x, y)$ is the total error obtained when the time series x and y are compared at each t_i and $i \in [1, N]$. For the investigations that follow, $N = 18 \times 10^3$ data points were used.

The effect of different levels of data smoothing is illustrated through the graphs of Figs. 3 and 4. While the top graphs show the error obtained when estimating $q'(t)$ and $q''(t)$ by smoothing the noisy observable $\eta(t)$, the lower graphs show the mutual information between various signals. Focusing on the results of Fig. 3, it is evident that smoothing will initially cause a drastic increase in the mutual information between $\eta(t)$ (or the smoothed version $s(t)$) and the smoothed estimates of $q'(t)$ and $q''(t)$, which have been labeled as $s'(t)$ and $s''(t)$. However, once a sufficient amount of noise has been removed, the mutual information curve essentially plateaus - thus indicating that over smoothing the data will not significantly alter the shared information; in addition, the error graphs show that over smoothing will increase the errors of estimating $q'(t)$ and $q''(t)$ with $s'(t)$ and $s''(t)$.

While the results of Fig. 3 used $\sigma = 0.1$, an increased noise level of $\sigma = 0.5$ was used for the analyses shown in Fig. 4. In addition to the multiple other cases investigated, these two figures demonstrate nearly identical trends. For instance, both figures show that different levels of smoothing are required to produce the most accurate derivative coordinates. The mutual information graphs also indicate that the best level of smoothing - the value that produces the most accurate derivative coordinates - occurs just after the bend in the mutual information curves.

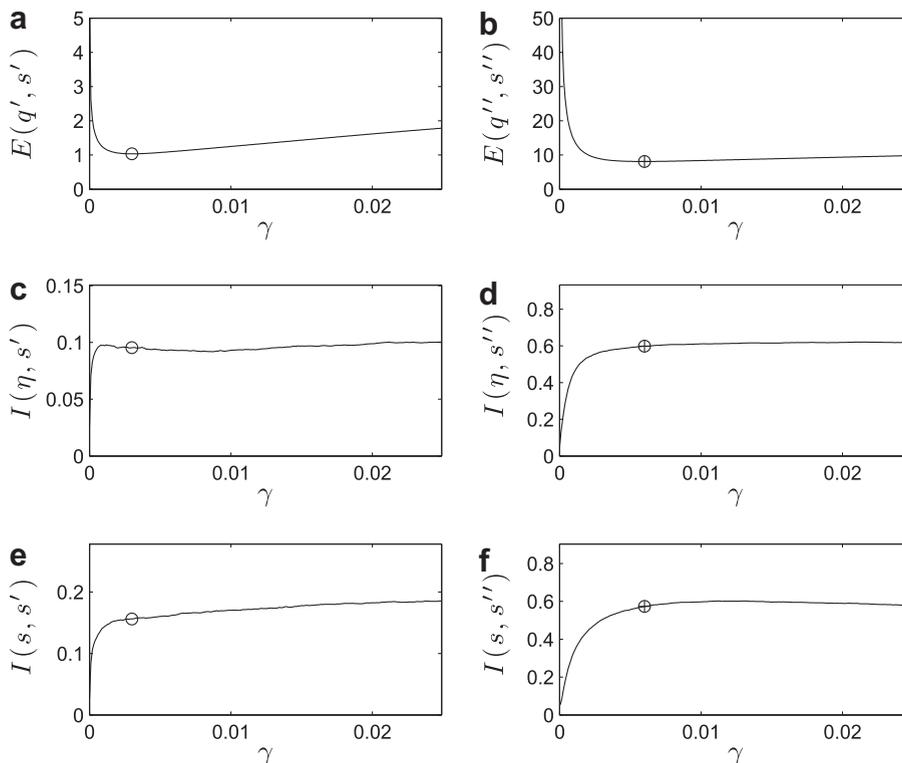


Fig. 4. Average mutual information and error results for a range of smoothing values and a fixed noise level $\sigma = 0.5$. Graphs (a) and (b) are error plots for the smoothed velocity and acceleration, respectively. Graphs (c)–(f) show average mutual information plots for the corresponding variables. The smoothing level that gives the smallest error is shown by a \circ on the plots with s' and by a \oplus on the plots with s'' .

