Singular Spectrum Analysis of Non-stationary Signals

Osinakachukwu G. Akaniro
Department of Computer Science
School of Science and Technology, Nottingham Trent University
Nottingham, NG11 8NS, United Kingdom
osinakachukwu.akaniro2018@my.ntu.ac.uk

Saeid Sanei
Electrical Engineering Department
Imperial College London
London, SW7 2BT, United Kingdom
s.sanei@imperial.ac.uk

***Abstract*— Singular spectrum analysis has been widely used for decomposition of single channel signals as well as prediction of time series. The technique is sensitive to the data non-stationarity and its performance rapidly deteriorates as the non-stationarity increases. In this paper a robust approach for adjusting the embedding dimension based on a measure of non-stationarity is proposed. The method has been applied to simulated and real data and its superior performance verified.**

***Index Terms*— Singular spectrum analysis, non-stationarity, embedding dimension, efficient SSA**

I. Introduction

The natural signals, particularly physiological signals,are often non-stationary and suffer from noise effect. Single channel or one-dimensional time series data often contain trends, cycles, anomalies, transient bursts and other components. For long time series, the desired patterns or trends which are usually buried in noise, are often difficult to visualize and discover. Singular spectrum analysis (SSA) [1] [2], as a subspace decomposition approach,applies non-parametric techniques that adapt the commonly used singular value decomposition (SVD) for decomposing (single channel) time series data.

Decomposition [3,4], restoration from noise [5-7], and prediction of time series [7] for single channel data has been successfully achieved using SSA. Moreover, the SSA concept has been used in developing a new adaptive line enhancer which can be applied at the presence of non-Gaussian noise and system non-linearity [9]. Recently, tensor SSA [10] and quaternion SSA [11] have been proposed for better analysis of sleep data and hypercomplex (a.k.a quaternion-valued) signals respectively. It soon found many applications in other fields such as biomedical signal processing and bioengineering [2] [12-16] . The traditional SSA however is very sensitive to non-stationary data and almost all the natural data, particularly physiological signals are non-stationary. Hence, this paper we focus on how to modify the SSA to make it suitable for non-stationary data analysis, decomposition, or prediction.

II. Singular Spectrum Analysis

The basic SSA method consists of two complementary stages: decomposition and reconstruction; each stage comprises of two separate stages. In the first stage the series is decomposed and in the second stage the original series is reconstructed and used for further analysis. The main concept in studying the properties of SSA is separability, which characterizes how well different source signals can be separated from their single channel mixtures. To cope with the absence of approximate separability there are different ways of modifying SSA leading to different versions such as SSA with single and double centering, Toeplitz SSA, and sequential SSA [4]. A brief description of the two SSA stages together with the corresponding mathematics is given in the following subsections.

A. Decomposition

This stage includes an embedding operation followed by singular value decomposition (SVD). The embedding operation maps a one-dimensional signal $f = [f_0 \ f_1 \ ... \ f_{L-1}]$ of length L into an $l \times M$ matrix $\mathbf{H}$ with rows of length $M < L$, called embedding dimension, as

$$\mathbf{H} = \begin{bmatrix} f_0 & f_1 & \cdots & f_{M-1} \\ f_1 & f_2 & \cdots & f_M \\ \vdots & \ddots & & \vdots \\ f_{l-1} & f_l & \cdots & f_{L-1} \end{bmatrix} \tag{1}$$

The window length M should be sufficiently large.The so-called trajectory matrix $\mathbf{H}$ is a Hankel matrix, which means that all the elements along the diagonals $i+j = const$ are the same. Then,

$$\mathbf{C}_f = \frac{1}{L-M+1}\mathbf{H}^T\mathbf{H} \tag{2}$$

which gives a square $M \times M$ matrix. The SVD decomposition of $\mathbf{C}_f$ leads to M eigenvalues and their corresponding eigenvectors. Moreover, $trace(\mathbf{C}_f) = \sigma_f^2$, which is the total variance of signal f. The SVD decomposition leads to sum of rank-1 matrices. Therefore, $\mathbf{H}_i = \sqrt{\lambda_i}\mathbf{U}_i\mathbf{V}_i^T$.

