Acta Appl Math (2010) 112: 205–221
DOI 10.1007/s10440-010-9560-z

A Nonlinear Generalization of Singular Value Decomposition and Its Applications to Mathematical Modeling and Chaotic Cryptanalysis

Prabhakar G. Vaidya · Sajini Anand P. S · Nithin Nagaraj

Received: 11 May 2009 / Accepted: 7 January 2010 / Published online: 23 January 2010

Abstract Singular Value Decomposition (SVD) is a powerful tool in linear algebra and has been extensively applied to Signal Processing, Statistical Analysis and Mathematical Modeling. We propose an extension of SVD for both the qualitative detection and quantitative determination of nonlinearity in a time series. The method is to augment the embedding matrix with additional nonlinear columns derived from the initial embedding vectors and extract the nonlinear relationship using SVD. The paper demonstrates an application of nonlinear SVD to identify parameters when the signal is generated by a nonlinear transformation. Examples of maps (Logistic map and Henon map) and flows (Van der Pol oscillator and Duffing oscillator) are used to illustrate the method of nonlinear SVD to identify parameters. The paper presents the recovery of parameters in the following scenarios: (i) data generated by maps and flows, (ii) comparison of the method for both noisy and noise-free data, (iii) surrogate data analysis for both the noisy and noise-free cases. The paper includes two applications of the method: (i) Mathematical Modeling and (ii) Chaotic Cryptanalysis.

Keywords Nonlinear time series analysis · Singular value decomposition · Chaotic Cryptanalysis · Mathematical Modeling

Mathematics Subject Classification (2000) 37N30 · 46N40 · 62J02

1 Introduction

Singular Value Decomposition (SVD) along with its related variations known as Principal Component Analysis and Independent Component Analysis are powerful techniques for

P.G. Vaidya · S. Anand P. S (✉)
National Institute of Advanced Studies, Indian Institute of Science Campus, Bangalore 560 012, India
e-mail: sachoo@gmail.com

P.G. Vaidya
e-mail: pgvaidya@yahoo.com

N. Nagaraj
Amrita School of Engg. Amrita Viswa Vidyapeetham, Kollam, Kerala 690525, India
e-mail: nithin@am.amritapuri.edu

data analysis in linear algebra which have found lot of applications in various fields such as signal processing, statistical analysis, biomedical engineering, genetics analysis, mathematical and statistical modeling, graph theory, psychology etc. [1–8]. Historically SVD has been used for finding the dimension of a linear system as it gives statistically independent set of variables which could span the state space. One SVD based method known as singular spectrum analysis are used for detecting nonlinearity in a qualitative manner [9]. An abrupt decrease in the profile of the singular spectrum is an indication of lower dimensional determinism. But this method fails to distinguish a chaotic time-series from its surrogates [10, 11]. Surrogates are stochastic counterparts of the data which has the same power spectrum.

Bhattacharya et al. proposed a method of quadratic scaling of singular values, that can distinguish between the data series and its surrogates [14]. Quadratic scaling gives more weightage to the decreasing singular values to highlight the deterministic and stochastic features.

The Grassberger–Procaccia (GP) algorithm is also used to show that a finite correlation dimension of an irregular time series is an indication of underlying deterministic nonlinearity [15]. GP algorithm has a few drawbacks as it fails when the data is noisy and is unable to distinguish stochastic processes with power-law power-spectra from chaos [16, 17]. The distribution of correlation coefficients is also used as a qualitative method to distinguish chaos from noise because the spectrum is flat for noise but gradually decays for chaos [18]. This method fails to distinguish between correlated noise and chaos [19, 20].

There are developments in the field of nonlinear auto regressive (NAR) and auto regressive moving average (NARMA) models for processing of nonlinear series. One method which is an alternative to linear regression, known as ACE (Alternating Conditional Expectation) algorithm is used to identify the optimal transformations to the data such that a linear regression model can be applied to the transformed data [21]. Many robust algorithms are used to find the parameters of appropriate models. Some them are: fast orthogonal search (FOS) method that can obtain correct model parameters irrespective of the model selection [22], one of its variation known as optimal parameter search (OPS) algorithm [23], various least square techniques: least square [24], total least square (TLS) [25] and minimizing the hypersurface distance (MHD) method [26]. These models get the estimates of parameters based on the best nonlinear polynomial fit. The main limitation of parametric models with respect to the processing of data generated by dynamical systems is that the parameter estimation is purely based on the statistical properties of the signal such as density and probability distribution, thereby ignoring the dynamical perspectives.

Chaotic signals are generated by lower dimensional deterministic nonlinear dynamical systems. They resemble noise though it is generated by an entirely deterministic system. Noise on the other hand is stochastic and very less is known about its origin. Noise is defined mainly based on its statistical properties (variance, mean, correlation, entropy) where as Chaotic signals can be defined based on its dynamic properties (initial condition, generating equation). Chaotic signals typically have small degrees of freedom compared to noise.

Embedding based techniques are useful for modeling as they can give qualitative information of the dynamics from the experimental time series. Data generated by nonlinear chaotic systems live in a lower dimensional manifold, unlike data from pure stochastic systems that tend to spread over the whole phase space. This property of nonlinear chaotic systems can be exploited for modeling and prediction of nonlinear chaotic signals. Embedding is defined as a smooth map Φ from a manifold M to a space U such that its image $\Phi(M) \subset U$ is a smooth manifold of U and Φ is a diffeomorphism between M and $\Phi(M)$. The existence theorem for embeddings in euclidean spaces was given by Whitney [27]. He proved that a smooth n-dimensional manifold can be embedded in R^{2n+1}. This theorem is

the basis of reconstruction techniques for phase portraits from time series measurements. Packard et al. suggested that phase portraits diffeomorphic to the underlying dynamical system could be reconstructed from time series [28]. Takens suggested reconstruction based on delay vectors of the time series and gave a firm theoretical foundation for both delay and derivative embedding [29].

Different Local Nonlinear Prediction (LNP) methods are available in the literature that use embedding-based techniques to model and predict bio-medical signals. Porta et al. suggest a method for LNP based on Takens embedding of data in a higher dimensional phase space and then subdividing the phase space into non-overlapping hypercubes [30, 31]. Prediction is based on the behavior of the median of the values given that the past samples belong to the same hypercube. Later Broomhead and King proposed SVD-based embedding to get an orthogonal basis such that as the chosen embedding dimension m increases beyond $2n + 1$ the manifold is confined to a subspace of fixed dimension [9]. SVD-based embeddings have some advantage over delay embeddings. Delay embedding chooses a basis for the embedding space in an arbitrary manner. Here the set of linearly independent orthonormal vectors given by SVD can form the local basis of the embedded manifold $\Phi(M)$. SVD-based embedding can be used for noise reduction by partitioning the embedding space and rejecting the *out-of-band* noise i.e. the less significant partition of the embedding space.

In this paper, we discuss a nonlinear extension of SVD to detect and exactly determine the nonlinearity. The proposed method of nonlinear SVD for maps is similar to the nonlinear AR and ARMA regression proposed by Lu et al. [23] and Marmarelis [32]. But the method is not limited to polynomial regression. Any deterministic nonlinearity present in the data could possibly be recovered. Nonlinear SVD method could be considered as functional regression in an extended phase space. Nonlinear SVD and time delayed embedding together with the method of finding derivatives from data [36] can be used to identify the nonlinearity of the system from the time series.

