Symplectic geometry spectrum analysis of nonlinear time series

Hong-Bo Xie\textsuperscript{1}, Tianruo Guo\textsuperscript{1}, Bellie Sivakumar\textsuperscript{2,3}, Alan Wee-Chung Liew\textsuperscript{4} and Socrates Dokos\textsuperscript{1}

\textsuperscript{1}Graduate School of Biomedical Engineering, and \textsuperscript{2}School of Civil and Environmental Engineering, The University of New South Wales, Sydney, New South Wales 2052, Australia
\textsuperscript{3}Department of Land, Air and Water Resources, University of California, Davis, CA 95616, USA
\textsuperscript{4}School of Information and Communication Technology, Griffith University, Gold Coast, Queensland 4222, Australia

Various time-series decomposition techniques, including wavelet transform, singular spectrum analysis, empirical mode decomposition and independent component analysis, have been developed for nonlinear dynamic system analysis. In this paper, we describe a symplectic geometry spectrum analysis (SGSA) method to decompose a time series into a set of independent additive components. SGSA is performed in four steps: embedding, symplectic QR decomposition, grouping and diagonal averaging. The obtained components can be used for de-noising, prediction, control and synchronization. We demonstrate the effectiveness of SGSA in reconstructing and predicting two noisy benchmark nonlinear dynamic systems: the Lorenz and Mackey-Glass attractors. Examples of prediction of a decadal average sunspot number time series and a mechanomyographic signal recorded from human skeletal muscle further demonstrate the applicability of the SGSA method in real-life applications.

1. Introduction

Dynamic behaviour of real-world systems can be represented by measurements along the temporal dimension, i.e. a time series. Such time series are typically collected over long periods of time and are usually a source of a large number of interesting behaviours that...
the system may have undergone in the past or evolve into in future. The general objective of
time-series analysis is to obtain the most essential information regarding the evolution
behaviours of a dynamic system [1–3]. Decomposition of a time series into a set of components is
particularly useful for such purposes, because reconstruction, prediction, phase synchronization
and control of nonlinear dynamic systems can be readily conducted using these components.
Over the past few decades, several time-series decomposition methods have been reported and
successfully applied to many problems in nonlinear dynamics, including wavelet transform,
independent component analysis (ICA), empirical mode decomposition (EMD) and singular
spectrum analysis (SSA).

The wavelet transform decomposes a time series into components in various frequency sub-
bands or scales with various resolutions [4]. It provides a framework for studying how frequency
content changes with time and, consequently, for detection and localization of short-duration
phenomena in a dynamic system [5–8]. Chandre et al. [5] described a wavelet-based method to
extract the ridges in the time–frequency landscape of a trajectory for analysing the phase space
structures of Hamiltonian systems. In particular, the method is able to detect resonance trappings
and transitions and allows for characterization of the notion of weak and strong chaos. Ozkurt &
Savaci [6,8] further applied the wavelet ridges to reconstruct the spiral and double-scroll attractors
of Chua’s circuit in phase space by time delay embedding. Han et al. [7] developed a wavelet soft
threshold technique to smooth the noisy chaotic time series generated by a Lorenz system as
well as the observed annual runoff from the Yellow River. Essentially, wavelet analysis is still
convolutional, and the ‘mother’ wavelet is phenomenon dependent [4]. The lack of orthogonality
and the limitation of finite length for some of the most commonly used ‘mother’ wavelets can
cause unwanted leakage between the different frequency modes [4,9].

ICA is an alternative linear decomposition-based method, deriving spatial filters to factorize
observed time series into a sum of temporally independent and spatially fixed components [10].
ICA is a computationally efficient blind source separation technique, widely used in many fields
noisy Gaussian environment by using a modified ICA. Chen et al. [12] presented a similar work to
extract real chaotic signals from a noisy source. They further developed a new scheme to combine
their improved ICA design with state feedback control to achieve chaos synchronization. Asano &
Nakao [13] used ICA components to analyse the complex dynamics of two types of spatio-
temporal chaos. For diffusively coupled complex Ginzburg–Landau oscillators, which exhibit
smooth amplitude patterns, ICA was able to extract localized one-humped basis vectors reflecting
the characteristic hole structures of the system. In addition, for non-locally coupled complex
Ginzburg–Landau oscillators with fractal amplitude patterns, ICA was able to extract localized
basis vectors with characteristic gap structures. ICA assumes a linear stationary mixing model in
which the mixing matrix consists of a set of constants independent of the changing structure of the
data over time. However, for many nonlinear systems, this is only true from certain observation
points or for very short lengths of time [14].

