

Variations of singular spectrum analysis for separability improvement: non-orthogonal decompositions of time series

NINA GOLYANDINA, ALEX SHLEMOV

Singular spectrum analysis (SSA) as a nonparametric tool for decomposition of an observed time series into sum of interpretable components such as trend, oscillations and noise is considered. The separability of these series components by SSA means the possibility of such decomposition. Two extensions of SSA, which weaken the separability conditions, are proposed. One of the proposed approaches considers inner products corresponding to oblique coordinate systems instead of the conventional Euclidean inner product. The other method changes contributions of the components by involving the series derivative to avoid component mixing. Performance of the suggested methods is demonstrated on simulated and real-life data.

AMS 2000 SUBJECT CLASSIFICATIONS: Primary 62M20, 62F10, 62F12; secondary 60G35, 65C20, 62G05.

KEYWORDS AND PHRASES: singular spectrum analysis, time series, time series decomposition, separability.

1. INTRODUCTION

Singular spectrum analysis [4, 6, 7, 11, 13, 33, 34] is a powerful method of time series analysis, which does not require a parametric model of the time series given in advance and therefore SSA is very well suitable for exploratory analysis. After an exploratory analysis has been performed, SSA enables to construct series models.

Singular spectrum analysis can solve very different problems in time series analysis which range from the series decomposition on interpretable series components to forecasting, missing data imputation, parameter estimation and many others. The main problem is the proper decomposition of the time series. For example, if one forecasts trend, then this trend should be extracted properly. For seasonal adjustment, the seasonality should be extracted correctly, and so on.

In [11, 24], the separability theory, which is responsible for the proper decomposition and proper component extraction, was developed. The separability of components means that the method is able to extract the time series components from the observed series that is the sum of many components. At the present time, there is a lot of publications

with theory of separability and applications where separability is important, see [2, 3, 8, 9, 15, 16, 18, 19, 25, 27, 29] among others.

For reasonable time series lengths and noise levels, trends, oscillations and noise are approximately separable by SSA [11, Sections 1.5 and 6.1]. However, the conditions of approximate separability can be restrictive, especially, for short time series.

The separability conditions are closely related to the properties of the singular value decomposition (SVD), which is the essential part of many statistical and signal processing methods: principal component analysis [20], low-rank approximations [22], different subspace-based methods [31] including singular spectrum analysis among many others. The main advantage of the SVD is its optimality features and bi-orthogonality; the drawback for approximation problems is the non-uniqueness of the SVD expansion if there are coinciding singular values.

In subspace-based methods, the SVD is applied to a trajectory matrix with rows and columns consisting of subseries of the initial series. In SSA, the obtained SVD components are grouped and the grouped matrices are transferred back to the series. Thus, we obtain a decomposition of the initial time series \mathbb{X} into a sum of series components, e.g., $\mathbb{X} = \tilde{\mathbb{S}} + \tilde{\mathbb{R}}$. If we deal with a series $\mathbb{X} = \mathbb{S} + \mathbb{R}$ containing two series components \mathbb{S} and \mathbb{R} , which we want to find, then (approximate) weak separability is by definition (approximate) orthogonality of subseries of \mathbb{S} and \mathbb{R} , which provides, due to the SVD bi-orthogonality, the existence of such a grouping that $\tilde{\mathbb{S}}$ and $\tilde{\mathbb{R}}$ are (approximately) equal to \mathbb{S} and \mathbb{R} correspondingly.

Non-uniqueness of the SVD in the case of coinciding singular values implies the condition of disjoint sets of singular values in the groups corresponding to different series components to avoid their possible mixture. This condition is necessary to obtain the so called strong separability, when any SVD of the trajectory matrix provides the proper grouping. In practice, the strong separability is needed (see for more details Section 2.2.1) and both conditions, orthogonality of component subseries and disjoint sets of singular values of component trajectory matrices, should be fulfilled.

The paper presents two methods, Iterative O-SSA and DerivSSA, which help to weaken the separability conditions

in SSA. For simplicity, we describe the methods for separation of two series components; separation of several components can be considered in analogous manner.

Orthogonality of subseries can be a strong limitation on the separated series. However, if we consider orthogonality with respect to non-standard Euclidean inner product, conditions of separability can be considerably weakened. This yields the first method called Oblique SSA (O-SSA) with the SVD step performed in a non-orthogonal coordinate system. The idea of Iterative Oblique SSA is similar to prewhitening that is frequently used in statistics as preprocessing: if we know covariances between components, then we can perform linear transformation and obtain uncorrelated components. Since the ‘covariances’ of the components are not known in advance, the iterative algorithm called Iterative Oblique SSA is suggested. Contribution of the components can be changed in a specific way during the iterations to improve separability.

The second method called DerivSSA helps to change the component contributions with no change of the structure of the separated series components. The approach consists in consideration of the series derivative together with the series itself. For example, two singular values produced by a sinusoid are determined by its amplitude. The derivative of a sine wave has the same frequency and changed amplitude, depending on frequency: $f(x) = \sin(2\pi\omega x + \phi)$ has amplitude 1, while its derivative has amplitude $2\pi\omega$. This is just a simple example; the method works with non-stationary series, not only with sinusoids. The use of derivatives helps to overcome the problem when the approximate orthogonality holds but the series components mix due to equal contributions. It seems that this approach is simpler and more general than the SSA-ICA (SSA with Independent Component Analysis) approach considered in [13, Section 2.5.4].

Since both suggested methods do not have approximating features, they cannot replace Basic SSA and therefore should be used in a nested manner. This means that Basic SSA extracts mixing series components (e.g. first we use Basic SSA for denoising) and then one of the suggested methods separates the mixing components. Let us demonstrate the nested use of extensions by an example. Let $\mathbb{X} = (x_1, \dots, x_N)$ be the series of length N , $\mathbb{X} = \mathbb{X}^{(1)} + \mathbb{X}^{(2)} + \mathbb{X}^{(3)}$. The result of Basic SSA is $\mathbb{X} = \tilde{\mathbb{X}}^{(1,2)} + \tilde{\mathbb{X}}^{(3)}$, the result of the considered extension is $\tilde{\mathbb{X}}^{(1,2)} = \tilde{\mathbb{X}}^{(1)} + \tilde{\mathbb{X}}^{(2)}$ and the final result is $\mathbb{X} = \tilde{\mathbb{X}}^{(1)} + \tilde{\mathbb{X}}^{(2)} + \tilde{\mathbb{X}}^{(3)}$.

The paper is organized as follows. We start with a short description of the algorithm of Basic SSA and standard separability notion (Section 2). The next two sections 3 and 4 are devoted to the extensions of singular spectrum analysis. In Section 3, Oblique SSA is considered. In Section 4, SSA involving series derivatives is investigated. Each section contains numerical examples of algorithm application. In Section 5, both methods are applied to real-life time series. The implementation of the algorithms is performed with the help of the RSSA package in R, see [10]. The implementation of

SSA in RSSA is very effective [21] and the considered extensions of SSA can be also efficiently implemented. Conclusions are contained in Section 6. Since the extensions are based on the use of inner products and decompositions in oblique coordinate systems, we put the necessary definitions and statements into Appendix A.

2. BASIC SSA

2.1 Algorithm

Consider a real-valued time series $\mathbb{X} = \mathbb{X}_N = (x_1, \dots, x_N)$ of length N . Let L ($1 < L < N$) be some integer called *window length* and $K = N - L + 1$.

For convenience, denote $\mathcal{M}_{L,K}$ the space of matrices of size $L \times K$, $\mathcal{M}_{L,K}^{(H)}$ the space of Hankel matrices of size $L \times K$, $X_i = (x_i, \dots, x_{i+L-1})^T$, $i = 1, \dots, K$, the L -lagged vectors and $\mathbf{X} = [X_1 : \dots : X_K]$ the L -trajectory matrix of the series \mathbb{X}_N . Define the embedding operator $\mathcal{T} : \mathbb{R}^N \mapsto \mathcal{M}_{L,K}$ as $\mathcal{T}(\mathbb{X}_N) = \mathbf{X}$.

Also introduce the projector \mathcal{H} (in the Frobenius norm) of $\mathcal{M}_{L,K}$ to $\mathcal{M}_{L,K}^{(H)}$, which performs the projection by the change of entries on auxiliary diagonals $i + j = \text{const}$ to their averages along the diagonals.

The algorithm of Basic SSA consists of four steps.

1st step: Embedding. Choose L . Construct the L -trajectory matrix: $\mathbf{X} = \mathcal{T}(\mathbb{X}_N)$.

2nd step: Singular value decomposition (SVD). Consider the SVD of the trajectory matrix:

$$(1) \quad \mathbf{X} = \sum_{i=1}^d \sqrt{\lambda_i} U_i V_i^T = \mathbf{X}_1 + \dots + \mathbf{X}_d,$$

where $\sqrt{\lambda_i}$ are singular values, U_i and V_i are the left and right singular vectors of \mathbf{X} , $\lambda_1 \geq \dots \geq \lambda_d > 0$, $d = \text{rank}(\mathbf{X})$. The number d is called L -rank of the series \mathbb{X} .

The triple $(\sqrt{\lambda_i}, U_i, V_i)$ is called i th *eigentriple* (abbreviated as ET).

3rd step: Eigentriple grouping. The grouping procedure partitions the set of indices $\{1, \dots, d\}$ into m disjoint subsets I_1, \dots, I_p . This step is less formal. However, there are different recommendations on grouping related to separability issues briefly described in Section 2.2.1.

Define $\mathbf{X}_I = \sum_{i \in I} \mathbf{X}_i$. The expansion (1) leads to the decomposition

$$(2) \quad \mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_p}.$$

If $p = d$ and $I_j = \{j\}$, $j = 1, \dots, d$, then the corresponding grouping is called *elementary*.

4th step: Diagonal averaging. Obtain the series by diagonal averaging of the matrix components of (2): $\tilde{\mathbb{X}}_N^{(k)} = \mathcal{T}^{-1} \mathcal{H} \mathbf{X}_{I_k}$.

Thus, the algorithm results in the decomposition of the observed time series

$$(3) \quad \mathbb{X}_N = \sum_{k=1}^p \tilde{\mathbb{X}}_N^{(k)}.$$

The reconstructed components produced by the elementary grouping will be called *elementary reconstructed series*.

2.2 Separability by Basic SSA

Notion of separability is very important to understand how SSA works. Separability of two time series $\mathbb{X}_N^{(1)}$ and $\mathbb{X}_N^{(2)}$ signifies the possibility of extracting $\mathbb{X}_N^{(1)}$ from the observed series $\mathbb{X}_N = \mathbb{X}_N^{(1)} + \mathbb{X}_N^{(2)}$. This means that there exists a grouping at Grouping step such that $\tilde{\mathbb{X}}_N^{(m)} = \mathbb{X}_N^{(m)}$.

Let us define the separability formally. Let $\mathbf{X}^{(m)}$ be the trajectory matrices of the considered series, $\mathbf{X}^{(m)} = \sum_{i=1}^{d_m} \sqrt{\lambda_{m,i}} U_{m,i} V_{m,i}^T$, $m = 1, 2$, be their SVDs. The column and row spaces of the trajectory matrices are called *column* and *row trajectory spaces* correspondingly.

Definition 1. Let L be fixed. Two series $\mathbb{X}_N^{(1)}$ and $\mathbb{X}_N^{(2)}$ are called weakly separable, if their column trajectory spaces are orthogonal and the same is valid for their row trajectory spaces, that is, $(\mathbf{X}^{(1)})^T \mathbf{X}^{(2)} = \mathbf{0}_{\mathbf{K}, \mathbf{K}}$ and $\mathbf{X}^{(1)} (\mathbf{X}^{(2)})^T = \mathbf{0}_{\mathbf{L}, \mathbf{L}}$.

Definition 2. Two series $\mathbb{X}_N^{(1)}$ and $\mathbb{X}_N^{(2)}$ are called strongly separable, if they are weakly separable and the sets of singular values of their L -trajectory matrices are disjoint, that is, $\lambda_{1,i} \neq \lambda_{2,j}$ for any i and j .

By definition, separability means orthogonality of the column and row spaces of the trajectory matrices of the series components $\mathbb{X}_N^{(1)}$ and $\mathbb{X}_N^{(2)}$. For approximate (asymptotic) separability with $\tilde{\mathbb{X}}_N^{(m)} \approx \mathbb{X}_N^{(m)}$ we need the condition of approximate (asymptotic) orthogonality of subseries of the considered components. Asymptotic separability is considered as $L, K, N \rightarrow \infty$.

For sufficiently long time series, SSA can approximately separate, for example, signal and noise, sine waves with different frequencies, trend and seasonality [11, 13].

Let us demonstrate the separability of two sinusoids with frequencies ω_1 and ω_2 : $x_n^{(i)} = A_i \cos(2\pi\omega_i n + \phi_i)$. These sinusoids are asymptotically separable, that is, their subseries are asymptotically orthogonal as their length tends to infinity. However, the rate of convergence depends on the difference between the frequencies. If they are close and the time series length is not long enough, the series can be far from orthogonal and therefore not separable.

Weak separability means that at SVD step there exists such an SVD that admits the proper grouping. The problem of possibility of a non-separating SVD expansion is related to non-uniqueness of the SVD in the case of equal singular values. Strong separability means that any SVD of the

series trajectory matrix admits the proper grouping. Therefore, we need strong (approximate) separability for the use in practice. For example, two sinusoids with equal amplitudes are asymptotically weakly separated, but asymptotically not strongly separated and therefore are mixed in the decomposition.

2.2.1 Separability measure

Very helpful information for detection of separability and group identification is contained in the so-called \mathbf{w} -correlation matrix. This matrix consists of weighted cosines of angles between the reconstructed time series components. The weights reflect the number of entries of the time series terms into its trajectory matrix.