The paper is organized as follows. Section 2 briefly reviews the conventional SVD technique. Section 3 describes the method of nonlinear SVD and Sect. 4 discuss the applicability of the method to data series. A numerical example of retrieval of Logistic map parameter from data is discussed in Sect. 5. Section 6 gives the comparison of nonlinear SVD and conventional SVD analysis of the data and its surrogates. Section 7 discusses the extension of the method for higher order maps. Section 8 shows the numerical results in the presence of two types of noises: uniform noise and gaussian noise. The paper includes two applications of the method. Section 9 demonstrates an application of the method to Mathematical Modeling. An application of the method to Chaotic Cryptanalysis is given in Sect. 10. Section 11 is a discussion on linear and nonlinear models and the conclusions are given in Sect. 12.

2 Singular Value Decomposition

Singular Value Decomposition can be considered as a generalization of the eigen decomposition of square matrices, to analyze rectangular matrices. SVD decomposes a rectangular matrix into three simple matrices: two orthogonal matrices and one diagonal matrix [24]. In general, SVD theorem can be stated as follows: any $m \times n$ matrix A, with $m \geq n$ can be factored into three matrices: U (column orthogonal, $m \times n$ matrix), W (diagonal, $n \times n$ matrix) and V (orthogonal, $n \times n$ matrix). When A is real, $A = UWV^T$ (where V^T is the transpose of V). For complex matrices, W remains real but U and V become unitary. The diagonal elements of W matrix are known as the singular values of A.

This decomposition is a technique that works well with matrices that are either singular or else numerically very close to singular. SVD is also used to calculate pseudo-inverses when the natural inverse of the matrix does not exist [37]. SVD and pseudo-inverses are generally used in statistics for solving least square problems. Data compression using SVD is one of the standard applications in image processing [24, 38].

3 Method of Nonlinear Singular Value Decomposition

Given a time series generated by any system $\{X\} = \{X_1, X_2, X_3, \ldots, X_N\}$ the aim of nonlinear system identification is (i) to detect if there is nonlinearity and (ii) to exactly determine the equation which generated the data. The current study is based on the data generated by nonlinear maps and flows. We will begin by using the standard Takens embedding: a method of reconstruction of the state space with time delayed data segments known as embedding vectors [29]. The embedding matrix E is created from the time delayed vectors as follows. A typical embedding vector Y^i is the m dimensional embedding vector generated from the given time series

$$\begin{aligned} Y^1 &= (X_1\ X_2\ \ldots\ X_m)^T, \\ Y^2 &= (X_2\ X_3\ \ldots\ X_{m+1})^T, \\ &\vdots \\ Y^i &= (X_i\ X_{i+1}\ \ldots\ X_{m+i})^T. \end{aligned}$$

Note that $(X_1\ X_2\ \ldots\ X_m)^T$ is a column vector and denotes the transpose of $(X_1\ X_2\ \ldots\ X_m)$. The collection $\{Y^i\}$ is the time delay embedding of the given data. Let the embedding matrix E be created from k embedding vectors as follows

$$E = \begin{bmatrix} (Y^1)^T \\ (Y^2)^T \\ \vdots \\ (Y^k)^T \end{bmatrix} = \begin{bmatrix} X_1 & X_2 & \ldots & X_m \\ X_2 & X_3 & \ldots & X_{m+1} \\ \vdots & \vdots & \vdots & \vdots \\ X_k & X_{k+1} & \ldots & X_{m+k} \end{bmatrix}. \tag{1}$$

For nonlinear SVD, we extend the embedding matrix E by adding nonlinear columns. Let F be the extended embedding matrix

$$F = [E : f_1\ f_2\ \ldots\ f_i]. \tag{2}$$

The last i columns of F matrix, $f_1, f_2, \ldots, f_i$ are functions of the columns of E matrix: $E^{\langle 1 \rangle}, E^{\langle 2 \rangle}, \ldots, E^{\langle m \rangle}$ where $E^{\langle i \rangle}$ denotes the ith column of E. In general, a nonlinear column refers to the square, cube, any other higher powers of the elements in a column, the element-wise product of two or more columns or any other kind of nonlinearity such as exponentiation and trigonometric functions of the entries in a column. If there is a nonlinear relationship between these column vectors, it could be interpreted as a linear relationship between the time delayed vectors and corresponding nonlinear columns. The dimension of F matrix is $k \times p$ where $p = (m + i)$. This extended embedding matrix can be considered as a higher dimensional linear system. The singular value decomposition can find the

linear relation between the embedding vectors and corresponding nonlinear columns, thus recovering the nonlinear relationship inherent in the data. SVD is performed on F to get,

$$F = UWV^T. \tag{3}$$

Therefore,

$$FV = UW. \tag{4}$$

Expanding (4)

$$F\left[V^{\langle 1\rangle}\ V^{\langle 2\rangle}\ \ldots\ V^{\langle p\rangle}\right] = \left[U^{\langle 1\rangle}\ U^{\langle 2\rangle}\ \ldots\ U^{\langle p\rangle}\right]\begin{bmatrix} W_{1,1} & 0 & 0 & \ldots & 0 \\ 0 & W_{2,2} & 0 & \ldots & 0 \\ 0 & 0 & W_{3,3} & \ldots & \ldots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & W_{p,p} \end{bmatrix}. \tag{5}$$

Using partitions of V and W and expanding along the last column of V and W we get,

$$\left[F^{\langle 1\rangle}\ F^{\langle 2\rangle}\ \ldots\ F^{\langle p\rangle}\right]\begin{bmatrix} V_{1,p} \\ V_{2,p} \\ \vdots \\ V_{p,p} \end{bmatrix} = \left[U^{\langle 1\rangle}\ U^{\langle 2\rangle}\ \ldots\ U^{\langle p\rangle}\right]\begin{bmatrix} 0 \\ 0 \\ \vdots \\ W_{p,p} \end{bmatrix}, \tag{6}$$

$$V_{1,p}(F^{\langle 1\rangle}) + V_{2,p}(F^{\langle 2\rangle}) + \cdots + V_{p,p}(F^{\langle p\rangle}) = 0(U^{\langle 1\rangle}) + 0(U^{\langle 2\rangle}) + \cdots + W_{p,p}(U^{\langle p\rangle}). \tag{7}$$

(Note that $U^{\langle i\rangle}$ stands for the ith column of U matrix. $W_{i,j}$ is the element on ith row and jth column of W matrix). If the pth singular value of W is zero,

$$W_{p,p} = 0,$$

then

$$V_{1,p}(F^{\langle 1\rangle}) + V_{2,p}(F^{\langle 2\rangle}) + \cdots + V_{p,p}(F^{\langle p\rangle}) = 0. \tag{8}$$

Or, in another notation,

$$\sum_p F_{j,p} V_{p,n} = 0 \quad \forall j. \tag{9}$$

Equations (8) and (9) can be seen as a statement that the columns of F including those formed by the nonlinear functions $\{f_1, f_2, \ldots, f_i\}$ now span a linear vector space. This equation is true for all rows of F and hence by exploiting this relation, the dependance between the linear and nonlinear columns of the F matrix can be recovered.