B. Reconstruction

Assuming $\mathbf{H}$ to be a rank-d matrix, $\mathbf{H}$ can be reconstructed as: $\widehat{\mathbf{H}} = \sum_{i=1}^{d} \mathbf{H}_i = \sum_{i=1}^{d} \sqrt{\lambda_i}\mathbf{U}_i\mathbf{V}_i^T$.
In many applications however, only one (which can be the main signal trend) or a group of underlying components (such as periodic components) may be of interest. In such cases a

978-1-7281-5707-8/20/$31.00 ©2020 IEEE

group of say p components will be used in the reconstruction from $\widehat{\mathbf{H}} = \{\hat{h}_{ij}\}$; i.e.

$$\widehat{\mathbf{H}} = \sum_{i=1}^{p} \mathbf{H}_i = \sum_{i=1}^{p} \sqrt{\lambda_i}\mathbf{U}_i\mathbf{V}_i^T \quad (3)$$

This is a lower-rank Hankel matrix and the signal (components) of interest $\hat{f}$ can be achieved by one-to-one correspondence or diagonal averaging, i.e.

$$\hat{f}_k = \frac{1}{2k+1}\sum_{j=0}^{k} \hat{h}_{k+j+1,j+1} \text{ for } k = 0, \ldots, M-1 \quad (4)$$

Although in using SSA for forecasting, often the first few eigentriples corresponding to the signal trend are used various adaptive and non-adaptive approaches for selection of particular eigentriples can be seen in the literature [15].

The choice of embedding dimension depends on few crucial factors. Firstly, the dimension of Hankel matrix corresponds to M and L. This means M cannot be too small, so the number of underlying components become more than $d = rank(\mathbf{C}_f)$. Secondly, for a large M, the size of Hankel matrix becomes large and the computation cost, particularly for SVD computation, becomes intensive. Finally and most importantly, the signal should remain stationary within an M-sample interval. The latter condition highly affects the reconstruction performance and the quality of the decomposed signals. Therefore, in the following section, an efficient SSA to cope with non-stationarity is proposed.

III. SSA FOR NON-STATIONARY SIGNALS

Considering that in the stationary case and a noiseless environment the rank d of original Hankel matrix $\mathbf{H}$ is known given that the rank is proportional to the underlying orthogonal components of the signal. In that case, the best value for $\mathbf{M}$ would be as low as d. However, in practice, the rank cannot be approximated before building up the Hankel matrix. So, one way is to set a M very high, say $L-1$, at the beginning, design the Hankel matrix, and then estimate the rank. Often the rank is very small for noiseless scenarios and very large for noisy cases. One way to overcome the problem for stationary cases is to have $M = d + M_0$, where M_0 is a constant and $M_0 < L - d$. This is mainly to accommodate the spurious or transient signal components as well as noise. For non-stationary (noiseless) signals or time series a large window size is not suitable as it highly deteriorates the reconstruction performance. In this case a measure of non-stationarity should be incorporated within the formulation. A signal is wide-sense stationary if the signal distribution is not time-varying. For a zero mean signal often stationary in variance (where the signal variance remains fixed) is used. More complex measures have been proposed mainly for segmentation of data into stationary segments such as [18-20].

Assume $0 \leq \gamma \leq 1$ indicates fully stationary ($\gamma = 0$) to fully non-stationary ($\gamma = 1$) signal distribution. In that case, we define a suitable window size (embedding dimension) equal to

$$M = d + \frac{1-\gamma}{1+\gamma}M_0 \quad (5)$$

For the noisy cases however, estimation of matrix rank is not helpful as the noise components significantly increase the estimated value for d. In this case, an information criterion-based method is often applied. In [21] the methods based on residual variance (RV), Akaike information criterion (AIC), Bayesian information criterion (BIC), and Wald tests on amplitudes (WA) and locations (WL) have been presented. Probably the most robust approach for detection of the number of sources is that developed by Bai and He [22]. In this approach an information criterion method in which the penalty functions are selected based on the spatio-temporal source model, has been developed to estimate the number of independent sources from the covariance matrix by

