Statistics and Its Interface Volume 10 (2017) 47–57

Semi-nonparametric singular spectrum analysis with projection

Nina Golyandina*,† and Alex Shlemov†

Singular spectrum analysis (SSA) is a technique of time series analysis. The Basic SSA method is nonparametric and constructs an adaptive decomposition based on the singular value decomposition (SVD). We propose a modification of Basic SSA which we call SSA with projection. This version of SSA is able to take into consideration a structure given in advance. SSA with projection includes preliminary projection of rows and columns of the series' trajectory matrix to given subspaces. One application of SSA with projection is the extraction of polynomial trends. It is demonstrated that SSA with projection can extract polynomial trends much better than Basic SSA, especially in the case of linear trends. Numerical examples, including comparison with the least-squares polynomial regression, are presented.

AMS 2000 subject classifications: Primary 62G05, 94A12; secondary 60G35, 37M10.
Keywords and phrases: Singular spectrum analysis, Time series, Time series decomposition, Separability, Regression.

1. INTRODUCTION

Singular spectrum analysis (SSA) is able to solve a wide range of problems in the time series analysis, from the series decomposition on interpretable series components to forecasting, missing data imputation, parameter estimation and many others, see, e.g., [5, 9, 11, 23] and references within. The key feature of SSA is that the basic method is model-free, does not need a priori information and therefore constructs an adaptive decomposition of a time series into a sum of e.g. a non-parametric trend, periodic components and noise (see [6, 24, 2, 19, 21, 3, 17] among others in application of SSA to the problem of trend extraction). This can be considered as a major advantage of the SSA-family methods relative to parametric methods of time series analysis. However, sometimes a priori information about the considered time series is available. For example, the trend can be expected to be linear or polynomial.

The Basic SSA method [9, Chapter 1] consists of trajectory matrix construction from the original time series, its decomposition into a sum of rank-one matrices by SVD, their grouping and then each group's transformation back to time series for obtaining a decomposition of the original time series into a sum of identifiable components. The grouping of the SVD components can be considered as a projection of the trajectory matrix columns on a subspace, which is adaptively constructed on the base of certain distinguished features of the SVD decomposition. SSA with projection starts with projections of the trajectory matrix columns and rows on subspaces chosen in advance and follows with a decomposition of the residual, in the same way as in Basic SSA. It appears that SSA with centering [9, Section 1.7.1] is a particular case of SSA with projection, where projections are performed on the subspaces spanning the vectors with elements equal to 1. It is shown in [9] that SSA with single and double centering serves for extracting constant or linear trends with better accuracy.

Projections involved into SSA with projection generalize the centering procedure. A natural application of SSA with projection, which is the main subject of this paper, serves for extraction of polynomial trends; however, the suggested method can be applied for a wider range of problems, e.g., for the use of information about a supporting series.

Let us explain the motivation for the suggested approach, which can be considered as a semi-nonparametric variation of SSA, in more detail.

In SSA, the separability theory is responsible for a proper decomposition and component extraction. The separability of a series component means that the method is able to extract this time series component from the observed series, which is a sum of many components. Basic SSA is able to approximately separate a trend (e.g., a linear trend) from oscillations. However, there is no series, which can be exactly separated from a linear trend. As a consequence, the separation accuracy is not high. It is shown in [9, Sections 1.7.1 and 6.3.2] that SSA with double centering weakens the separability conditions and therefore improves the accuracy in the situation of approximate separability. Thus, it is expected that, within the SSA-family methods, SSA with projection can improve separability for components of a specific structure, which is in accordance with the projection subspaces.

Let us compare the least squares approach to the estimation of regression parameters and SSA with double centering. The linear regression method minimizes the prediction error using the least-squares method, while SSA tries

*Corresponding author.
†The work was partially supported by the RFBR grant 16-04-00821.

to reconstruct values of the series components using their orthogonality. For example, for a series with common term $x_n = t_n + s_n$, where $t_n = an + b$ and $s_n = A\sin(2\pi\omega n + \phi)$, the least-squares approach generally cannot estimate the linear trend t_n without error, while in the conditions of separability SSA with double centering is able to find the exact linear trend. This advantage of SSA is more visible for short time series; for long time series, linear regression and SSA yield close estimates of the linear trend. Note that for the case of approximate separability the linear-trend estimate found by SSA with double centering will be only close to a straight line, while the linear regression always provides a linear function as a trend estimation.