Let $w_n = \#\{(i, j) : 1 \leq i \leq L, 1 \leq j \leq K, i + j = n + 1\}$. Define the \mathbf{w} -scalar product of time series of length N as $(\mathbb{Y}_N, \mathbb{Z}_N)_{\mathbf{w}} = \sum_{n=1}^N w_n y_n z_n = \langle \mathbf{Y}, \mathbf{Z} \rangle_{\mathbf{F}}$. Then

$$\rho_{\mathbf{w}}(\mathbb{Y}_N, \mathbb{Z}_N) = (\mathbb{Y}_N, \mathbb{Z}_N)_{\mathbf{w}} / (\|\mathbb{Y}_N\|_{\mathbf{w}} \|\mathbb{Z}_N\|_{\mathbf{w}}).$$

Well separated components in (3) have small correlation whereas poorly separated components generally have large correlation. Therefore, looking at the matrix of \mathbf{w} -correlations between elementary reconstructed series $\tilde{\mathbb{X}}_N^{(i)}$ and $\tilde{\mathbb{X}}_N^{(j)}$ one can find groups of correlated series components and use this information for the consequent grouping. One of the rules is not to include the correlated components into different groups. Also, \mathbf{w} -correlations can be used for checking the grouped decomposition.

It is convenient to depict in absolute magnitude the matrix of \mathbf{w} -correlations between the series components graphically in the white-black scale, where small correlations are shown in white, while correlations with moduli close to 1 are shown in black.

2.2.2 Scheme of Basic SSA application

Let us briefly describe the general scheme of Basic SSA application, thoroughly described in [11, 13]:

- Choice of window length L in accordance with a-priori recommendations (see, in addition, [9]).
- Execution of Embedding and Decomposition steps.
- Analysis of the eigentriples and the \mathbf{w} -correlation matrix to perform grouping of eigentriples. The main principle is: eigenvectors repeat the form of a series component that produces these eigentriples. \mathbf{w} -Correlations also provide a guess for proper grouping.
- Execution of Grouping and Reconstruction steps to obtain the desired series decomposition.
- If separability does not take place for the given L and the obtained decomposition is not appropriate, then the change of the window length L is recommended.

Note that the proper grouping to obtain a suitable series decomposition can be impossible if the signal components (described, as a rule, by a number of leading SVD

components) are mixed. For example, if a signal eigenvector contains both periodic and slowly varying components, this means that the trend and periodic components are not separable, at least for the chosen window length L . If we see the mixture of two sine-waves with different frequencies, this means that these sine-waves are not separable for this L .

If it appears that for the chosen L there is no separability (weak or strong), the attempt to obtain separability is performed with other choices of L . For example, a possible lack of strong separability between a trend of complex form and a seasonality can be overcome by means of the use of small window lengths. However, weak separability can be weakened by this trick and Sequential SSA should be used to obtain an accurate decomposition of the residual after the trend extraction.

For the majority of time series, SSA with a proper choice of window length is able to separate series components and to obtain a desirable series decomposition. However, sometimes Basic SSA cannot separate components, e.g. short sine wave series with close frequencies or sine waves with equal amplitudes.

2.2.3 Identification of separated sinusoids

Separation of sine-wave components is of special interest. Each sine-wave component generates two elementary series components, which have correlation close to 1. If a sinusoid is separated from the residual, maybe, approximately, then two elementary components produced by it are almost not correlated with the other elementary components and therefore we will see a black square 2×2 on the \mathbf{w} -correlation matrix of elementary components.

To find two SVD components corresponding to a sine-wave, scatterplots of eigenvectors (which are approximately sine and cosine) can be also used. If the period value is integer, the scatterplot of sine vs cosine looks like a regular polygon, where the number of vertices is equal to the period.

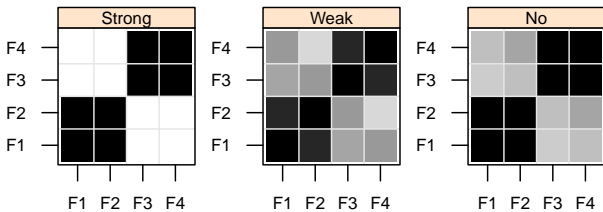


Figure 1. Sum of two sinusoids: \mathbf{w} -correlation matrices for different types of separability

For example, consider the series \mathbb{X}_N , where $x_n = x_n^{(1)} + x_n^{(2)}$, $x_n^{(k)} = A_k \sin(2\pi\omega_k n)$, the series length $N = 119$, with three different sets of parameters:

(A) ‘strong separability’, $A_1 = 2$, $A_2 = 1$, $\omega_1 = 1/12$, $\omega_2 = 1/10$;

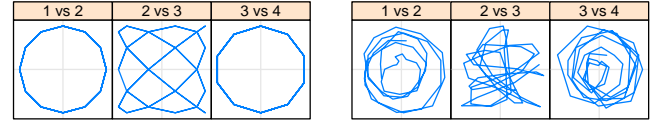


Figure 2. Sum of two sinusoids: scatterplots of eigenvectors with good (left) and bad (right) separability

(B) ‘weak separability, no strong separability’, $A_1 = A_2 = 1$, $\omega_1 = 1/12$, $\omega_2 = 1/10$;

(C) ‘no weak separability’, $A_1 = 2$, $A_2 = 1$, $\omega_1 = 1/12$, $\omega_2 = 1/13$, the series is corrupted by Gaussian white noise with standard deviation 4.

The difference between good and bad separability is clearly seen in Fig. 1 and 2. One can see that the matrices of weighted correlations for the examples (B) and (C) are very similar, although in general weighted correlations for the example (B) can be arbitrary. Figure 2 shows the scatterplots of eigenvectors for the examples (A) and (C). The pairs of eigenvectors produced by exactly separated sinusoids form regular polygons.

2.3 Series of finite rank and series governed by linear recurrence relations

Let us describe the class of series of finite rank, which is natural for SSA. Note that only such time series can be exactly separated and exactly continued by SSA [11, Section 2.2 and Chapter 5].

We define L -rank of a series \mathbb{X}_N as the rank of its L -trajectory matrix. Series with rank-deficient trajectory matrices are of special interest. A time series is called *time series of finite rank r* if its L -trajectory matrix has rank r for any $L \geq r$ (it is convenient to assume that $L \leq K$).

Under some not restrictive conditions, a series \mathbb{S}_N of finite rank r is governed by a linear recurrence relation (LRR) of order r , that is

$$(4) \quad s_{i+r} = \sum_{k=1}^r a_k s_{i+r-k}, \quad 1 \leq i \leq N-r, \quad a_r \neq 0.$$

The LRR (4) is called minimal, since it is unique and has minimal order among LRRs governing \mathbb{S}_N . Let us describe how we can restore the form of the time series by means of the minimal LRR.

Definition 3. A polynomial $P_r(\mu) = \mu^r - \sum_{k=1}^r a_k \mu^{r-k}$ is called a characteristic polynomial of the LRR (4).

Let the time series $\mathbb{S}_\infty = (s_1, \dots, s_n, \dots)$ satisfy the LRR (4) for $i \geq 1$. Consider the characteristic polynomial of the LRR (4) and denote its different (complex) roots by μ_1, \dots, μ_p , where $p \leq r$. All these roots are non-zero, since

$a_r \neq 0$. Let the multiplicity of the root μ_m be k_m , where $1 \leq m \leq p$ and $k_1 + \dots + k_p = r$. We will call the set $\{\mu_j\}_{m=1}^p$ *characteristic (or signal) roots* of the series governed by an LRR. Note that in the framework of SSA non-minimal LRRs, which have so called extraneous roots in addition to the signal ones, are considered and the extraneous roots are studied ([30]); however, here we will deal only with characteristic roots to describe the signal model.

It is well-known that the time series $S_\infty = (s_1, \dots, s_n, \dots)$ satisfies the LRR (4) for all $i \geq 0$ if and only if

$$(5) \quad s_n = \sum_{m=1}^p \left(\sum_{j=0}^{k_m-1} c_{m,j} n^j \right) \mu_m^n.$$

for some $c_{m,j} \in \mathbb{C}$. For real-valued time series, (5) implies that the class of time series governed by LRRs consists of sums of products of polynomials, exponentials and sinusoids.

The important advantage of SSA is that although the model (5) of signals is involved in theoretical results, the SSA algorithm does not perform explicit estimation of the model parameters for reconstruction and forecasting. This provides the possibility to deal with signals that are locally approximated by the model; in particular, to extract slowly-varying trends and modulated sine waves. The indicated feature of the SSA approach holds for the extensions considered below.

3. OBLIQUE SSA

Although many interpretable series components like trend (a slowly varying component) and seasonality are asymptotically orthogonal, for the given time series length the orthogonality can be not reached even approximately. Therefore, it would be helpful to weaken the orthogonality condition. The suggested approach consists in using an orthogonality, which still means the equality of an inner product to 0, but this is a non-ordinary inner product which is adapted to time series components, which we want to separate.

It is well-known that any inner product in Euclidean space is associated with a symmetric positive-definite matrix \mathbf{A} and is defined as $\langle X_1, X_2 \rangle_{\mathbf{A}} = (\mathbf{A}X_1, X_2)$. The standard inner product is given by the identity matrix. Inner product implies \mathbf{A} -orthogonality of the vectors if $\langle X_1, X_2 \rangle_{\mathbf{A}} = 0$. If the matrix \mathbf{A} is semi-definite, then it produces the inner product given in its column (or row, it is the same due to symmetry) space. Below, considering $\langle X_1, X_2 \rangle_{\mathbf{A}}$, we will always assume that the vectors X_i , $i = 1, 2$, belong to the column space of \mathbf{A} .

Thus, non-standard Euclidean inner products induce such notions as oblique coordinate systems, orthogonality of vectors, which are oblique in ordinary sense, and so on.

Let us present an elementary example. Let $X = (1, 2)^T$ and $Y = (1, 1)^T$. Certainly, these vectors are not orthogonal

in the usual sense: $(X, Y) = 3$. However, if we define

$$(6) \quad \mathbf{A} = \begin{pmatrix} 5 & -3 \\ -3 & 2 \end{pmatrix},$$

then $\langle X, Y \rangle_{\mathbf{A}} = (\mathbf{A}X, Y) = 0$ and $(\mathbf{O}_{\mathbf{A}}X, \mathbf{O}_{\mathbf{A}}Y) = 0$ for any $\mathbf{O}_{\mathbf{A}}$ such that $\mathbf{O}_{\mathbf{A}}^T \mathbf{O}_{\mathbf{A}} = \mathbf{A}$, e.g.

$$\mathbf{O}_{\mathbf{A}} = \begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix}.$$

This means that $\{X, Y\}$ is an orthogonal basis with respect to the \mathbf{A} -inner product $\langle \cdot, \cdot \rangle_{\mathbf{A}}$ and $\mathbf{O}_{\mathbf{A}}$ corresponds to an orthogonalizing map. The matrix \mathbf{A} can be chosen such that X and Y have any \mathbf{A} -norm. The choice (6) corresponds to \mathbf{A} -orthonormality.

To describe a so called Oblique SSA, let us introduce the SVD of a matrix \mathbf{X} produced by two oblique bases, \mathbf{L} -orthonormal and \mathbf{R} -orthonormal correspondingly, in the row and column spaces (Definition 7). We say that $\mathbf{X} = \sum_{i=1}^d \sigma_i P_i Q_i^T$ is the (\mathbf{L}, \mathbf{R}) -SVD, if $\{P_i\}_{i=1}^d$ is an \mathbf{L} -orthonormal system and $\{Q_i\}_{i=1}^d$ is an \mathbf{R} -orthonormal system, that is, the decomposition is (\mathbf{L}, \mathbf{R}) -biorthogonal. This kind of SVD is called Restricted SVD (RSVD) given by the triple $(\mathbf{X}, \mathbf{L}, \mathbf{R})$, see [23] for details. Mathematics related to inner products $\langle \cdot, \cdot \rangle_{\mathbf{A}}$ with positive-semidefinite matrix \mathbf{A} and the corresponding RSVD is shortly described in Appendix A from the viewpoint of decompositions into a sum of elementary matrices. We formulate the necessary definitions and propositions in a convenient form to make the suggested algorithms clearer.

Oblique SSA (O-SSA) is the modification of the Basic SSA algorithm described in Section 2, where the SVD step is changed by the (\mathbf{L}, \mathbf{R}) -SVD for some matrices \mathbf{L} and \mathbf{R} consistent with \mathbf{X} (see Definition 6). We will use the notions introduced in the algorithm of Basic SSA also for its oblique modification.

Proposition 4 provides the algorithm which reduces the (\mathbf{L}, \mathbf{R}) -SVD to the ordinary SVD.

Algorithm 1. (\mathbf{L}, \mathbf{R}) -SVD.

Input: \mathbf{Y} , (\mathbf{L}, \mathbf{R}) consistent with \mathbf{Y} .

Output: The (\mathbf{L}, \mathbf{R}) -SVD in the form (8).

1. Calculate $\mathbf{O}_{\mathbf{L}}$ and $\mathbf{O}_{\mathbf{R}}$, e.g., by Cholesky decomposition.
2. Calculate $\mathbf{O}_{\mathbf{L}} \mathbf{Y} \mathbf{O}_{\mathbf{R}}^T$.
3. Find the ordinary SVD decomposition (10).
4. $\sigma_i = \sqrt{\lambda_i}$, $P_i = \mathbf{O}_{\mathbf{L}}^\dagger U_i$ and $Q_i = \mathbf{O}_{\mathbf{R}}^\dagger V_i$, where \dagger denotes pseudo-inverse.

