Nonlinear Analysis: Real World Applications 12 (2011) 2752–2766

Contents lists available at ScienceDirect

Nonlinear Analysis: Real World Applications

journal homepage: www.elsevier.com/locate/nonrwa

# Singular spectrum analysis based on the perturbation theory

Hossein Hassani [a,b,*], Zhengyuan Xu [c], Anatoly Zhigljavsky [a]

[a] Statistics Group, School of Mathematics, Cardiff University, CF24 4AG, UK

[b] Statistical Research and Training Center (SRTC), Tehran, 1413717911, Islamic Republic of Iran

[c] Department of Electronic Engineering, Tsinghua University, Beijing 100084, PR China

ARTICLE INFO

*Article history:*
Received 14 September 2010
Accepted 27 March 2011

*Keywords:*
Singular spectrum analysis
Perturbation theory
Reconstruction
Forecasting

ABSTRACT

Singular Spectrum Analysis (SSA) has been exploited in different applications. It is well known that perturbations from various sources can seriously degrade the performance of the methods and techniques. In this paper, we consider the SSA technique based on the perturbation theory and examine its performance in both reconstructing and forecasting noisy series. We also consider the sensitivity of the technique to different window lengths, noise levels and series lengths. To cover a broad application range, various simulated series, from dynamic to chaotic, are used to verify the proposed algorithm. We then evaluate the performance of the technique using two real well-known series, namely, monthly accidental deaths in the USA, and the daily closing prices of several stock market indices. The results are compared with several classical methods namely, Box–Jenkins SARIMA models, the ARAR algorithm, GARCH model and the Holt–Winter algorithm.



## 1. Introduction

Consider a noisy signal vector $\widetilde{Y}_N = (\widetilde{y}_1, \ldots, \widetilde{y}_N)^T$ of length $N$, where $^T$ denotes the transpose. Let us add the additive noise $\delta Y_N$ to the noise free series (signal) $Y_N$ and assume that the noise $\delta Y_N$ is uncorrelated with the signal:

$$\widetilde{Y}_N = Y_N + \delta Y_N. \tag{1}$$

Let $K = N - L + 1$, where $L$ is some integer called the window length (we can assume $L \leq N/2$). Define the so-called 'trajectory matrix' $\widetilde{\mathbf{X}} = (\widetilde{x}_{ij})_{i,j=1}^{L,K}$, where $\widetilde{x}_{ij} = \widetilde{y}_{i+j-1}$. Note that $\widetilde{\mathbf{X}}$ is a Hankel matrix (by the definition, these are the matrices such that their $(i, j)$-th entries depend only on the sum $i + j$).

We then consider $\widetilde{\mathbf{X}}$ as a multivariate data with $L$ rows and $K$ columns. The columns $\widetilde{X}_j$ of $\widetilde{\mathbf{X}}$, considered as vectors, lie in an $L$-dimensional space $\mathbb{R}^L$. It is obvious that:

$$\widetilde{\mathbf{X}} = \mathbf{X} + \delta\mathbf{X}, \tag{2}$$

where $\mathbf{X}$ and $\delta\mathbf{X}$ represent Hankel matrices of the signal $Y_N$ and noise $\delta Y_N$, respectively. We stress that the perturbation term $\delta\mathbf{X}$ is not assumed to be small and does not tend to be zero. We assume that the perturbation term $\delta\mathbf{X}$ is small relative to $\mathbf{X}$.

The existence of noise in the series can seriously limit the performance of the methods and techniques (see, for example, [1]). In general, there are two main approaches for either fitting a model or forecasting new data points of noisy time series $\widetilde{Y}_N$. According to the first one, ignoring $\delta Y_N$ (or $\delta\mathbf{X}$), we fit a model directly from noisy data (such as ARIMA type models [2]) and use the fitted model for forecasting. According to the second approach, we start by filtering the noisy time

* Corresponding author at: Statistics Group, School of Mathematics, Cardiff University, CF24 4AG, UK.
*E-mail addresses:* HassaniH@cf.ac.uk (H. Hassani), xuzy@tsinghua.edu.cn (Z. Xu), ZhigljavskyAA@cf.ac.uk (A. Zhigljavsky).

1468-1218/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.nonrwa.2011.03.020

series in order to reduce the noise level $\delta Y_N$ (or $\delta \mathbf{X}$) (such as SSA [3]) and fit a model to noise-reduced data and then use the fitted model for forecasting new data points. In the case of the former approach, the fitted model and therefore the forecasting performance are often poor if the noise level is relatively high, especially for the economics and financial time series.

In the ideal case of noise reduction we completely remove $\delta \mathbf{X}$ from the perturbed matrix $\widetilde{\mathbf{X}}$. In this case, the fitted model on the noise-reduced time series should be optimal, as we have removed the noise term $\delta \mathbf{X}$ from the noisy matrix $\widetilde{\mathbf{X}}$. If the noise has been significantly reduced, then the latter approach is expected to give better results than the former approach.

There are several noise reduction methods. Having a method for decomposing the vector space of the noisy time series into a subspace that is generated by the noise free series and a subspace for the noise series, we can construct the noise free time series. An approximate decomposition of the vector space of the noisy time series into noise free time series and noise subspace can be performed with, for example, the orthogonal matrix factorization technique such as singular value decomposition (SVD). It has been shown that SVD based methods and signal subspace (SS) methods are more effective than many others for noise reduction and forecasting in financial and economics time series (see, for example, [4]).

The idea to perform SS method and modified SVD was used in [5] for reconstruction of noise free series. A general framework for recovering noise free series has been presented in [6]. The method forms the basis for a very general class of subspace-based noise reduction algorithms, is based on the assumption that the original time series exhibits some well-defined properties or follows a certain model. Noise free series is therefore obtained by mapping the original time series onto the space of series that possess the same structure as the noise free series.

In this context, the SSA technique, which is a SVD and SS based method, can be considered as a proper method for noise reduction and forecasting time series data sets. The SSA technique includes the elements of classical time series analysis, multivariate statistics, multivariate geometry, dynamical systems and signal processing. The aim of SSA is to make a decomposition of the original series into the sum of a small number of independent and interpretable components such as a slowly varying trend, oscillatory components and a structureless noise [3].

Any seemingly complex series with a potential structure could provide an example of a successful application of SSA [3]. Possible application areas of SSA are diverse: from mathematics and physics to economics and financial mathematics, from metrology and oceanology to social science and market research (see, for example, [7–20] and references therein). A thorough description of the theoretical and practical foundations of the SSA technique (with several examples) can be found in [3,20].

All the aforementioned research is based on the standard SSA technique which is based on the least squares (LS) estimate. The LS estimate of the noise free series is obtained by truncating the singular values. In fact, the LS estimator projects the noisy time series onto the perturbed signal (noise + signal) subspace. The reconstructed series using LS estimator has the lowest possible (zero) signal distortion and the highest possible noise level. Note that the results of noise reduced series using SSA technique (based on the LS estimate), are better than those obtained by classical methods which use noisy series (for comparison between SSA and classical time series methods see, for example, [1,10,12,19]). But as we mentioned above the reconstructed series still has some part of the initial noise level $\delta Y_N$ due to the nature of LS estimate. In fact, the reconstructed series is a mixture of signal and part of the noise. The ideal situation is that we remove, whatever we can, this part of the noise from the reconstructed series. In this paper, we consider an alternative method which is based on the perturbation theory to overcome this problem. The proposed algorithm enables us to remove the part of the noise from reconstructed series which has been obtained using the LS estimate. The similarities and dissimilarities between SSA and the subspace-based methods of signal processing, and application of perturbation theory, with different point of view, have been discussed in [21].

The structure of the paper is as follows. The next section briefly describes perturbation theory and its application for subspace methods. The SSA technique is presented in Section 3 and the improvement of the technique based on the perturbation theory is considered in this section. The empirical results are then presented and described in Section 4 and some conclusions are given in Section 5.

## 2. Perturbation theory

### 2.1. Some theorems

Let us now consider the problem of separation of an additive noise component from the perturbation theory point of view. First we consider some useful theorems.

**Theorem 1.** *Let $\mathbf{X}$ and $\delta \mathbf{X}$ be Hermitian matrices and $\widetilde{\mathbf{X}} = \mathbf{X} + \delta \mathbf{X}$. Let the eigenvalues of $\mathbf{X}$ be $\lambda_1 \geqslant \cdots \geqslant \lambda_L$, and let the eigenvalues of $\widetilde{\mathbf{X}}$ be $\widetilde{\lambda}_1 \geqslant \cdots \geqslant \widetilde{\lambda}_L$. If $\mu_L$ is the smallest eigenvalue of $\delta \mathbf{X}$, then* [22]

$$\widetilde{\lambda}_i \geqslant \lambda_i + \mu_L \quad i = 1, \ldots, L. \tag{3}$$

*There are two useful characteristics about the above theorem; it restricts the location of the eigenvalues of the perturbed matrix $\widetilde{\mathbf{X}}$, but there is no restriction on the size of the perturbation $\delta \mathbf{X}$. Some perturbation bounds of the singular values have been considered in* [23] *as follows.*

**Theorem 2.** *Perturbation bounds for the singular values of* $L \times K$ *matrix* $\widetilde{\mathbf{X}} = \mathbf{X} + \delta\mathbf{X}$ *are*

$$\begin{aligned} &|\widetilde{\lambda}_i - \lambda_i| \le \|\delta\mathbf{X}\|_2 \\ &\sum_{i=1}^{L} (\widetilde{\lambda}_i - \lambda_i) \le \|\delta\mathbf{X}\|_F^2. \end{aligned} \tag{4}$$

The above conditions indicate that the singular values of the matrix $\mathbf{X}$ are well-conditioned with respect to perturbations. That is, perturbations of $y_t$ produce similar or smaller perturbations in the singular values [23].

