C. R. Acad. Sci. Paris, Ser. I 349 (2011) 987–990

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Statistics

Separability and window length in singular spectrum analysis

Séparabilité et longueur de fenêtre dans l'analyse d'un spectre singulier

Hossein Hassani [a,b], Rahim Mahmoudvand [c], Mohammad Zokaei [c,*]

[a] *Statistics Group, School of Mathematics, Cardiff University, CF24 4AG, UK*
[b] *Institute for International Studies (I.I.E.S), 1967743711, Tehran, IR, Iran*
[c] *Faculty of Mathematical Sciences, Shahid Beheshti University, GC, Iran*

ARTICLE INFO

Article history:
Received 9 March 2011
Accepted after revision 18 July 2011
Available online 25 August 2011

Presented by Paul Deheuvels

ABSTRACT

To find the optimal value of window length in singular spectrum analysis (SSA), we consider the concept of separability between the signal and noise component. The theoretical results confirm that for a wide class of time series, the suitable value of this parameter is $median\{1,\ldots,T\}$ with the series of length T. The theoretical results obtained here coincide with those obtained previously from the empirical point of view.

RÉSUMÉ

Pour déterminer la valeur optimale de la longueur de fenêtre dans l'analyse d'un spectre singulier (SSA) on utilise le concept de séparabilité entre le signal et la composante du bruit. Les résultats théoriques confirment, que pour une classe importante de séries temporelles, la valeur la mieux adaptée de ce paramètre est la médiane de $\{1,\ldots,N\}$ pour des séries de longueur N. Les résultats théoriques obtenus dans cette Note coïncident avec ceux qui sont utilisés à partir de méthodes empiriques.

1. Introduction

Singular Spectrum Analysis (SSA) is a relatively novel but powerful technique in time series analysis that has been developed and applied to many practical problems (see, for example, [1–5] and references therein).

The whole procedure of the SSA technique depends upon two basic, but very important, parameters i) the window length L and ii) the number of eigenvalues r, that one needs to select for reconstructing noise free series from a noisy series of length T. Certainly, the choice of parameter L depends on the data we have and the analysis we aim to perform. The improper choice of L would imply an inferior decomposition.

Elsner and Tsonis [6] give some discussion and remark that choosing $L = T/4$ is a common practice. It has been recommended that L should be large enough but not larger than $T/2$ [1]. Large values of L allow longer period oscillations to be resolved, but choosing L too large leaves too few observations from which to estimate the covariance matrix of the L variables. Although considerable attempt and various techniques have been considered for selecting the optimal value of L, there is inadequate theoretical justification for choosing L. In order to find the optimal value of L we investigate the concept of separability, which is the main methodological concept in SSA.

* Corresponding author.
E-mail addresses: hassanih@cf.ac.uk (H. Hassani), r.mahmodvand@gmail.com (R. Mahmoudvand), zokaei@sbu.ac.ir (M. Zokaei).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2011.07.012

Throughout the Note, we consider a time series $Y_T = S_T + \epsilon_T$ of length T, where S_T is the component of interest (usually called signal) and ϵ_T is the noise component (can be a random noise or a deterministic component). The aim of the SSA reconstruction stage is to find an estimate for the signal components S_T, $\hat{S}_T$. In the ideal situation we completely remove the noise component ϵ_T; i.e., $S_T = \hat{S}_T$. However, in the real situation we cannot reconstruct S_T completely, but we try to find $\hat{S}_T$ close to S_T with respect to different criteria.

2. Theoretical results

2.1. Singular spectrum analysis: SSA

A short description of the SSA technique is given below (for more information see [1]).

Stage I. Decomposition

Step I: Embedding. Embedding can be considered as a mapping that transfers a one-dimensional time series $Y_T = (y_1, \dots, y_T)$ into the multi-dimensional series $X_1, \dots, X_K$ with vectors $X_i = (y_i, \dots, y_{i+L-1})^T \in \mathbf{R}^L$, where L $(2 \leqslant L \leqslant T-1)$ is the *window length* and $K = T - L + 1$. The result of this step is the trajectory matrix

$$\mathbf{X} = [X_1, \dots, X_K] = (x_{ij})_{i,j=1}^{L,K}. \tag{1}$$

Note that the trajectory matrix $\mathbf{X}$ is a Hankel matrix, which means that all the elements along the second diagonal $i + j = const$ are equal.