Note that if \mathbf{L} and \mathbf{R} are the identity matrices, then Oblique SSA coincides with Basic SSA, $\sigma_i = \sqrt{\lambda_i}$, $P_i = U_i$ and $Q_i = V_i$.

3.1 Separability

The notion of weak and strong (\mathbf{L}, \mathbf{R}) -separability, which is similar to conventional separability described in Section 2.2.1, can be introduced. Again, let $\mathbb{X} = \mathbb{X}^{(1)} + \mathbb{X}^{(2)}$, \mathbf{X} be its trajectory matrix, $\mathbf{X}^{(m)}$ be the trajectory matrices of the series components, $\mathbf{X}^{(m)} = \sum_{i=1}^{r_m} \sigma_{m,i} P_{m,i} Q_{m,i}^T$ be their (\mathbf{L}, \mathbf{R}) -SVDs, $m = 1, 2$. We assume that \mathbf{L} and \mathbf{R} are consistent with \mathbf{X} , $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$.

Definition 4. Let L be fixed. Two series $\mathbb{X}_N^{(1)}$ and $\mathbb{X}_N^{(2)}$ are called weakly (\mathbf{L}, \mathbf{R}) -separable, if their column trajectory spaces are \mathbf{L} -orthogonal and their row trajectory spaces are \mathbf{R} -orthogonal, that is, $(\mathbf{X}^{(1)})^T \mathbf{L} \mathbf{X}^{(2)} = \mathbf{0}_{\mathbf{K}, \mathbf{K}}$ and $\mathbf{X}^{(1)} \mathbf{R} (\mathbf{X}^{(2)})^T = \mathbf{0}_{\mathbf{L}, \mathbf{L}}$.

Definition 5. Two series $\mathbb{X}_N^{(1)}$ and $\mathbb{X}_N^{(2)}$ are called strongly (\mathbf{L}, \mathbf{R}) -separable, if they are weakly (\mathbf{L}, \mathbf{R}) -separable and $\sigma_{1,i} \neq \sigma_{2,j}$ for any i and j .

The (\mathbf{L}, \mathbf{R}) -separability of two series components means \mathbf{L} -orthogonality of their subseries of length L and \mathbf{R} -orthogonality of the subseries of length $K = N - L + 1$.

The following theorem shows that the (\mathbf{L}, \mathbf{R}) -separability is in a sense much less restrictive than the ordinary one.

Theorem 1. Let $\mathbb{X} = \mathbb{X}^{(1)} + \mathbb{X}^{(2)}$ be the series of length N , L be the window length and the L -rank of \mathbb{X} be equal to r . Let $\mathbb{X}^{(m)}$ be the series of L -rank r_m , $m = 1, 2$, $r_1 + r_2 = r$. Then there exist separating matrices $\mathbf{L} \in \mathcal{M}_{L,L}$ and $\mathbf{R} \in \mathcal{M}_{K,K}$ of rank r such that the series $\mathbb{X}^{(1)}$ and $\mathbb{X}^{(2)}$ are strongly (\mathbf{L}, \mathbf{R}) -separable.

Proof. Denote $\{P_i^{(m)}\}_{i=1}^{r_m}$ a basis of the column space of $\mathbf{X}^{(m)}$ and $\{Q_i^{(m)}\}_{i=1}^{r_m}$ a basis of the row space of $\mathbf{X}^{(m)}$, $m = 1, 2$; e.g., $P_i^{(m)} = P_{m,i} \in \mathbb{R}^L$, $Q_i^{(m)} = Q_{m,i} \in \mathbb{R}^K$. Define

$$\begin{aligned} \mathbf{P} &= [P_1^{(1)} : \dots : P_{r_1}^{(1)} : P_1^{(2)} : \dots : P_{r_2}^{(2)}], \\ \mathbf{Q} &= [Q_1^{(1)} : \dots : Q_{r_1}^{(1)} : Q_1^{(2)} : \dots : Q_{r_2}^{(2)}]. \end{aligned}$$

By the theorem conditions, the matrices \mathbf{P} and \mathbf{Q} are of full rank. Since \mathbf{P}^\dagger and \mathbf{Q}^\dagger orthonormalize the columns of the matrices \mathbf{P} and \mathbf{Q} (Proposition 2), then the trajectory matrices $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are (\mathbf{L}, \mathbf{R}) bi-orthogonal for $\mathbf{L} = (\mathbf{P}^\dagger)^T \mathbf{P}^\dagger$ and $\mathbf{R} = (\mathbf{Q}^\dagger)^T \mathbf{Q}^\dagger$. Therefore the series $\mathbb{X}^{(1)}$ and $\mathbb{X}^{(2)}$ are (\mathbf{L}, \mathbf{R}) -separable.

Proposition 5 shows that it is possible to change $\sigma_{m,i}$ keeping bi-orthogonality, that is, it explains how to get strong separability not corrupting weak one. \square

Remark 1. Consider two time series governed by minimal LRRs of orders r_1 and r_2 , $r_1 + r_2 \leq \min(L, K)$. The conditions of Theorem 1 fulfill if and only if the sets of characteristic roots of the series are disjoint. Really, the sets of characteristic roots are disjoint if and only if the column and row spaces of L -trajectory matrices intersect only in $\{0\}$, that is, \mathbf{P} and \mathbf{Q} are of full rank.

Remark 2. Theorem 1 together with Remark 1 shows that any two times series governed by LRRs with different characteristic roots can be separated by some (\mathbf{L}, \mathbf{R}) -SVD for sufficiently large series and window lengths.

Note that Theorem 1 is not constructive, since the trajectory spaces of the separated series should be known for exact separation. However, we can try to estimate these spaces and thereby to improve the separability.

Measures of oblique separability. If Oblique SSA does not separate the components exactly, a measure of separability is necessary. We can consider the analogue of \mathbf{w} -correlations described in Section 2.2.1, since they are defined through the Frobenius inner products of trajectory matrices and therefore can be generalized; see Appendix A.3 for definition of $\rho_{\mathbf{L}, \mathbf{R}}$ in (12). Define (\mathbf{L}, \mathbf{R}) \mathbf{w} -correlation between the reconstructed series $\tilde{\mathbb{X}}^{(1)}$ and $\tilde{\mathbb{X}}^{(2)}$ as $\rho_{(\mathbf{L}, \mathbf{R})}(\tilde{\mathbb{X}}^{(1)}, \tilde{\mathbb{X}}^{(2)})$. Note that due to diagonal averaging, the column and row spaces of $\tilde{\mathbf{X}}^{(m)}$ do not necessarily belong to the column spaces of \mathbf{L} and \mathbf{R} correspondingly, that is, matrices \mathbf{L} and \mathbf{R} can be not consistent with $\tilde{\mathbf{X}}^{(m)}$, $m = 1, 2$. Therefore, $\rho_{\mathbf{L}, \mathbf{R}}$ takes into consideration only projections of columns and rows of $\tilde{\mathbf{X}}^{(1)}$ and $\tilde{\mathbf{X}}^{(2)}$ on the column spaces of \mathbf{L} and \mathbf{R} (Remark 7). This means that $\rho_{\mathbf{L}, \mathbf{R}}$ can overestimate the separability accuracy.

For Oblique SSA, when only one of coordinate systems (left or right) is oblique, the conventional \mathbf{w} -correlations between series are more appropriate measures of separability, since in the case of exact oblique separability we have orthogonal (in the Frobenius inner product) matrix components (Corollary 1).

Other important measure of proper separability is the closeness of the reconstructed series components to time series of finite rank. This can be measured by the contribution of the leading $r_m = |I_m|$ eigentriples into the SVD of the trajectory matrix $\tilde{\mathbf{X}}^{(m)}$ of the m th reconstructed series component $\tilde{\mathbb{X}}^{(m)}$. If we denote $\tilde{\lambda}_{m,i}$ the eigenvalues of the ordinary SVD of $\tilde{\mathbf{X}}^{(m)}$, then $\tau_{r_m}(\tilde{\mathbf{X}}^{(m)}) = 1 - \sum_{i=1}^{r_m} \tilde{\lambda}_{m,i} / \|\tilde{\mathbf{X}}^{(m)}\|^2$ reflects the closeness of the m th series to the series of rank r_m .

3.2 Nested Oblique SSA

Rather than the ordinary SVD, the SVD with respect to non-orthogonal coordinate systems provides approximation in an inappropriate way. That is why Oblique SSA cannot be used for extraction of the leading components, in particular, for extraction of the signal and for denoising.

Therefore, the nested way of using Oblique SSA is suggested. The approach is somewhat similar to factor analysis, where a factor space can be estimated by principal component analysis and then interpretable factors are extracted from the factor space.

Suppose that Basic SSA can extract the signal but cannot separate the signal components. For example, let the time series consist of a noisy sum of two sinusoids. Then Basic

SSA can perform denoising but probably cannot separate these sinusoids, if their frequencies are close. Thus, Basic SSA is used for estimation of the subspace of the sum of sinusoids and then some other method can be used to separate the sinusoids themselves. The choice of parameters for better separation is thoroughly investigated in [9]. Note that the nested approach is similar to the refined grouping used in [13, Section 2.5.4] for the SSA-ICA algorithm.

Thus, let us apply Basic SSA with proper parameters and let a matrix decomposition $\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_p}$ be obtained at Grouping step of Basic SSA; each group corresponds to a separated time series component. Let the s th group $I = I_s$ be chosen for a refined decomposition. Denote $\mathbf{Y} = \mathbf{X}_I$, $r = \text{rank } \mathbf{Y}$, $\mathbb{Y} = \mathcal{T}^{-1}\mathcal{H}\mathbf{Y}$ the series obtained from \mathbf{Y} by diagonal averaging.

Algorithm 2. (Nested Oblique SSA.)

Input: The matrix \mathbf{Y} , matrices (\mathbf{L}, \mathbf{R}) , which are consistent with \mathbf{Y} (see Definition 6).

Output: a refined series decomposition $\mathbb{Y} = \tilde{\mathbb{Y}}^{(1)} + \dots + \tilde{\mathbb{Y}}^{(l)}$.

1. Construct an (\mathbf{L}, \mathbf{R}) -SVD of \mathbf{Y} by Algorithm 1 in the form

$$\mathbf{Y} = \sum_{i=1}^r \sigma_i P_i Q_i^T.$$

2. Partition the set $\{1, \dots, r\} = \bigsqcup_{m=1}^l J_m$ and perform grouping to obtain a refined matrix decomposition $\mathbf{Y} = \mathbf{Y}_{J_1} + \dots + \mathbf{Y}_{J_l}$.
3. Obtain a refined series decomposition $\mathbb{Y} = \tilde{\mathbb{Y}}^{(1)} + \dots + \tilde{\mathbb{Y}}^{(l)}$, where $\tilde{\mathbb{Y}}^{(m)} = \mathcal{T}^{-1}\mathcal{H}\mathbf{Y}_{J_m}$.

Thus, after application of Algorithm 2 to the group I_s , we obtain the following decomposition of the series \mathbb{X} :

$$\mathbb{X} = \tilde{\mathbb{X}}^{(1)} + \dots + \tilde{\mathbb{X}}^{(p)}, \text{ where } \tilde{\mathbb{X}}^{(s)} = \tilde{\mathbb{Y}}^{(1)} + \dots + \tilde{\mathbb{Y}}^{(l)}.$$

For simplicity, below we will consider the case $l = 2$.

3.3 Iterative O-SSA

Let us describe an iterative version of Algorithm 2, that is, an iterative algorithm for obtaining appropriate matrices \mathbf{L} and \mathbf{R} for the (\mathbf{L}, \mathbf{R}) -SVD of \mathbf{X}_I . For proper use of nested decompositions, we should expect that the matrix \mathbf{X}_I is close to a rank-deficient trajectory matrix of rank r .

To explain the main point of the method, assume that $\mathbf{X}_I = \mathbf{Y}$ is the trajectory matrix of \mathbb{Y} . Let $\mathbb{Y} = \mathbb{Y}^{(1)} + \mathbb{Y}^{(2)}$ and the trajectory matrices \mathbf{Y}_1 and \mathbf{Y}_2 be of ranks r_1 and r_2 , $r_1 + r_2 = r$. Then by Theorem 1 there exist r -rank separating matrices \mathbf{L}^* , \mathbf{R}^* of sizes $L \times L$ and $K \times K$ correspondingly and a partition $\{1, \dots, r\} = J_1 \sqcup J_2$ such that we can perform the proper grouping in the $(\mathbf{L}^*, \mathbf{R}^*)$ -SVD and thereby obtain $\mathbf{Y}_{J_1} = \mathbf{Y}_1$ and $\mathbf{Y}_{J_2} = \mathbf{Y}_2$.

Unfortunately, we do not know \mathbf{L}^* and \mathbf{R}^* , since they are determined by unknown trajectory spaces of $\mathbb{Y}^{(1)}$ and $\mathbb{Y}^{(2)}$.

Therefore, we want to construct the sequence of (\mathbf{L}, \mathbf{R}) -SVD decompositions (8), which in some sense converges to the separating decomposition.