EMD is an emerging novel signal analysis technique able to decompose any time series into
a set of spectrally independent oscillatory modes, known as intrinsic-mode functions. Compared
with wavelets, EMD has the inherent advantage that the basic decomposition functions are
determined by the properties of the data itself. That is, EMD decomposes a signal in a natural
way without prior knowledge of the signal of interest embedded in the data series [4,15]. Several
recent studies have demonstrated that EMD is particularly useful in phase synchronization
analysis of complex oscillators with several spectral components as well as for non-stationary
oscillator prediction [16–19]. Unfortunately, the EMD lacks a firm theoretical framework and
suffers from drawbacks, such as extrema locations, extrema interpolation, end effects and sifting
stopping criterion [20]. Although some studies have attempted to alleviate the aforementioned
problems, the theoretical basis of EMD still needs to be reinforced to achieve more precise and
reliable results.

Distinct from the above techniques, SSA uses singular value decomposition (SVD) to reshape
a time series into a sum of independent and interpretable components, such as a slowly
varying trend, oscillatory components and noise [3]. It is now a popular tool for detecting

trends in different resolutions, smoothing, extracting seasonality components, finding structure

in short time series, detecting chaos, determining noise level (NL) and testing causality [3,21–23].

However, SSA is based on SVD, which is by nature a linear technique. Several studies have

indicated that it can lead to misleading results when nonlinear structures are analysed [24–26].

Distinct from Euclidean geometry, symplectic geometry is even-dimensional, dependent on

a bilinear antisymmetric non-singular cross product—the symplectic cross product. Symplectic

geometry originated from and is widely used in the study of Hamiltonian dynamic systems

[26–30]. It has been rapidly expanded to describe singular differential equations, partial
differential equations and other dynamic systems [31]. Its applications have expanded from
classical mechanics to control theory, geometrical optics, thermodynamics, electromagnetic fields,

quantum systems and biomechanics. It can also be used to characterize dissipative dynamical

systems [32,33]. For example, Ivancevic [32] presented a symplectic geometry approach for use in
dissipative biological systems of human motion.

Several studies have shown that symplectic geometry-based methods are superior to SVD-

based techniques in the detection of chaos [26,27], estimation of the embedding dimension

of a nonlinear dynamic system [29] and de-noising nonlinear systems [30]. In this paper, we

present a method, termed symplectic geometry spectrum analysis (SGSA), which is parallel to

wavelets, EMD, ICA and SSA, to decompose a time series into the sum of a small number of

independent and interpretable components. The idea of SGSA originated from the application

of symplectic geometry to resolve the eigenvalue problem for optimal control [34–36]. The key

difference between SGSA and the former approaches is that SGSA is based on a symplectic

QR decomposition in symplectic space. The derived components can be used to resolve many

problems, such as de-noising, reconstruction of an attractor, nonlinear prediction, extraction of

seasonality components and detection of change-points. In order to illustrate its applicability,

we apply SGSA to reconstruct Lorenz and Mackey-Glass attractors from their noisy series, as

well as to assess the prediction performance of SGSA on them. We further predict the decadal

average sunspot data and a mechanomyographic (MMG) signal collected from human skeletal

muscle by SGSA to test its effectiveness for real systems. Compared with an improved local linear

approximation (ILLA) method, lower normalized mean square error (NMSE) values obtained

from SGSA indicate its potential value for practical application in different fields.

2. Symplectic geometry spectrum analysis

The SGSA method first builds a trajectory matrix from the original time series in a process
called embedding. This matrix consists of vectors obtained by means of a sliding window

that traverses the series. The trajectory matrix is transformed into a symmetric form and then

into a Hamiltonian matrix in symplectic space. The Hamiltonian matrix is further subjected

to a symplectic QR decomposition to obtain its eigenvalues and eigenvectors. Each of these
eigenvectors can be transformed into a reconstructed time series, collectively termed symplectic

reconstructed components (SRC). The sum of all these SRC is equal to the original time series.