The analogous relation between the parametric regression and SSA with projection is expected for the general case of polynomial trends. In particular, we can suppose that for time series with seasonality the 'SSA with projection' method will be able to extract linear and polynomial trends more accurately than the parametric regression approach.

It is important that the use of projection on a fixed basis does not contradict the non-parametric nature of SSA. If the basis for projection is chosen incorrectly, the trend estimate by SSA with projection will not have a considerable bias, since it can be accomplished by components of the adaptive part of the whole decomposition. This is not the case for the parametric approach.

The structure of the paper is as follows. We start with a brief description of the Basic SSA algorithm and standard separability notion (Section 2). Section 3 is devoted to the generalization of centering used in SSA and contains the underlying theory, including the proof of the algorithm and the separability conditions. Section 4 demonstrates the examples of the algorithm application for trend extraction. The real-life examples are studied in Sections 4.1 and 4.2 to show the relation between Basic SSA, SSA with projection and the linear regression (least-squares) approach. Numerical comparison is performed in Section 4.3. The paper is summarized and conclusions are drawn in Section 5.

Finally, let us briefly comment the title of this paper. Approaches, which deal with a combination of parametric and nonparametric models, are sometimes called semiparametric if the parametric part of the model is of interest and semi-nonparametric if both parts are important (see the references [4] and [13] as examples of such approaches to statistical econometric problems). SSA with projection is exactly the latter.

2. BACKGROUND

2.1 Algorithm of Basic SSA

Consider a real-valued time series $\mathbb{X} = \mathbb{X}_N = (x_1, \ldots, x_N)$ 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 $L \times K$ and $\mathcal{M}^{(H)}_{L,K}$ the space of Hankel matrices of size $L \times K$. Consider the *lagged vectors* $X_i = (x_i, \ldots, x_{i+L-1})^{\mathrm{T}}$, $i = 1, \ldots, K$, and the *trajectory matrix* $\mathbf{X} = [X_1 : \ldots : X_K] \in \mathcal{M}^{(H)}_{L,K}$ of the series $\mathbb{X}_N$.

Define the one-to-one embedding operator $\mathcal{T} : \mathsf{R}^N \mapsto \mathcal{M}^{(H)}_{L,K}$ as $\mathcal{T}(\mathbb{X}_N) = \mathbf{X}$. Also introduce the projector $\mathscr{H}$ (in Frobenius norm) of $\mathcal{M}_{L,K}$ to $\mathcal{M}^{(H)}_{L,K}$. Projection is performed by replacing the entries on antidiagonals $i + j = \text{const}$ to their averages along the antidiagonals.

The Basic SSA algorithm consists of two stages and four steps.

2.1.1 Decomposition stage

1st step: Embedding. Let L be chosen. At this step the L-trajectory matrix is composed: $\mathbf{X} = \mathcal{T}(\mathbb{X}_N)$.

2nd step: Singular Value Decomposition (SVD). The SVD of the trajectory matrix is constructed:

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

where λ_i are the eigenvalues of $\mathbf{X}\mathbf{X}^{\mathrm{T}}$, $\sqrt{\lambda_i}$ are called singular values of $\mathbf{X}$, U_i and V_i are left and right singular vectors of $\mathbf{X}$, $\lambda_1 \geq \ldots \geq \lambda_d > 0$, $d = \operatorname{rank}(\mathbf{X})$.

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

2.1.2 Reconstruction stage

3rd step: Eigentriple grouping. The grouping procedure partitions the set of indices $\{1, \ldots, d\}$ into m disjoint subsets $I_1, \ldots, I_m$.

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

$$\mathbf{X} = \mathbf{X}_{I_1} + \ldots + \mathbf{X}_{I_m}. \quad (2)$$

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

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

2.1.3 The resultant decomposition and projections

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

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

A typical example of (3) is the decomposition into a sum of a trend, oscillations and noise.