Let us have an initial $(\mathbf{L}^{(0)}, \mathbf{R}^{(0)})$ -SVD decomposition of \mathbf{Y} and group its components to obtain some initial estimates $\tilde{\mathbb{Y}}^{(1,0)}$ and $\tilde{\mathbb{Y}}^{(2,0)}$ of $\mathbb{Y}^{(1)}$ and $\mathbb{Y}^{(2)}$. Then we can use the trajectory spaces of $\tilde{\mathbb{Y}}^{(1,0)}$ and $\tilde{\mathbb{Y}}^{(2,0)}$ to construct the new inner product which is expected to be closer to the separating one. Therefore, we can expect that $\tilde{\mathbb{Y}}^{(1,1)}$ and $\tilde{\mathbb{Y}}^{(2,1)}$ will be closer to $\mathbb{Y}^{(1)}$ and $\mathbb{Y}^{(2)}$ and therefore we take their trajectory spaces to construct a new inner product; and so on. Certainly, if the initial decomposition is strongly separating, then we obtain that $\tilde{\mathbb{Y}}^{(m,1)} = \tilde{\mathbb{Y}}^{(m,0)} = \mathbb{Y}^{(m)}$, $m = 1, 2$.

3.3.1 Basic algorithm

We call the iterative version of Algorithm 2 (Nested Oblique SSA) as *Iterative Oblique SSA* or *Iterative O-SSA*.

Algorithm 3. (Scheme of Iterative O-SSA.)

Input: The matrix \mathbf{Y} of rank r , which is the input matrix for Algorithm 2, a partition $\{1, \dots, r\} = J_1 \sqcup J_2$, $r_m = |J_m|$, the accuracy ε and the maximal number of iterations M . Also we should choose a pair of matrices $(\mathbf{L}^{(0)}, \mathbf{R}^{(0)})$, consistent with \mathbf{Y} as initial data.

Together with the partition, the matrices provide the decompositions $\mathbf{Y} = \mathbf{Y}_{J_1}^{(0)} + \mathbf{Y}_{J_2}^{(0)}$ and $\mathbb{Y} = \tilde{\mathbb{Y}}^{(1,0)} + \tilde{\mathbb{Y}}^{(2,0)}$.

Output: $\mathbb{Y} = \tilde{\mathbb{Y}}^{(1)} + \tilde{\mathbb{Y}}^{(2)}$.

1. $k = 1$.
2. Call of Algorithm for calculation of $(\mathbf{L}^{(k)}, \mathbf{R}^{(k)})$ consistent with \mathbf{Y} .
3. Construct the $(\mathbf{L}^{(k)}, \mathbf{R}^{(k)})$ -SVD of \mathbf{Y} :

$$(7) \quad \mathbf{Y} = \sum_{i=1}^r \sigma_i^{(k)} P_i^{(k)} (Q_i^{(k)})^T = \mathbf{Y}_{J_1}^{(k)} + \mathbf{Y}_{J_2}^{(k)}.$$

4. Obtain the decomposition of the series $\mathbb{Y} = \tilde{\mathbb{Y}}^{(1,k)} + \tilde{\mathbb{Y}}^{(2,k)}$, where $\tilde{\mathbb{Y}}^{(m,k)} = \mathcal{T}^{-1}\mathcal{H}\mathbf{Y}_{J_m}^{(k)}$, $m = 1, 2$.
5. If $k \geq M$ or $\max(\|\tilde{\mathbb{Y}}^{(m,k)} - \tilde{\mathbb{Y}}^{(m,k-1)}\|^2 / N, m = 1, 2) < \varepsilon^2$, then $\tilde{\mathbb{Y}}^{(m)} \leftarrow \tilde{\mathbb{Y}}^{(m,k)}$, $m = 1, 2$, and STOP; else $k \leftarrow k + 1$ and go to step 2.

Remark 3. Note that the initial matrices $(\mathbf{L}^{(0)}, \mathbf{R}^{(0)})$ can be chosen such that the initial decomposition (7) for $k = 0$ is a part of the SVD (1) and thereby coincides with the ordinary SVD of \mathbf{Y} , that is, $\mathbf{L}^{(0)}$ and $\mathbf{R}^{(0)}$ are the identity matrices. Then the partition can be performed as follows. In the decomposition (1), we can choose two sets of eigentriple numbers and consider their union as I . The chosen sets of numbers automatically generate the partition $J_1 \sqcup J_2$. For example, if two groups, ET2,8 and ET3–6, are chosen, then $I = \{2, 3, 4, 5, 6, 8\}$, $r = 6$, $J_1 = \{1, 6\}$, $J_2 = \{2, 3, 4, 5\}$.

To finalize Algorithm 3, we should present the algorithm for step 2. Define Π_{col} the orthogonal projection operator (in the ordinary sense) on the column space of \mathbf{Y} , Π_{row} the projection operator on the row space of \mathbf{Y} .

Algorithm 4. (Calculation of $(\mathbf{L}^{(k)}, \mathbf{R}^{(k)})$.)

Input: The partition $\{1, \dots, r\} = J_1 \sqcup J_2$, $r_m = |J_m|$, the pair of matrices $(\mathbf{L}^{(k-1)}, \mathbf{R}^{(k-1)})$.

Output: The pair of matrices $(\mathbf{L}^{(k)}, \mathbf{R}^{(k)})$.

1. Calculate $\tilde{\mathbf{Y}}_m = \mathcal{H}\mathbf{Y}_{J_m}^{(k-1)}$, $m = 1, 2$.
2. Construct the ordinary SVDs:

$$\tilde{\mathbf{Y}}_m = \sum_{i=1}^{d_m} \sqrt{\lambda_i^{(m)}} U_i^{(m)} (V_i^{(m)})^T, \quad m = 1, 2,$$

(we need the first r_m terms only).

3. Find the projections $\hat{U}_i^{(m)} = \Pi_{\text{col}} U_i^{(m)}$ and $\hat{V}_i^{(m)} = \Pi_{\text{row}} V_i^{(m)}$ for $i = 1, \dots, r_m$, $m = 1, 2$. Denote

$$\hat{\mathbf{U}}^{(m)} = [\hat{U}_1^{(m)} : \dots : \hat{U}_{r_m}^{(m)}], \quad \hat{\mathbf{V}}^{(m)} = [\hat{V}_1^{(m)} : \dots : \hat{V}_{r_m}^{(m)}].$$

4. Calculate $\mathbf{L}^{(k)} = (\hat{\mathbf{U}}^\dagger)^T \hat{\mathbf{U}}^\dagger$ and $\mathbf{R}^{(k)} = (\hat{\mathbf{V}}^\dagger)^T \hat{\mathbf{V}}^\dagger$, where $\hat{\mathbf{U}} = [\hat{\mathbf{U}}^{(1)} : \hat{\mathbf{U}}^{(2)}]$ and $\hat{\mathbf{V}} = [\hat{\mathbf{V}}^{(1)} : \hat{\mathbf{V}}^{(2)}]$.

Note that we assume that the matrices $\hat{\mathbf{U}}^{(m)}$ and $\hat{\mathbf{V}}^{(m)}$ obtained at step 3 are of full rank; otherwise, the algorithm does not work.

For the constructed iterative Algorithm 3, the convergence $\tilde{\mathbf{Y}}^{(1,k)}$ and $\tilde{\mathbf{Y}}^{(2,k)}$ to some series $\mathbb{Y}^{(1)}$ and $\mathbb{Y}^{(2)}$ is not proved theoretically; however, numerical experiments confirm the convergence for the most of reasonable examples.

Let us shortly discuss why one can expect the convergence of the iterations to the proper decomposition. First, note that Iterative O-SSA does not change the separating decomposition, that is, the separating decomposition is a fixed point of the algorithm. Then, the separating decomposition $\mathbf{Y} = \mathbf{Y}_1 + \mathbf{Y}_2$ should satisfy the following properties:

- (1) \mathbf{Y}_1 and \mathbf{Y}_2 are Hankel;
- (2) $\text{rank } \mathbf{Y}_1 = r_1$, $\text{rank } \mathbf{Y}_2 = r_2$;
- (3) the column and row spaces of \mathbf{Y}_1 and \mathbf{Y}_2 lie in the column and row spaces of \mathbf{Y} ;
- (4) \mathbf{Y}_1 and \mathbf{Y}_2 are (\mathbf{L}, \mathbf{R}) bi-orthogonal for $\mathbf{L} = \mathbf{L}^*$ and $\mathbf{R} = \mathbf{R}^*$.

Each iteration consequently tries to meet these properties:

- (1) hankelization at step 1 is the orthogonal projection on the set of Hankel matrices;
- (2) taking the r_m leading components in the SVDs of series (step 2) performs the low-rank projections;
- (3) there is the step 3 of projection on the row and column spaces of \mathbf{Y} ;
- (4) the choice of (\mathbf{L}, \mathbf{R}) -inner products at step 4 makes the matrices bi-orthogonal.

3.3.2 Modification with sigma-correction

If the initial point for iterations is not far from the separating pair $(\mathbf{L}^*, \mathbf{R}^*)$, we can expect that the convergence will take place, since we are close to the fixed-point value and we can expect that $\sigma_i^{(k)}$ are changed slightly. However, in general, a possible reordering of the decomposition components between iterations of Iterative O-SSA can interfere convergence. The case of $J_1 = \{1, \dots, r_1\}$, when the minimal singular value σ_{r_1} of the first series is kept essentially larger than the maximal singular value σ_{r_1+1} of the second series, would provide safety.

Let us describe the modification of Iterative O-SSA that provides reordering of the components, moves them apart and thereby relaxes the problem of mixture of components. Modification consists in an adjustment of calculation of $\hat{\mathbf{U}}^{(2)}$ and $\hat{\mathbf{V}}^{(2)}$ at step 3 of Algorithm 4.

Algorithm 5. (Modification of Algorithm 4.)

Input and *Output* are the same as in Algorithm 4 except for an additional parameter $\varkappa > 1$ called the separating factor.

The algorithm is the same except for an additional step 3a after step 3.

3a: If $\lambda_{r_1}^{(1)} < \varkappa^2 \lambda_1^{(2)}$ at step 2 of Algorithm 4, then define $\mu = \varkappa \sqrt{\lambda_1^{(2)} / \lambda_{r_1}^{(1)}}$ and change $\hat{\mathbf{U}}^{(2)} \leftarrow \sqrt{\mu} \hat{\mathbf{U}}^{(2)}$, $\hat{\mathbf{V}}^{(2)} \leftarrow \sqrt{\mu} \hat{\mathbf{V}}^{(2)}$. Due to reordering, put $J_1 = \{1, \dots, r_1\}$, $J_2 = \{r_1+1, \dots, r\}$.

Note that the adjustment implicitly leads to the change of the order of matrix components in (7), since they are ordered by $\sigma_i^{(k)}$. Thereby we force an increase of the matrix components related to the first series component. Proposition 5 explains this adjustment.

Remark 4. *The reordering procedure is fulfilled by sequential adjustment of the component weights and therefore depends on the component enumeration.*

Note that the described correction can help to provide the strong separability if the weak one takes place.

3.4 Separability of sine waves with close frequencies

3.4.1 Noise-free cases

Let us consider the sum of two sinusoids $x_n = \sin(2\pi\omega_1 n) + A \sin(2\pi\omega_2 n)$, $n = 1, \dots, N$, $N = 150$, with close frequencies $\omega_1 = 0.065$ and $\omega_2 = 0.06$ and unequal amplitudes, 1 and $A = 1.2$. Let the window length $L = 70$. Since sinusoids with such close frequencies are far from being orthogonal for the considered window and series lengths, Basic SSA cannot separate them, see Fig. 3 (top) where the result of the Basic SSA decomposition is depicted.

To separate the sinusoids we apply the Iterative O-SSA algorithm (Algorithm 3) with no sigma-correction, $\varepsilon = 10^{-5}$ and two groups ET1–2 and ET3–4. The maximal number

M of iterations was taken very large and therefore was not reached. Decomposition after Iterative O-SSA is depicted in Fig. 3 (bottom).

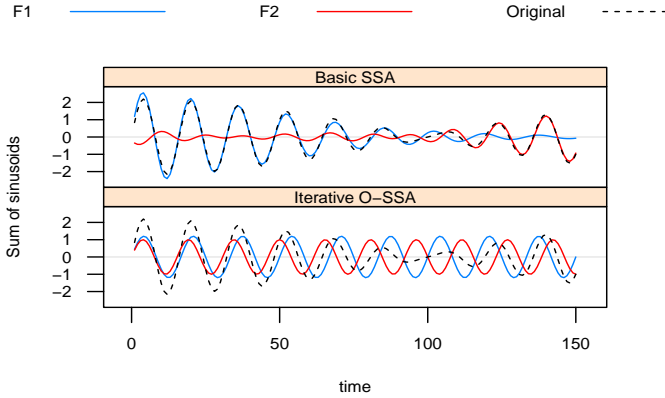


Figure 3. Sum of two sinusoids with close frequencies: decomposition by Basic SSA and Iterative O-SSA

Let us apply the measures of separability described in Section 3.1. Note that the conventional \mathbf{w} -correlations do not reflect the quality of decomposition. For the initial decomposition we have 0.08. After Iterative O-SSA the \mathbf{w} -correlation becomes to be equal to -0.44 , while (\mathbf{L}, \mathbf{R}) \mathbf{w} -correlation is almost 0. The last result confirms that the method separates harmonics exactly. Other measure of true decomposition is the closeness of the components to series of finite ranks. Since the ranks should be equal to the number of the components in the chosen groups, we can calculate the proportion of the corresponding number of the leading components in their SVD decompositions. The mean proportion $(0.5(\tau_{r_1}(\mathbb{X}^{(1)}) + \tau_{r_2}(\mathbb{X}^{(2)})))$ is changed from 0.06 to almost 0.

Let us fix $\omega_2 = 0.06$. Then for $\omega_1 = 0.065$ the algorithm stops after 113 iterations, for $\omega_1 = 0.07$ the number of iterations is equal to 26, for $\omega_1 = 0.08$ it is equal to just 6; see blue line in Fig. 5 (top).

Note that we do not need to use the sigma-correction, since the sinusoids have different amplitudes.