### 2.2. Application in the subspace method

Consider the following matrix

$$\widetilde{\mathbf{X}} = \mathbf{X} + \delta\mathbf{X} \tag{5}$$

where $\widetilde{\mathbf{X}}$ is a perturbed version of $\mathbf{X}$ with perturbation $\delta\mathbf{X}$. The SVD of the matrix $\mathbf{X}$ can be written as:

$$\mathbf{X} = [\mathbf{U}_s \quad \mathbf{U}_n] \begin{bmatrix} \boldsymbol{\Sigma}_s & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{V}_s^T \\ \mathbf{V}_n^T \end{bmatrix} = \mathbf{U}_s \boldsymbol{\Sigma}_s \mathbf{V}_s^T \tag{6}$$

where $\mathbf{U}_s \in \mathbb{R}^{L\times r}$, $\boldsymbol{\Sigma}_s \in \mathbb{R}^{r\times r}$ and $\mathbf{V}_s \in \mathbb{R}^{K\times r}$. The matrices $\mathbf{U}_s$ and $\mathbf{V}_s$ span the column spaces of $\mathbf{X}$ and $\mathbf{X}^T$, respectively, whereas $\mathbf{U}_n$ and $\mathbf{V}_n$ span their orthogonal spaces. Similarly, the SVD of the matrix $\widetilde{\mathbf{X}}$ can be written as:

$$\widetilde{\mathbf{X}} = [\widetilde{\mathbf{U}}_s \quad \widetilde{\mathbf{U}}_n] \begin{bmatrix} \widetilde{\boldsymbol{\Sigma}}_s & 0 \\ 0 & \widetilde{\boldsymbol{\Sigma}}_n \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{V}}_s^T \\ \widetilde{\mathbf{V}}_n^T \end{bmatrix} = \widetilde{\mathbf{U}}_s \widetilde{\boldsymbol{\Sigma}}_s \widetilde{\mathbf{V}}_s^T + \widetilde{\mathbf{U}}_n \widetilde{\boldsymbol{\Sigma}}_n \widetilde{\mathbf{V}}_n^T. \tag{7}$$

It is clear that the SVD of the matrix $\widetilde{\mathbf{X}}$ is completely different from the SVD of the matrix $\mathbf{X}$ due to the perturbation term $\delta\mathbf{X}$. Next the aim is to derive general expressions for approximations to the perturbed terms up to the second order of $\delta\mathbf{X}$. Assume the perturbed terms are as follows:

$$\begin{aligned} \widetilde{\mathbf{U}}_s &= \mathbf{U}_s + \delta\mathbf{U}_s = \mathbf{U}_s + \mathbf{U}_n\mathbf{P}_1 + \mathbf{U}_s\mathbf{P}_2 \\ \widetilde{\mathbf{V}}_s &= \mathbf{V}_s + \delta\mathbf{V}_s = \mathbf{V}_s + \mathbf{V}_n\mathbf{P}_3 + \mathbf{V}_s\mathbf{P}_4 \end{aligned} \tag{8}$$

$$\begin{aligned} \widetilde{\mathbf{U}}_n &= \mathbf{U}_n + \delta\mathbf{U}_n = \mathbf{U}_n + \mathbf{U}_s\mathbf{Q}_1 + \mathbf{U}_n\mathbf{Q}_2 \\ \widetilde{\mathbf{V}}_n &= \mathbf{V}_n + \delta\mathbf{V}_n = \mathbf{V}_n + \mathbf{V}_s\mathbf{Q}_3 + \mathbf{V}_n\mathbf{Q}_4 \end{aligned} \tag{9}$$

$$\widetilde{\boldsymbol{\Sigma}}_s = \boldsymbol{\Sigma}_s + \delta\boldsymbol{\Sigma}_s. \tag{10}$$

Note that each of the perturbed terms in (8) and (9) consists of two parts: the first part captures the perturbation in its orthogonal space and the second part considers perturbation in each subspace. Let us, for example, consider the perturbation term $\mathbf{U}_s$. The perturbation term $\delta\mathbf{U}_s$ consists of two parts: $\mathbf{U}_n\mathbf{P}_1$ captures the perturbation in its orthogonal space $\mathbf{U}_n$ and $\mathbf{U}_s\mathbf{P}_2$ considers perturbation in subspace $\mathbf{U}_s$. Now one needs to determine a set of unknowns; $\{\mathbf{P}_i\}_{i=1}^4$, $\{\mathbf{Q}_i\}_{i=1}^4$ and $\delta\boldsymbol{\Sigma}_s$ in order to remove all perturbations or refine the series. Let the following assumptions hold according to the SVD of the matrices $\mathbf{X}$ and $\widetilde{\mathbf{X}}$:

$$\mathbf{U}_s^T\mathbf{U}_s = \mathbf{I}, \quad \mathbf{V}_s^T\mathbf{V}_s = \mathbf{I}, \quad \mathbf{U}_n^T\mathbf{U}_n = \mathbf{I}, \quad \mathbf{V}_n^T\mathbf{V}_n = \mathbf{I}, \quad \mathbf{U}_s^T\mathbf{U}_n = \mathbf{0}, \quad \mathbf{V}_s^T\mathbf{V}_n = \mathbf{0} \tag{11}$$

$$\widetilde{\mathbf{U}}_s^T\widetilde{\mathbf{U}}_s = \mathbf{I}, \quad \widetilde{\mathbf{V}}_s^T\widetilde{\mathbf{V}}_s = \mathbf{I}, \quad \widetilde{\mathbf{U}}_n^T\widetilde{\mathbf{U}}_n = \mathbf{I}, \quad \widetilde{\mathbf{V}}_n^T\widetilde{\mathbf{V}}_n = \mathbf{I}, \quad \widetilde{\mathbf{U}}_s^T\widetilde{\mathbf{U}}_n = \mathbf{0}, \quad \widetilde{\mathbf{V}}_s^T\widetilde{\mathbf{V}}_n = \mathbf{0}. \tag{12}$$

Let $\boldsymbol{\Delta}_s = (\boldsymbol{\Sigma}_s\boldsymbol{\Sigma}_s^T)^{-1}$ and consider different projections of $\delta\mathbf{X}$ as:

$$\mathbf{E}_{ss} = \mathbf{U}_s^T\delta\mathbf{X}\mathbf{V}_s, \qquad \mathbf{E}_{sn} = \mathbf{U}_s^T\delta\mathbf{X}\mathbf{V}_n, \qquad \mathbf{E}_{ns} = \mathbf{U}_n^T\delta\mathbf{X}\mathbf{V}_s, \qquad \mathbf{E}_{nn} = \mathbf{U}_n^T\delta\mathbf{X}\mathbf{V}_n. \tag{13}$$

Using the assumptions considered above, the unknowns are, up to second order of $\delta\mathbf{X}$, as follows [24]:

$$\mathbf{Q}_1 = \boldsymbol{\Delta}_s(\boldsymbol{\Sigma}_s\mathbf{E}_{ss}^T\boldsymbol{\Delta}_s\boldsymbol{\Sigma}_s\mathbf{E}_{ns}^T - \mathbf{E}_{sn}\mathbf{E}_{nn}^T) + \mathbf{F}_1 \tag{14}$$

$$\mathbf{Q}_3 = -\boldsymbol{\Delta}_s\mathbf{E}_{ns}^T\mathbf{E}_{nn} + \boldsymbol{\Sigma}_s^{-1}\mathbf{E}_{ss}\boldsymbol{\Sigma}_s^{-1}\mathbf{E}_{sn} + \mathbf{F}_2 \tag{15}$$

where

$$\mathbf{F}_1 = -\boldsymbol{\Delta}_s\boldsymbol{\Sigma}_s\mathbf{E}_{ns}^T, \qquad \mathbf{F}_2 = -\boldsymbol{\Sigma}_s^{-1}\mathbf{E}_{sn}. \tag{16}$$

The other unknowns can be found based on the $\mathbf{Q}_1$, $\mathbf{Q}_3$, $\mathbf{F}_1$ and $\mathbf{F}_2$ as follows:

$$\begin{aligned} &\mathbf{P}_1 = -\mathbf{Q}_1^T, \qquad \mathbf{P}_2 = -\frac{1}{2}\mathbf{F}_1^T\mathbf{F}_1, \qquad \mathbf{P}_3 = -\mathbf{Q}_3^T, \qquad \mathbf{P}_4 = -\frac{1}{2}\mathbf{F}_2^T\mathbf{F}_2, \\ &\mathbf{Q}_2 = -\frac{1}{2}\mathbf{F}_1^T\mathbf{F}_1, \qquad \mathbf{Q}_4 = -\frac{1}{2}\mathbf{F}_2^T\mathbf{F}_2, \\ &\delta\mathbf{\Sigma}_s = \mathbf{E}_{ss} - \mathbf{E}_{sn}\mathbf{F}_2^T - \frac{1}{2}\mathbf{\Sigma}_s\mathbf{F}_2\mathbf{F}_2^T + \frac{1}{2}\mathbf{F}_1\mathbf{F}_1^T\mathbf{\Sigma}_s. \end{aligned} \tag{17}$$

The proof of this, can be found in [24] with some substitution. To do this, one needs to substitute $\mathbf{\Lambda}_n$ with zero (thus $\alpha$ therein) in the theorem represented in [24]. For more clarification we have presented the proof using the notations used here and the case considered above in Appendix.

## 3. SSA based on the perturbation theory

### 3.1. SSA: A brief description

The main purpose of SSA is to decompose the original series into a sum of series, so that each component in this sum can be identified as either a trend, periodic or quasi-periodic (perhaps, amplitude-modulated), or noise. This is followed by a reconstruction of the original series. The main idea of the Basic SSA is as follows.

Consider the real-valued nonzero time series $Y_N = (y_1, \dots, y_N)$ of sufficient length $N$. Let $K = N - L + 1$, where $L$ ($L \le N/2$) is some integer called the window length. Define the matrix

$$\mathbf{X} = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} y_1 & y_2 & y_3 & \cdots & y_K \\ y_2 & y_3 & y_4 & \cdots & y_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \cdots & y_N \end{pmatrix} \tag{18}$$

and call it the trajectory matrix. Obviously $x_{ij} = y_{i+j-1}$ so that the matrix $\mathbf{X}$ has equal elements on the diagonals $i+j = \text{const}$.

Define the matrix $\mathbf{X}\mathbf{X}^T$. Singular value decomposition (SVD) of $\mathbf{X}\mathbf{X}^T$ provides us with the collections of $L$ eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_L \ge 0$ and the corresponding eigenvectors $P_1, P_2, \dots, P_L$, where $P_i$ is the normalized eigenvector corresponding to the eigenvalue $\lambda_i$ ($i = 1, \dots, L$). Note that one can apply SVD to a variety of matrices; for example, in addition to $\mathbf{X}\mathbf{X}^T$, it is customary to use either the covariance or correlation matrix computed from $\mathbf{X}$, treated as a multivariate data matrix [3].