Remark 1. *Columns of a grouped matrix* $\mathbf{X}_I$ *are the orthogonal projections of columns of the trajectory matrix* $\mathbf{X}$ *to* $\operatorname{span}(U_i,\, i \in I)$ *with respect to the Euclidean norm. Rows of* $\mathbf{X}_I$ *are the orthogonal projections of rows of* $\mathbf{X}$ *to* $\operatorname{span}(V_i,\, i \in I)$.

2.2 Separability by Basic SSA

For understanding how SSA works, the notion of separability is very important. Separability of two time series $\mathbb{X}_N^{(1)}$ and $\mathbb{X}_N^{(2)}$ signifies the possibility of extracting $\mathbb{X}_N^{(1)}$ and $\mathbb{X}_N^{(2)}$ from the observed sum $\mathbb{X}_N = \mathbb{X}_N^{(1)} + \mathbb{X}_N^{(2)}$. This means that there exists a grouping at Grouping step such that $\widetilde{\mathbb{X}}_N^{(k)} = \mathbb{X}_N^{(k)}$, $k = 1, 2$.

Definition 1. The column and row spaces of the trajectory matrix of a series are called column and row spaces of the series respectively.

Properties of the SVD give rise to the following definition of separability.

Definition 2. The Basic SSA (weak) separability is defined as the orthogonality of the column and row spaces of the series $\mathbb{X}_N^{(1)}$ and $\mathbb{X}_N^{(2)}$.

In the case of approximate (asymptotic) Basic SSA separability, where $\widetilde{\mathbb{X}}_N^{(k)} \approx \mathbb{X}_N^{(k)}$, we obtain the condition of approximate (asymptotic) orthogonality.

For sufficiently long time series, Basic SSA can approximately separate, for example, a signal and noise, sine waves with different frequencies, a trend and a seasonality [9, 11].

The separability introduced through orthogonality of the column and row spaces of separated series is called weak separability; it means that at the SVD step there exists such an SVD that allows the proper grouping. Strong separability means that each SVD decomposition allows the proper grouping. Several nonparametric modifications of SSA for improvement of the weak and strong separability are considered in [10]. In this paper we will improve the separability by a semi-nonparametric variation of Basic SSA.

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. Following [9, Chapter 5], we define the L-rank of a series $\mathbb{X}_N$ as the rank of its L-trajectory matrix. Rank-deficiency of trajectory matrices of exactly separated series arises from the condition of orthogonality of their column and row spaces. Therefore, 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 $K \geq L \geq r$.

Under some unrestrictive conditions [9, Section 5.2], series $\mathbb{S}_N$ of finite rank r is governed by a linear recurrence relation (LRR) of order r, that is,

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

The LRR (4) is called *minimal* and r is called the dimension of the series. Note that there can be many LRRs governing the time series; however, the minimal LRR is unique. Let us describe how we can restore the form of the time series by means of the minimal LRR.

Definition 3. The 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 an infinite time series $\mathbb{S}_\infty = (s_1, \ldots, s_n, \ldots)$ satisfy the LRR (4) for $i \geq 1$, that is, without an upper limit. Consider the characteristic polynomial of the LRR (4) and denote its different (complex) roots by $\mu_1, \ldots, \mu_p$, 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 $k_1 + \ldots + k_p = r$. We will call μ_j *characteristic roots* of the series governed by an LRR.

It is well-known that the time series $\mathbb{S}_\infty = (s_1, \ldots, s_n, \ldots)$ satisfies the LRR (4) for all $i \geq 0$ if and only if

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

where the coefficients c_{mj} are determined by the first r series terms [9, Section 5.2]. 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.

Rank of the series is determined by the number of nonzero terms and degrees of polynomials in (5). For example, an exponentially-modulated sinusoid $s_n = Ae^{\alpha n} \sin(2\pi\omega n + \phi)$ is constructed from two conjugate complex roots $\mu_{1,2} = e^{\alpha \pm i2\pi\omega} = \rho e^{\pm i2\pi\omega}$ if its frequency ω belongs to $(0, 0.5)$. Therefore, the rank of this exponentially-modulated sinusoid is equal to 2. The rank of an exponential is equal to 1, the rank of a linear function corresponding to the root 1 of multiplicity 2 equals 2, and so on. The paper [22] contains conditions of separability expressed in terms of characteristic roots.