If we consider equal amplitudes with $A = 1$ and take $\varkappa = 2$ (Algorithm 5), then Iterative O-SSA still converges even for $\omega_2 = 0.065$ (191 iterations) to the true solution.

3.4.2 Nested separability in presence of noise

Let us add noise to the sum of two sinusoids and take $x_n = \sin(2\pi\omega_1 n) + A \sin(2\pi\omega_2 n) + \delta\varepsilon_n$ with close frequencies $\omega_1 = 0.07$ and $\omega_2 = 0.06$ and unequal amplitudes, 1 and $A = 1.2$. Here ε_n is white Gaussian noise with variance 1, $\delta = 1$. Let again $N = 150$, $L = 70$.

Basic SSA well separates the sinusoids from noise, but cannot separate these sinusoids themselves. Thus, Iterative O-SSA applied to the estimated signal subspace should be

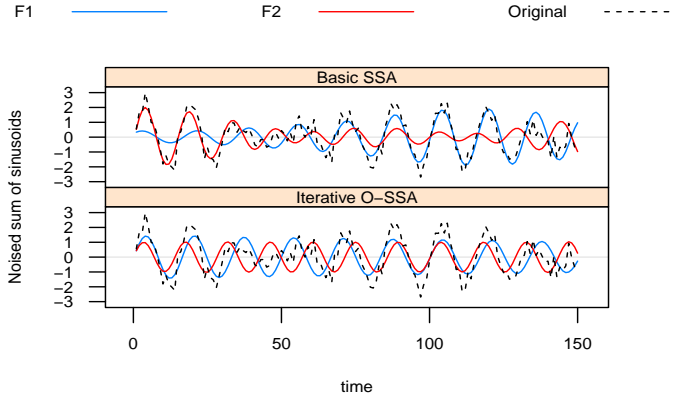


Figure 4. Noisy sum of two sinusoids with close frequencies: decomposition by Basic SSA and Iterative O-SSA

used. We use the sigma-correction with $\varkappa = 2$, since the difference between amplitudes, 1 and 1.2, appears to be small for strong separability in presence of noise. As before, we set the initial grouping ET1–2 and ET3–4.

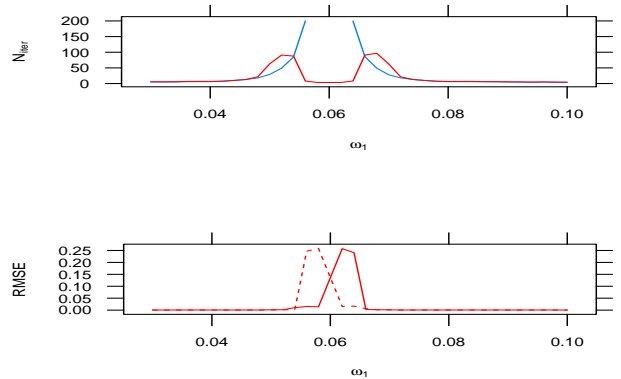


Figure 5. Dependence of number of iterations (top) and RMSE errors of frequency estimations (bottom) on ω_1 for $\omega_2 = 0.6$

The decomposition by Basic SSA at top and by Iterative O-SSA at bottom is depicted in Fig. 4. The number of iterations is equal to 32, what is just slightly larger than 26 in the noiseless case.

Let us investigate the dependence of number of iterations on ω_1 with the fixed $\omega_2 = 0.06$. We change ω_1 from 0.03 to 0.059 and from 0.061 to 0.1. Fig. 5 (top) shows the number of iterations for noiseless signal (blue line) and the estimated mean number of iterations for the noisy signal (red line); the number of repetitions equals 1000, 5% winsorized estimates of means were calculated. Note that the number of iterations was limited by 200, although for the pure signal convergence held for each ω_1 from the considered set.

A surprisingly small number of iterations for the noisy signal and close frequencies is explained by convergence to an wrong limit, see Fig. 5 (bottom) with root mean square errors of LS-ESPRIT estimates for ω_1 and ω_2 based on the subspaces spanned by eigenvectors from ET1–2 and ET3–4 (see, e.g., [28] or [13, Section 2.8.2] for the ESPRIT algorithms). Since we use the nested decomposition, the noise slightly influences the reconstruction accuracy for frequencies that are quite different (ω_1 smaller than 0.048 and larger than 0.072).

4. SSA WITH DERIVATIVES. EXTENSION FOR STRONG SEPARABILITY

In this section we describe an extension of SSA that helps to overcome the problem of lack of strong separability if weak separability holds.

Recall that the lack of strong separability of two series components is caused by equal singular values in the sets of the singular values generated by each of time series. In turn, the singular values depends on coefficients A_1 and A_2 before the series components in the sum $A_1 s_n^{(1)} + A_2 s_n^{(2)}$. The question is how to change the coefficients A_1 and A_2 in conditions of unknown $s_n^{(1)}$ and $s_n^{(2)}$ to make the singular values different.

It seems that the most natural approach is to use the derivative of the time series in order to change the coefficients and not to change the component subspaces. For example, if $x_n = A \sin(2\pi\omega n + \phi)$, then $x'_n = 2\pi\omega A \cos(2\pi\omega n + \phi)$, that is, the coefficient $A' = 2\pi\omega A$. If we take two sinusoids with different frequencies, then derivation changes their amplitudes differently. For $x_n = Ae^{\alpha n}$, derivation also changes the coefficient before the exponential, since $x'_n = \alpha Ae^{\alpha n}$, and preserves the rate. For the most of series of finite rank, the derivative subspace coincides with the series subspace. The exception is polynomial series, when the derivative subspace is a subset of the initial subspace.

Certainly, since we deal with discrete time, we consider $\varphi_n(\mathbb{X}) = x_{n+1} - x_n$ instead of derivative. However, the approach of taking differences works. For example, for series $\mathbb{X} = \mathbb{X}_N$ of length N with $x_n = A \sin(2\pi\omega n + \phi)$, we obtain the series $\Phi_{N-1}(\mathbb{X}) = (\varphi_1(\mathbb{X}), \dots, \varphi_{N-1}(\mathbb{X}))$ of length $N - 1$ with $\varphi_n(\mathbb{X}) = 2 \sin(\pi\omega) A \cos(2\pi\omega n + \pi\omega + \phi)$; for $x_n = Ae^{\alpha n}$, we obtain $\varphi_n(\mathbb{X}) = (e^\alpha - 1)Ae^{\alpha n}$.

Thus, we can combine the initial series and its derivative to imbalance the component contribution and therefore to obtain their strong separability. For sinusoids, the smaller the period, the larger the increase of the sinusoid amplitude. Therefore, derivation increases the contribution of high frequencies. This effect can increase the level of the noise component, if the series is corrupted by noise. Hence, the nested version of the method implementation should be produced; in particular, the noise component should be removed by Basic SSA in advance.

Remark 5. *The approach involving derivatives (that is, sequential differences) can be naturally extended to considering an arbitrary linear filtration φ instead of taking sequential differences. In this paper we deal with derivatives, since this particular case is simple and has very useful applications.*

In Section 4.1 we consider the initial series and its derivative together as two series, regulating the contribution of the derivative, and apply then the multivariate version of SSA. Section 4.2 transforms this approach to a special nested version of Oblique SSA called DerivSSA.

4.1 SSA with derivatives as MSSA

Let us consider the system of two time series $(\mathbb{X}_N, \gamma\Phi_{N-1}(\mathbb{X}))$ and apply Multivariate SSA (MSSA).

The MSSA algorithm can be found, for example, in [6, 12] for time series of equal lengths. However, it is naturally extended to different lengths. In particular, MSSA for time series of different lengths is described in [5, Section III.2] and [14].

In MSSA, the embedding operator \mathcal{T} transfers two time series $(\mathbb{X}_{N_1}, \mathbb{Y}_{N_2})$ to the stacked L -trajectory matrix $[\mathbf{X} : \mathbf{Y}]$. That is, the only difference with Basic SSA consists in the construction of the embedding operator \mathcal{T} .

Let $\mathbb{X}_N = \mathbb{X}_N^{(1)} + \mathbb{X}_N^{(2)}$ and $\mathbb{X}_N^{(1)}$ and $\mathbb{X}_N^{(2)}$ be of finite rank and approximately separable. Therefore their row and column trajectory spaces are approximately orthogonal. Then the same is valid for $\Phi_{N-1}(\mathbb{X}^{(1)})$ and $\Phi_{N-1}(\mathbb{X}^{(2)})$ in view of the fact that their column spaces belongs to the column spaces of $\mathbb{X}_N^{(1)}$ and $\mathbb{X}_N^{(2)}$, while their row spaces are spanned by vectors of the same structure that the vectors constituting bases of the row spaces of $\mathbb{X}_N^{(1)}$ and $\mathbb{X}_N^{(2)}$, except for these basis vectors has length $K - 1$ instead of K . Therefore, approximate orthogonality still hold. Since $\Phi_{N-1}(\mathbb{X}) = \Phi_{N-1}(\mathbb{X}^{(1)}) + \Phi_{N-1}(\mathbb{X}^{(2)})$, MSSA applied to $(\mathbb{X}_N, \gamma\Phi_{N-1}(\mathbb{X}))$ will approximately separate the time series $\mathbb{X}_N^{(1)}$ and $\mathbb{X}_N^{(2)}$. Certainly, we will not have exact separability; however, it is not so important for practice.

As it was mentioned before, a drawback of the described approach is that the method cannot be applied to noisy series, since it intensifies high-frequency harmonics and therefore strengthens noise. Therefore, denoising should be applied as preprocessing. Also, SSA involving derivatives changes component contributions (this is what we want) but simultaneously the method loses approximation features. These reasons lead to the necessity to use the nested way of decomposition introduced in Section 3.2.

4.2 Nested SSA with derivatives (DerivSSA)

Let us formulate the nested version of SSA with derivatives called DerivSSA. As well as in Section 3.2, let $\mathbf{Y} = \mathbf{X}_I$ be one of matrices in the decomposition $\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_p}$ obtained at Grouping step of Basic SSA; each group corresponds to a separated time series component and we want to construct a refined decomposition of \mathbf{Y} . As before, denote $r = \text{rank } \mathbf{Y}$, $\mathbb{Y} = \mathcal{T}^{-1}\mathcal{H}\mathbf{Y}$.

Algorithm 6. (DerivSSA.)

Input: The matrix \mathbf{Y} , the weight of derivative $\gamma > 0$.

Output: a refined series decomposition $\mathbb{Y} = \tilde{\mathbb{Y}}^{(1)} + \dots + \tilde{\mathbb{Y}}^{(l)}$.

1. Denote $\Phi(\mathbf{Y}) = [Y_2 - Y_1 : \dots : Y_N - Y_{N-1}]$. Construct the matrix $\mathbf{Z} = [\mathbf{Y} : \gamma\Phi(\mathbf{Y})]$.
2. Perform the SVD of \mathbf{Z} : $\mathbf{Z} = \sum_{i=1}^r \sqrt{\lambda_i} U_i V_i^T$.
3. Construct the following decomposition of $\mathbf{Y} = \mathbf{X}_J$ into the sum of elementary matrices: $\mathbf{Y} = \sum_{i=1}^r U_i U_i^T \mathbf{Y}$.
4. Partition the set $\{1, \dots, r\} = \bigsqcup_{m=1}^l J_m$ and perform grouping to obtain a refined matrix decomposition $\mathbf{Y} = \mathbf{Y}_{J_1} + \dots + \mathbf{Y}_{J_l}$.
5. Obtain a refined series decomposition $\mathbb{Y} = \tilde{\mathbb{Y}}^{(1)} + \dots + \tilde{\mathbb{Y}}^{(l)}$, where $\tilde{\mathbb{Y}}^{(m)} = \mathcal{T}^{-1} \mathcal{H} \mathbf{Y}_{J_m}$.

Note that steps 2 and 3 of algorithm are correct, since the column space of \mathbf{Z} coincides with the column space of \mathbf{Y} . Therefore, $\text{rank } \mathbf{Z} = r$ and $\{U_i\}_{i=1}^r$ is the orthonormal basis of the column space of \mathbf{Y} .

The following proposition shows that Algorithm 6 is exactly Algorithm 2 with a specific pair of matrices (\mathbf{L}, \mathbf{R}) , where $P_i = U_i$, Q_i are normalized vectors $\mathbf{Y}^T U_i$ in (7).

Proposition 1. *The left singular vectors of the ordinary SVD of \mathbf{Z} coincide with the left singular vectors of the (\mathbf{L}, \mathbf{R}) -SVD of the input matrix \mathbf{Y} , where $\mathbf{L} \in \mathcal{M}_{L,L}$ is the identity matrix and \mathbf{R} is defined by the equality $\mathbf{R} = \mathbf{E} + \gamma^2 \mathbf{F}^T \mathbf{F}$, where $\mathbf{E} \in \mathcal{M}_{K,K}$ is the identity matrix and*

$$\mathbf{F} = \begin{pmatrix} -1 & 1 & 0 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & -1 & 1 & 0 \\ 0 & \dots & 0 & 0 & -1 & 1 \end{pmatrix} \in \mathcal{M}_{K-1,K}.$$

Proof. Note that the standard inner product in the row space of \mathbf{Z} can be expressed as $(Z_1, Z_2)_{2K-1} = (Q_1, Q_2)_K + \gamma^2 (\Phi(Q_1), \Phi(Q_2))_{K-1}$, where Q_1 and Q_2 consist of the first K components of Z_1 and Z_2 , $\Phi(Q) \in \mathbb{R}^{K-1}$ applied to a vector $Q = (q_1, \dots, q_K)^T$ consists of successive differences of vector components $q_{i+1} - q_i$. Thus, if we introduce the inner product $\langle Q_1, Q_2 \rangle_{\mathbf{R}} = (\mathbf{R} Q_1, Q_2)_K$, then the ordinary SVD of \mathbf{Z} can be reduced to the (\mathbf{L}, \mathbf{R}) -SVD of \mathbf{Y} with the corresponding matrices \mathbf{L} and \mathbf{R} . \square

Remark 6. *If \mathbf{Y} is the trajectory matrix of a series \mathbb{Y}_N , then the nested SSA with derivatives is equivalent to the MSSA implementation described in Section 4.1. Indeed, the trajectory matrix of the derivative time series $\Phi_{N-1}(\mathbb{Y})$ coincides with the matrix $\Phi(\mathbf{Y})$. Although, if \mathbf{Y} is not Hankel, there is no MSSA analogue.*

4.3 Separation of sine waves with equal amplitudes

Consider the series $x_n = \sin(2\pi n/10) + \sin(2\pi n/15)$, $n = 1, \dots, N$, $N = 150$, $L = 70$, which is depicted in Fig. 6.