Any eigenvectors which negligibly contribute to the norm of the original matrix can be neglected.

In order to approximate the original time series, a diagonal averaging technique is applied to

every symplectic principal component matrix. Figure 1 summarizes the block diagram of SGSA

technique, and the above description may be expressed in formal terms as follows.

(a) Embedding

Consider a real-valued non-zero time series \( S = x_1, x_2, \ldots, x_n \), where \( n \) is the number of samples. Takan’s embedding theorem states that a topologically equivalent representation of the original multi-dimensional system behaviour can be reconstructed from the one-dimensional signal by means of the method of delay. That is, the original time series can be mapped into a
Figure 1. Block diagram of SGSAs. (Online version in colour.)

multi-dimensional state space as

$$X = \begin{bmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_m^T \end{bmatrix} = \begin{bmatrix} x_1 & x_1+\tau & \cdots & x_1+(d-1)\tau \\ x_2 & x_2+\tau & \cdots & x_2+(d-1)\tau \\ \vdots & \vdots & \ddots & \vdots \\ x_m & x_m+\tau & \cdots & x_m+(d-1)\tau \end{bmatrix},$$

where $d$ is the embedding dimension, $\tau$ is the delay time for the phase space reconstruction and $m = n - (d - 1)\tau$ is the number of points in the $d$-dimensional attractor. If $\tau = 1$, the matrix $X$ is a Hankel matrix, because it has equal elements on positive-slope skew-diagonals where the sum of the row and column subscripts is equal to a constant. If $\tau > 1$, equal elements in $X$ are not restricted to positive-slope skew-diagonals.

(b) Symplectic QR decomposition

If matrix $A = X^TX$ is the auto-correlation of the trajectory matrix $X_{m \times d}$, we can construct a Hamiltonian matrix

$$M = \begin{bmatrix} A & 0 \\ 0 & -A^T \end{bmatrix}.$$ 

After the real Hamiltonian matrix, $M$ is squared to form $N = M^2$, a symplectic orthogonal matrix $Q$ is constructed such that

$$P^TNP = \begin{bmatrix} B & R \\ 0 & B^T \end{bmatrix},$$

where $B$ is upper Hessenberg matrix. Various methods can be used to construct the symplectic orthogonal matrix $P$\textsuperscript{[29,34]}. Given a Householder matrix $Q$, it can be shown that the matrix $H = [Q \ 0; \ 0 \ Q]$ is also a Householder matrix and that, furthermore, $H$ is symplectic \textsuperscript{[34]}. In order to simplify the computation, we can construct matrix $Q$ by Schmidt orthogonalization, using $H$ to replace $P$ to obtain the upper Hessenberg matrix. The QR algorithm is used to compute the eigenvalues $\sigma(B) = \{\sigma_1, \sigma_2, \ldots, \sigma_d\}$ of $B$. If $A$ is real and symmetric, then the eigenvalues of $A$ are equal to those of $B$. Furthermore, the eigenvalues $\lambda(X)$ of $X$ can be obtained from the positive
square roots of $\sigma(B)$ as
\[ \lambda_j = \sqrt{\sigma_j}, \quad j = 1, 2, \ldots, d, \] (2.4)
with eigenvalues sorted in descending order
\[ \sigma_1 > \sigma_2 > \cdots > \sigma_d. \] (2.5)

The distribution of $\sigma_j$ is the symplectic geometry spectra of $A$ with relevant symplectic orthonormal bases. Low values of $\sigma_j$ are often related to the noise component in the data. The corresponding matrix $Q$ represents the symplectic eigenvectors of $A$, such that the $i$th column of $Q (Q_i)$ is equal to the $i$th eigenvector of $A$. The transformed coefficients matrix $S$ is then formed such that the $i$th row, $S_i$, is given by
\[ S_i = Q_i^T X \quad (i = 1, 2, \ldots, d). \] (2.6)

We then obtain the corresponding reconstructed matrix
\[ X_i = Q_i S_i \quad (i = 1, 2, \ldots, d). \] (2.7)

The trajectory matrix can be written as
\[ X = X_1 + X_2 + \cdots + X_d. \] (2.8)