Also, the representation (5) helps to easily construct bases of column spaces of complex time series governed by LRRs: the basis is constructed from the linearly independent vectors $\left(1^j \mu_m^1, 2^j \mu_m^2, \ldots, L^j \mu_m^L\right)^{\mathrm{T}}$, $j = 0, \ldots, k_m - 1$. For linear series, the basis consists of $(1, 1, \ldots, 1)^{\mathrm{T}}$ and $(1, 2, \ldots, L)^{\mathrm{T}}$.

2.4 SSA with centering

There are modifications of SSA called SSA with centering [9, Sections 1.7 and 6.3]. They serve for better separation of constant (SSA with single centering) and linear (SSA with double centering) trends. Initially, SSA with single centering was created by analogy with Principal Component Analysis, where the columns of a data matrix (which can be interpreted as the rows of the trajectory matrix of a series) are centered before the application of the SVD. Let us describe the approach with centering.

Let us consider a time series $\mathbb{X}$ of length N, a window length L, $K = N - L + 1$, the trajectory matrix $\mathbf{X}$ of the series $\mathbb{X}$.

A general form of the considered modification of Decomposition stage can be expressed as

1. Calculation of a special matrix $\mathbf{C}^{(\text{center})} = \mathbf{C}(\mathbf{X})$ based on a priori information.
2. Computation of $\mathbf{X}^\star = \mathbf{X} - \mathbf{C}^{(\text{center})}$.
3. Construction of the SVD: $\mathbf{X}^\star = \sum_{i=1}^{d^\star} \sqrt{\lambda_i^\star} U_i^\star (V_i^\star)^{\mathrm{T}}$.

Thus, we have the decomposition $\mathbf{X} = \mathbf{C}^{(\text{center})} + \sum_{i=1}^{d^\star} \sqrt{\lambda_i^\star} U_i^\star (V_i^\star)^{\mathrm{T}}$.

Denote $E_M = (1, \ldots, 1)^{\mathrm{T}} \in \mathsf{R}^M$ the vector of units. Centering is considered in the following forms:

- *Single row centering* when $\mathbf{C}^{(\text{center})}_{\text{row}}(\mathbf{X}) = (\mathbf{X} E_K / K) E_K^{\mathrm{T}}$ corresponds to averaging by rows, that is, each element of a row of $\mathbf{C}^{(\text{center})}_{\text{row}}$ consists of the average of the corresponding row of the trajectory matrix.
- *Single column centering* when $\mathbf{C}^{(\text{center})}_{\text{col}}(\mathbf{X}) = E_L (E_L^{\mathrm{T}} \mathbf{X} / L)$ corresponds to averaging by columns.
- *Double centering* when $\mathbf{C}^{(\text{center})}_{\text{both}} = \mathbf{C}^{(\text{center})}_{\text{row}} + \mathbf{C}^{(\text{center})}_{\text{col}}\big(\mathbf{X} - \mathbf{C}^{(\text{center})}_{\text{row}}(\mathbf{X})\big)$ corresponds to averaging by both rows and columns.

Note that the centering can be considered as a projection of rows and/or columns of $\mathbf{X}$ on $\operatorname{span}(E_K)$ or $\operatorname{span}(E_L)$ respectively, since $E_K E_K^{\mathrm{T}}$ and $E_L E_L^{\mathrm{T}}$ are exactly the matrices of the projection operators. Therefore, centering in SSA can be considered as a preliminary projection of the trajectory matrix on a given subspace; the residual matrix $\mathbf{X}^\star$ is subsequently expanded by SVD or any other decomposition.

3. SSA WITH PROJECTION

Let us generalize the approach described in Section 2.4 involving projections to arbitrary spaces.