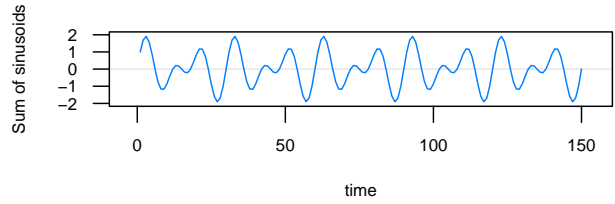


Figure 6. Sum of two sinusoids with equal amplitudes

Sinusoids with periods 10 and 15 are approximately separable for such series and window lengths. However, since the sinusoid amplitudes are equal, there is no strong separability and therefore after Basic SSA we obtain an unsatisfactory decomposition, an arbitrary mixture of the sinusoids (top picture of Fig. 7) with \mathbf{w} -correlation between reconstructed by ET1–2 and ET3–4 series equal to 0.92.

The decomposition performed by DerivSSA with $\gamma = 10$ applied to the group ET1–4 with $J_1 = \{1, 2\}$ and $J_2 = \{3, 4\}$ (Algorithm 6) is depicted in the bottom graph of Fig. 7 and demonstrates the very accurate separability, \mathbf{w} -correlation is equal to 0.01. The second measure, the mean proportion $0.5(\tau_{r_1}(\mathbb{X}^{(1)}) + \tau_{r_2}(\mathbb{X}^{(2)}))$, is diminished from 0.3266 to 0.0003. For this example, the obtained decomposition practically does not depend on γ for all $\gamma > 2$.

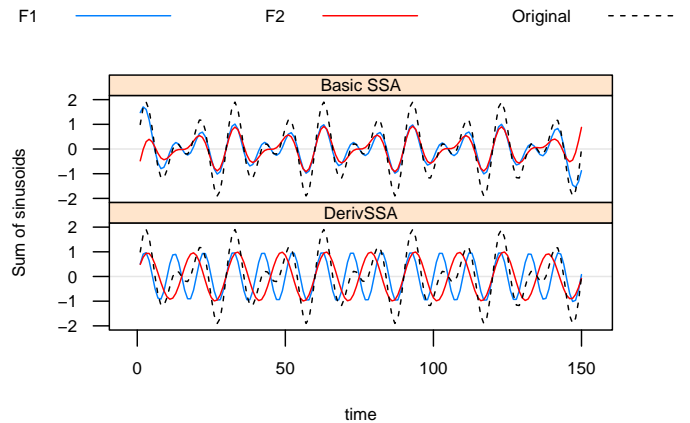


Figure 7. Sum of two sinusoids with equal amplitudes: reconstruction by Basic SSA (top) and DerivSSA (bottom)

5. REAL-LIFE TIME SERIES

In this section we apply Iterative O-SSA (Algorithm 3 and 4 with possible modification provided by Algorithm 5) and DerivSSA (Algorithm 6) to real-life time series. The role of the methods for separability of sine-waves was demonstrated in Sections 3.4 and 4.3 with the help of simulated

data. The obtained conclusions are generally valid for real-life series: DerivSSA adds to Basic SSA the ability to separate sine waves with close amplitudes, while Iterative O-SSA can help in separation of sine waves, which are not orthogonal, that is, their frequencies are insufficiently far one from another. Note that since in real-life series with seasonality there are no close frequencies, DerivSSA can be very useful for seasonality decomposition.

In this section we consider the problem of trend extraction. The choice of examples is explained by the following considerations.

If a time series is long enough, then the oscillations are well weakly separated from the trend and only strong separability is under question. Therefore, we expect that DerivSSA will work for trends of complex forms.

For short series, the trend can be not orthogonal to a periodic component like seasonality; therefore, DerivSSA can even worsen the separability; moreover, derivation suppresses low-frequency components. On the other hand, Iterative O-SSA is specially designed to separate non-orthogonal series components.

We will take only one iteration in Iterative O-SSA method, since it is sufficient to obtain good decomposition in the considered examples and also makes the methods comparable by computational cost.

5.1 Improving of strong separability

Let us consider US Unemployment data (monthly, 1948-1981, thousands) for male (20 years and over). Data are taken from [1], the series length N is equal to 408, see Fig.8. Since the series is long, we can expect weak separability of the trend and the seasonality. For better weak separability we choose the window length equal to $L = N/2 = 204$, which is divisible by 12.

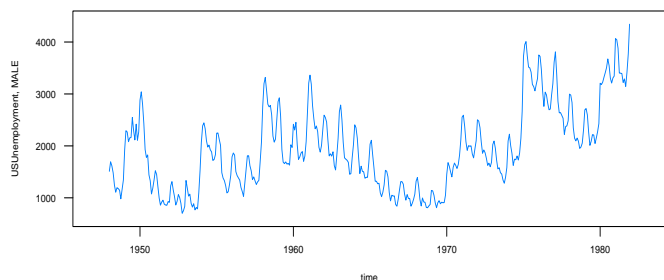


Figure 8. US unemployment: initial series

Basic SSA does not separate the trend and seasonality (see Fig. 9 and Fig. 13 (left)) for this time series, likely due to lack of strong separability. This is the typical situation when the trend has a complex form, trend components are mixed with the seasonality components and therefore the so

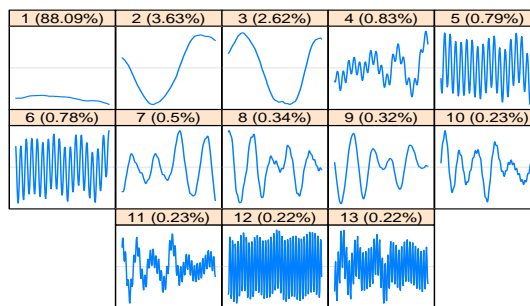


Figure 9. US unemployment: eigenvectors obtained by Basic SSA

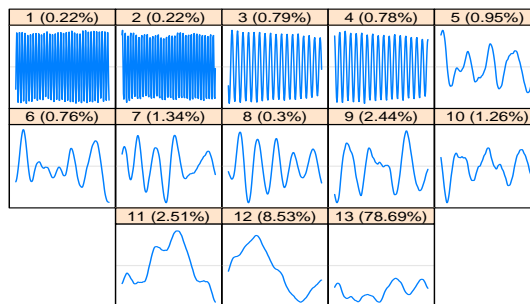


Figure 10. US unemployment: eigenvectors obtained by DerivSSA

called Sequential SSA was recommended [11, Section 1.7.3]. However, this is the case when DerivSSA should help.

We apply DerivSSA to the group ET1–13 that can be related to the signal. DerivSSA separates different frequencies so that components with higher frequencies become leading ones. Since the low-frequency components in the considered series have large contribution, the weight of derivatives should be large to make the seasonal components leading; we take $\gamma = 1000$.

The resulting eigenvectors are depicted in Fig. 10. One can see that the first 4 components contain seasonality, while the eigenvectors 5–13 contains components of the trend. The mixture of the components within the trend group is not important. Fig. 10 demonstrates that the seasonal components are now separated from the residual. Fig. 11 depicting the DerivSSA reconstructions of the trend and the seasonality confirms that DerivSSA visibly improves the reconstruction accuracy, especially at the ends of the series.

Since Iterative O-SSA has possibility of sigma-correction, it also can help to move apart the decomposition components, and therefore we can apply Iterative O-SSA to the group ET1–13 with the refined groups ET1–4,7–11 (trend)

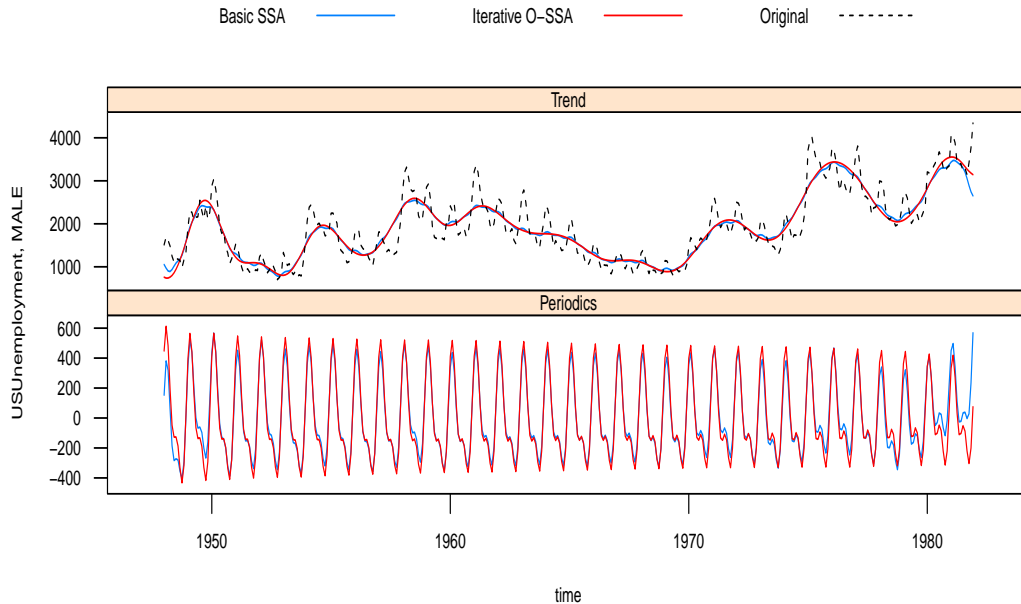


Figure 11. US unemployment: Decompositions by Basic SSA and Iterative O-SSA, which coincides with that by DerivSSA.

and ET5,6,12,13 (seasonality). Since the components of the Basic SSA decomposition are mixed, we refer the components that contain mostly trend and slow cycles to the first group and the components that contain mostly seasonality to the second group. As eigenvectors reflect forms of the corresponding time series components, we can use the graph of eigenvectors shown in Fig. 9 for the initial grouping. For example, the fourth eigenvector looks like slow oscillations corrupted by seasonality and therefore we refer it to the trend group, while the fifth eigenvector looks like seasonal component corrupted by something slow varying and we refer it to the seasonality group. We apply one iteration with sigma-correction, taking $\varkappa = 2$. After reordering caused by the sigma-correction, the first trend group consists of the first eight components 1–8, while the second seasonality group consists of 9–13 components, see Fig. 12.

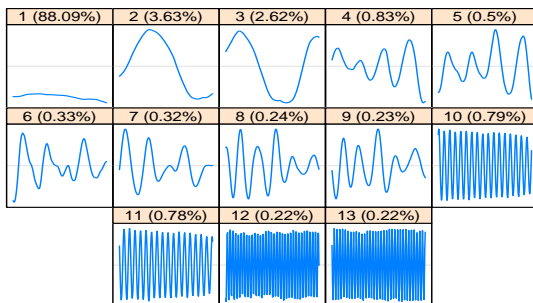


Figure 12. US unemployment: Iterative O-SSA eigenvectors

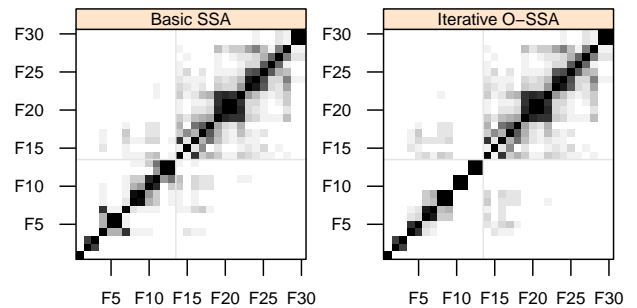


Figure 13. US unemployment: w -correlations before (left) and after (right) Iterative O-SSA

The trend eigenvectors of the DerivSSA decomposition (Fig. 10, ET5–13) differ from that of the O-SSA decomposition (Fig. 12, ET1–8), the seasonality components are almost the same. Nevertheless, the result of Iterative O-SSA reconstruction is visibly the same as that of DerivSSA shown in Fig. 11 and therefore we do not depict this reconstruction.

Fig. 13 contains the w -correlations between the elementary components provided by Basic SSA (left) and the w -correlations between the elementary components reconstructed by Iterative O-SSA (right). The figure confirms the improving of separability. For trend extraction, it is important that correlations between trend and seasonality groups

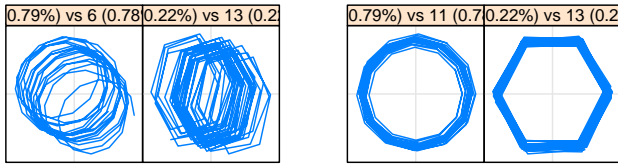


Figure 14. US unemployment: 2D plots of periodic eigenvectors before (left) and after (right) Iterative O-SSA

are close to zero. Really, correlations between ET1–8 and ET9–13 are small. Mixture of the components within the trend group is not important. One can see that the trend components are still slightly mixed with the noise components. However, we had a mixture with the residual before iterations (left) and this cannot be corrected by Iterative O-SSA (right), since the nested version is used. Fig. 14 shows the improvement of separability with the help of scatter-plots of seasonal eigenvectors. After one iteration, plots of seasonal eigenvectors form almost regular polygons.