Step II: Singular Value Decomposition, SVD. In this step we perform the SVD of $\mathbf{X}$. Denote by $\lambda_1, \dots, \lambda_L$ the eigenvalues of $\mathbf{XX}^T$ arranged in the decreasing order $(\lambda_1 \geqslant \cdots \geqslant \lambda_L \geqslant 0)$ and by $U_1, \dots, U_L$ the corresponding eigenvectors. The SVD of $\mathbf{X}$ can be written as $\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_L$, where $\mathbf{X}_i = \sqrt{\lambda_i} U_i {V_i}^T$ and $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$ (if $\lambda_i = 0$ we set $\mathbf{X}_i = 0$).

Stage II. Reconstruction

Step I: Grouping. The grouping step corresponds to splitting the elementary matrices into several groups and summing the matrices within each group. Let $I = \{i_1, \dots, i_p\}$ $(p < L)$ be a group of indices $i_1, \dots, i_p$. Then the matrix $\mathbf{X}_I$ corresponding to the group I is defined as $\mathbf{X}_I = \mathbf{X}_{i_1} + \cdots + \mathbf{X}_{i_p}$. The split of the set of indices $\{1, \dots, L\}$ into disjoint subsets $I_1, \dots, I_m$ corresponds to the representation $\mathbf{X} = \mathbf{X}_{I_1} + \cdots + \mathbf{X}_{I_m}$. In our case, we have only two groups; $m = 2$. I_1 and I_2 are related to the noise and signal components, respectively.

Step II: Diagonal averaging. The purpose of diagonal averaging is to transform a matrix to the form of a Hankel matrix, which can be subsequently converted to a time series.

2.2. Properties of Hankel matrix

Let us now consider the properties of Hankel matrix as the second step of both stages of the SSA algorithm are based on the properties of the Hankel matrix $\mathbf{X}$.

Theorem 2.1. *Let* $\mathbf{B}$ *denote the Hankelized form of the arbitrary* $L \times K$ *matrix* $\mathbf{A}$*. Then*: $T^L_{\mathbf{A}-\mathbf{B}} = T^L_{\mathbf{A}} - T^L_{\mathbf{B}}$ *where,* $T^L_{\mathbf{A}} = \mathrm{tr}(\mathbf{AA}^T)$.

Proof. It is sufficient to show that $\mathrm{tr}(\mathbf{AB}^T) = \mathrm{tr}(\mathbf{BB}^T)$.

$$\mathrm{tr}\big(\mathbf{AB}^T\big) = \sum_{s=1}^{T+1} \sum_{l=s_1}^{s_2} a_{l,s-l} b_{l,s-l} = \sum_{s=1}^{T+1} \sum_{l=s_1}^{s_2} a_{l,s-l} \bar{a}_s = \sum_{s=1}^{T+1} w^L_{s-1} \bar{a}^2_s = \mathrm{tr}\big(\mathbf{BB}^T\big)$$

where, $s_1 = \max\{1, s - T + L\}$, $s_2 = \min\{L, s\}$ and $w^L_s = \min\{s, L, T - s + 1\}$. □

Corollary 2.1.1. *Let* $\mathbf{A}$ *be an arbitrary* $L \times K$ *matrix and* $\mathbf{B}$ *be its corresponding Hankelized form. Then*:

$$\mathrm{tr}\big(\mathbf{AA}^T\big) \geqslant \mathrm{tr}\big(\mathbf{BB}^T\big). \tag{2}$$

Corollary 2.1.2. *Matrix* $\mathbf{B}$ *is the nearest matrix to* $\mathbf{A}$ *among all Hankel matrices of dimension* $L \times K$ *with respect to* $T^L_{\mathbf{A}}$.