4 Nonlinear SVD of Data

We begin with our assumption that the underlying equation is a function of the delay vectors. In case the data is noise free and the conventional SVD of data does not result in a small enough singular value then we try different F's of the form $F = [E : f_1\ f_2\ \ldots\ f_i]$ as shown in Sect 3. The method is based on trial and error, of different choices of F. The criterion is to

try different f_i's for F matrix and select the F which gives a 'zero' or 'nearly zero' singular value. One could get a zero singular value only when the data is noise free. If the data is contaminated with noise we will have to go for the best fit by selecting the F matrix which gives the least singular values. Next section shows a numerical example of data generated by Logistic map. We made a simple guess for the nonlinear function and it worked for that case. If it did not, we would have tried other functions. The numerical results for noisy data is discussed in Sect. 8. It is found that the method breaks down when the noise level is high. In short when the data is noisy, the singular value $W_{p,p} \neq 0$ even if we get the right F matrix of nonlinear functions. Hence we try different nonlinear functions and choose the F that gives the lowest value of $W_{p,p}$. Ideally we want $\frac{W_{p,p}}{W_{1,1}}$ (the ratio of pth singular value to the 1st singular value) to be below some preset criterion. As the noise level in the data increases chances are higher that the method of nonlinear SVD fails. Therefore our confidence in the estimated model equation goes down with the increase in noise.

The choice of the nonlinear functions is one which involves trial and error. In this paper we have illustrated a choice of quadratic and cubic polynomials and also sinusoidal functions. These three have one thing in common. They are all solutions of low order differential equations. We have found in practice that solutions of differential equations are good candidates for this purpose. It turns out that these three are solutions of linear equations with constant coefficients. But even solutions of equations with variable coefficients (e.g. the Bessel function) or nonlinear low dimensional equations prove to be good candidates.

5 Numerical Example

Consider the data generated by a Logistic map $X_{n+1} = \lambda X_n(1 - X_n)$ where $0 \leq X < 1$ and λ is an unknown parameter. We show in this section how λ could be retrieved from the data. Let the data series sampled at a chosen time delay τ be $\{X_n\} = \{X_1, X_2, \ldots, X_N\}$. The data can be embedded in three dimensions using embedding vectors as shown in the following matrix:

$$E = \begin{bmatrix} X_1 & X_2 & X_3 \\ X_2 & X_3 & X_4 \\ & \cdots & \\ X_5 & X_6 & X_7 \end{bmatrix}.$$

The dimension of E matrix is 5×3. After SVD operation on the embedding matrix E, we observed that none of the singular values of E go to zero indicating no linear dependance present in the data. Now E matrix is extended to F matrix as follows,

$$\begin{aligned} F &= [E \ : \ f_1] \\ &= [E^{\langle 1 \rangle} \ E^{\langle 2 \rangle} \ E^{\langle 3 \rangle} \ f_1] \\ &= [E^{\langle 1 \rangle} \ E^{\langle 2 \rangle} \ E^{\langle 3 \rangle} \ E^{\langle 1 \rangle 2}]. \end{aligned}$$

So that a typical row of F is seen as $[X_p \ X_{p+1} \ X_{p+2} \ X_p^2]$ where p goes from 1 to 4. SVD of F gives a zero singular value $W_{4,4} = 0$ indicating a linear dependence between the first column X_n and the added nonlinear column X_n^2. Equation (8) for this particular numerical example is,

$$V_{1,4}(X_n) + V_{2,4}(X_{n+1}) + V_{3,4}(X_{n+2}) + V_{4,4}(X_n^2) = 0. \tag{10}$$

Table 1 Estimated values of estimates $\hat{a}_1, \hat{a}_2$ for the Logistic map: $X_{n+1} = a_1 X_n - a_2 X_n^2$ where $0 \le X_n \le 1$ for the parameter values $a_1 = 4$ and $a_2 = 4$ in the presence of noise using nonlinear SVD. Peak Signal to Noise Ratio (PSNR) is defined as $20 \log 10\{\max(\text{Signal})/\sqrt{MSE}\}$ and its unit is decibel (dB)

Parameters	Noise (uniform distribution)			Noise (gaussian distribution)		
	$\hat{a}_1$	$\hat{a}_2$	PSNR (dB)	$\hat{a}_1$	$\hat{a}_2$	PSNR (dB)
	4.0367	3.9941	44.2314	4.099	4.0879	43.4502
$a_1 = 4$	4.1427	4.0206	38.7308	4.106	4.0919	41.8398
$a_2 = 4$	3.8757	3.8473	37.7963	4.247	4.1568	37.095
	4.2358	4.1516	36.4626	4.299	4.1484	34.5231
	4.3895	4.2506	33.6095	4.1563	3.9952	32.8561
	4.4809	4.2783	30.7117	4.0301	3.824	28.8170
	4.429	3.8507	24.6316	3.8041	2.5627	24.151

Substituting the numerical values, the exact equation can be recovered from (10) as follows

$$\begin{aligned} -0.696311 X_n + 0.174078 X_{n+1} + 0 X_{n+2} + 0.696311 X_n^2 &= 0, \\ 4X_n - X_{n+1} - 4X_n^2 &= 0, \\ 4X_n(1 - X_n) &= X_{n+1}. \end{aligned}$$

For the numerical example discussed above, the initial condition that generated the trajectory was 0.2 and the selected parameter value λ of the Logistic map was 4. Figure 4(i) shows the phase space of the Logistic map and Fig. 4(ii) is the time series generated using this initial condition. We have observed that the method of nonlinear SVD works reasonably well even in the presence of noise, provided the noise is below some threshold value. Let the data $\{X_n\}$ be contaminated by additive noise $\{P_n\}$ which is either gaussian or uniform, $\{\widehat{X}_n\} = \{X_n\} + \{P_n\}$. The same procedure can be done on the embedding matrix F created from the noisy data $\{\widehat{X}_n\}$ to extract the nonlinearity. Table 1 shows the estimated values of parameter λ for the data generated by Logistic family of maps under the presence of different types of noises. The preset criterion for the noisy case was 10^{-6}. For recovering the parameters, we made an assumption that the singular values smaller than this can be considered zero and the underlying equation is extracted as explained in the noise-free case.

6 Comparison of Non-linear SVD with Standard SVD and Surrogates

Figure 1 shows the quadratically scaled singular value spectrum: $n^2\sigma_n$ versus n; where σ_n is the nth singular value generated by the standard SVD on the embedding matrix E. The time series $\{X\}$ generated from the Logistic map: $X_{n+1} = \lambda X_n(1 - X_n)$ where $0 \le X_n \le 1$ and $\lambda = 4$ is used to create an embedding matrix E with unity delay as explained in Sect. 3. The dimension of the embedding matrix is selected as 21×21 for this particular example. Standard SVD operation gives 21 non-zero singular values: note that the profile gradually increases and slowly comes down. Similarly for the nonlinear SVD operation, the embedding matrix F is generated as explained in Sect. 4. The dimension of the embedding matrix is kept same as that of E i.e. 21×21 of which the last 10 columns are squares of the first 10 columns. Now SVD operation gives 21 singular values, out of which the last 10 are zero.