1) Applying SVD to decompose $\mathbf{C}_f$ to obtain the eigenvalues λ_i, where $\lambda_1 > \cdots > \lambda_m$
2) Calculating the information criterion value using the eigenvalues. The information criterion (IC) for the noisy case with unknown noise can be calculated as [22]

$$IC = \left(k + 1 - w + \frac{2(n_e-k)^2 + n_e - k + 2}{6(n_e-k)} - \sum_{i=1}^{k}\frac{\bar{\lambda}_{n_e-k}^2}{(\lambda_i - \bar{\lambda}_{n_e-k})^2}\right)\cdot \log\left(\bar{\lambda}_{n_e-k}^{-(n_e-k)}\prod_{i=k+1}^{n_e}\lambda_i\right) + 2g(k,n_e)\beta(w-1) \quad (6)$$

where in this equation $g(k, n_e) = k(n_e - k + 2)(n_e - k + 2)/2$, and $\bar{\lambda}_{n_e-k}$ is the average of the $n_e - k$ smallest eigenvalues. In this equation n_e is the number of rows in the Hankel matrix $\mathbf{H}$, k is the number of actual underlying sources to be estimated, w is the number of time points to form a spatio-temporal data matrix, which in this case is equal to M_0, and $\beta(w)$ is the penalty function which can have constant or logarithmic values.
3) According to the rule of information criterion [22] the number of sources with minimum IC is selected as the estimated number of sources.
To proceed, we set $d = IC$. It has been shown that in moderate noise environment, the accuracy of the method is above 80%, which is reasonable and more robust compared to those can be achieved by AIC, BIC, WA, or WL.
4) In the following experiments we assume $M_0 = round\left(\frac{L-d}{2}\right) \approx round(\frac{L}{2})$ for $L \gg d$. Then, the value of d is estimated as discussed above.

IV. EXPERIMENTS

A. Simulated Signals

In the first experiment a signal trend comprising of three segments each sampled from three different distributions and noise added to the signal. Here, the criterion for selection of the eigentriples is to retain first two dominant eigentriples to be able to reconstruct the trend. The signal, its decomposition using stationary SSA and its separation using the proposed non-stationary SSA have been depicted in Figure 1.

B. Real EMG contaminated with ECG

Electromyogram (EMG) signals provide valuable information relating to peripheral and central motor functions and have been widely adopted in the study of motor function and movement disorders. Surface EMGs represent a superposition of electrical activities from motor unit action potentials located subcutaneous to the detecting electrodes. Surface EMGs have previously been applied to assess muscular activity quantitatively and diagnose various muscular disorders. These signals however are non-stationary and often available using only single-channel recordings. The EMGs are often affected by electrocardiogram (ECG) mainly because the heartbeat is propagated through the blood vessels.

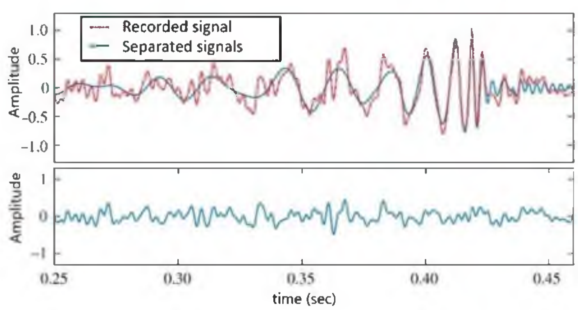

Fig. 1. A simulated noisy non-stationary signal; top the original signal in Red and the trend (Blue) obtained by applying non-stationary SSA; bottom is the residual signal (noise component).

Restoration of EMG signals from artefacts is extremely difficult due to their randomness and being noisy in nature. Unfortunately, having more than one channel recording is not always possible and if accessible the mixing system would be highly nonlinear. Moreover, generally the conditions of independency and stationarity of the signal components are not fully satisfied. In an attempt by Lu et al. [23] an adaptive noise canceller based on recursive-least squares algorithm was developed for removing the ECG artefact from surface EMGs recorded in patients with cervical dystonia. Even in this approach an ECG signal recorded from a separate channel was used as a reference signal. In our attempt, however, no reference signal has been used.Here, the criterion for selection of the eigentriples is to select those belonging to periodic ECG. So, couples of eigenvalues of the same power are selected.

In Figure 2 the result of application of the method to an EMG signal contaminated with strong ECG can be viewed. The signals were recorded from the human forearm using surface EMG. The sampling frequency was 2 KHz and the subject was relatively relaxed during slow arm movements. The overall recorded data length was 70 seconds (140K data samples), divided into 20K sample segments (10 sec each). In these experiments no pre-processing was performed. Therefore,the noisy EMG remained un-altered and the ECG was completely separated.