2.3. Approximate smoothing criterion

While it was initially tempting to search for an exact smoothing level that yields the absolute minimum error, we found this approach impractical, given the fact that the noisy time series is the only observable. Instead, we have focused on an approximate criteria that leverages the trends in the mutual information curves. For instance, Fig. 5 shows unit-normalized curvature and slope plots of the mutual information curves of Figs. 3 and 4; these were obtained by fitting the function $f(\gamma) = (c_0 + c_1\gamma)(1 + e^{-c_2\gamma})$ to the $I(x, y)$ curves, where c_0 – c_2 are fitted constants, and taking the derivatives with respect γ . Here, $\tilde{I}_{x,y}$ represents the normalized slope of $I(x, y)$ and $\kappa(I_{x,y})$ denotes the normalized curvature of $I(x, y)$. The correlation coefficient obtained when fitting $\tilde{f}(\gamma)$ to the mutual information curves of Figs. 3 and 4 are (a) 0.96, (b) 0.99, (c) 0.97, and (d) 0.96, where the letters (a)–(d) correspond to the letters assigned to the graphs appearing in Fig. 5.

The shown in Fig. 5 indicate that the slope is initially large, but it rapidly decays to a constant value - indicating a leveling off in the shared information between the smoothed signals for further levels of smoothing. Thus one idea would be to consider a small value ϵ and to choose the smoothing level after the normalized slope dropped below ϵ . The problem with this approach was that the choice of ϵ was arbitrary. As an alternative, a threshold value could be adopted using the normalized curvature. For instance, a range of accurate smoothing levels could be identified when the curvature exceeded a threshold ϵ . However, once again, the choice of ϵ is rather arbitrary - unless the two ideas are combined. Specifically, we contend that the first intersection of the normalized curvature and slope, where $\tilde{I}_{x,y} = \kappa(I_{x,y})$, should be used as the threshold ϵ ; the acceptable range for γ is then defined by the normalized curvature that exceeds this threshold $\kappa(I_{x,y}) \geq \epsilon$. This was the approach taken to successfully smooth the data shown in Fig. 1.

To summarize, the proposed criterion considers a range of acceptable smoothing levels as opposed to a single value. The idea stems from the notion that data smoothing can remove noise and therefore recover information; this is confirmed by the shape of the $I(x, y)$ curves and quantified through the normalized slope and curvature.

3. Conclusions

This paper described an approach to recover the derivative coordinates from a noisy time series. We have assumed the underlying system was deterministic and continuous. Although a data smoothing technique was shown to successfully recover the derivative coordinates, a criterion was needed to ascertain the appropriate level of data smoothing. To investigate