(c) Grouping

The purpose of this step is to identify components with different periods as well as the noise. The grouping procedure splits the set of indices $(i = 1, 2, \ldots, d)$ into $p$ disjoint subsets $I_1, I_2, \ldots, I_p$, so that the elementary matrix in equation (2.8) is regrouped into $p$ groups. Let $I = \{i_1, i_2, \ldots, i_l\}$, then the resultant matrix corresponding to the group $i$ is defined as $X_i = X_{i_1} + X_{i_2} + \cdots + X_{i_l}$. The partition of the set $\{1, 2, \ldots, d\}$ into disjoint subsets $I_1, I_2, \ldots, I_p$ corresponds to the new expansion
\[ X = X_{I_1} + X_{I_2} + \cdots + X_{I_p}. \] (2.9)

(d) Diagonal averaging

The final step of SGSA is the transformation of each regrouped matrix in equation (2.9) into a new series with the same length $n$. The aim of diagonal averaging is to generate reconstructed elements of each resultant matrix by averaging over positive-slope skew diagonal groupings. The new element has the same position as the corresponding elements in the original series. For $\tau > 1$, the diagonal averaging can be carried out using the following algorithm. Consider matrix $Y_{m \times d}$ with elements $y_{ij}, 1 \leq i \leq m, 1 \leq j \leq d$. Let $d^* = \min(m, d)$, $m^* = \max(m, d)$ and $n = m + (d - 1)\tau$. Let $y_{ij}^* = y_{ij}$ if $m < d$ and $y_{ij}^* = y_{ji}$ otherwise. Diagonal averaging transfers matrix $Y$ to a series $y_1, y_2, \ldots, y_n$ according to the following:

\[ y_k = \begin{cases} 
\frac{1}{k} \sum_{p=1}^{k} y_{p,k-p+1} & \text{for } 1 \leq k < d^* \\
\frac{1}{d^*} \sum_{p=1}^{d^*} y_{p,k-p+1} & \text{for } d^* \leq k \leq m^* \\
\frac{1}{n-k+1} \sum_{p=k-m^*+1}^{n-m^*+1} y_{p,k-p+1} & \text{for } m^* < k \leq n.
\end{cases} \] (2.10)

Equation (2.10) corresponds to the averaging of the matrix elements over the ‘diagonals’ $i + j = k + 1$. Such diagonal averaging, applied to any sub-matrix in equation (2.9), generates an $n$-length
time series, and thus the original series $S$ is decomposed into a sum of $p$ series
\[ S = S_1 + S_2 + \cdots + S_p. \] (2.11)

For the case $\tau > 1$, we can insert a $d^* \times (\tau - 1)$ zero matrix into every two adjacent columns of $Y_{d^* \times m^*}$ to form a new matrix
\[
Y' = \begin{bmatrix}
  y_{11} & 0 & \cdots & 0 & y_{12} & 0 & \cdots & 0 & y_{13} & \cdots & y_{1m^*} \\
  y_{21} & 0 & \cdots & 0 & y_{22} & 0 & \cdots & 0 & y_{23} & \cdots & y_{2m^*} \\
  \vdots & 0 & \cdots & 0 & \vdots & 0 & \cdots & 0 & \vdots & \cdots & \vdots \\
  y_{d^*1} & 0 & \cdots & 0 & y_{d^*2} & 0 & \cdots & 0 & y_{d^*3} & \cdots & y_{d^*m^*}
\end{bmatrix}. \] (2.12)

The diagonal averaging formula (equation (2.10)) is then applied to the new matrix $Y'$ to generate the time series. However, the number of elements belonging to every $d^* \times (\tau - 1)$ zero matrix should be subtracted from the total number of diagonal elements when carrying out the averaging.

With proper choices of $d$ and the sets $I_1, I_2, \ldots, I_p$, these SRC can be associated with the trend, oscillations or noise of the original time series. If grouping is skipped, the trajectory matrix can be decomposed into $d$ SRCs.

### 3. Symplectic geometry spectrum analysis applied to synthetic chaotic data

State space reconstruction and prediction of chaotic time series are two fundamental and important problems in chaotic dynamics. To demonstrate the usefulness of the proposed SGSA method, we apply it to reconstruct and predict two benchmark chaotic systems embedded in noise: the Lorenz system and the Mackey-Glass system.