A group of $r$ (with $1 \le r < L$) eigenvectors determine an $r$-dimensional hyperplane in the $L$-dimensional space $\mathbb{R}^L$ of vectors $X_j$. The distance between vectors $X_j$ ($j = 1, \dots, K$) and this $r$-dimensional hyperplane can be rather small (it is controlled by the choice of the eigenvalues) meaning that the projection of $\mathbf{X}$ onto this hyperplane approximates the original matrix $\mathbf{X}$ well. If we choose the first $r$ eigenvectors $P_1, \dots, P_r$, then the squared $L_2$-distance between this projection and $\mathbf{X}$ is equal to $\sum_{j=r+1}^{L} \lambda_j$. According to the Basic SSA algorithm, the $L$-dimensional data is projected onto this $r$-dimensional subspace and the subsequent averaging over the diagonals allows us to obtain an approximation to the original series.

### 3.2. Proposed algorithm

In order to apply the perturbation theory in the SSA technique we need to have a priori information about the noise component $\delta Y_N$ or $\delta\mathbf{X}$. However, the noise series $\delta Y_N$ is unknown in practice and usually there is no a priori information. One way to overcome this problem is to have an estimate of $\delta\mathbf{X}$. Here we use $\widetilde{\mathbf{X}} - \hat{\mathbf{X}}$ as an estimate of $\delta\mathbf{X}$, where $\hat{\mathbf{X}}$ is obtained using basic SSA. That is, we first apply the basic SSA technique to the noisy time series to find an initial estimate of $\delta\mathbf{X}$. We then reconstruct $\mathbf{X}$ using its estimate based on the perturbation theory. Therefore, the proposed algorithm consists of two complementary stages: an initial estimate of $\delta\mathbf{X}$ and reconstruction $\mathbf{X}$ both of which include several separate steps. At the first stage we decompose the series using a basic SSA technique and at the second stage we reconstruct the original series using SSA based on the perturbation theory and use the reconstructed series (which is without noise) for forecasting new data points. Let us now formally describe this algorithm next.

*Formal description of the proposed algorithm.*

Let us have a noisy time series $\tilde{Y}_N = (\tilde{y}_1, \dots, \tilde{y}_N)$. Fix $L$ ($L \le N/2$), the window length, and let $K = N - L + 1$.

1. (*Computing the trajectory matrix*): transfers a one-dimensional time series $\tilde{Y}_N = (\tilde{y}_1, \dots, \tilde{y}_N)$ into the multi-dimensional series $\tilde{X}_1, \dots, \tilde{X}_K$ with vectors $\tilde{X}_i = (\tilde{y}_i, \dots, \tilde{y}_{i+L-1})^T$, where $K = N - L + 1$. The result of this step is the trajectory matrix $\tilde{\mathbf{X}} = \left[\tilde{X}_1, \dots, \tilde{X}_K\right]$.
2. (*Constructing a matrix for applying SVD*): compute the matrix $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$.

3. (*SVD of the matrix* $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$): compute the eigenvalues and eigen-vectors of the matrix $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$ and represent it in the form $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T = \tilde{P}\tilde{\Lambda}\tilde{P}^T$. Here $\tilde{\Lambda} = \mathrm{diag}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_L)$ is the diagonal matrix of eigenvalues of $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$ ordered so that $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \cdots \geq \tilde{\lambda}_L \geq 0$ and $\tilde{P} = (\tilde{P}_1, \tilde{P}_2, \ldots, \tilde{P}_L)$ is the corresponding orthogonal matrix of eigenvectors of $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$.
4. (*Selection of eigenvectors*): select a group of $r$ $(1 \leq r \leq L)$ eigenvectors $\tilde{P}_{i_1}, \tilde{P}_{i_2}, \ldots, \tilde{P}_{i_r}$.
   The grouping step corresponds to splitting the elementary matrices $\tilde{\mathbf{X}}_i$ into several groups and summing the matrices within each group.
5. Compute the matrix $\hat{\mathbf{X}} = \|\hat{x}_{i,j}\| = \sum_{k=1}^{r} \tilde{P}_{i_k}\tilde{P}_{i_k}^T\tilde{\mathbf{X}}$.
6. Estimating the noise matrix $\delta\mathbf{X}$: to estimate $\delta\mathbf{X}$, we use the difference between the initial estimate of signal matrix $\mathbf{X}$, $\hat{\mathbf{X}}$, and the noisy matrix $\tilde{\mathbf{X}}$; $\delta\mathbf{X} \approx \tilde{\mathbf{X}} - \hat{\mathbf{X}}$.
7. Estimating signal matrix $\mathbf{X}$ using the perturbation theory: set $\mathbf{X} = \mathbf{U}_s\mathbf{\Sigma}_s\mathbf{V}_s^T$, where $\mathbf{U}_s$, $\mathbf{\Sigma}_s$ and $\mathbf{V}_s^T$ are the refined version of the noisy matrices $\tilde{\mathbf{U}}_s$, $\tilde{\mathbf{\Sigma}}_s$ and $\tilde{\mathbf{V}}_s^T$, respectively, and can be obtained using (8) and (10). Note that performing a SVD of the estimated noise matrix $\delta\mathbf{X}$ in step 6, enables us to estimate $\mathbf{U}_n$ and $\mathbf{V}_n$.
8. Transition to the one-dimensional series can now be achieved by averaging over the diagonals of the matrix $\mathbf{X}$. Thus, the results of this step is an approximation of $Y_N$.
9. The refined series $Y_N$ can now be used for forecasting.

*Forecasting procedure.*

The SSA technique can be applied to the time series that approximately satisfies linear recurrent formulae.[1] The class of time series governed by linear recurrent formulae is rather wide; it includes harmonics, polynomial, and exponential time series. Let us briefly describe the so-called SSA recurrent forecasting algorithm (for more information see [3]).

Define the original noisy series $\tilde{Y}_N = (\tilde{y}_1, \ldots, \tilde{y}_N)$ and the reconstructed series $Y_N = (y_1, \ldots, y_N)$. For an eigenvector $U \in \mathbb{R}^L$ we denote the vector of the first $L-1$ components of the vector $U$ as $U^\triangledown \in \mathbb{R}^{L-1}$. Set $\upsilon^2 = \pi_1^2 + \cdots + \pi_r^2 < 1$, where $\pi_i$ is the last component of the eigenvector $U_i$ $(i = 1, \ldots, r)$. It can be proved that the last component $y_L$ of any vector $Y = (y_1, \ldots, y_L)^T$ is a linear combination of the earlier components $(y_1, \ldots, y_{L-1})$; that is, $y_L = a_1y_{L-1} + \cdots + a_{L-1}y_1$ where the vector of coefficients $A = (a_1, \ldots, a_{L-1})$ can be expressed as $A = \sum_{i=1}^{r} \pi_i U_i^\triangledown/(1-\upsilon^2)$. The forecasts $\hat{y}_{T+1}, \ldots, \hat{y}_{N+h}$ are then obtained as

$$\hat{y}_i = \begin{cases} y_i & \text{for } i = 1, \ldots, N \\ \sum_{j=1}^{L-1} a_j\hat{y}_{i-j} & \text{for } i = N+1, \ldots, N+h. \end{cases}$$

## 4. Empirical results

### 4.1. Simulated data

We shall consider two types of time series: real and simulated time series. The capability of the SSA technique based on the perturbation theory ($\text{SSA}_{\text{PT}}$), in reconstructing and forecasting, is initially assessed by applying it to simple simulated sinusoidal series:

$$\begin{aligned} S012 &= \beta_0 + \beta_1\sin(2t\pi/12) + \beta_2\sin(2t\pi/7) + \beta_3\sin(2t\pi/5) + \epsilon_t \\ S01 &= \beta_0 + \beta_1\sin(2t\pi/12) + \beta_2\sin(2t\pi/7) + \epsilon_t \\ S1 &= \beta_1\sin(2t\pi/12) + \beta_2\sin(2t\pi/7) + \epsilon_t \end{aligned} \tag{19}$$

where $\epsilon_t$ is a white noise series. In total 300 data were generated and we added different normally distributed noise to each point of the original series. The simulation was repeated 1000 times. The first 200 observations were considered as in-sample (reconstruction) and the rest as out-of-sample (forecasting). Note that usually every harmonic component with a different frequency produces two eigentriples with close singular values (except for frequency 0.5 which provides one eigentriple with a saw-tooth singular vector). For example, one needs to select the first five eigenvalues for reconstruction of the series *S012*, and the first three for the series *S01*. It should be noted that we need to consider one eigentriple for the intercept, which is the first one in this particular example. To calculate the precision we use the ratio of Root Mean Square Error (RMSE):

---

[1] We say that the time series $Y_N$ satisfies an linear recurrent formulae of order $L-1$ if there are numbers $a_1, \ldots, a_{L-1}$ such that

$$y_{N-i} = \sum_{k=1}^{L-1} a_k y_{N-i-k}, \quad 0 \leq i \leq N-L.$$

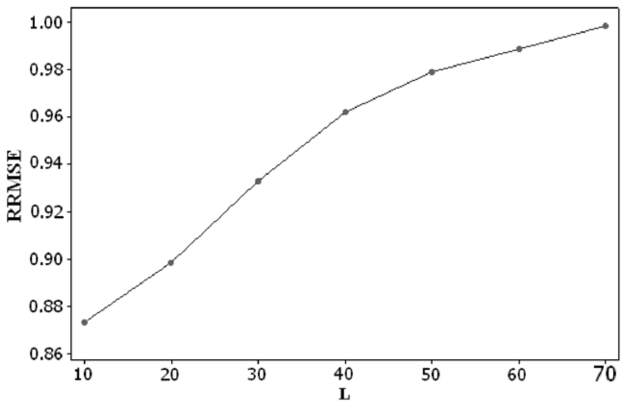


**Fig. 1.** The value of RRMSE in reconstructing noisy series *S012* for different window lengths.

$$\mathrm{RRMSE} = \left( \frac{\sum_{i=1}^{n} (y_{N+i} - \widetilde{y}_{N+i})^2}{\sum_{i=1}^{n} (y_{N+i} - \widehat{y}_{N+i})^2} \right)^{1/2} \tag{20}$$

where $n$ represents the number of forecasted points, $\widetilde{y}_{N+i}$ are the forecasted values of $y_{N+i}$ obtained by the SSA technique based on the perturbation theory ($\mathrm{SSA_{PT}}$) and $\widehat{y}_{N+i}$ are the forecasted values of $y_{N+i}$ obtained by other techniques such as SSA based on the least squares estimate ($\mathrm{SSA_{LS}}$). If RRMSE $< 1$, then the $\mathrm{SSA_{PT}}$ procedure outperforms the alternative prediction method. Alternatively, RRMSE $> 1$ would indicate that the performance of the corresponding $\mathrm{SSA_{PT}}$ procedure is worse than the predictions of the competing method. We also use RRMSE in considering reconstruction accuracy. The only difference is we use reconstructed values, obtained by each technique, rather than predicted values in (20).