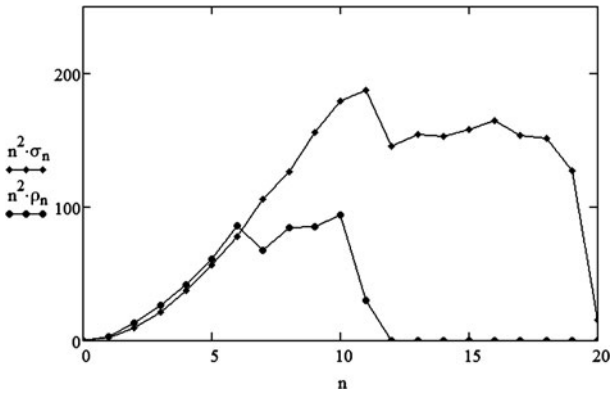

Fig. 1 Quadratically scaled spectrum of singular values for the case of standard SVD ($n^2\sigma_n$) and non-linear SVD ($n^2\rho_n$)

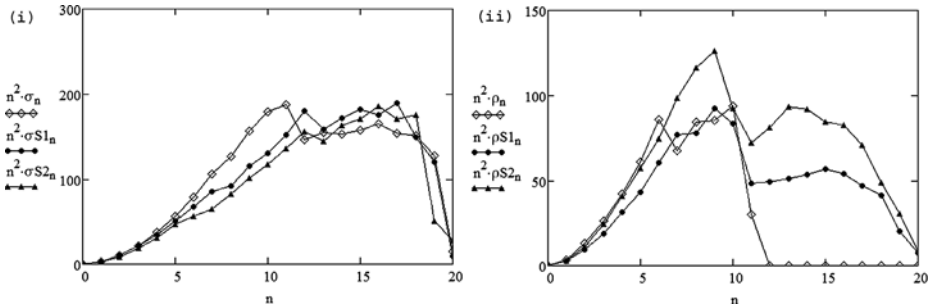

Fig. 2 (**i**) Quadratically scaled spectrum of singular values by standard SVD on the data ($n^2\sigma_n$) and its surrogates ($n^2\sigma_{S1_n}$) and ($n^2\sigma_{S2_n}$). (**ii**) Similar spectrum by nonlinear SVD for the data ($n^2\rho_n$) and its surrogates ($n^2\rho_{S1_n}$) and ($n^2\rho_{S2_n}$). S1 and S2 are two surrogates generated from the Fourier Transform of the original data by randomizing the phases

Observe that the nonlinear SVD profile is significantly different from that of conventional SVD case as the former is 'flat' towards the end. Figure 1, the spectrum for nonlinear SVD shows zero singular values compared to the standard SVD. There is a significant qualitative change in the spectrum. The profile of nonlinear SVD case drops to zero rapidly compared to the standard SVD case.

Similar analysis is done for the surrogates. Surrogates are generated from the Fourier Transform of the data by randomizing the phases; and thereby maintaining the probability distribution and power-spectrum. K surrogate data series $\{Xsurr(k)\}$ are generated from $\{X\}$ such that $\{Xsurr(k)\}$ and $\{X\}$ have the same power spectrum [39, 40]. Hence $\{Xsurr(k)\}$ is considered as the nondeterministic counterpart of $\{X\}$. Figures 2(i) and (ii) show the quadratically scaled spectra of singular values of the data and its surrogates for the standard SVD and nonlinear SVD respectively under noise free conditions. Many surrogate data were generated and only two of them are used for plotting the scaled singular spectra in Fig. 2 for the sake of clarity. We found that all the surrogate data have qualitatively similar spectra. It is clear that the selection of nonlinear columns has worked since the nonlinear SVD has given identically zero singular values as shown Fig. 2(ii). Moreover the method is able to clearly distinguish between the data and the surrogates. The figures also contain the spectrum for the original data $\{X\}$ for a comparison with its surrogates. Similar analysis for noisy data is included in Sect. 8.

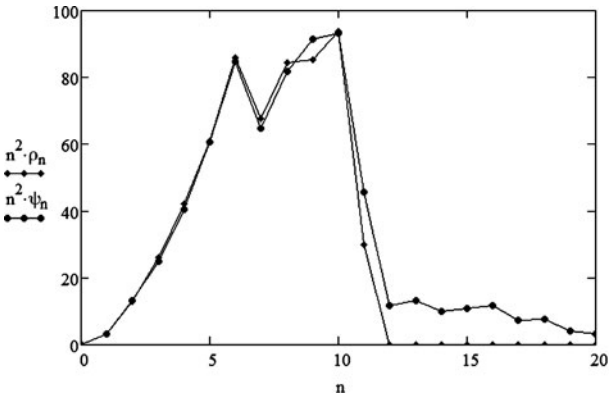

Fig. 3 Quadratically scaled singular value spectra by nonlinear SVD for different choices of F matrices on data generated by Logistic map: $n^2\rho_n$ versus n (where ρ_n is the nth singular value of F with quadratic columns) and $n^2\psi_n$ versus n (where ψ_n is the nth singular value of F with cubic columns)

Assume that we have varied the number and types of the nonlinear terms present in the embedding matrix F. For the case of data from quadratic map, we have added cubic columns in the F matrix instead of quadratic columns. The singular value spectrum looks qualitatively the similar to the quadratic case as shown in Fig. 3. But the exact relationship cannot be retrieved quantitatively, as none of the singular values go to zero. We need to try different embedding matrices and select the one which gives at least one nearly zero singular value.

7 Recovering Non-linearity: Higher Order Maps

Let us now discuss the problem of recovering the non-linear equation when the data is generated from a higher order map of the following form,

$$X_{n+2} = f(X_n, X_{n+1}). \tag{11}$$

The well known Henon map, falls in this category

$$X_{n+1} = c - aX_n^2 + Y_n, \tag{12}$$

$$Y_{n+1} = bX_n. \tag{13}$$

We generated some data using this map and later assumed that only the X data is available. In that case it is more convenient to list this in the form of (11). With that form in mind we set up F to be [1 X_{n+2} X_{n+1} X_n X_n^2 X_{n+1}^2 $(X_{n+1}X_n)$]. Performing SVD on the embedding matrix F we get the following relationship between the iterates

$$0 = 0.496904 - 0.6956656X_{n+1}^2 + 0.1490712X_n - 0.496904X_{n+2},$$
$$0 = 1 - 1.4X_{n+1}^2 + 0.3X_n - X_{n+2}.$$

Therefore,

$$X_{n+2} = 1 - 1.4X_{n+1}^2 + 0.3X_n. \tag{14}$$

This can be seen as equivalent to (12) and (13) with parameter values $a = 1.4$, $b = 0.3$ and $c = 1$. Thus using the proposed method the parameters are retrieved from the X data.

Similarly consider the case of the Logistic map again. We are once again required to find the parameters λ, but all the odd iterates of the time series are suppressed. In this case we could consider the map of the form,

$$X_{n+2} = g(X_n). \tag{15}$$

The non-linear SVD on the modified data can retrieve a quartic nonlinearity. The predicted equation, for this particular example is of the form $g(x) = Bx + Cx^2 + Dx^3 + Ex^4$ where B, C, D, E are functions of the parameter λ. SVD on $F = [X_n\ X_{n+2}\ X_n^2\ X_n^3\ X_n^4]$ gives a zero singular value corresponding to the following relationship between the columns

$$0 = 14.7609X_n - X_{n+2} - 71.4725X_n^2 + 113.4232X_n^3 - 56.7116X_n^4.$$

Therefore,

$$X_{n+2} = 14.7609X_n - 71.4725X_n^2 + 113.4232X_n^3 - 56.7116X_n^4. \tag{16}$$

But for the Logistic map $X_{n+1} = \lambda X_n(1 - X_n)$. Hence the second iterate can be written as a function of its first iterate as follows,

$$X_{n+2} = (\lambda^2)X_n - (\lambda^2 + \lambda^3)X_n^2 + (2\lambda^3)X_n^3 - (\lambda^3)X_n^4. \tag{17}$$

Comparing (16) and (17) we recover the value of the parameter $\lambda = 3.842$; which is in exact agreement with the parameter value we used to generate the data.