Figures for the decomposition of DerivSSA analogous to Fig. 13 and 14 are very similar and we do not present them in the paper. Note that in DerivSSA we group components after their separation, what is easier than to group mixing components for Iterative O-SSA before separation. That is, in the considered example the resultant decomposition is the same, but application of DerivSSA is easier.

5.2 Improving of weak separability

Let us consider the series ‘Fortified wine’ (fortified wine sales, Australia, monthly, from January 1980, thousands of litres) taken from [17]. The first 120 points of the series are depicted in Fig. 15.

The series length is long enough to obtain weak separability; therefore, we will consider short subseries to demonstrate the advantage of Iterative O-SSA for improving of weak separability.

We take here the window length $L = 18$ to make the difference between methods clearly visible on the figures, although the relation between accuracies of the considered methods is the same for other choices of window lengths. Let us consider two subseries, from 30th to 78th points and from 36th to 84th points. The difference consists in behavior of the seasonality at the ends of the subseries.

As well as in the previous example, we start with Basic SSA. ET1 is identified as corresponding to trend, other components are produced by seasonality and noise (we do not

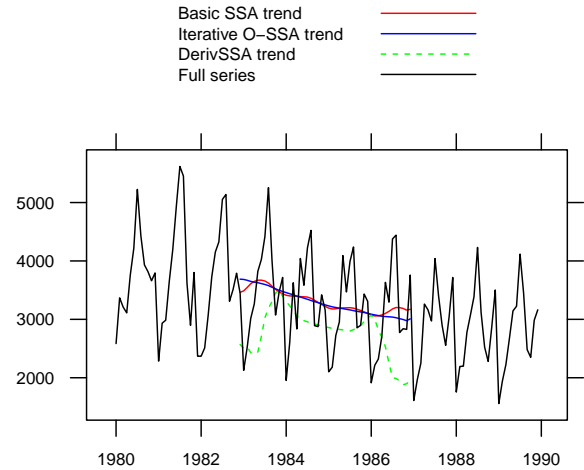


Figure 15. Fortified wines: trend reconstruction by DerivSSA and Iterative O-SSA for subseries of points 30–78.

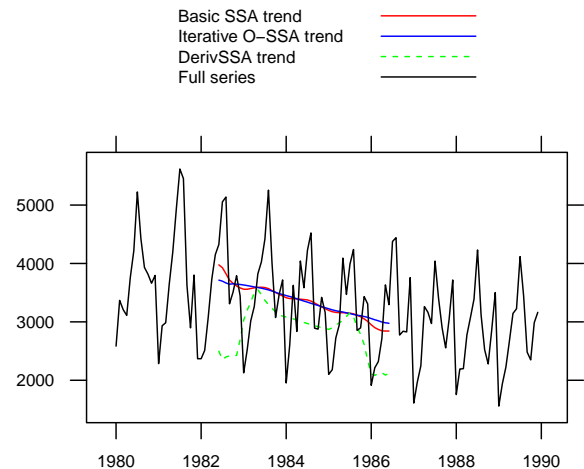


Figure 16. Fortified wines: trend reconstruction by DerivSSA and Iterative O-SSA for subseries of points 36–84.

include their pictures). One can see in Fig. 15 and 16 (red line) that the reconstructed trend is slightly mixed with the seasonality and steps after the seasonality at the ends of the series.

To apply Iterative O-SSA, we should choose a group of elementary components containing the trend components and approximately separated from the residual. Let it be ET1–7. Thus, we apply one iteration of O-SSA to the refined groups ET1 and ET2–7. Since the trend has the contribution much larger than the residual, we consider Iterative O-SSA with

no sigma-correction. The result of reconstruction is much more relevant, see Fig. 15 and 16 (blue line). Green line in the same figures shows that DerivSSA gives more poor reconstruction than Basic SSA in this example.

6. CONCLUSION

We suggested two modifications of SSA, which can considerably improve the separability and thereby the reconstruction accuracy. Iterative O-SSA shows its advantage dealing with separation of sine waves with close frequencies and with extraction of trend for short series. DerivSSA shows its advantage in conditions of weak separability dealing with long enough series with complex-form trends and sine waves with equal amplitudes.

We demonstrated that for separation of trend even one iteration of Iterative O-SSA can improve the separability, while DerivSSA works only in conditions of approximate weak separability. On the other hand, for separability of weakly separable sine waves with equal amplitudes DerivSSA works more effectively than Iterative O-SSA.

The important aspect of both methods is that they should be applied to the estimated signal subspace (more general, to the estimated subspace of the sum of components that we want to separate), that is, they work in a nested manner. We can consider the methods as refining of the decomposition obtained by Basic SSA (generally, the subspace estimation can be performed by any method, not necessarily by SSA). Despite the both methods have the underlying model of series governed by linear recurrence relations, the methods do not use the model directly. This allows one to apply the methods even if the signal satisfies the model only locally. For example, the trend usually does not satisfy an LRR exactly; however, it can be extracted by SSA and its considered extensions.

The common part of the methods is the generalized SVD (so called Restricted SVD), which provides decompositions that are not bi-orthogonal with respect to the conventional inner product. These methods do not use the optimality properties of the generalized SVD; however, this is not essential for their success in the signal decomposition.

The further development of the considered methods can consist in their combination for effective solution of the problem of lack of both weak and strong separability and in the use of the obtained improved non-orthogonal decomposition for forecasting.

APPENDIX A. INNER PRODUCTS AND RELATED MATRIX DECOMPOSITIONS

Here we provide the necessary information about matrix decompositions with respect to given inner products in the row and column spaces (see e.g. [32, Th.3]), which are called in [23] Restricted SVD (RSVD).

A.1 Inner products

Usually, orthogonality of vectors in \mathbb{R}^M is considered in a conventional manner: X_1 and X_2 in \mathbb{R}^M are orthogonal if their Euclidean inner product is equal to 0, i.e. $(X_1, X_2)_M = 0$, where $(\cdot, \cdot)_M$ is the standard inner product in \mathbb{R}^M . Sometimes we will omit the dimension in denotation if it is clear from the context. It is well-known that any inner product in \mathbb{R}^M can be defined as $\langle X_1, X_2 \rangle_{\mathbf{A}} = X_1^T \mathbf{A} X_2$ for a symmetric positive-definite matrix \mathbf{A} . For any $\mathbf{O}_{\mathbf{A}}$ such that $\mathbf{O}_{\mathbf{A}}^T \mathbf{O}_{\mathbf{A}} = \mathbf{A}$ we have $\langle X_1, X_2 \rangle_{\mathbf{A}} = (\mathbf{O}_{\mathbf{A}} X_1, \mathbf{O}_{\mathbf{A}} X_2)_M$. Evidently, $\mathbf{O}_{\mathbf{A}}$ is defined up to multiplication by an orthogonal matrix.

The inner product yields the notion of orthogonality. We will say that two vectors are *\mathbf{A} -orthogonal* if $\langle X_1, X_2 \rangle_{\mathbf{A}} = 0$.

Let the matrix \mathbf{A} be symmetric positive semi-definite, $\text{rank}(\mathbf{A}) = r$. Then \mathbf{A} can be decomposed as $\mathbf{A} = \mathbf{O}_{\mathbf{A}}^T \mathbf{O}_{\mathbf{A}}$ with $\mathbf{O}_{\mathbf{A}} \in \mathcal{M}_{r, M}$ and $\langle X_1, X_2 \rangle_{\mathbf{A}} = (\mathbf{O}_{\mathbf{A}} X_1, \mathbf{O}_{\mathbf{A}} X_2)_r$. Note that the row space of $\mathbf{O}_{\mathbf{A}}$ is the same for any choice of $\mathbf{O}_{\mathbf{A}}$ and coincides with the column space of \mathbf{A} . If the matrix \mathbf{A} is not positive definite, then we obtain a degenerate inner product, that is, if $\langle X, X \rangle_{\mathbf{A}} = 0$, then it is not necessary that $X = 0_M$. However, for vectors belonging to the column space of \mathbf{A} the equality $\langle X, X \rangle_{\mathbf{A}} = 0$ yields $X = 0_M$. Thus, if we consider inner product generated by a rank-deficient matrix \mathbf{A} , then we should consider it only on the column space of \mathbf{A} . In particular, we can correctly define \mathbf{A} -orthogonality of vectors from the column space of \mathbf{A} .

The following evident proposition shows that any basis can be considered as \mathbf{A} -orthonormal for some choice of $\mathbf{O}_{\mathbf{A}}$.

Proposition 2. *Let P_1, \dots, P_r be a set of linearly independent vectors in \mathbb{R}^M . Then P_1, \dots, P_r are \mathbf{A} -orthonormal for $\mathbf{O}_{\mathbf{A}} = \mathbf{P}^\dagger$, where $\mathbf{P} = [P_1 : \dots : P_r]$.*

Note that the column space of \mathbf{P} coincides with the row space of $\mathbf{O}_{\mathbf{A}}$. We call a matrix $\mathbf{O}_{\mathbf{A}}$ that makes a set P_1, \dots, P_r \mathbf{A} -orthonormal *orthonormalizing matrix* of this set. Certainly, the orthonormalizing matrix is not uniquely defined.

A.2 Oblique decompositions

Let us consider a minimal decompositions of $\mathbf{Y} \in \mathcal{M}_{L, K}$ of rank r in the form

$$(8) \quad \mathbf{Y} = \sum_{i=1}^r \sigma_i P_i Q_i^T,$$

where $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$, $\{P_i\}_{i=1}^r$ and $\{Q_i\}_{i=1}^r$ are linearly independent (therefore, $\{P_i\}_{i=1}^r$ is a basis of the column space of \mathbf{Y} , $\{Q_i\}_{i=1}^r$ is a basis of the row space of \mathbf{Y}). It is convenient to write (8) in the matrix form: $\mathbf{Y} = \mathbf{P} \mathbf{\Sigma} \mathbf{Q}^T$, where $\mathbf{P} = [P_1 : \dots : P_r]$, $\mathbf{Q} = [Q_1 : \dots : Q_r]$ and $\mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_r)$.

Proposition 3. Let \mathbf{O}_L be an orthonormalizing matrix of $\{P_i\}_{i=1}^r$ and \mathbf{O}_R be an orthonormalizing matrix of $\{Q_i\}_{i=1}^r$. Then

$$(9) \quad \mathbf{O}_L \mathbf{Y} \mathbf{O}_R^T = \sum_{i=1}^r \sigma_i (\mathbf{O}_L P_i) (\mathbf{O}_R Q_i)^T$$

is an SVD of $\mathbf{O}_L \mathbf{Y} \mathbf{O}_R^T \in \mathcal{M}_{r,r}$ with the left singular vectors $\mathbf{O}_L P_i \in \mathbb{R}^r$ and the right singular vectors $\mathbf{O}_R Q_i \in \mathbb{R}^r$.

This proposition follows from the fact that any bi-orthogonal decomposition is an SVD.

Definition 6. If the column space of \mathbf{L} contains the column space of \mathbf{Y} and the column space of \mathbf{R} contains the row space of \mathbf{Y} , then we will call such a pair (\mathbf{L}, \mathbf{R}) consistent with the matrix \mathbf{Y} .

Definition 7. For (\mathbf{L}, \mathbf{R}) consistent with \mathbf{Y} , we say that (8) is an (\mathbf{L}, \mathbf{R}) -SVD, if the system $\{P_i\}_{i=1}^r$ is \mathbf{L} -orthonormal and the system $\{Q_i\}_{i=1}^r$ is \mathbf{R} -orthonormal.

In a matrix statement of problem [23], the (\mathbf{L}, \mathbf{R}) -SVD is called Restricted SVD of \mathbf{Y} with respect to (\mathbf{L}, \mathbf{R}) .

It follows from Definition 7 that (9) is an SVD if and only if (8) is an (\mathbf{L}, \mathbf{R}) -SVD, where $\mathbf{L} = \mathbf{O}_L^T \mathbf{O}_L$ and $\mathbf{R} = \mathbf{O}_R^T \mathbf{O}_R$, \mathbf{O}_L and \mathbf{O}_R are orthonormalizing.

Proposition 3 says that any minimal decomposition into a sum of matrices of rank 1 in the form (8) is the (\mathbf{L}, \mathbf{R}) -SVD for some matrices \mathbf{L} and \mathbf{R} .

Proposition 4. Let

$$(10) \quad \mathbf{O}_L \mathbf{Y} \mathbf{O}_R^T = \sum_{i=1}^r \sqrt{\lambda_i} U_i V_i^T$$

be the ordinary SVD of $\mathbf{O}_L \mathbf{Y} \mathbf{O}_R^T$. Then the decomposition (8) with $\sigma_i = \sqrt{\lambda_i}$, $P_i = \mathbf{O}_L^\dagger U_i$ and $Q_i = \mathbf{O}_R^\dagger V_i$ is the (\mathbf{L}, \mathbf{R}) -SVD.

Proposition 4 follows from Proposition 3 and provides the method how the (\mathbf{L}, \mathbf{R}) -SVD can be calculated (see Algorithm 1).

Let us show how we can change the set of σ_i in the (\mathbf{L}, \mathbf{R}) -SVD (8) without change of directions of P_i and Q_i , that is, of $P_i/\|P_i\|$ and $Q_i/\|Q_i\|$.