2.3. Separability

The main concept in studying SSA properties is 'separability', which characterizes how well different components can be separated from each other. The following quantity (called the weighted correlation or *w-correlation*) is a natural measure of similarity between two series $Y_T^{(1)}$ and $Y_T^{(2)}$ [1]:

$$\rho_{12}^{(w)} = \frac{(Y_T^{(1)}, Y_T^{(2)})_w}{\|Y_T^{(1)}\|_w \|Y_T^{(2)}\|_w}$$

where $\|Y_T^{(i)}\|_w = \sqrt{(Y_T^{(i)}, Y_T^{(i)})_w}$, $(Y_T^{(i)}, Y_T^{(j)})_w = \sum_{p=1}^{T} w_p^L y_p^{(i)} y_p^{(j)}$ $(i, j = 1, 2)$, $w_p^L = \min\{p, L, T - p + 1\}$.

If the absolute value of the w-correlations is small, then the corresponding series are almost w-orthogonal, but, if it is large, then the two series are far from being w-orthogonal and are therefore weakly separable. Assume we only have two components signal and noise. Therefore, the value of w-correlations shows that how the reconstructed signal has been separated from the noise component. In the following we will show that the minimum value of w-correlations is obtained at $L = [\frac{T+1}{2}]$.

Theorem 2.2. *Let $\widetilde{S}_L^r$ be the reconstructed series based on the first r singular values of the trajectory matrix* $\mathbf{X}$. *Then*:

1. $\widetilde{S}_L^r = \widetilde{S}_K^r$,
2. $\widetilde{N}_L^r = \widetilde{N}_K^r$, *where* $\widetilde{N}_L^r = Y_T - \widetilde{S}_L^r$.

Proof. Below, we only provide the proof for the first equality. The second equality is easily obtained by the first equality. Recall that $\widetilde{S}_L^r$ and $\widetilde{S}_K^r$ are constructed by diagonal averaging of the matrices $\mathbf{X}_L^r$ and $\mathbf{X}_K^r$, respectively, where:

$$\mathbf{X}_L^r = \sum_{i=1}^{r} \sqrt{\lambda_i} U_i V_i^T, \qquad \mathbf{X}_K^r = \sum_{i=1}^{r} \sqrt{\lambda_i} V_i U_i^T.$$

Thus, the results can be obtained by equality $\mathbf{X}_L^r = (\mathbf{X}_K^r)^T$. The vectors $\widetilde{N}_L^r$ and $\widetilde{N}_K^r$ are usually called noise vector. The separability between $\widetilde{N}_L^r$ and $\widetilde{S}_L^r$ (or between $\widetilde{N}_K^r$ and $\widetilde{S}_K^r$) is a very important issue for i) reconstruction stage, and ii) for forecasting procedure. □

Corollary 2.2.1. *Let $\rho_{L,r}^w$ denote the w-correlation between $\widetilde{S}_L^r$ and $\widetilde{N}_L^r$. Then $\rho_{L,r}^w = \rho_{K,r}^w$. This confirms that we only need to consider $L \in \{2, \ldots, [\frac{T+1}{2}]\}$.*

Theorem 2.3. *Let* $\mathbf{X}$ *be the trajectory matrix defined as* (1) *and* $\mathbf{X} = \mathbf{S} + \mathbf{N} = \widetilde{\mathbf{S}} + \widetilde{\mathbf{N}}$ *where,* $\mathbf{S} = \sum_{i=1}^{r} \sqrt{\lambda_i} U_i V_i^T$, $\mathbf{N} = \sum_{j=r+1}^{L} \sqrt{\lambda_j} U_j V_j^T$, *and matrices* $\widetilde{\mathbf{S}}$ *and* $\widetilde{\mathbf{N}}$ *are Hankelized matrices of* $\mathbf{S}$ *and* $\mathbf{N}$, *respectively. Then* $T_{\widetilde{\mathbf{S}}\widetilde{\mathbf{N}}}^L = \operatorname{tr}(\widetilde{\mathbf{S}}\widetilde{\mathbf{N}}^T) > 0$.