*The effect of window length.*

Let us first consider the effect of noise reduction with respect to different window lengths $L$ which is the single parameter in the decomposition stage. Certainly, the choice of parameter $L$ depends on the data we have and the analysis we aim to perform. An improper choice of $L$ would imply an inferior decomposition. It should be noted that variations in $L$ may influence separability feature of the SSA technique; the orthogonality and closeness of the singular values. Here we consider $L$ between 10 and 70 which is approximately $N/3$.

Figs. 1–3 show the RRMSE of reconstructed series for different simulated series. As it appears from these figures, $\mathrm{SSA_{PT}}$ has a better performance in reconstruction of noisy series, particulary for small window length. The performances of both methods are similar for a large window length.

As the figures show, RRMSE approaches 1 as the window length increases confirming that both methods have similar performance for a large window length. The graphs also show that there is a gradual increase in RRMSE with window length. For example for window length 10, the performance of $\mathrm{SSA_{PT}}$ is up to 15% better than $\mathrm{SSA_{LS}}$ in reconstruction noisy series *S01*. However, there is not a significant discrepancy between the performance of $\mathrm{SSA_{PT}}$ and $\mathrm{SSA_{LS}}$ for window length greater than 50.

Note that the minimum value of RMSE for both $\mathrm{SSA_{PT}}$ and $\mathrm{SSA_{LS}}$ occurs for a large window length. Let us, for example, consider the RMSE of $\mathrm{SSA_{PT}}$ and $\mathrm{SSA_{LS}}$ in reconstructing *S012* in more detail. Fig. 4 shows the RMSE of $\mathrm{SSA_{PT}}$ and $\mathrm{SSA_{LS}}$. As can be seen from the figure, there is a gradual decrease in RMSE with window length. In fact, the maximum accuracy in reconstruction, using both methods, occurs for a large window length. The figure also shows that the RMSE of $\mathrm{SSA_{PT}}$ is smaller than those obtained using $\mathrm{SSA_{LS}}$. Moreover, the figure indicates that the discrepancy between $\mathrm{SSA_{PT}}$ and $\mathrm{SSA_{LS}}$ reduces as the window length increases. In the rest of this work, we only consider the RRMSE as considering two RMSEs and the RRMSE gives equal information, but the RRMSE is more informative.

*The effect of noise level.*

To gain a better understanding of the effect of noise reduction with respect to different window lengths $L$, we also consider different signal to noise ratios (SNR). Here the SNR is the ratio of standard deviation of the noise free series (signal) to standard deviation of noise. Figs. 5–7 show RRMSE for different values of SNR. For example, Fig. 5 shows RRMSE for the series *S012* where we have an intercept and two different harmonic components. As it appears from the figure, there is a gradual increase in RRMSE with SNR. In fact, the minimum RRMSE occurs for a high noise level or lowest SNR. This result confirms that the new SSA algorithm works better for a situation where the series is a mixture of low signal level and high noise level. For example, for $L = 10$ and SNR $= 0.3$, the results indicate that the performance of the $\mathrm{SSA_{PT}}$ is up to 15% better than the basic $\mathrm{SSA_{LS}}$, while this is approximately 4% for SNR $= 15$. However, there is no significant discrepancy between two methods for a series with a high SNR. A similar results can be seen for $L = 40$ and $L = 70$, but the RRMSE tends to 1 faster

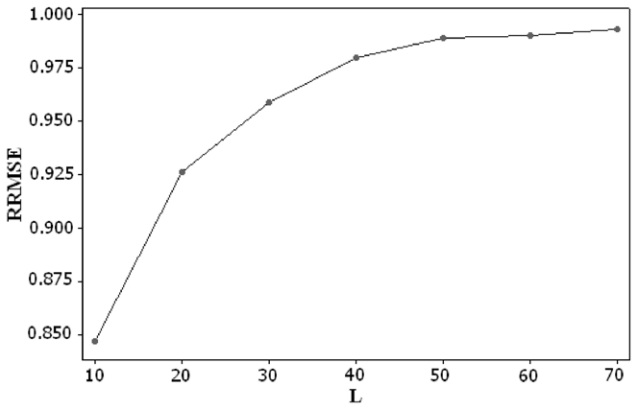


**Fig. 2.** The value of RRMSE in reconstructing noisy series *S01* for different window lengths.

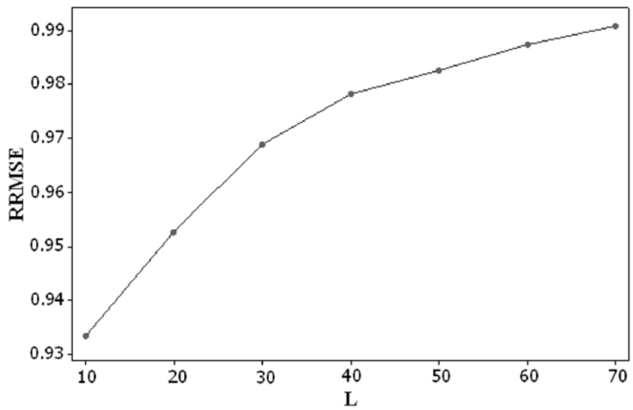


**Fig. 3.** The value of RRMSE in reconstructing noisy series *S1* for different window lengths.

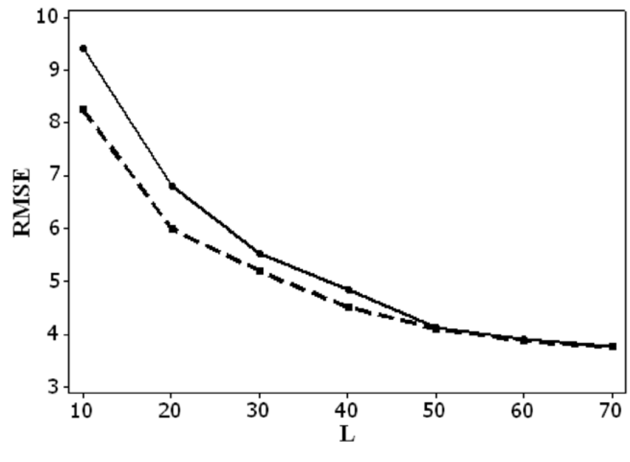


**Fig. 4.** The value of RMSE in reconstructing noisy sine series for different window lengths using $SSA_{PT}$ (dashed line) and $SSA_{LS}$ (thick line).

than for $L = 10$. These results confirm our previous discussion about separability and window length; a larger window length provides better separability.

Let us now consider the problem of separability briefly. For a fixed length $L$, consider a certain SVD of the noisy series $\widetilde{Y}_N$ of length $N$, and assume that the series $\widetilde{Y}_N$ is a sum of two series $Y_N$ and $\delta Y_N$; $\widetilde{Y}_N = Y_N + \delta Y_N$. In this case, separability of the series $Y_T$ and $\delta Y_N$ means that we can split the matrix terms of the SVD of the trajectory matrix $\delta\mathbf{X}$ into two different groups, so that the sums of terms within the groups give the trajectory matrices $\mathbf{X}$ and $\delta\mathbf{X}$ of the series $Y_N$ and $\delta Y_N$, respectively (for more information see [3]).

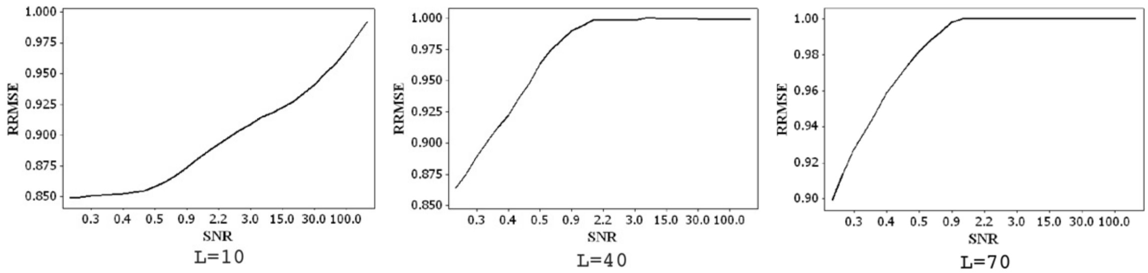


**Fig. 5.** The values of RRMES for different noise levels for the series *S012*.

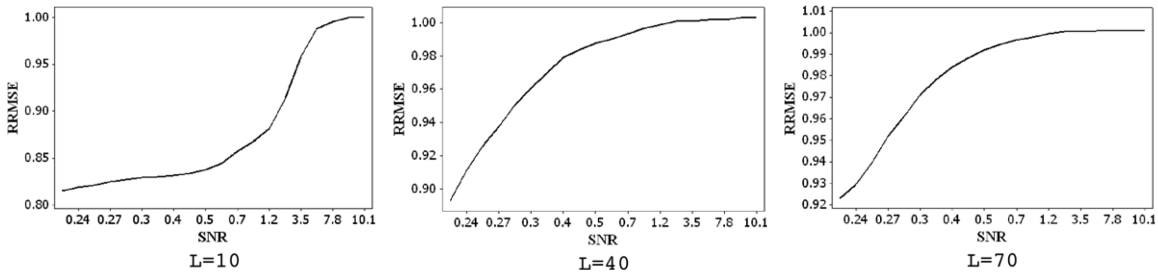


**Fig. 6.** The values of RRMES for different noise levels for the series *S01*.

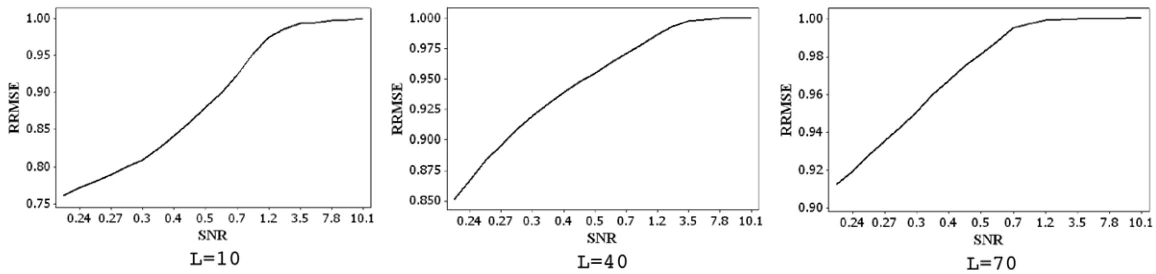


**Fig. 7.** The values of RRMES for different noise levels for the series *S1*.

Figs. 6 and 7 show the results for series *S01* and *S1*. As the figures show, a similar interpretation, such as those concluded for series *S012*, can be stated for these series. It should be, however, noted that the RRMSE for more complex series is greater than for a simple series. For example for $L = 10$, the RRMSE is approximately 85% for series *S012* while this is about 80% and 75% for series *S01* and *S1*, respectively.