The Lorenz system is defined by the following three-state nonlinear ordinary differential equations:
\[
\begin{align*}
  \frac{dx}{dt} &= \sigma(y - x), \\
  \frac{dy}{dt} &= -xz + rx - y \\
  \frac{dz}{dt} &= xy - bz,
\end{align*}
\] (3.1)

with parameter values $\sigma = 10$, $b = 8/3$ and $r = 28$.

The Mackey-Glass system is given by the delay differential equation
\[
\frac{dx}{dt} = -bx(t) + \frac{ax(t - \tau_0)}{1 + x^c(t - \tau_0)},
\] (3.4)

where $a = 0.2$, $b = 0.1$, $c = 10$ and $\tau_0 = 17$.

We use fourth-order explicit and implicit Runge–Kutta numerical integration algorithms for the Lorenz and Glass-Mackey systems, respectively, to simulate the original noise-free time series. The integration step for the Lorenz and Mackey-Glass equations were set to 0.02 and 0.5, respectively. The first 5000 points are discarded to eliminate initial transients.

#### (a) Reconstruction of noisy chaotic attractors

The embedding dimension for both systems is set to eight. Noise is then superimposed onto the time series through the addition of independent and identically distributed (i.i.d.) Gaussian white noise with various NLs. Figures 2 and 3 illustrate the SGSA decomposition results for the noisy Lorenz and Mackey-Glass series, respectively, with 10% NL. The top panel of both figures shows the original noisy data, whereas the lower panels illustrate the eight reconstructed components by SGSA. The first SRC, showing the main trend, is similar to the original series. Compared with this, the amplitudes of the remaining SRCs are much lower. A simple Fourier transform could
indicate the different frequency band distributions of each SRC. These results demonstrate that SGSA can reliably extract various sub-band components from the original series, highlighting its ability in various applications, including time-series de-noising, compression, prediction and trend detection.

Figure 4 illustrates the symplectic geometry spectrum of the same Lorenz and Mackey-Glass series but with various NLs ranging from 2 to 10% in steps of 2%. Although the NLs are different, the first two SRCs of both dynamic systems are almost the same for each series. We can, therefore, deduce that the first and the second SRCs contain the most information on this system. Figures 5 and 6, respectively, show the reconstructed attractors of the clean Lorenz and Mackey-Glass series (red), along with that constructed from the sum of various SRCs under 10% NL (blue), in which the first sub-panel shows the attractor of the first SRC, until the eighth sub-panel which shows the attractor of the sum of all SRCs. From visual inspection, the clean attractor’s envelope (red) exceeds that of the first SRC (blue) in the first sub-panel of both figures, indicating an information loss for using only the first SRC to reconstruct the noisy attractor. However, we also observe that

Figure 2. (a) $x$ component of noisy Lorenz series with 10% NL. (b) Eight reconstructed components by SGSA. (Online version in colour.)
the blue envelopes of sub-panels 3–8, corresponding to the sum of the first three to eight SRCs, exceed that of the clean data (red). This implies that redundant SRCs do not contribute to the accuracy of reconstructed attractors due to the noise component they represent. However, the clean data attractors for both series are best matched to the sum of the first two SRCs shown in sub-panel 2 for each figure. This is consistent with the fact that the first and the second SRCs contain the most essential information of both systems indicated in figure 4. Therefore, it can be concluded that the dynamic trajectory of a noisy dynamic system could be recovered by selecting an appropriate number of SRCs obtained from SGSA.

(b) Prediction of noisy chaotic time series

A popular method for chaotic time-series prediction is the local approximation (LA) in phase space, proposed by Farmer & Sidorowich [37]. A time series with a single dynamical variable is first embedded into a \( d \)-dimensional state space similar to that of SGSA. The dynamical system generating the time series can be interpreted in the form of a nonlinear function \( F_\tau \),

---

**Figure 3.** (a) Noisy Mackey-Glass series with 10% NL. (b) Eight reconstructed components by SGSA. (Online version in colour.)
\[ x_{i+\tau} = \hat{F}_\tau X_i \]