Proof. The above definition of matrices $\mathbf{S}$ and $\mathbf{N}$, and also using the orthogonality feature of eigenvectors confirm that $\operatorname{tr}(\mathbf{S}\mathbf{N}^T) = 0$. Moreover, the equality

$$\operatorname{tr}\big((\mathbf{S} + \mathbf{N})(\mathbf{S} + \mathbf{N})^T\big) = \operatorname{tr}\big((\widetilde{\mathbf{S}} + \widetilde{\mathbf{N}})(\widetilde{\mathbf{S}} + \widetilde{\mathbf{N}})^T\big)$$

follows that:

$$\operatorname{tr}\big(\widetilde{\mathbf{S}}\widetilde{\mathbf{N}}^T\big) = \frac{[\operatorname{tr}(\mathbf{S}\mathbf{S}^T) - \operatorname{tr}(\widetilde{\mathbf{S}}\widetilde{\mathbf{S}}^T)] + [\operatorname{tr}(\mathbf{N}\mathbf{N}^T) - \operatorname{tr}(\widetilde{\mathbf{N}}\widetilde{\mathbf{N}}^T)]}{2}. \tag{3}$$

Applying Corollary 2.1.1 confirms that the right-hand side of Eq. (3) is non-negative which completes the proof. □

Notation: The superscript on matrices in the subsequent theorems stands for the number of rows of them.

Theorem 2.4. *$T_{\widetilde{\mathbf{S}}^L}^L$ is an increasing function of L on $L \in \{2, \ldots, [\frac{T+1}{2}]\}$, provided that there exists a Hankel matrix* $\mathbf{C}$ *of dimension $L \times K$ such that*:

$$T_{\mathbf{S}^L - \mathbf{C}}^L \leqslant T_{\mathbf{S}^L}^L - T_{\widetilde{\mathbf{S}}^{L-m}}^{L-m} \tag{4}$$

where $m \in \{1, \ldots, L-2\}$. Therefore, the maximum values of these functions are attained at $L = [\frac{T+1}{2}]$.

Proof. Below, we only consider the proof for $T_{\widetilde{\mathbf{S}}^L}^L$. Similar discussion and proof can be obtained for $T_{\widetilde{\mathbf{N}}^L}^L$. Recall from Theorem 2.1 and Corollary 2.1.2 that for every Hankel matrix $\mathbf{C}$ of dimension $L \times K$:

$$T_{\mathbf{S}^L}^L - T_{\widetilde{\mathbf{S}}^L}^L = T_{\mathbf{S}^L - \widetilde{\mathbf{S}}^L}^L \leqslant T_{\mathbf{S}^L - \mathbf{C}}^L. \tag{5}$$

Now, assume $\mathbf{C}$ is a Hankel matrix that satisfies Eq. (4). The proof is then completed using Eqs. (4) and (5). □

Theorem 2.5. *$T^L_{\mathbf{S}^L-\widetilde{\mathbf{S}}^L}$ is a decreasing function of L on $L \in \{2, \ldots, [\frac{T+1}{2}]\}$, provided that there exists a Hankel matrix $\mathbf{C}$ of dimension $L \times K$ such that:*

$$T^L_{\mathbf{S}^L-\mathbf{C}} \leqslant T^{L-m}_{\mathbf{S}^{L-m}} - T^{L-m}_{\widetilde{\mathbf{S}}^{L-m}} \tag{6}$$

where $m \in \{1, \ldots, L-2\}$. Therefore, the minimum value of this function is attained at $L = [\frac{T+1}{2}]$.