*The effect of time-series length*.

Let us now consider the influence of the time-series length in decomposition and reconstruction of noisy series. In order to examine this we used series *S012*, *S01* and *S1* with different length $N$ (varies between 100 and 1000). Fig. 8 shows the value of RRMSE in reconstructing the series *S012* (thick line), *S01* (dashed line) and *S1* (thin line) for different values of $N$. As the results show there is no change in RRMSE as $N$ increases. This is because the series considered here have a structure which can be described via a deterministic component. This means the series has a clear structure and this structure is captured very well by the SSA. In this context, Hassani et al. [12] showed that in the ideal situation, when we have a series which is a sum of a deterministic component (fully recovered by SSA) and a random noise, the error of the SSA forecast will be exactly the same at all horizons. Here the same results are obtained for reconstruction of a series with deterministic components. Therefore, we can conclude that for a series which is a sum of a deterministic component and a random noise, the error of the SSA forecast (for $h$ step ahead) and reconstruction (for different series length $N$) remains stable.

*The effect of non-stationary noise*.

So far, we considered the situation where the noise component $\epsilon_t$ is stationary. Let us now consider the situation where $\epsilon_t$ is not stationary. One of the most common instances of non-stationary behaviour is heteroscedasticity, i.e., the variance of noise is proportional to the amplitude of the underlying signal. In the following we examine the capability of $SSA_{PT}$ to detect heteroscedastic noise and reconstruct noise free series. Fig. 9 (left) shows a realization of the series *S012* corrupted with a heteroscedasticity noise. Fig. 9 (right) shows the values of RRMES for different heteroscedasticity noise levels. Here we only represent the results for $L = 10$, but the results are similar for $L = 40$ and $L = 70$. Again, similar to the results obtained for

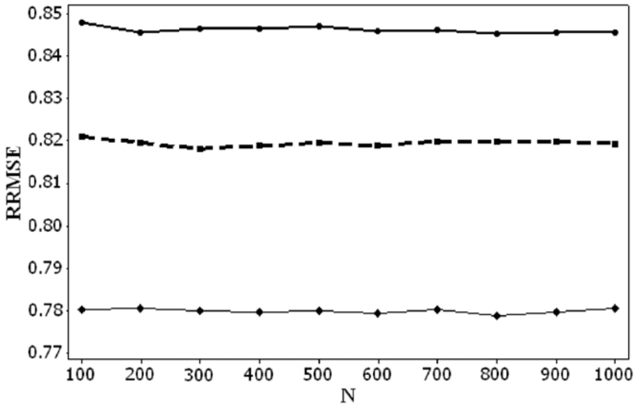


**Fig. 8.** The value of RRMSE in reconstructing of noisy series for different $N$; *S012* (thick line), *S01* (dashed line) and *S1* (thin line).

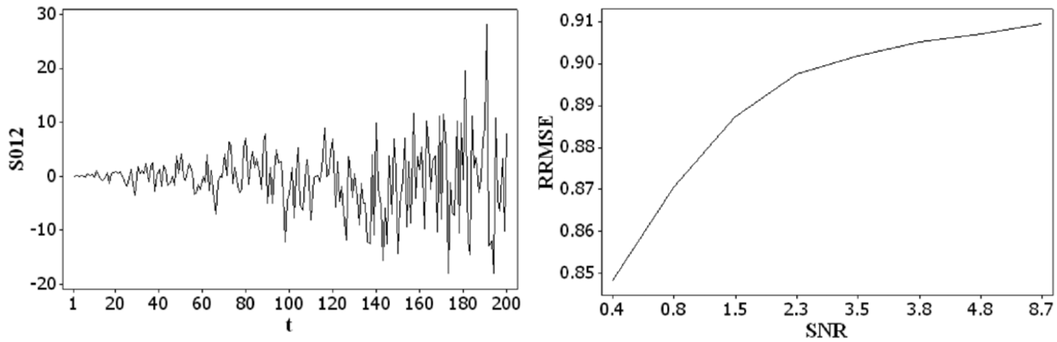


**Fig. 9.** Left: a realization of the series *S012* corrupted with a heteroscedasticity noise. Right: the values of RRMES for different heteroscedasticity noise levels.

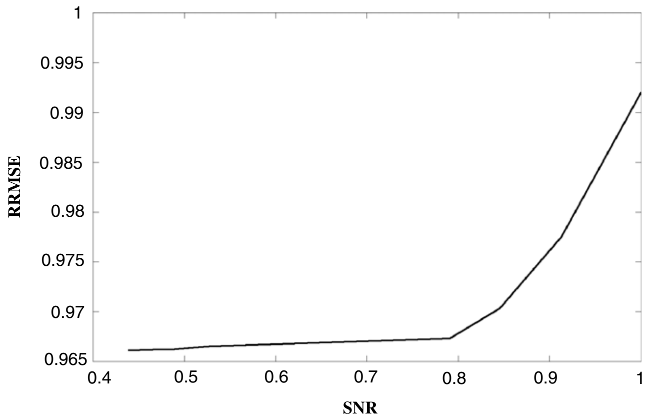


**Fig. 10.** The values of RRMES in reconstructing a Hénon map.

stationary noise, the results indicate that the performance of $SSA_{PT}$ is much better than those obtained by $SSA_{LS}$. Therefore, we can conclude that $SSA_{PT}$ works well for detection of a series corrupted with either stationary or non-stationary noise.

### *4.2. Chaotic time series*

In this section, we show that the proposed technique can be used to obtain the embedding dimension of the chaotic time series. The capability of the SSA technique as a noise reduction method for chaotic time series was initially tested by applying the technique to the Hénon map with usual parameter values: $A = 1.4$ and $B = 0.3$. In total 1895 data are generated and we add different normally distributed noise to each point of the original series.

**Table 1**
Descriptive statistics of several stock indices returns series before and after filtering.

| Statistics | Method | DAX 30 | CAC 40 | FTSE 100 | IBEX 35 | S&P 500 | PSI 20 | ASE |
|---|---|---|---|---|---|---|---|---|
| Mean $\times 10^{-3}$ | Original | 0.24 | 0.28 | 0.24 | 0.36 | 0.35 | 0.21 | 0.58 |
| | GARCH | −0.24 | −0.21 | −0.18 | −0.30 | −0.17 | −0.21 | 0.84 |
| | $SSA_{LS}$ | 0.23 | 0.28 | 0.24 | 0.36 | 0.35 | 0.21 | 0.57 |
| | $SSA_{PT}$ | 0.24 | 0.27 | 0.25 | 0.36 | 0.35 | 0.21 | 0.58 |
| S.D. $\times 10^{-1}$ | Original | 0.11 | 0.11 | 0.09 | 0.11 | 0.09 | 0.08 | 0.16 |
| | GARCH | 0.11 | 0.11 | 0.09 | 0.11 | 0.09 | 0.08 | 0.16 |
| | $SSA_{LS}$ | 0.09 | 0.09 | 0.07 | 0.09 | 0.09 | 0.07 | 0.13 |
| | $SSA_{PT}$ | 0.08 | 0.07 | 0.06 | 0.07 | 0.08 | 0.07 | 0.11 |
| Kurtosis | Original | 4.57 | 3.39 | 3.44 | 3.83 | 4.33 | 8.58 | 6.94 |
| | GARCH | 4.46 | 3.38 | 3.44 | 3.76 | 4.33 | 8.44 | 6.91 |
| | $SSA_{LS}$ | 3.81 | 3.44 | 3.64 | 3.77 | 4.30 | 6.93 | 6.10 |
| | $SSA_{PT}$ | 3.57 | 3.31 | 3.48 | 3.51 | 4.12 | 6.13 | 5.90 |

Fig. 10 shows the values of RRMES in reconstructing a Hénon map for different noise levels. The first two eigenvalues were selected in reconstructing a noisy Hénon map. The results indicate that the performance of $SSA_{PT}$ is slightly better than those obtained by $SSA_{LS}$. The results indicate that the discrepancy between $SSA_{PT}$ and $SSA_{LS}$ in reconstructing Hénon map is smaller than those obtained for a sine series. The performance of $SSA_{LS}$ for filtering of a noisy Hénon map was studied by Hassani et al. [1]. They showed that the $SSA_{LS}$ technique can be used as a powerful noise reduction method for filtering either chaotic series or financial time series. They also showed that the $SSA_{LS}$ performance is much better than the considered linear and non-linear models for a noisy Hénon map. The new SSA based method presented here can be therefore used as a noise reduction technique for financial time series.

### *4.3. Real data*

#### *4.3.1. Financial time series*

Hassani et al. [12] considered the daily closing prices of several stock market indices to examine whether noise reduction matters in measuring dependencies of the financial series. Here we also use the same series. Next we examine the effect of noise reduction on the daily closing prices of several stock market indices: ASE (Greece), CAC 40 (France), DAX 30 (Germany), FTSE 100 (UK), PSI 20 (Portugal), IBEX 35 (Spain) and S&P 500 (USA). These data sets have been used by many authors. For example, Szpiro [25] in studying the S&P 500 Index, found an increasing presence of noise. Davis and Mikosch [26] consider plots of the sample autocorrelation function (ACF) of the squares of the S&P index for different periods and found that either the process is non-stationary or that the process exhibits heavy tails.

The BDS test [27] for nonlinearity was used to test whether the series are Independent and Identically Distributed (IID). The BDS test can detect many types of departures from being IID, and can serve as a general model specification test, especially in the presence of nonlinear dynamics. The results of the BDS test indicate a significant dependence in all series confirming the existing results of dependencies in the stock market literature [28].

Table 1 represents a summary of descriptive statistics for the series before and after filtering. The rows related to Kurtosis shows the value of Kurtosis of the series. A positive value typically indicates that the distribution has a sharper peak, thinner shoulders, and fatter tails than the normal distribution. As it appears from Table 1, all series have fatter tails than the normal distribution. Thus, the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model was considered as a noise reduction method for filtering the series. Note that GARCH models are widely used in various financial applications such as risk management, option pricing, foreign exchange, and the term structure of interest rates [29].