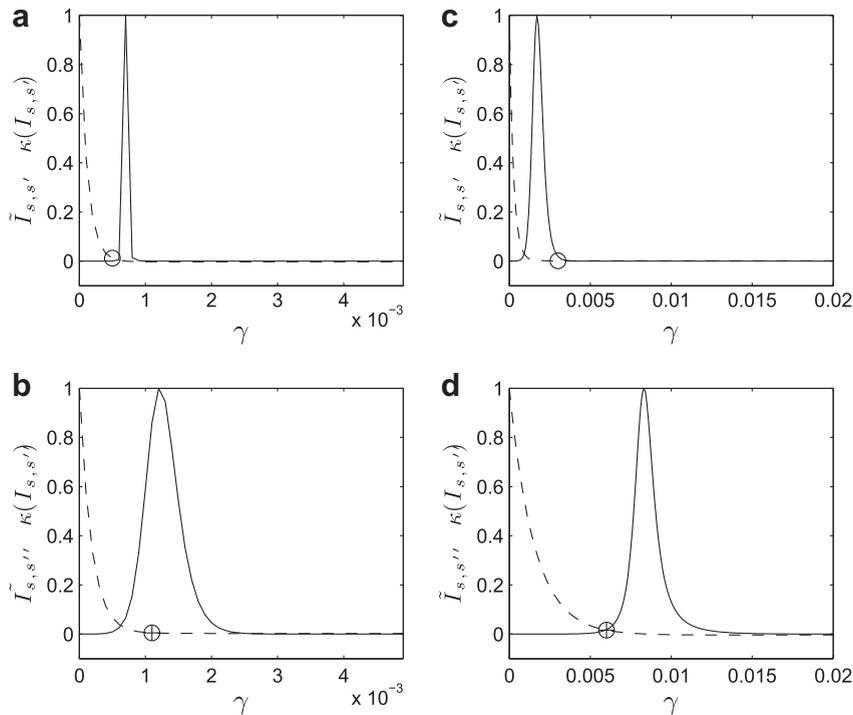


Fig. 5. Unit-normalized slope (solid line) and curvature (dotted line) plots of the mutual information curves of Figs. 3 and 4. Graphs give the slope and curvature for: (a) $I(s, s')$ and $\sigma = 0.1$, (b) $I(s, s'')$ and $\sigma = 0.1$, (c) $I(s, s')$ and $\sigma = 0.5$, and (d) $I(s, s'')$ and $\sigma = 0.5$. The smoothing level that gives the smallest error is shown by a \circ on the plots with s' and by a \oplus on the plots with s'' .

the best level of smoothing, we applied an information theoretic to show information loss for increased levels of noise intensity. Conversely, we also showed that data smoothing can recover information.

An approximate criterion was then developed that balanced the notion of recovering information from data smoothing with the observation that nearly negligible information changes occur once the time series has been sufficiently smoothed. In particular, the slopes and curvatures of the mutual information curves were used to describe an approximate criterion for an appropriate range of data smoothing - the range that provides the most accurate recovery of the derivative coordinates.

Acknowledgement

Support from the National Science Foundation (CMMI-0900266) is gratefully acknowledged.

References

- [1] Packard NH, Crutchfield JP, Farmer JD, Shaw RS. Geometry from time series. *Phys Rev Lett* 1980;45:712–6.
- [2] Takens F. Detecting strange attractors in turbulence. In: Rand A, Young LS, editors. *Dynamical Systems and Turbulence*. Lecture Notes in Mathematics. New York: Springer-Verlag; 1981. p. 366–81. Ch. 898.
- [3] Arbaranel HD. *Analysis of Observed Chaotic Data*. 2nd ed. New York, NY: Springer; 1996.
- [4] Nayfeh AH, Balachandran B. *Applied Nonlinear Dynamics*. 1st ed. New York, NY: John Wiley & Sons, Inc.; 1995.
- [5] Kennel MB, Abarbanel HD. False neighbors and false strands: A reliable minimum embedding dimension algorithm. *Phys Rev E* 2002;66:026209.
- [6] Fraser AM. Reconstructing attractors from scalar time series: A comparison of singular and redundancy criteria. *Physica D* 1989;34:391–404.
- [7] Fraser AM. Independent coordinates for strange attractors from mutual information. *Phys Rev A* 1986;33:1134–40.
- [8] Garcia SP, Almeida JS. Nearest neighbor embedding with different time delays. *Phys Rev E* 2005;71:037204.
- [9] Kreyszig E. *Advanced Engineering Mathematics*. 8th ed. New York, NY: John Wiley and Sons, Inc.; 1999.
- [10] Boor CD. *A Practical Guide to Splines*. New York, NY: Springer-Verlag; 1978.
- [11] Nichols JM. Inferences about information flow and dispersal for spatially extended population systems using time-series data. *Proc R Soc London B* 2005;272:871–6.
- [12] Nichols JM, Seaver M, Trickey ST, Salvino LW, Pecora DL. Detecting impact damage in an experimental composite structures: an information-theoretic approach. *Smart Mater Struct* 2006;15:424–34.