In order to objectively test the results a correlation factor such as that used in [9] is used here. It is expected that the decomposed signals are perfectly orthogonal. Therefore, lower correlation between the decomposed signals is desired. Table 1 shows the correlations for both fixed M and the one with an adaptive estimation of M based on the signal non-stationarity.

TABLE I
AVERAGE CORRELATION COEFFICIENT BETWEEN THE ESTIMATED ECG AND EMG FOR 42 DATA SEGMENTS FROM 21 SUBJECTS USING M_0 AND ADAPTIVELY ESTIMATED M.

	fixed M	*Adaptive* M
Correlation between the decomposed components as in [9]	0.0667	0.22×10^{-3}

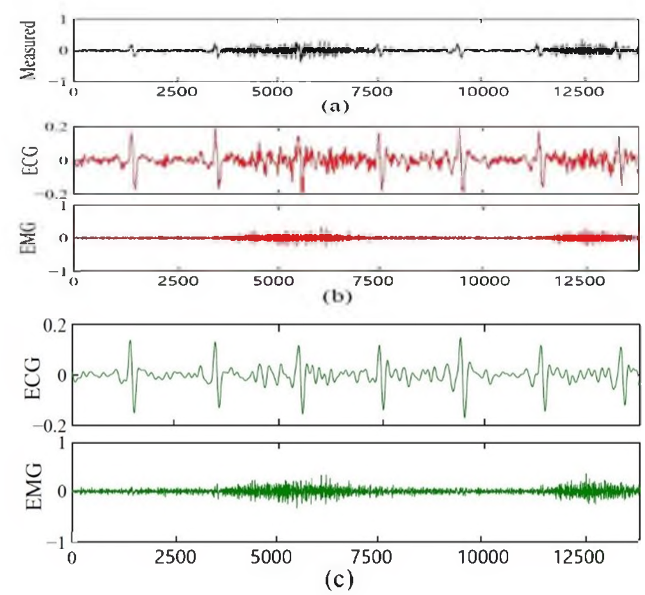

Fig. 2. The result of application of the proposed method to an EMG signal recorded from the flexor carpi muscle of a monkey's hand; (a) the non-stationary EMG signal which is corrupted by ECG artefact, (b) the outcome of stationary SSA application, and (c) the result after adjusting the embedding dimension according to the signal non-stationarity.

V. CONCLUSIONS

In this article the SSA embedding dimension is adaptively estimated and used to enable more accurate (or even efficient) single channel decomposition. In the experiments the method has been applied to both synthetic non-stationary noisy signal and real EMG signals corrupted with ECG. The result has been compared with the case where the signal is wrongly assumed stationary and its superior performance verified. The proposed method can be used for separation of intermittent or transient components in the signal which makes it useful in communication signal processing such as anti-jamming or avoiding cosmic interference.

REFERENCES

[1] J. B. Elsner and A. A. Tsonis, Singular Spectral Analysis. A New Tool in Time Series Analysis. New York: 1996, Plenum Press.

[2] V. Golyandina, N. Nekrutkin and A. Zhigljavsky, Analysis of Time Series Structure: SSA and Related Techniques. Boca Raton, Florida: Chapman and Hall/CRC, 2001

[3] S. Sanei, M. Ghodsi, and H. Hassani, "A constrained singular spectrum analysis approach to murmur detection from heart sounds," Elsevier Journal of Medical Engineering and Physics, vol. 33, no. 3, pp. 362-367, 2011

[4] M. Ghodsi, H. Hassani, S. Sanei, and Y. Hicks, "The use of noise information for detection of temporo-mandibular disorder," Elsevier Journal of Biomedical Signal Processing and Control, vol. 4, no. 2, 2008, ISSN 17468094

[5] M. Ghodsi, H. Hassani, and S. Sanei, "Extracting foetal heart signal from noisy maternal ECG by singular spectrum analysis," Journal of Statistics and its Interface, Special Issue on the Application of SSA, vol. 3, pp. 399–411, 2010

[6] S. Mahvash Mohammadi, M. Ghavami, S. Kouchaki, and S. Sanei, "Improving time-frequency domain sleep EEG classification via singular spectrum analysis," Journal of Neuroscience Methods, vol. 273, pp. 96-106, August 2016, doi:10.1016/j.jneumeth.2016.08.00

[7] M. Ghodsi, S. Sanei, Y. Hicks, T. Lee, and S. Dunne, "A facial pattern recognition approach for detection of temporo-mandibular disorders," Proc. European Signal Processing Conference, EUSIPCO 2007.