The results in Table 1 indicate that the filtered series based on the SSA, for all cases have a smaller standard deviation, S.D., than those values obtained by the GARCH model. Again, as the results show, the performance of $SSA_{PT}$ is slightly better than $SSA_{LS}$. The same results can also be seen for the values of the maximum and minimum of the series, but we omit them here.

Table 2 shows the values of the ACF at lag-1 and $\lambda = \left(1 - \exp[-2I(X, Y)]\right)^{\frac{1}{2}}$ of several stock indices returns series before and after filtering, where $I(X, Y)$ is the mutual information between two series $X$ and $Y$. As it appears from Table 2, the values of the ACF are changed after filtering. In fact, the values were immediately affected by filtering. It should also be noted that the sign of the ACF of the series IBEX 35, PSI 20 and ASE was changed from positive to negative after filtering by the GARCH model indicating that the performance of the GARCH model is not very good for filtering the series. It seems that the results obtained for $\lambda$ after filtering, are more robust than those for the ACF.

We also used Ljung–Box $Q$-statistics to test whether the values obtained for the ACF, before and after filtering, are significantly different from zero; $*$ indicates significant results at the 1% level of significance. The results indicate that the values of the ACF of the original series and those obtained after filtering by the SSA (except for S&P) are statistically significant.

We also considered the significance test for $\lambda$. In order to perform the test we followed the method which has been introduced in [30]. The critical values have been simulated for the null distribution and found through simulation of critical

**Table 2**
The values of the ACF at lag-1 and λ of several stock indices returns series before and after filtering.

| | DAX 30 | CAC 40 | FTSE 100 | IBEX 35 | S&P 500 | PSI 20 | ASE |
|---|---|---|---|---|---|---|---|
| ACF | | | | | | | |
| Original | 0.0519* | 0.0344* | 0.0234* | 0.0524* | 0.0147 | 0.137* | 0.147* |
| GARCH | 0.0001 | 0.0000 | 0.0235 | −0.0007 | 0.0145 | −0.0001 | −0.0001 |
| $SSA_{LS}$ | 0.1790* | 0.1680* | 0.1516* | 0.2383* | 0.0147 | 0.4406* | 0.4505* |
| $SSA_{PT}$ | 0.1921* | 0.1834* | 0.1857* | 0.2511* | 0.0581 | 0.5437* | 0.4728* |
| λ | | | | | | | |
| Original | 0.3079* | 0.2358* | 0.1508* | 0.2564* | 0.1540* | 0.3502* | 0.3157* |
| GARCH | 0.2799* | 0.1171* | 0.1508* | 0.5382* | 0.1540* | 0.7951* | 0.2909* |
| $SSA_{LS}$ | 0.2921* | 0.2425* | 0.2326* | 0.2855* | 0.1475* | 0.4977* | 0.5263* |
| $SSA_{PT}$ | 0.3142* | 0.2713* | 0.2678* | 0.2911* | 0.1876* | 0.5216* | 0.5419* |

values based on a white noise. Dionisio et al. [31] have presented the critical values for a number of sample sizes and different significant levels. Again, $*$ indicates the results at the 1% level of significance; the values of λ, before and after filtering, are statistically significant.

The results indicate that the dynamics of the system are approximated better using SSA. In fact, using the SSA technique enables us to adjust the series to better satisfy the approximations to the dynamics. It leads to noise reduction more effective than traditional filtering such as GARCH, specifically for a complex time series.

*4.3.2. Monthly accidental deaths in the USA*

Let us now consider the performance of the SSA technique based on the perturbation theory by applying it to a well-known time series data set, namely, monthly accidental deaths in the USA. The Death series shows the monthly accidental deaths in the USA between 1973 and 1978. This data have been used by many authors (see, for example, [32]) and can be found in many time series data libraries.

In this section we compare the SSA technique with several well-known methods namely, the traditional Box–Jenkins SARIMA models, the ARAR Algorithm and the Seasonal Holt–Winters Algorithm. Brockwell and Davis [32] applied these methods on the Death series to forecast the six future data points. Below, these methods are described shortly and the results of their forecasting are compared with the SSA technique.

*SARIMA model.*

Box and Jenkins [2] provide a methodology for fitting a model to an empirical series. This systematic approach identifies a class of models appropriate for the empirical data sequence at hand and estimates its parameters. A general class of Box and Jenkins models includes ARIMA and SARIMA models that can model a large class of autocorrelation functions. We use the models below for forecasting the six future data as are described in [32]: Model I:

$$\nabla\nabla_{12}y_t = 28.831 + (1 - 0.478B)(1 - 0.588B^{12})Z_t, \quad Z_t \sim WN(0, 94390). \tag{21}$$

Model II:

$$\nabla\nabla_{12}y_t = 28.831 + Z_t - 0.596Z_{t-1} - 0.407Z_{t-6} - 0.685Z_{t-12} + 0.460Z_{t-13} \quad Z_t \sim WN(0, 94390) \tag{22}$$

where the backward shift operator $B$ is: $B^jZ_t = Z_{t-j}$ and $\nabla_j = 1 - B^j$. Note that the seasonal difference of a time series is the series of changes from one season to the next. For monthly accidental deaths in the USA, in which there are 12 periods in a season, the seasonal difference of the series at period $t$ is $\nabla_{12}y_t = y_t - y_{t-12}$. In the forecasting the series, we see that the first difference of $y_t$ is far from random (it is still strongly seasonal), and the seasonal difference is far from stationary (it resembles a random walk). Therefore, both kinds of differencing are needed to render the series stationary and to account for the gross pattern of seasonality. It should be noted that the first difference of the seasonal difference of a monthly time series at period $t$ is equal to $\nabla\nabla_{12}y_t$. This is the amount by which the change from the previous period to the current period is different from the change that was observed exactly one year earlier. Thus, for example, the first difference of the seasonal difference in May 1978 is equal to the April-to-May change in 1978 minus the April-to-May change in 1977.

*ARAR algorithm.*

The ARAR algorithm is an adaption of ARARMA algorithm in which the idea is to apply automatically selected ‘memory-shortening’ transformations (if necessary) to the data and then to fit an ARMA model to the transformed series. The ARAR algorithm used here is a version of this in which the ARMA fitting step is replaced by the fitting of the subset AR model to the transformed data.

*Holt–Winters Seasonal algorithm (HWS).*

The Holt–Winters (HW) algorithm uses a set of simple recursions that generalize the exponential smoothing recursions to generate forecasts of series containing a locally linear trend. The Holt–Winters seasonal algorithm (HWS) extends the HW algorithm to handle data in which there are both trend and seasonal variation of known period.

**Table 3**
The value RRMSE of the post-sample forecasts.

| Method | RRMSE ($SSA_{PT}$/Other methods) | | | |
|---|---|---|---|---|
| | $r = 12$ | $r = 13$ | $r = 14$ | $r = 15$ |
| Model I | 0.60 | 0.56 | 0.59 | 0.60 |
| Model II | 0.62 | 0.63 | 0.63 | 0.65 |
| H-W | 0.72 | 0.68 | 0.71 | 0.73 |
| ARAR | 0.91 | 0.86 | 0.89 | 0.91 |
| $SSA_{LS}$ | 0.98 | 0.91 | 0.86 | 0.74 |

*Forecasting results*.

The window $L = 24$ and the first 12 singular values ($r = 12$) have been used in reconstructing and forecasting the series $y_t$ and singular values 13–24 have been considered as noise components (for more information about parameters selection, for this series, see [10]).

The results are presented in Table 3. The values of RMSE show performance of forecasting. The last four columns, labeled RRMSE, show the ratios of RMSEs SSA/other methods. As it appears from Table 3, the forecasting performance using $SSA_{PT}$ is much better than other forecasting methods; the $SSA_{PT}$ is the best among the methods considered here. For example for $r = 12$, the value of RMSE for the $SSA_{PT}$ is 40% less than the first one (model I) and almost 10% less than the ARAR algorithm. From the table, one can see that the $SSA_{PT}$ performance is better that the $SSA_{LS}$.

We also considered the performance of the SSA forecasting results with respect to different values of selected eigenvalues $r$. We choose the same window length $L$ but different eigenvalues $r$. The results are presented in Table 3, for the first 13, 14 and 15 eigenvalues. As the table shows, again, the $SSA_{PT}$ technique outperforms the other classical methods and also indicates that $SSA_{PT}$ is less sensitive than $SSA_{LS}$ for this particular example. Note that the quality of the forecast is changed when one changes the number of eigenvalues in the reconstruction step. Of course, forecasting accuracy and reconstruction quality are related. By selecting a group of eigenvalues, and considering other eigenvalues as noise, some frequencies may be filtered out completely. This destroys the signal structure and then gives a poorer reconstruction. In general, a high signal to noise ratio will result in good forecasting and vice-versa.

## 5. Conclusion

The results of this paper confirm that techniques such as the SSA which is based on the signal subspace method and SVD can be applied as a powerful technique for noise reduction and also for forecasting future data points of a noisy series. We considered the perturbations in subspace decomposition on a perturbed matrix. We studied how perturbed subspaces and singular values can be refined using perturbation theory. It was shown that the perturbations can be derived as functions of the perturbation in the matrix up to the second order. According to the results obtained based on the perturbation theory of a noisy matrix, we introduced the SSA technique based on the perturbation theory ($SSA_{PT}$). The results illustrate that $SSA_{PT}$ performs very well in reconstructing noisy series. The comparison of the forecasting results showed that $SSA_{PT}$ is much more accurate than several well-known classical methods, in forecasting of a real time series. We also consider the sensitivity of the technique with respect to different window lengths, noise levels and series lengths. The results show that $SSA_{PT}$ works very well even for a small window length. This future of SSA is very important in deal with short time series.

To cover a broad application range, various simulated series, from dynamic to chaotic and from stationary to non-stationary, are used to verify the proposed algorithm. The results with strong evidence confirm that the performance of the $SSA_{PT}$ is not sensitive to these conditions. It should be noted that in the SSA many probabilistic and statistical concepts are employed; however, as was stated earlier, the technique is non-parametric and does not make any statistical assumptions such as stationarity concerning either signal or noise in the data. One may consider this as one of the advantages of the technique compared to other classical methods which usually rely on some restricted assumptions such as normality or stationarity of the series.