Proof. To proof this, we can employ the similar approach used in Theorem 2.4. □

Corollary 2.5.1. *The minimum value of w-correlation attains at $L = [\frac{T+1}{2}]$, provided that length of series is large enough and there exists a Hankel matrix $\mathbf{C}$ of dimension $L \times K$ such that inequality (6) is fulfilled.*

Proof. To proof this assertion, it is enough to show that $T^L_{\mathbf{S}^L}$ is an increasing function of L. For a large value of T, we have:

$$T^L_{\mathbf{S}^L} = \sum_{j=1}^{r} \lambda_j^{(L,T)} \geqslant \sum_{j=1}^{r} \lambda_j^{(L-1,T-1)} \approx T^{L-1}_{\mathbf{S}^{L-1}}. \tag{7}$$

Moreover, note that every Hankel matrix $\mathbf{C}$ of dimension $L \times K$ that satisfies inequality (6) also satisfies inequality (4). This completes the proof. □

Corollary 2.5.1 indicates that the reconstructed signal and noise using the first r eigen triples are almost w-orthogonal, if we choose $L = [\frac{T+1}{2}]$. Finding a Hankel matrix $\mathbf{C}$ that satisfies inequality (6) is not easy. However, we can find some equivalent conditions.

Theorem 2.6. *Let $\sigma_l^2(\mathbf{S}^L)$ is the lth secondary diagonal variance of the matrix $\mathbf{S}^L$. If $\sigma_l^2(\mathbf{S}^L) \leqslant \sigma_l^2(\mathbf{S}^{L-m})$, then Theorem 2.5 is satisfied and inequality (6) has infinite solutions with respect to $\mathbf{C}$.*

Proof. The first part of theorem is satisfied by using Theorems 2.1.1 and 2.1. For the second part, using inequality (6) we have:

$$\sum_{j=2}^{T+1} \sum_{i=s_1}^{s_2} \left(s^L_{i,j-i} - c_{j-1}\right)^2 \leqslant \sum_{j=2}^{T+1} \sum_{i=s_1}^{s_2} \left(s^{L-m}_{i,j-i} - \bar{s}^{L-m}_{j-1}\right)^2 = \sum_{j=2}^{T+1} \sigma^2_{j-1}\left(\mathbf{S}^{L-m}\right). \tag{8}$$

Now note that if the following inequality satisfies for $j = 2, \ldots, T+1$:

$$\sum_{i=s_1}^{s_2} \left(s^L_{i,j-i} - c_{j-1}\right)^2 - \sigma^2_{j-1}\left(\mathbf{S}^{L-m}\right) \leqslant 0, \tag{9}$$

then inequality (8) is fulfilled. But, the left-hand side of the inequality (9) is a quadratic form of c_{j-1} and has the following discriminant:

$$\Delta_{j-1} = \frac{\sigma^2_{j-1}(\mathbf{S}^{L-m}) - \sigma^2_{j-1}(\mathbf{S}^L)}{w^L_{j-1}} \geqslant 0. \tag{10}$$

Therefore, we can find infinite real c_{j-1} that is satisfied inequality (9) that completes the proof. □

Remark 1. Theorem 2.6 provides a sufficient condition as we know by Corollaries 2.1.1 and 2.1 that $T^L_{\mathbf{S}^L-\widetilde{\mathbf{S}}^L} \leqslant T^{L-m}_{\mathbf{S}^{L-m}-\widetilde{\mathbf{S}}^{L-m}}$ is equivalent to $\sum_{j=2}^{T+1} \sigma^2_{j-1}(\mathbf{S}^L) \leqslant \sum_{j=2}^{T+1} \sigma^2_{j-1}(\mathbf{S}^{L-m})$.

Acknowledgement

This research was supported by a grant from Shahid Beheshti University, Iran.

References

[1] N. Golyandina, V. Nekrutkin, A. Zhigljavsky, Analysis of Time Series Structure: SSA and Related Techniques, Chapman & Hall/CRC, New York, London, 2001.
[2] H. Hassani, Singular spectrum analysis: Methodology and comparison, Journal of Data Science 5 (2) (2007) 239–257.
[3] H. Hassani, S. Heravi, A. Zhigljavsky, Forecasting European industrial production with singular spectrum analysis, International Journal of Forecasting 25 (1) (2009) 103–118.
[4] A. Zhigljavsky, Singular spectrum analysis for time series: Introduction to this special issue, Statistics and Its Interface 3 (3) (2010) 255–258.
[5] 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.
[6] J.B. Elsner, A.A. Tsonis, Singular Spectrum Analysis: A New Tool in Time Series Analysis, Plenum, 1996.