In practice, a least-square criterion is often used to determine the local \( \hat{F} \) for the first-order approximation. The value of predicted \( x_{i+\tau} \) is obtained from equation (3.5). In order to make the prediction more than one step ahead, earlier predictions are used in constructing a new LA function \( \hat{F} \) [37]. In the development of the LA approach, it is assumed that the prediction accuracy is determined by the approximation quality, which, in turn, depends on the volume of the available data. The method is, therefore, effective for a time series consisting of an infinite sequence of noise-free observations [38]. However, in real life, experimental observations are inevitably corrupted by various sources of noise. In such cases, the probability that false neighbours appear and true neighbours are rejected increases. The prediction accuracy is naturally limited by the erroneous choice of phase space neighbours, even when increasing the number of observations. Moreover, the predicted values are typically iterated to obtain further future values in multi-step ahead prediction. Hence, any prediction error at a given time will propagate and accumulate in later predictions. To improve the prediction quality, it therefore seems useful to combine SGSA and LA into a unified method. Here, we use SGSA to filter the initial time series before making a local prediction. For the above SRCs obtained from noisy Lorenz and Mackey-Glass series, we use only the sum of the first two SRCs as input to construct the local prediction model. Two thousand points are used for training the model and the remaining 1000 points are used for testing. The prediction quality is evaluated in terms of the normalized mean squared error (NMSE).

We first evaluate the NMSEs for the SGSA method on the Lorenz and Mackey-Glass series from one- to 20-step predictions when the NL is 2 and 5%. Figures 7 and 8 show the effect of embedding dimension on NMSE while predicting the Lorenz and Mackey-Glass series, respectively, while figures 9 and 10 show the effect of time delay. The prediction lead time spans from 1 to 20. The results shown in these figures demonstrate that appropriate SGSA embedding parameters result in low NMSEs in prediction. For the Lorenz series, the optimal embedding dimensions are 14 and 15 for 2% and 5% NLs, respectively. For the Mackey-Glass series, it is 12 and 13 for the same two NLs, respectively. These values of optimal embedding dimensions for noisy series are consistent
with the C-C-1 method proposed by Cai et al. [39] for the selection of embedding dimension. Therefore, we recommend their method for determining the embedding dimension in the first step of SGSA. Figure 9 indicates the optimal delay for Lorenz series prediction is 15 and 16 for 2% and 5% NLs, respectively, and the value is 14 and 15 for Mackey-Glass series, as shown in figure 10. Similarly, the optimal time delay parameters for prediction of these noisy series are also in agreement with the C-C-1 method [39]. In summary, our results indicate that the error in SGSA prediction could be significantly decreased by selecting an appropriate number of SRCs, embedding dimension and time delay.

4. Symplectic geometry spectrum analysis applied to real-life data

To test the applicability of SGSA to real-life data, we first apply it to decompose a compound signal comprising two sinusoids, and subsequently using it for prediction of two real-life

---

**Figure 5.** Reconstructed attractor for the clean Lorenz series (red) and the sum of its various SRCs (blue) with 10% Gaussian noise. (Online version in colour.)
Figure 6. Reconstructed attractor for the clean Mackey-Glass series (red) and the sum of its various SRCs (blue) with 10% Gaussian noise. (Online version in colour.)

(a) Decomposition of deterministic signal

A simple compound signal $S$ comprising two sinusoids is used in this simulation

$$S = \sin(2\pi t) + 5\sin(20\pi t).$$ (4.1)

Figure 11a,b shows the compound signal and its two sinusoid components, respectively. With similar decomposition and reconstruction procedures used for the Lorenz and Mackey-Glass series, SGSA can accurately recover two different frequency components (figure 11c), demonstrating the effectiveness of SGSA for signal reconstruction. However, there is a reconstructed error at each end of each reconstructed component. This is the so-called ‘end effect’, also happening in other time-series decomposition methods, for example in HHT [20]. Parameter optimization or mirror extending methods can be adopted to effectively eliminate or at least alleviate the ‘end effect’ in signal reconstruction [20].

data. The decadal sunspot data series, a benchmark dataset in time-series prediction, and a mechanomyographical signal collected from human skeletal muscle, are chosen as examples.
(b) Prediction of sunspot data