## Appendix

The following proof can be found in general form in [24]. Let us consider the projections of $\widetilde{\mathbf{X}}$ onto different perturbed subspaces using the assumptions stated in (12):

$$\widetilde{\mathbf{X}}^T\widetilde{\mathbf{U}}_s = \widetilde{\mathbf{V}}_s\widetilde{\mathbf{\Sigma}}_s, \qquad \widetilde{\mathbf{X}}^T\widetilde{\mathbf{U}}_n = \widetilde{\mathbf{V}}_n\widetilde{\mathbf{\Sigma}}_n, \qquad \widetilde{\mathbf{X}}\widetilde{\mathbf{V}}_s = \widetilde{\mathbf{U}}_s\widetilde{\mathbf{\Sigma}}_s, \qquad \widetilde{\mathbf{X}}\widetilde{\mathbf{V}}_n = \widetilde{\mathbf{U}}_n\widetilde{\mathbf{\Sigma}}_n \tag{23}$$

and in a similar from the projections of $\mathbf{X}$ onto different perturbed subspaces is:

$$\mathbf{X}^T\mathbf{U}_s = \mathbf{V}_s\mathbf{\Sigma}_s, \qquad \mathbf{X}^T\mathbf{U}_n = \mathbf{V}_n\mathbf{\Sigma}_n, \qquad \mathbf{X}\mathbf{V}_s = \mathbf{U}_s\mathbf{\Sigma}_s, \qquad \mathbf{X}\mathbf{V}_n = \mathbf{U}_n\mathbf{\Sigma}_n. \tag{24}$$

Now let us now consider $\widetilde{\mathbf{X}}^T\widetilde{\mathbf{U}}_s = \widetilde{\mathbf{V}}_s\widetilde{\mathbf{\Sigma}}_s$.

$$\begin{aligned}
&\widetilde{\mathbf{X}}^T\widetilde{\mathbf{U}}_s = \widetilde{\mathbf{V}}_s\widetilde{\mathbf{\Sigma}}_s \\
&(\mathbf{X}+\delta\mathbf{X})^T(\mathbf{U}_s+\delta\mathbf{U}_s) = (\mathbf{V}_s+\delta\mathbf{V}_s)(\mathbf{\Sigma}_s+\delta\mathbf{\Sigma}_s) \qquad (25)\\
&(\mathbf{X}+\delta\mathbf{X})^T(\mathbf{U}_s+\mathbf{U}_n\mathbf{P}_1+\mathbf{U}_s\mathbf{P}_2) = (\mathbf{V}_s+\mathbf{V}_n\mathbf{P}_3+\mathbf{V}_s\mathbf{P}_4)(\mathbf{\Sigma}_s+\delta\mathbf{\Sigma}_s).
\end{aligned}$$

Eq. (25), using Eqs. (6) and (24), is simplified to:

$$\delta\mathbf{X}^T\mathbf{U}_s + \delta\mathbf{X}^T\mathbf{U}_n\mathbf{P}_1 + \delta\mathbf{X}^T\mathbf{U}_s\mathbf{P}_2 + \mathbf{V}_s\mathbf{\Sigma}_s\mathbf{P}_2 = \mathbf{V}_s\delta\mathbf{\Sigma}_s^T + \mathbf{V}_n\mathbf{P}_3\mathbf{\Sigma}_s^T + \mathbf{V}_n\mathbf{P}_3\delta\mathbf{\Sigma}_s^T + \mathbf{V}_s\mathbf{P}_4\mathbf{\Sigma}_s^T + \mathbf{V}_s\mathbf{P}_4\delta\mathbf{\Sigma}_s^T. \qquad (26)$$

Let us now premultiply both sides of (26) by $\mathbf{V}_n^T$ and $\mathbf{V}_s^T$, respectively, and use the assumption stated in (11), then we find two new equations which can be useful to find unknowns. Note that these equation are obtained using $\widetilde{\mathbf{X}}^T\widetilde{\mathbf{U}}_s = \widetilde{\mathbf{V}}_s\widetilde{\mathbf{\Sigma}}_s$:

$$\mathbf{E}_{ss}^T + \mathbf{E}_{ns}^T\mathbf{P}_1 + \mathbf{\Sigma}_s^T\mathbf{P}_2 + \mathbf{E}_{ss}^T\mathbf{P}_2 = \delta\mathbf{\Sigma}_s^T + \mathbf{P}_4\mathbf{\Sigma}_s^T + \mathbf{P}_4\delta\mathbf{\Sigma}_s^T \qquad (27)$$

$$\mathbf{E}_{sn}^T + \mathbf{E}_{nn}^T\mathbf{P}_2 + \mathbf{E}_{sn}^T\mathbf{P}_2 = \mathbf{P}_3\mathbf{\Sigma}_s^T + \mathbf{P}_3\delta\mathbf{\Sigma}_s^T. \qquad (28)$$

Similar to those obtained in (27) and (28), the following equations can be obtained using other equalities in (23):

$$\mathbf{E}_{nn}^T + \mathbf{E}_{sn}^T\mathbf{Q}_1 + \mathbf{E}_{nn}^T\mathbf{Q}_2 = \mathbf{0} \qquad (29)$$

$$\mathbf{E}_{ns}^T + \mathbf{E}_{ss}^T\mathbf{Q}_1 + \mathbf{\Sigma}_s^T\mathbf{Q}_1 + \mathbf{E}_{ns}^T\mathbf{Q}_2 = \mathbf{0} \qquad (30)$$

$$\mathbf{E}_{ss} + \mathbf{E}_{sn}\mathbf{P}_3 + \mathbf{\Sigma}_s\mathbf{P}_4 + \mathbf{E}_{ss}\mathbf{P}_4 = \delta\mathbf{\Sigma}_s + \mathbf{P}_2\mathbf{\Sigma}_s + \mathbf{P}_2\delta\mathbf{\Sigma}_s \qquad (31)$$

$$\mathbf{E}_{ns} + \mathbf{E}_{nn}\mathbf{P}_3 + \mathbf{E}_{ns}\mathbf{P}_4 = \mathbf{P}_1\mathbf{\Sigma}_s + \mathbf{P}_1\delta\mathbf{\Sigma}_s \qquad (32)$$

$$\mathbf{E}_{nn} + \mathbf{E}_{ns}\mathbf{Q}_3 + \mathbf{E}_{nn}\mathbf{Q}_4 = \mathbf{0} \qquad (33)$$

$$\mathbf{E}_{sn} + \mathbf{E}_{ss}\mathbf{Q}_3 + \mathbf{\Sigma}_s\mathbf{Q}_3 + \mathbf{E}_{sn}\mathbf{Q}_4 = \mathbf{0}. \qquad (34)$$

The unknowns can be obtained using the above equations. It should be noted that $\mathbf{Q}_2$ is a Hermitian matrix; $\mathbf{Q}_2 = \mathbf{Q}_2^T$. After some simplifications the following equality holds between $\mathbf{Q}_1$ and $\mathbf{Q}_2$:

$$\mathbf{Q}_2 \approx \frac{1}{2}\mathbf{Q}_1^T\mathbf{Q}_1. \qquad (35)$$

In addition to the above equality, the following equalities hold:

$$\mathbf{Q}_4 \approx -\frac{1}{2}\mathbf{Q}_3^T\mathbf{Q}_3, \qquad \mathbf{P}_2 \approx -\frac{1}{2}\mathbf{P}_1^T\mathbf{P}_1, \qquad \mathbf{P}_4 \approx -\frac{1}{2}\mathbf{P}_3^T\mathbf{P}_3. \qquad (36)$$

Let us first show that $\mathbf{P}_1 = -\mathbf{Q}_1^T$. To prove this, we need the following lemma.

*Lemma.*

Let $\widetilde{\mathbf{X}} = \mathbf{X} + \delta\mathbf{X}$ with SVD's of $\mathbf{X}$ and $\widetilde{\mathbf{X}}$ be given in (6) and (7), respectively. Assume that $\|\delta\mathbf{X}\|_2$ is less than the smallest nonzero singular value of $\mathbf{X}$. Let the $r$ dimensional subspace spanned by the columns of $\widetilde{\mathbf{U}}_s$, the perturbed signal subspace, be defined by $\widetilde{\mathbf{S}}_s = \text{span}(\widetilde{\mathbf{U}}_s)$ and the $K-r$ dimensional subspace spanned by the columns of $\widetilde{\mathbf{U}}_n$, the perturbed orthogonal subspace, be defined by $\widetilde{\mathbf{S}}_n = \text{span}(\widetilde{\mathbf{U}}_n)$. Then, $\widetilde{\mathbf{S}}_n$ is spanned by the columns of $\mathbf{U}_n + \mathbf{U}_s\mathbf{Q}_1$ and $\widetilde{\mathbf{S}}_s$ is spanned by the columns of $\mathbf{U}_s + \mathbf{U}_n\mathbf{P}_1$ where $\mathbf{Q}_1$ and $\mathbf{Q}_1$ are matrices whose norms are of the order of $\delta\mathbf{X}$ [24]. The lemma above gives bases for the perturbed signal and orthogonal subspaces. For the orthogonal subspace we have:

$$(\mathbf{U}_n^T + \mathbf{Q}_1^T\mathbf{U}_s^T)(\mathbf{U}_n + \mathbf{Q}_1\mathbf{U}_s) = \mathbf{I} + \mathbf{Q}_1^T\mathbf{Q}_1. \qquad (37)$$

The above equation shows how the basis for the perturbed orthogonal subspace can be normalized. Therefore, an orthonormal basis for the perturbed orthogonal subspace is given by

$$(\mathbf{U}_n + \mathbf{U}_s\mathbf{Q}_1)(\mathbf{I} + \mathbf{Q}_1^T\mathbf{Q}_1)^{-\frac{1}{2}}. \qquad (38)$$

A similar equation holds for the perturbed signal's subspace. An orthonormal basis for the perturbed signal's subspace is given by

$$(\mathbf{U}_s + \mathbf{U}_n\mathbf{P}_1)(\mathbf{I} + \mathbf{P}_1^T\mathbf{P}_1)^{-\frac{1}{2}}. \qquad (39)$$

We know that the perturbed signal and orthogonal subspaces are orthogonal to each other. Thus the unnormalized basis vectors given in the lemma are orthogonal. That is,

$$(\mathbf{U}_n^T + \mathbf{Q}_1^T\mathbf{U}_s^T)(\mathbf{U}_s + \mathbf{U}_n\mathbf{P}_1) = \mathbf{0} \qquad (40)$$

$$\Rightarrow \mathbf{P}_1 + \mathbf{Q}_1^T = \mathbf{0}, \qquad \Rightarrow \mathbf{P}_1 = -\mathbf{Q}_1^T. \qquad (41)$$

Therefore, we only need to obtain $\mathbf{Q}_1$ as others can be obtained based on $\mathbf{Q}_1$. Let us now consider $\mathbf{Q}_1$.