8 Numerical Results for Noisy Data

In this section we show some numerical results for noisy data. We will see that the nonlinear SVD continues to help us uncover the underlying deterministic dynamics even when some noise is present.

For each problem, it would be helpful to develop a criterion of the upper bound on acceptable noise levels up to which we can continue to have some confidence in the method. We have found in the examples cited below that this upper bound can be empirically determined in a given problem by using the singular value spectrum as a guide. We looked at the ratio $\frac{W_{p,p}}{W_{1,1}}$ and found it useful to set the upper bound as 10^{-6}. In practice the spectral criterion translates into a criterion for the signal to noise ratio (PSNR). PSNR is defined as $20 \log 10\{\max(\text{Signal})/\sqrt{MSE}\}$ where MSE is the Mean Squared Error, the average of the square of the noise added to the signal. For the examples cited below we found that for PSNR above 28 dB for gaussian noise and 33 dB for uniform noise the spectral criterion mentioned above was satisfied. As the noise level increases the ratio of $W_{p,p}/W_{1,1}$ goes higher than 10^{-6} and our confidence in the method is reduced.

Figures 4(iii) and (iv) show the state space created from the noisy data for both the uniform and gaussian noises. Conventional SVD on the noisy data gives the value of the ratio of the last singular value to the first in the range $(10^{-1}, 10^{-3})$ for different noise levels of the data from Logistic and Henon maps shown in Tables 1 and 2.

At this point we would need to distinguish between two possibilities: (i) the linear model is correct but there is a lot of noise present or (ii) there are nonlinearities, with or without lesser noise present. For these two cases, we can explore some simple nonlinear alternatives and find that the ratio goes below the preset criterion, confirming the second possibility.

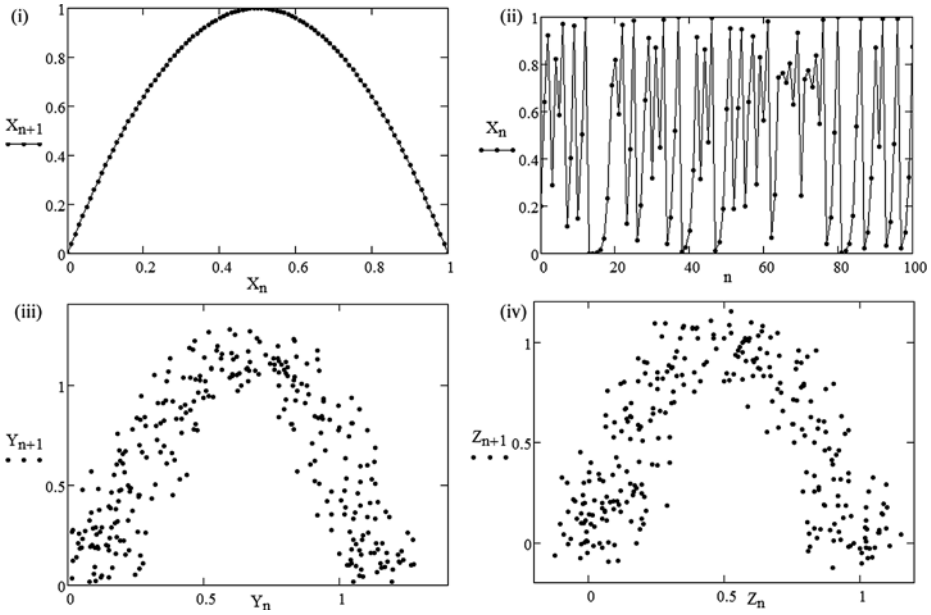

Fig. 4 (**i**) The phase space of the Logistic map: $X \mapsto 4X(1-X)$ where $0 \leq X \leq 1$. (**ii**) A chaotic time series X, generated from initial condition 0.2 using the Logistic map. (**iii**) Phase space for noisy data Y, with added uniform noise in the range $(0, 1)$ with noise level 28.166%. (**iv**) Phase space for noisy data Z with added gaussian noise in the range $(0, 1)$ with noise level 28.289%. Noise level is defined as the ratio of the maximum noise value to the maximum signal value

Table 2 Estimated values of parameters $\hat{a}, \hat{b}, \hat{c}$ for the Henon map $X_{n+1} = c - aX_n^2 + Y_n;\ Y_{n+1} = bX_n$ in the presence of noise using nonlinear SVD

Parameters	Noise (uniform distribution)				Noise (gaussian distribution)			
	$\hat{a}$	$\hat{b}$	$\hat{c}$	PSNR (dB)	$\hat{a}$	$\hat{b}$	$\hat{c}$	PSNR (dB)
	1.389	0.2896	1.003	46.863	1.4001	0.2998	1.003	44.5158
$a = 1.4$	1.3812	0.2923	1.010	40.764	1.3983	0.2964	1.005	40.93
$b = 0.3$	1.357	0.2769	1.009	36.178	1.4	0.3167	0.9881	35.93
$c = 1.0$	1.3717	0.262	1.0204	32.975	1.4376	0.3258	1.008	34.113
	1.3665	0.2758	1.067	30.4932	1.396	0.3279	1.04	31.336
	1.2969	0.2680	1.0349	26.549	1.5206	0.3401	1.08	28.3245
	1.3268	0.1566	1.1489	20.8353	1.2512	0.2917	0.9547	22.714

For nonlinear SVD, we have set value of the criterion $\frac{W_{p,p}}{W_{1,1}}$ as 10^{-6}. If the ratio $\frac{W_{p,p}}{W_{1,1}}$ is below the preset value 10^{-6} the parameters are retrieved from the data. Table 1 shows the estimated values of parameter $\hat{a}, \hat{b}$ for the data generated by Logistic family of maps. $X_{n+1} = a_1 X_n - a_2 X_n^2$ where $0 \leq X_n \leq 1$ for the values of $a = 4$ and $b = 4$ for different Peak Signal to Noise Ratio (PSNR) using the nonlinear SVD algorithm. Similarly Table 2 shows the estimated values for parameter $\hat{a}, \hat{b}, \hat{c}$ for the Henon map $X_{n+1} = c - aX_n^2 + Y_n;\ Y_{n+1} = bX_n$ in the presence of noise using nonlinear SVD.