[8] N. Golyandina and A. Korobeynikov, "Basic singular spectrum analysis and forecasting with R," Computational Statistics & Data Analysis, vol. 71, pp. 934–954, 2014.

[9] S. Sanei, T. Lee, and V. Abolghasemi, "A new adaptive line enhancer based on singular spectrum analysis," IEEE Transactions on Biomedical Engineering, vol. 59, no. 2, pp. 428-434, 2012.

[10] S. Kouchaki, S. Sanei, E. L. Arbon, and D. J. Dijk, "Tensor based singular spectrum analysis for automatic scoring of sleep EEG," IEEE Transactions on Neural Systems & Rehabilitation Engineering, vol. 23, no. 1, pp. 1-9, January 2015. doi.org/10.1109/TNSRE.2014.2329557

[11] S. Enshaeifar, S. Kouchaki, C. Cheong Took, and S. Sanei, "Quaternion singular spectrum analysis of electroencephalogram with application to sleep analysis," IEEE Transactions on Neural Systems & Rehabilitation Engineering, vol. 24, no. 1, pp. 57 -67, 2016.

[12] S. Sanei, C. Cheong Took, and S. Enshaeifar "Quaternion adaptive line enhancer based on singular spectrum analysis," Proceedings of IEEE ICASSP 2018, Calgary, Canada.

[13] S. Sanei, M. Ghodsi, and H. Hassani, "A constrained singular spectrum analysis approach to murmur detection from heart sounds," Elsevier Journal of Medical Engineering and Physics, vol. 33, no. 3, pp. 362–367, 2011.

[14] F. Ghaderi, H. Mohseni, and S. Sanei, "Localizing heart sounds in respiratory signals using singular spectrum analysis," IEEE Transactions on Biomedical Engineering, vol. 59, no. 12, pp. 3360-3367, 2011.

[15] S. Sanei and H. Hassani, Singular Spectrum Analysis of Biomedical Signals, CRC Press, 2015, ISBN-10: 1466589272.

[16] S. Wang, H. L. Tang, L. I. Al Turk, Y. Hu, S. Sanei, G. M. Saleh and T. Peto, "Localising micro-aneurisms in fundus images through singular spectrum analysis," IEEE Transactions on Biomedical Engineering, vol. 64, no. 5, pp. 990-1002, 2016.

[17] T. K. M. Lee, J. G. Lim, S. Sanei, and S. S. W. Gan, "Advances on singular spectrum analysis of rehabilitative assessment data," Journal of Medical Imaging and Health Informatics, vol. 5, no. 2, pp. 350-358, April 2015.

[18] H. Azami, H. Hassanpour, J. Escudero, and S. Sanei, "An intelligent approach for variable size segmentation of non-stationary signals," Elsevier Journal of Advanced Research, vol. 6, pp. 687–698, 2014.

[19] H. Azami, S. E. Arnold, S. Sanei, Z. Chang, G. Sapiro, J. Escudero, and A. S. Gupta, "Multiscale fluctuation-based dispersion entropy and its applications to neurological diseases," IEEE Access, vol. 7, no. 1, pp. 68718-68733, Print ISSN: 2169-3536, 2019.

[20] S. Das and G. P. Nason, "Measuring the degree of non-stationarity of a time series," Stat, vol. 5, no. 1, pp. 295-305, 2016.

[21] L. J. Waldorp, H. M. Huizenga, A. Nehorai, and R. P. Grasman, "Model selection in electromagnetic source analysis with an application to VEFs," IEEE Transactions on Biomedical Engineering, vol. 49, no. 10, pp. 1121-1129, March 2002.

[22] X. Bai and B. He, "Estimation of number of independent brain electric sources from the scalp EEGs," IEEE Transactions on Biomedical Engineering, vol. 53, no. 10, pp. 1883-1892, 2006.

[23] G. Lu, J. S. Brittain, P. Holland, J. Yianni, A. L. Green, J. F. Stein, T. Z. Aziz, and S. Wang, "Removing ECG noise from surface EMG signals using adaptive filtering," Neuroscience Letters, vol. 462, no. 1, pp. 14–19, 2009.