The following equality is obtained between $\mathbf{Q}_1$ and $\mathbf{Q}_3$ using (27)–(34).

$$\mathbf{E}_{sn} + \mathbf{E}_{ss}\mathbf{Q}_3 + \mathbf{\Sigma}_s\mathbf{Q}_3 = \mathbf{Q}_1\mathbf{E}_{nn}. \tag{42}$$

The above equation can be written as the following form:

$$\mathbf{Q}_3 = \mathbf{\Sigma}_s^{-1}\mathbf{E}_{sn} - \mathbf{\Sigma}_s^{-1}\mathbf{E}_{ss}\mathbf{Q}_3 + \mathbf{\Sigma}_s^{-1}\mathbf{Q}_1\mathbf{E}_{nn}. \tag{43}$$

Now we need to express $\mathbf{Q}_3$ by $\mathbf{Q}_1$. Substituting on the right-hand side of (43) and neglecting higher order terms, (43) is simplified to

$$\mathbf{Q}_3 = \mathbf{\Sigma}_s^{-1}\mathbf{E}_{sn} + \mathbf{\Sigma}_s^{-1}\mathbf{E}_{ss}\mathbf{\Sigma}_s^{-1}\mathbf{E}_{ss} + \mathbf{\Sigma}_s^{-1}\mathbf{Q}_1\mathbf{E}_{nn}. \tag{44}$$

In a similar way, the following equality is obtained between $\mathbf{Q}_1$ and $\mathbf{Q}_3$ using (27)–(34):

$$\mathbf{E}_{ns}^T + \mathbf{E}_{ss}^T\mathbf{Q}_1 + \mathbf{\Sigma}_s^T\mathbf{Q}_1 = \mathbf{Q}_3\mathbf{E}_{nn}^T. \tag{45}$$

Substituting $\mathbf{Q}_3$ in (44) into (47) and discarding higher order terms, we obtain an equation for $\mathbf{Q}_1$ as follows

$$\mathbf{\Sigma}_s^T\mathbf{Q}_1 = -\mathbf{E}_{ns}^T - \mathbf{\Sigma}_s^{-1}\mathbf{E}_{sn}\mathbf{E}_{nn}^T - \mathbf{E}_{ss}^T\mathbf{Q}_1. \tag{46}$$

The above equation shows that it is not easy to obtain a closed form for $\mathbf{Q}_1$ in the current matrix-form equation. However, we can use the recursive technique. Note that we are only interested in the expression of up to the second-order of $\delta\mathbf{X}$. Multiplying both sides of (47) by $\mathbf{\Sigma}_s$, and introducing a new definition $\mathbf{\Delta}_s = (\mathbf{\Sigma}_s\mathbf{\Sigma}_s^T)^{-1}$, (47) becomes

$$\mathbf{Q}_1 \approx -\mathbf{\Delta}_s\mathbf{\Sigma}_s\mathbf{E}_{ns}^T - \mathbf{\Delta}_s\mathbf{E}_{sn}\mathbf{E}_{nn}^T - \mathbf{\Delta}_s\mathbf{\Sigma}_s\mathbf{E}_{ss}^T\mathbf{Q}_1. \tag{47}$$

Now, we use recursive method and keeping terms only up to the second-order perturbations, we then use the following matrix form to obtain $\mathbf{Q}_1$,

$$\mathbf{Q}_1 \approx -\mathbf{\Delta}_s\mathbf{\Sigma}_s\mathbf{E}_{ns}^T - \mathbf{\Delta}_s\mathbf{E}_{sn}\mathbf{E}_{nn}^T + \mathbf{\Delta}_s\mathbf{\Sigma}_s\mathbf{E}_{ss}^T\mathbf{\Delta}_s\mathbf{\Sigma}_s\mathbf{E}_{ns}^T. \tag{48}$$

Now, rearranging all terms in (48) and new definition $\mathbf{F}_1 = -\mathbf{\Delta}_s\mathbf{\Sigma}_s\mathbf{E}_{ns}^T$, (48) becomes (14),

$$\mathbf{Q}_1 = \mathbf{\Delta}_s(\mathbf{\Sigma}_s\mathbf{E}_{ss}^T\mathbf{\Delta}_s\mathbf{\Sigma}_s\mathbf{E}_{ns}^T - \mathbf{E}_{sn}\mathbf{E}_{nn}^T) + \mathbf{F}_1. \tag{49}$$

## References


[1] H. Hassani, A. Dionisio, M. Ghodsi, The effect of noise reduction in measuring the linear and nonlinear dependency of financial markets, Nonlinear Analysis: Real World Applications 11 (1) (2010) 492–502.
[2] G.E.P. Box, G.M. Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, 1970.
[3] N. Golyandina, V. Nekrutkin, A. Zhigljavsky, Analysis of Time Series Structure: SSA and Related Techniques, Chapman & Hall, CRC, 2001.
[4] A. Soofi, L. Cao, Nonlinear forecasting of noisy financial data, in: Soofi, Cao (Eds.), Modeling and Forecasting Financial Data: Techniques of Nonlinear Dynamics, Kluwer Academic Publishers, Boston, 2002.
[5] D.W. Tufts, R. Kumaresan, I. Kirsteins, Data adaptive signal estimation by singular value decomposition of a data matrix, Proceedings of the IEEE 70 (6) (1982) 684–685.
[6] J.A. Cadzow, Signal enhancement composite property mapping algorithm, IEEE Transactions on Acoustics, Speech, and Signal Processing 36 (1) (1988) 49–62.
[7] D.S. Broomhead, G.P. King, Extracting qualitative dynamics from experimental data, Physica D 20 (1986) 217–236.
[8] R. Vautard, P. Yiou, M. Ghil, Singular-spectrum analysis: a toolkit for short, noisy chaotic signal, Physica D 58 (1992) 95–126.
[9] M. Ghil, C. Taricco, Advanced spectral analysis methods, in: G.C. Castagnoli, A. Provenzale (Eds.), Past and Present Variability of the Solar-Terrestrial System: Measurement, Data Analysis and Theoretical Model, IOS Press, 1997, pp. 137–159.
[10] H. Hassani, Singular spectrum analysis: methodology and comparison, Journal of Data Science 5 (2) (2007) 239–257.
[11] H. Hassani, A. Zhigljavsky, Singular spectrum analysis: methodology and application to economics data, Journal of Systems Science and Complexity (JSSC) 22 (3) (2009) 372–394.
[12] H. Hassani, S. Heravi, A. Zhigljavsky, Forecasting European industrial production with singular spectrum analysis, International Journal of Forecasting 25 (1) (2009) 103–118.
[13] H. Hassani, A. Soofi, A. Zhigljavsky, Predicting daily exchange rate with singular spectrum analysis, Nonlinear Analysis: Real World Applications 11 (3) (2010) 2023–2034.
[14] M. Ghodsi, H. Hassani, S. Sanei, Extracting fetal heart signal from noisy maternal ECG by singular spectrum analysis, Statistics and its Interface 3 (3) (2010) 399–411.
[15] H. Hassani, M. Zokaei, D. von Rosen, S. Amiri, M. Ghodsi, Does noise reduction matter for curve fitting in growth curve models? Computer Methods and Programs in Biomedicine 96 (1) (2009) 173–181.
[16] H. Hassani, R. Mahmoudvand, M. Yarmohammadi, Filtering and denoising in the linear regression model, Fluctuation and Noise Letters 9 (4) (2010) 343–358.
[17] M. Ghodsi, H. Hassani, S. Sanei, Y. Hicks, The use of noise information for detection of temporomandibular disorder, Journal of Biomedical Signal Processing and Control 4 (2) (2009) 79–85.
[18] H. Hassani, D. Thomakos, A review on singular spectrum analysis for economic and financial time series, Statistics and its Interface 3 (3) (2010) 377–397.
[19] H. Hassani, Singular spectrum analysis based on the minimum variance estimator, Nonlinear Analysis: Real World Applications 11 (3) (2009) 2065–2077.
[20] D. Danilov, A. Zhigljavsky (Eds.), Principal Components of Time Series: The 'Caterpillar' Method, University of St. Petersburg Press, 1997, (in Russian).
[21] V. Nekrutkin, Perturbation expansions of signal subspaces for long signals, Statistics and its Interface 3 (2010) 297–319.
[22] G.W. Stewart, J. Sun, Matrix Perturbation Theory, Academic, San Diego, CA, 1990.
[23] G. Golub, C. van Loan, Matrix Computations, third ed., The Johns Hopkins University Press, London, 1996.

[24] Z. Xu, Perturbation analysis for subspace decomposition with applications in subspace-based algorithms, IEEE Transactions on Signal Processing 50 (11) (2002) 2820–2830.
[25] G. Szpiro, Noise in unspecified, non-linear time series, Journal of Econometrics 78 (1997) 229–255.
[26] R.A. Davis, T. Mikosch, The sample autocorrelations of financial time series models, in: W.J. Fitzgerald, R.L. Smith, A.T. Walden, P. Young (Eds.), Nonlinear and Nonstationary Signal Processing, Cambridge University Press, Cambridge, England, 2000, pp. 247–274.
[27] W.A. Brock, W.D. Dechert, J. Scheinkman, A test for independence based on the correlation dimension, Department of Economics, University of Wisconsin, University of Houston and University of Chicago, 1987 (Revised Version, Brock, W.A., W.D. Dechert, J. Scheinkman and B. LeBaron), 1991.
[28] A. Abhyankar, L.S. Copeland, W. Wong, Nonlinear dynamics in real-time equity market indices: evidence from the United Kingdom, The Economic Journal 105 (1995) 864–880.
[29] J. Fan, Q. Yao, Non Linear Time Series: Nonparametric and Parametric Methods, Springer, NY, 2005.
[30] G. Granger, J. Lin, Using the mutual information coefficient to identify lags in nonlinear models, Journal of Time Series Analysis 15 (1994) 371–384.
[31] A. Dionisio, R. Menezes, D.A. Mendes, Entropy-based independence test, Nonlinear Dynamics 44 (2006) 351–357.
[32] P.J. Brockwell, R.A. Davis, Introduction to Time Series and Forecasting, 2nd ed., Springer, 2002.