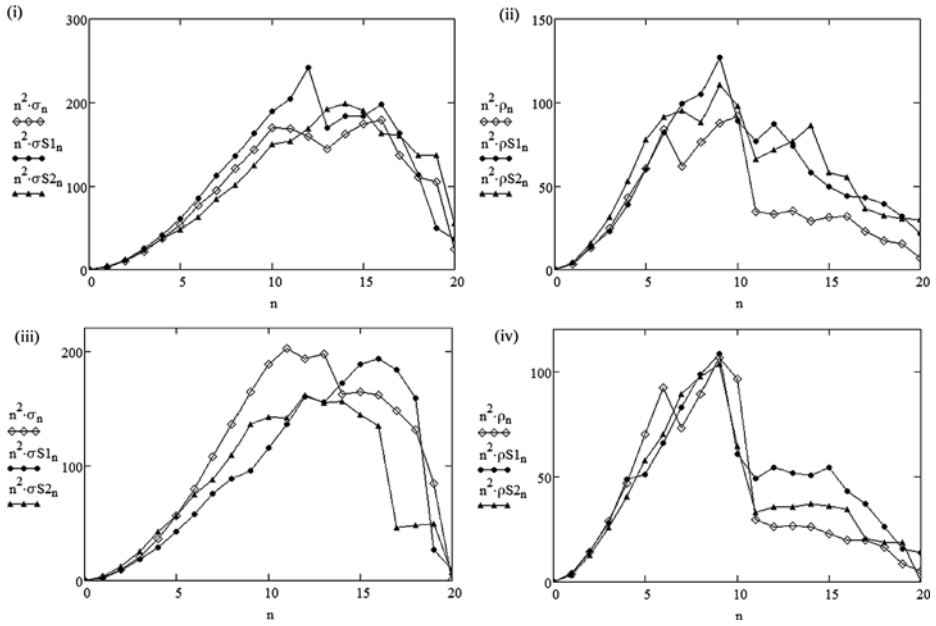

Fig. 5 Quadratically scaled spectrum of singular values by standard SVD on the logistic data ($n^2\sigma_n$) and its surrogates ($n^2\sigma_{S1_n}$) and ($n^2\sigma_{S2_n}$) along with similar spectrum by nonlinear SVD for the data ($n^2\rho_n$) and its surrogates ($n^2\rho_{S1_n}$) and ($n^2\rho_{S2_n}$). (**i**) Standard SVD and (**ii**) nonlinear SVD for the case of gaussian noise in the range (0, 1) with added noise level 28.313%. (**iii**) Standard SVD and (**iv**) nonlinear SVD and for the case of uniform noise in the range (0, 1) with added noise level 28.17%

Figure 5 shows the quadratically scaled singular spectra by standard SVD on the logistic data ($n^2\sigma_n$) and its surrogates ($n^2\sigma_{S1_n}$) and ($n^2\sigma_{S2_n}$) along with similar spectrum by nonlinear SVD for the data ($n^2\rho_n$) and its surrogates ($n^2\rho_{S1_n}$) and ($n^2\rho_{S2_n}$). Figures 4(i) and (ii) show the case of Gaussian noise N(0, 1) with noise level 28.123% in the data. Figures 4(iii) and (iv) show noise added from uniform distribution (0, 1) with noise level 28.089%. Noise level is defined as the ratio of the maximum noise value to the maximum signal value. The size of the embedding matrix is kept constant for all the data and surrogate matrices for both the standard and nonlinear SVD operation. The surrogate data spectrum is added for the comparative analysis. It is clear that nonlinear SVD distinguishes the original data from its surrogates even under the presence of noise, provided the noise level is low.

As we discussed in Sect. 1, if the goal is to detect the nonlinearity but not to determine the underlying model, one could use the method suggested by Porta et al. In [30] the case of the logistic model itself is discussed with parameter value $\lambda = 3.7$ under various noise levels. They have shown that an error function dips much further with actual data than with its surrogates. A comparison of this method and various other methods are discussed in [31].

9 Application to Mathematical Modeling

Let us consider how this method can be useful to fit a nonlinear differential equation to an experimental time series. Assume that the series is taken from a lower dimensional deterministic dynamical system. We want to show the possibility of modeling i.e. fitting a differential equation to the data.

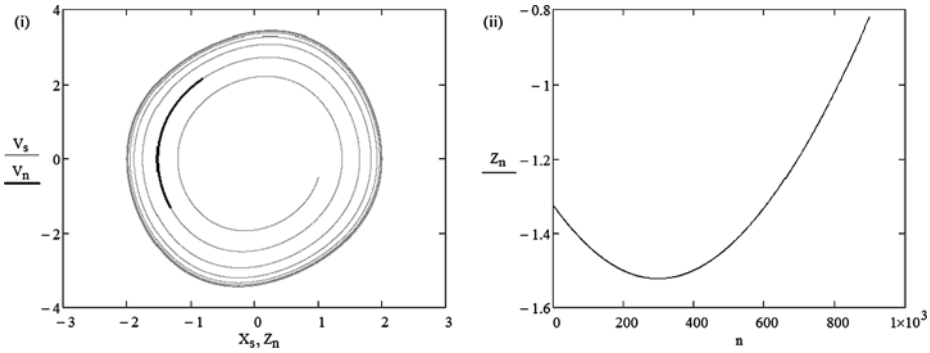

Fig. 6 (**i**) Phase space of the original data (X, V) generated by Team A; the highlighted portion shows the section of a short data segment Z of length 1000 which was sent to Team B. (**ii**) The temporal signal Z

This section shows to identify a specific differential equation from the time series based on two assumptions (i) the sampling is frequent (ii) noise level is very low. We demonstrate here how this could be done by a combination of nonlinear SVD and the method of finding accurate derivative from data [36].

Assume that we are given an experimental time series say Z of length 1000 (accurate up to 10^{-16}) shown in Fig. 6(ii) along with an information that the step size was 0.001. We illustrate the method and the issues involved by showing how we find the unknown differential equation from this time series. In this example, the time series is relatively short, which is clear especially when we see the state space picture Fig. 6(i). (If the method works for a short time series, then if we are dealing with a longer time series, the easiest option available to us is to break the longer series as a sequence of shorter series. In practice, we might be able to do better because for a longer time series, some noise reduction methods might work better. Therefore, we focused on finding a procedure for shorter time series, because this is usually the tougher challenge and in practice the longer version may not be available.)

Let Z be denoted as $Y(t)$. We calculate n derivatives of $Y(t)$ $(Y_2, Y_3, \ldots, Y_{n+1})$ from $Y(t)$ as mentioned in [36]. Y_n is the $(n-1)$th derivative of $Y(t)$. $Y(t)$ is denoted as Y_1 {the '0th' derivative of $Y(t)$} for the rest of the section.

Initially we assumes that data is generated by a differential equation of the following form

$$\frac{d}{dt}X_1 = X_2, \tag{18}$$

$$\frac{d}{dt}X_2 = G(X_1, X_2), \tag{19}$$

where G is a relatively simple multinomial function we need to determine experimentally.

As the first step we assume that G is a nonlinear function consisting of linear and quadratic terms of Y_1 and Y_2 as shown below

$$G = c_1Y_1 + c_2Y_2 + c_3Y_1^2 + c_4Y_1.Y_2 + c_5Y_2^2.$$

Hence the columns of the embedding matrix F were chosen as $[Y_3\ Y_1\ Y_2\ Y_1^2\ Y_2^2\ (Y_1Y_2)]$.