Proposition 5. Let (8) be the (\mathbf{L}, \mathbf{R}) -SVD with $\mathbf{O}_L = \mathbf{P}^\dagger$ and $\mathbf{O}_R = \mathbf{Q}^\dagger$. Then

$$(11) \quad \mathbf{Y} = \sum_{i=1}^r \tilde{\sigma}_i \tilde{P}_i \tilde{Q}_i^T,$$

where $\tilde{\sigma}_i = \sigma_i/(\mu_i \nu_i)$, $\tilde{P}_i = \mu_i P_i$ and $\tilde{Q}_i = \nu_i Q_i$, is (after reordering of $\tilde{\sigma}_i$) the $(\tilde{\mathbf{L}}, \tilde{\mathbf{R}})$ -SVD with $\mathbf{O}_{\tilde{\mathbf{L}}} = \tilde{\mathbf{P}}^\dagger$ and $\mathbf{O}_{\tilde{\mathbf{R}}} = \tilde{\mathbf{Q}}^\dagger$.

The case of one-side non-orthogonal decompositions, when one of the matrices, \mathbf{R} or \mathbf{L} , is identical, is of special concern. It is shown in [23] that then Restricted SVD is Quotient SVD (often called Generalized SVD [26]). If \mathbf{L} is the identity matrix, then P_i , $i = 1, \dots, r$, are orthonormal in the conventional sense and form an orthonormal basis of the column space of \mathbf{Y} . If \mathbf{R} is the identity matrix, then Q_i , $i = 1, \dots, r$, are orthonormal and constitute an orthonormal basis of the row space.

A.3 Matrix scalar products and approximations

Let $\mathbf{X}, \mathbf{Y} \in \mathcal{M}_{L,K}$, (\mathbf{L}, \mathbf{R}) be consistent with both \mathbf{X} and \mathbf{Y} , $\mathbf{L} = \mathbf{O}_L^T \mathbf{O}_L$ and $\mathbf{R} = \mathbf{O}_R^T \mathbf{O}_R$.

Define the induced Frobenius inner product as

$$\langle \mathbf{X}, \mathbf{Y} \rangle_{F,(\mathbf{L},\mathbf{R})} = \langle \mathbf{O}_L \mathbf{X} \mathbf{O}_R^T, \mathbf{O}_L \mathbf{Y} \mathbf{O}_R^T \rangle_F.$$

Note that the definition does not depend on the choice of \mathbf{O}_L and \mathbf{O}_R , since $\langle \mathbf{C}, \mathbf{D} \rangle_F = \text{tr}(\mathbf{C}^T \mathbf{D}) = \text{tr}(\mathbf{C} \mathbf{D}^T)$.

For two matrices \mathbf{C} and \mathbf{D} we say that they

1. (\mathbf{L}, \mathbf{R}) F-orthogonal if $\langle \mathbf{C}, \mathbf{D} \rangle_{F,(\mathbf{L},\mathbf{R})} = 0$,
2. (\mathbf{L}, \mathbf{R}) -left orthogonal if $\mathbf{C}^T \mathbf{L} \mathbf{D} = \mathbf{0}$,
3. (\mathbf{L}, \mathbf{R}) -right orthogonal if $\mathbf{C} \mathbf{R} \mathbf{D}^T = \mathbf{0}$,
4. (\mathbf{L}, \mathbf{R}) bi-orthogonal if the left and right orthogonalities hold.

Left or right orthogonality is the sufficient condition for F-orthogonality. The matrix components of an (\mathbf{L}, \mathbf{R}) -SVD are (\mathbf{L}, \mathbf{R}) bi-orthogonal and therefore (\mathbf{L}, \mathbf{R}) F-orthogonal.

The measure of (\mathbf{L}, \mathbf{R}) orthogonality is

$$(12) \quad \rho_{(\mathbf{L},\mathbf{R})}(\mathbf{X}, \mathbf{Y}) = \frac{\langle \mathbf{X}, \mathbf{Y} \rangle_{F,(\mathbf{L},\mathbf{R})}}{\|\mathbf{X}\|_{F,(\mathbf{L},\mathbf{R})} \|\mathbf{Y}\|_{F,(\mathbf{L},\mathbf{R})}}.$$

Let $\mathbf{X} = \sum_i \mathbf{X}_i$, where $\mathbf{X}_i = \sigma_i P_i Q_i^T$, be the (\mathbf{L}, \mathbf{R}) -SVD. Then $\|\mathbf{X}_i\|_{F,(\mathbf{L},\mathbf{R})} = \sigma_i$ and $\|\mathbf{X}\|_{F,(\mathbf{L},\mathbf{R})}^2 = \sum_i \sigma_i^2$. The contribution of \mathbf{X}_k is equal to $\sigma_k^2 / \sum_i \sigma_i^2$.

The following proposition follows from the representation of the Frobenius scalar product through the trace of matrix multiplication.

Proposition 6. If \mathbf{X} and \mathbf{Y} are (\mathbf{L}, \mathbf{R}) left-orthogonal, then \mathbf{X} and \mathbf{Y} are $(\mathbf{L}, \tilde{\mathbf{R}})$ F-orthogonal for any $\tilde{\mathbf{R}}$.

Corollary 1. Let \mathbf{L} be the identity matrix and \mathbf{X} and \mathbf{Y} be (\mathbf{L}, \mathbf{R}) left-orthogonal for some matrix \mathbf{R} . Then the conventional F-orthogonality of \mathbf{X} and \mathbf{Y} holds and $\|\mathbf{X} + \mathbf{Y}\|_F^2 = \|\mathbf{X}\|_F^2 + \|\mathbf{Y}\|_F^2$.

Corollary 1 shows that if at least in either row or column matrix spaces the conventional inner product is given, that is, vectors are orthogonal in the ordinary sense, then the conventional F-orthogonality can be considered and F-norm and F-inner product can be used to measure the approximation accuracy and the component orthogonality.

Remark 7. The introduced definitions and statements are appropriate if \mathbf{L} and \mathbf{R} are consistent with the matrices \mathbf{X} and \mathbf{Y} (see Definition 6). Otherwise, e.g., (12) can be formally calculated, but this measure will reflect only the correlation between projections of columns and rows of \mathbf{X} and \mathbf{Y} on the row spaces of \mathbf{L} and \mathbf{R} correspondingly.

Let us remark that the conventional Frobenius norm is an interpretable characteristic of approximation, while the norm based on $\langle \cdot, \cdot \rangle_{F(\mathbf{L}, \mathbf{R})}$ is much worse interpretable, since it is equivalent to approximation by the Frobenius norm of the matrix $\mathbf{O}_L \mathbf{X} \mathbf{O}_R^T$.

ACKNOWLEDGMENTS

The authors are grateful to the anonymous reviewers and the editor for their useful comments and suggestions, and to Konstantin Usevich for fruitful discussions which helped to considerably improve the paper.

Received 27 October 2013

REFERENCES

- [1] ANDREWS, D. F. and HERZBERG, A. M. (1985). *Data. A Collection of Problems from Many Fields for the Student and Research Worker*. Springer, New York.
- [2] ATIKUR RAHMAN KHAN, M. and POSKITT, D. S. (2013). A note on window length selection in Singular Spectrum Analysis. *Aust. New Zeal. J. Stat.* **55** 87–108.
- [3] AWICHI, R. O. and MÜLLER, W. (2013). Improving SSA Predictions by inverse distance weighting. *REVSTAT* **13** 105–119.
- [4] BROOMHEAD, D. and KING, G. (1986). Extracting qualitative dynamics from experimental data. *Physica D* **20** 217–236.
- [5] DANILOV, D. and ZHIGLJAVSKY, A., eds. (1997). *Principal components of time series: the “Caterpillar” method*. St.Petersburg Press (<http://www.gistatgroup.com/gus/book1/>, in Russian).
- [6] ELSNER, J. B. and TSONIS, A. A. (1996). *Singular Spectrum Analysis: A New Tool in Time Series Analysis*. Plenum Press.
- [7] GHIL, M., ALLEN, R. M., DETTINGER, M. D., IDE, K., KONDRASHOV, D., MANN, M. E., ROBERTSON, A., SAUNDERS, A., TIAN, Y., VARADI, F. and YIOU, P. (2002). Advanced spectral methods for climatic time series. *Rev. Geophys.* **40** 1–41.
- [8] GILLARD, J. and KNIGHT, V. (2013). Using Singular Spectrum Analysis to obtain staffing level requirements in emergency units. *J. Oper. Res. Soc.* <http://dx.doi.org/10.1057/jors.2013.41>.
- [9] GOLYANDINA, N. (2010). On the choice of parameters in Singular Spectrum Analysis and related subspace-based methods. *Stat. Interface* **3** 259–279.
- [10] GOLYANDINA, N. and KOROBAYNIKOV, A. (2014). Basic Singular Spectrum Analysis and forecasting with R. *Comput. Stat. Data Anal.* **71** 934–954.
- [11] GOLYANDINA, N., NEKRUTKIN, V. and ZHIGLJAVSKY, A. (2001). *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman&Hall/CRC.
- [12] GOLYANDINA, N. and STEPANOV, D. (2005). SSA-based approaches to analysis and forecast of multidimensional time series. In *Proceedings of the 5th St.Petersburg Workshop on Simulation, June 26-July 2, 2005, St. Petersburg State University, St. Petersburg* 293–298.
- [13] GOLYANDINA, N. and ZHIGLJAVSKY, A. (2013). *Singular Spectrum Analysis for Time Series. Springer Briefs in Statistics*. Springer.
- [14] GOLYANDINA, N., KOROBAYNIKOV, A., SHLEMOV, A. and USEVICH, K. (2013). Multivariate and 2D extensions of Singular Spectrum Analysis with the Rssa package. *arXiv:1309.5050*.
- [15] HASSANI, H. and ZHIGLJAVSKY, A. (2009). Singular spectrum analysis: methodology and application to economics data. *J. Syst. Sci. Complex.* **22** 372–394.
- [16] HUDSON, I. L. and KEATLEY, M. R. (2010). Singular Spectrum Analysis: Climatic Niche Identification. In *Phenological Research* (I. L. Hudson and M. R. Keatley, eds.) 393–424. Springer Netherlands.
- [17] HYNDMAN, R. J. (2013). Time Series Data Library. Accessed on 10/08/2013.
- [18] ITOH, N. and KURTHS, J. (2011). Forecasting of EEG data by singular spectrum analysis. *Pacific Science Review* **13** 131–139.
- [19] ITOH, N. and MARWAN, N. (2013). An extended singular spectrum transformation (SST) for the investigation of Kenyan precipitation data. *Nonlinear Processes in Geophysics* **20** 467–481.
- [20] JOLLIFFE, I. T. (2002). *Principal Component Analysis*, 2nd ed. *Springer Series in Statistics*. Springer-Verlag, New York.
- [21] KOROBAYNIKOV, A. (2010). Computation- and space-efficient implementation of SSA. *Stat. Interface* **3** 357–368.
- [22] MARKOVSKY, I. (2012). *Low Rank Approximation: Algorithms, Implementation, Applications*. Springer.
- [23] DE MOOR, B. L. R. and GOLUB, G. H. (1991). The Restricted Singular Value Decomposition: Properties and applications. *SIAM J. Matrix Anal. Appl.* **12** 401–425.
- [24] NEKRUTKIN, V. (1997). Decompositions of time series. In *Principal components of time series: the Caterpillar method* (D. Danilov and A. Zhigljavsky, eds.) 194–227. St.Petersburg Press (<http://www.gistatgroup.com/gus/book1/>, in Russian).
- [25] NEKRUTKIN, V. (2010). Perturbation expansions of signal subspaces for long signals. *Stat. Interface* **3** 297–319.
- [26] PAIGE, C. and SAUNDERS, M. (1981). Towards a generalized singular value decomposition. *SIAM J. Numer. Anal.* **18** 398–405.
- [27] PATTERSON, K., HASSANI, H., HERAVI, S. and ZHIGLJAVSKY, A. (2011). Multivariate singular spectrum analysis for forecasting revisions to real-time data. *J. of Appl. Stat.* **38** 2183–2211.
- [28] ROY, R. and KAILATH, T. (1989). ESPRIT: estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust.* **37** 984–995.
- [29] RUCH, F. and BESTER, D. (2013). Towards a Measure of Core Inflation using Singular Spectrum Analysis. *S. Afr. J. Econ.* **81** 307–329.
- [30] USEVICH, K. (2010). On signal and extraneous roots in Singular Spectrum Analysis. *Stat. Interface* **3** 281–295.
- [31] VAN DER VEEN, A.-J., DEPRETTERE, E. F. and SWINDLEHURST, A. L. (1993). Subspace Based Signal Analysis using Singular Value Decomposition. In *Proceedings IEEE* **81** 1277–1308.
- [32] VAN LOAN, C. F. (1976). Generalizing the Singular Value Decomposition. *SIAM J. Numer. Anal.* **13** 76–83.
- [33] VAUTARD, R., YIOU, P. and GHIL, M. (1992). Singular-Spectrum Analysis: A toolkit for short, noisy chaotic signals. *Physica D* **58** 95–126.
- [34] ZHIGLJAVSKY, A. (Guest Editor, 2010). Special issue on the singular spectrum analysis in time series. *Stat. Interface* **3**.

Nina Golyandina
 Department of Statistical Modeling
 St. Petersburg State University
 Universitetsky pr 28, Peterhof 198504
 St. Petersburg, Russia
 E-mail address: nina@gistatgroup.com

Alex Shlemov
 Department of Statistical Modeling
 St. Petersburg State University
 Universitetsky pr 28, Peterhof 198504
 St. Petersburg, Russia
 E-mail address: shlemovalex@gmail.com