The forecasting of solar activity plays an important role in various scientific areas, including the operation of low-Earth orbiting satellites, long-term forcing in climate models and geophysical applications [40]. The solar cycle is very difficult to predict due to noise contamination, high dispersion level and high variability both in phase and amplitude at different time scales. The decadal sunspot number, representing the average number of sunspots per annum over a 10-year interval centred on a given year, considered here is composed of three back-to-back series covering the period from 9455 BC to AD 2007. The first dataset of the historical period 9455 BC to AD 1895 is covered by the sunspot number based on dendrochronologically dated radiocarbon
concentrations (SNRC) [41]. The second dataset, grouping sunspot number (GSN), spans from AD 1885 to 1995 [42]. GSN is a new updated sunspot series presented by Hoyt & Schatten [42], which contains 80% more raw information and is more homogeneous than the Wolf sunspot number [42,43]. Since January 1981, the Royal Observatory of Belgium has operated the Sunspot Index Data Center (SIDC), the World Data Center for the Sunspot Index. The SIDC collects monthly observations from worldwide stations in order to calculate the International Sunspot Number (ISN). The centre broadcasts the daily, monthly, yearly sunspot numbers, with middle-range predictions (up to 12 months) [44]. The third dataset, the ISN during the period 1995–2007, is connected to GSN dataset [40]. Volobuev & Makarenko [40] used a smoothing spline technique to adjust the GSN smoothness to SNRC to construct the composite dataset. They also presented

Figure 9. The effect of time delay on NMSE for short-term prediction of noisy Lorenz series with 2 and 5% Gaussian noise using SGSA. (Online version in colour.)

Figure 10. The effect of time delay on NMSE for short-term prediction of noisy Mackey-Glass series with 2 and 5% Gaussian noise using SGSA. (Online version in colour.)
Figure 11. Time-series signal: (a) composed of two sinusoids in different frequency and (b) is decomposed using SGSA method. Two different frequency components are accurately reconstructed (c). (Online version in colour.)

Figure 12. Decadal average sunspot series over AD 1000 to 2000 and its 10-step prediction values by SGSA and ILLA. (Online version in colour.)

The method to predict this dataset. Figure 12 shows the actual decadal sunspot values from AD 1007 to 2007, as well as the 10-step predicted values by SGSA and ILLA. As evidenced by the predicted curves and NMSE values (0.0332 for ILLA and 0.0107 for SGSA), the predicted values using ILLA deviate much more from the actual values when compared with the one obtained using SGSA.

(c) Prediction of mechanomyographic signal

There are multiple scenarios in which short-term prediction can be a valuable tool in medicine. Seizure prediction via analysis of the electroencephalogram before the onset of physical manifestations can improve therapeutic possibilities and the quality of life for epilepsy patients [45]. Ventricular fibrillation (VF) is the most common mode of death among cardiac patients. In
order to effectively offer high-energy defibrillation therapy for VF, the development of an accurate and rapid algorithm for VF short-term prediction using the electrocardiogram is of paramount importance [46]. Electromyographic (EMG) and MMG signals, recordings of electrical and mechanical activities detectable on the body surface during muscle contraction, have been applied to control different human–machine interfaces (HMI)s [47–49]. A real-time signal prediction scheme is needed in these systems in order to compensate for the effect of muscle fatigue, which can degrade the robustness and accuracy of HMI{s} [50,51]. In this study, we apply the current and preceding 2000 MMG data points as a training sample to perform a real time 32-step prediction. The MMG data are at a sample rate of 1000 Hz. A detailed description of the experimental protocol and data recording procedures can be found in Xie et al. [48]. The reason for setting this step length is that the minimal length of a moving window satisfying real-time control in MMG or EMG-based HMI is 32 in most cases [48]. Figure 13 shows the actual and predicted values using SGSA and ILLA on a 500-point MMG segment. The NMSEs of the SGSA and ILLA prediction are 0.0437 and 0.0896, respectively. Compared with ILLA, this example further indicates the superiority of SGSA for nonlinear time-series short-term prediction.

5. Conclusion

We have presented a new theoretical framework for time-series decomposition based on symplectic geometry theory. The framework can be used as an alternative method for wavelet transform, ICA, Hilbert–Huang transform and SSA in various applications. In this study, we have used it to reconstruct the attractor from noisy dynamic systems, as well as recover the sinusoidal signal components. We have also used the reconstructed SGSA components to predict both simulated and experimentally recorded noisy time series. The performance of the method on two benchmark problems in astrophysics and medicine shows its superiority over the existing ILLA method. Efforts into investigating the applicability of the SGSA method for studying several other problems associated with nonlinear dynamic systems, including classification, missing data estimation, are currently underway, details of which will be reported elsewhere.

References