The F matrix is the time delay embedding of data vectors with additional nonlinear columns generated from their derivatives, similar to the case for maps as shown in Sects. 4

and 5. The time delay for embedding is kept as the sampling interval and we got the singular values of F as (201.33, 162.24, 83.95, 3.95, 0.44, 0.03). We get the ratio of 6th singular value to the 1st singular value, $\frac{W_{6,6}}{W_{1,1}}$ as 1.5×10^{-4}. Since this value is above the chosen threshold 10^{-6}, we decides to go to the next step of adding cubic terms to the embedding matrix.

Now the assumed nonlinear function is,

$$G = c_1Y_1 + c_2Y_2 + c_3Y_1^2 + c_4Y_2^2 + c_5Y_1Y_2 + c_6Y_1^3 + c_7Y_2^3 + c_8(Y_1^2Y_2) + c_9(Y_2^2Y_1)]$$

and the corresponding embedding matrix is

$$F = [Y_3\ Y_1\ Y_2\ Y_1^2\ Y_2^2\ (Y_1Y_2)\ Y_1^3\ Y_2^3\ (Y_1^2Y_2)\ (Y_2^2Y_1)].$$

SVD is performed and the singular values of F are (181.0635, 115.3297, 44.2941, 2.1510, 22.1215, 0.1768, 0.0247, 0.0044, 0.0003, (2.995×10^{-11})). Since the last singular value is very small and the ratio $\frac{W_{10,10}}{W_{1,1}}$ is below the chosen threshold value 10^{-6} we assume it to be zero and recover the linear relationship corresponding to that singular value. Hence the retrieved coefficient array is,

$$\hat{C} = [1, 2.895, -0.237, 0, 0, 0, 0, 0, 0.237, 0]^T.$$

And we have recovered the following equation,

$$0 = Y_3 - 0.237Y_2 + 0.237Y_2Y_1^2 + 2.895Y_1, \tag{20}$$

$$Y_3 = 0.237Y_2 - 0.237Y_2Y_1^2 - 2.895Y_1, \tag{21}$$

$$Y_3 = 0.237Y_2(1 - Y_1^2) - 2.895Y_1. \tag{22}$$

Y_2 and Y_3 are the 1st and 2nd derivatives of Y_1. Rewriting (22) we get,

$$\frac{d}{dt}X_1 = X_2,$$

$$\frac{d}{dt}X_2 = cX_2(1 - X_1^2) - kX_1.$$

This is the Van der Pol equation with parameter values $k = 2.895$ and $c = 0.237$.

The estimated values $k = 2.895$ and $c = 0.237$ from the data series are in agreement with the values used in the Van der Pol to generate the data series. The form of the equation and the parameter values are exactly predicted using the proposed method.

This is the case where the data is noise free. We needs the information about the sampling interval in order to calculate the derivatives as mentioned in [36]. Figure 6(i) shows the phase space of Van der Pol oscillator that generated the data and Fig. 6(ii) is the temporal signal Z of length 1000 which was used to find the equation.

10 Cryptanalysis of Chaotic Data

This section explains how the method of nonlinear SVD is used for cryptanalysis. Here the goal is to identify the underlying system equation from the time series based on two assumptions (i) the sampling is frequent, (ii) noise level is very low. We demonstrate how

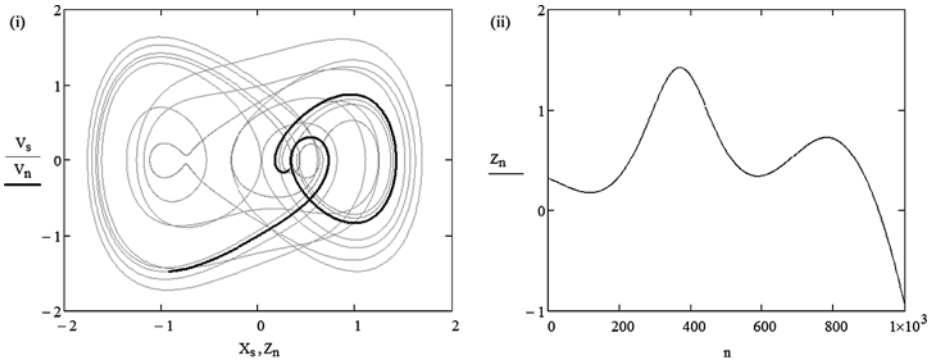

Fig. 7 Phase space of the original data (X, V) generated the duffing oscillator under chaos; the highlighted portion shows the section of a short data segment of length 1000 (Z) which was used to identify the system. The temporal signal Z is shown on the *right side*

this could be done by a combination of nonlinear SVD and the method of finding accurate derivative from data [36].

Consider the following narrative of cryptanalysis. Two Teams A and B are using a communication channel for sending information across each other. The sender, A is generating data from a differential equation of the form

$$\frac{d}{dt}X_1 = X_2, \tag{23}$$

$$\frac{d}{dt}X_2 = -kX_1 - cX_2 - \delta X_1^2 + A\cos(\omega t + \Phi). \tag{24}$$

This is the well known Duffing equation. The parameter values were chosen as $k = 0.01$, $c = 0.04496$, $\delta = 1$, $A = 1.02$, $\omega = 0.44964$ and $\Phi = 0$ such that the system exhibit chaos. The temporal signal and phase space of the duffing oscillator is displayed in Fig. 7. A time series of length 10,000 was generated using Runge–Kutta method from an initial condition $(0.656, 0.172)$ with sampling step size 0.01; from which a portion of X of length 1000 (shown as the short segment Z in Fig. 7) was taken and send to Team B.

Team B is left with a small time series $Y(t) = Z$ (accurate up to 10^{-16}) along with an information that the step size was 0.01. The Team B is also informed that the data is generated using an equation of the form,

$$\frac{d}{dt}X_1 = X_2, \tag{25}$$

$$\frac{d}{dt}X_2 = G(X_1, X_2) + A\cos(\omega t + \Phi), \tag{26}$$

where G is a low order polynomial function. We illustrate the method and the issues involved by showing how Team B works for the unknown differential equation with a short time series (assuming that Team A have ensured adherence to the conditions (i) the sampling was frequent and (ii) noise level was very low).

B calculates n derivatives $(Y_2, Y_3, \ldots, Y_{n+1})$ from $Y(t)$ as mentioned in [36] (where Y_n is the $(n-1)$th derivative of $Y(t)$ and $Y(t) = Y_1$). Now B assume that G is a nonlinear function consisting of linear, quadratic and cubic terms of Y_1 and Y_2 along with the sinusoidal

functions,

$$G = c_{10}Y_1 + c_{01}Y_2 + c_{20}Y_1^2 + c_{11}Y_1Y_2 + c_{02}Y_2^2 + c_{30}Y_1^3 + c_{03}Y_2^3 + c_{21}(Y_1^2Y_2) + c_{12}(Y_2^2Y_1)] + P\sin(\omega t) + Q\cos(\omega t).$$

The corresponding embedding matrix is

$$F = [Y_3\ Y_1\ Y_2\ Y_1^2\ Y_2^2\ (Y_1Y_2)\ Y_1^3\ Y_2^3\ (Y_1^2Y_2)\ (Y_2^2Y_1)\ \sin(\omega t + \Phi)\ \cos(\omega t + \Phi)].$$

SVD operation on F gives the set of singular values (46.1294, 28.0900, 20.9784, 18.8628, 12.5940, 6.8899, 4.8714, 4.2835, 2.0742, 0.8881, 0.5682, 0). Since the last singular value is zero we can recover the linear relationship corresponding to that singular value. The coefficient array $\hat{C}$ is,

$$\hat{C} = [1, 0.01, 0.04496, 0, 0, 0, 1, 0, 0, 0, -0.98958, 0.24723]^T.$$

Now Team B has recovered the following equation from $\hat{C}$

$$Y_3 + 0.01Y_1 + 0.04496Y_2 + 1Y_1^3 - \sqrt{0.98958^2 + 0.24723}\cos(\omega t) = 0,$$
$$Y_3 + 0.01Y_1 + 0.04496Y_2 + 1Y_1^3 - 1.02\cos(\omega t) = 0,$$

where Y_2 and Y_3 are the 1st and 2nd derivatives of Y_1. Rewriting this Team B get the Duffing equation with parameter values $k = 0.01$, $c = 0.04496$, $\delta = 1$, $A = 1.02$, $\omega = 0.44964$ and $\Phi = 0$.

When Teams A and B get together B finds that the estimated parameter values and the form of the equation are in exact agreement with the Duffing equation from which the data was generated.

11 Linear Versus Nonlinear Models

A second order linear system $X_n = aX_{n-1} + bX_{n-2}$ was simulated with parameter values $a = -1.5$ and $b = -1$. In the absence of noise the linear SVD had a sharper fall off than the nonlinear SVD. In the case of surrogate data neither linear SVD nor nonlinear SVD had a sharp fall off. In such cases on the grounds of parsimony alone the linear model would be chosen. In the presence of small amount of noise the same situation continues. However when the noise is beyond a certain value neither linear nor nonlinear SVD will show a substantial qualitative difference with the surrogate data to have any degree of confidence in either of the models.

12 Conclusion

We have proposed a non-linear extension of the singular value decomposition (SVD) technique by means of appending additional columns to the trajectory matrix which are non-linearly derived from the existing columns. We propose nonlinear SVD as a method which is useful for the qualitative detection and quantitative determination of nonlinearity from a short time series. We have demonstrated the utility of non-linear SVD for recovering the non-linear relationship from time series generated by discrete and continuous dynamical

systems. As an example, we have demonstrated the results for the data from Logistic map, Henon map, Van der Pol Oscillator and Duffing oscillator. The proposed method works quiet well in the presence of noise, for both Gaussian and uniform noise (provided the noise level is not high). In principle, the method can work with any type of non-linearity. The paper contains a comparative analysis of the results for the data and its surrogates. It also includes two applications of the method; one to Mathematical Modeling and other to Chaotic Cryptanalysis.

Acknowledgements We thank Department of Science and Technology for supporting the Ph.D. programme at National Institute of Advanced Studies, Bangalore. We thank the reviewers and the editor for their comments which we have found to be very helpful.

References

1. Le Bihan, N., Mars, J.: Signal Proc. **84**, 1177 (2004)
2. Yeung, M.K.S., Tegner, J., Collins, J.J.: Proc. Natl. Acad. Sci. **99**(9), 6163 (2002)
3. De Lathauwer, L., De Moor, B., Vandewalle, J.: IEEE Trans. Biomed. Eng. **47**(5), 567 (2000)
4. Alter, O., Brown, P.O., Botstein, D.: Proc. Natl. Acad. Sci. **100**, 6 (2003)
5. Hu, J., Wright, F.A., Zou, F.: J. Am. Stat. Assoc. **101**(473), 41 (2006)
6. Alter, O., Brown, P.O., Botstein, D.: Proc. Natl. Acad. Sci. **97**(18), 10101 (2000)
7. Luo, B., Hancock, E.R.: IEEE Trans. Pattern Anal. Mach. Intell. **23**(10), 1120 (2001)
8. Takane, Y.: In: Yanai, H., Okada, A., Shigemasu, K., Kano, Y., Meulman, J. (eds.) New Developments in Psychometrics. Springer, Tokyo (2002)
9. Broomhead, D.S., King, G.P.: Physica D **20**, 217 (1986)
10. Sahay, A., Sreenivasan, K.R.: Philos. Trans. R. Soc. Lond. A **354**, 1715 (1996)
11. Abarbanel, H.D.I., Brown, R., Sidorowitch, J.J., Tsimring, L.S.: Rev. Mod. Phys. **65**, 1331 (1993)
12. Kugiumtzis, D.: Physica D **95**, 13 (1996)
13. Buzug, T., Pfister, G.: Physica D **58**, 127 (1992)
14. Bhattacharya, J., Kanjilal, P.P.: Physica D **132**, 100 (1999)
15. Grassberger, P., Procaccia, I.: Physica D **9**, 189 (1983)
16. Osborne, A.R., Provenzale, A.: Physica D **35**, 357 (1989)
17. Ruelle, D.: Proc. R. Soc. Lond. **427**, 241 (1989)
18. Sugihara, G., May, R.: Nature **344**, 734 (1990)
19. Cazellus, B., Ferrare, B.: Nature **25**, 355 (1992)
20. Ensminger, P.A., Vinson, M.: J. Theor. Biol. **170**, 259 (1994)
21. Breiman, L., Friedman, J.: Computer Science and Statistics: Interface, pp. 121–134 (1985)
22. Korenberg, M.J.: Biol. Cybern. **60**, 267 (1989)
23. Lu, S., Ju, K.H., Chon, K.H.: IEEE Trans. Biomed. Eng. **48**(10), 1116 (2001)
24. Strang, G.: Linear Algebra and Its Applications. Brooks Cole, Florence (2003)
25. Nievergelt, Y.: SIAM Rev. **36**(2), 258–264 (1994)
26. Lu, S., Ju, K.H., Chon, K.H.: IEEE Trans. Signal Process. **51**(12) (2003)
27. Whitney, H.: Ann. Math. **37**, 645 (1936)
28. Packard, N.H., Crutchfield, J.P., Farmer, J.D., Shaw, R.S.: Phys. Rev. Lett. **45**, 712 (1980)
29. Takens, K.: Lectures Notes in Mathematics, vol. 898 (1981)
30. Porta, A., Baselli, G., Guzzetti, S., Pagani, M., Malliani, A., Cerutti, S.: IEEE Trans. Biomed. Eng. **47**(12), 1555 (2000)
31. Porta, A., Guzzetti, S., Furlan, R., Gnecchi-Ruscone, T., Montano, N., Malliani, A.: IEEE Trans. Biomed. Eng. **54**(1), 94 (2007)
32. Marmarelis, V.Z.: Ann. Biomed. Eng. **21**, 573 (1993)
33. Vaidya, P.G., Angadi, S.: Chaos Solitons Fractals **17**, 379 (2003)
34. Lia, S., Álvarezb, G., Chena, G.: Chaos Solitons Fractals **25**(1), 109 (2005)
35. Wu, X., Hu, H., Zhang, B.: Chaos Solitons Fractals **22**(2), 367 (2004)
36. Vaidya, P.G.: Chaos Solitons Fractals **17**, 433 (2003)
37. Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T.: Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press, Cambridge (2004)
38. Gonzales, R.C., Woods, R.E.: Digital Image Processing. Addison-Wesley, Reading (1992)
39. Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., Farmer, J.D.: Physica D **58**, 77 (1992)
40. Kennel, M.B., Isabelle, S.: Phys. Rev. A **46**, 3111 (1992)