Notation. Let $\Pi_{\text{col}} : \mathsf{R}^L \to \mathcal{L}_{\text{col}}$ and $\Pi_{\text{row}} : \mathsf{R}^K \to \mathcal{L}_{\text{row}}$ be orthogonal projectors, where $\mathcal{L}_{\text{col}} \in \mathsf{R}^L$ is called the column projection space and $\mathcal{L}_{\text{row}} \in \mathsf{R}^K$ is called the row projection space. For any $\mathbf{Y} \in \mathcal{M}_{L,t}$, denote $\Pi_{\text{col}}(\mathbf{Y})$ the matrix consisting of the columns, which result from projections of the columns of $\mathbf{Y}$, while for any $\mathbf{Y} \in \mathcal{M}_{t,K}$ denote $\Pi_{\text{row}}(\mathbf{Y})$ the matrix consisting of the rows, which result from projections of the rows of $\mathbf{Y}$.

Denote a basis of the column projection space $(P_i, i = 1, \ldots, p)$ and a basis of the row projection space $(Q_i, i = 1, \ldots, q)$, $\mathbf{P} = [P_1 : \ldots : P_p]$, $\mathbf{Q} = [Q_1 : \ldots : Q_q]$. Without loss of generality we assume that $\{P_i, i = 1, \ldots, p\}$ and $\{Q_i, i = 1, \ldots, q\}$ are orthonormal bases of $\mathcal{L}_{\text{col}}$ and $\mathcal{L}_{\text{row}}$ (otherwise, we can perform ortho-normalization).

In SSA with projection, the scheme of SSA with centering, which is described in Section 2.4, is extended to arbitrary projections, that is, $\mathbf{C} = \Pi_{\text{col}}(\mathbf{X})$ for column projection, $\mathbf{C} = \Pi_{\text{row}}(\mathbf{X})$ for row projection and $\mathbf{C} = \Pi_{\text{both}}(\mathbf{X})$ for double projection, where

$$
\begin{aligned}
\Pi_{\text{both}}(\mathbf{X}) &= \Pi_{\text{row}}(\mathbf{X}) + \Pi_{\text{col}}(\mathbf{X} - \Pi_{\text{row}}(\mathbf{X})) \\
&= \Pi_{\text{col}}(\mathbf{X}) + \Pi_{\text{row}}(\mathbf{X} - \Pi_{\text{col}}(\mathbf{X})) \\
&= \Pi_{\text{row}}(\mathbf{X}) + \Pi_{\text{row}}(\mathbf{X}) - (\Pi_{\text{col}} \circ \Pi_{\text{row}})(\mathbf{X}).
\end{aligned} \tag{6}
$$

If either the column or row basis is absent (that is, the corresponding projection should not be performed), then we formally set the corresponding projector to be the zero operator implying $\mathbf{C} = \Pi_{\text{both}}(\mathbf{X})$ for any mode.

Decomposition into elementary matrices. A general form of the decomposition provided by SSA with projection is

$$
\mathbf{X} = \mathbf{C} + \sum_{i=1}^{d^\star} \sqrt{\lambda_i^\star} U_i^\star (V_i^\star)^{\mathrm{T}}, \tag{7}
$$

where $\mathbf{C} = \Pi_{\text{both}}(\mathbf{X})$ and $\sum_{i=1}^{d^\star} \sqrt{\lambda_i^\star} U_i^\star (V_i^\star)^{\mathrm{T}}$ is the SVD of $\mathbf{X} - \mathbf{C}$. Let us demonstrate that the matrix $\mathbf{C}$ can be presented as a sum of elementary matrices of rank 1. Then (7) will be similar to the result of decomposition (1) performed by Basic SSA and therefore the reconstruction stage will be also similar to that of Basic SSA.

Note that $\Pi_{\text{col}}(\mathbf{Y}) = \mathbf{P}\mathbf{P}^{\mathrm{T}}\mathbf{Y} = \sum_{i=1}^{p} P_i (\mathbf{Y}^{\mathrm{T}} P_i)^{\mathrm{T}}$ and $\Pi_{\text{row}}(\mathbf{Y}) = \mathbf{Y}\,\mathbf{Q}\mathbf{Q}^{\mathrm{T}} = \sum_{i=1}^{q} (\mathbf{Y} Q_i) Q_i^{\mathrm{T}}$ are decompositions into sums of elementary rank-one matrices. Therefore, $\mathbf{C} = \Pi_{\text{both}}(\mathbf{X})$ also can be expanded to a sum of elementary matrices, since Π_{both} can be expressed as a sequential application of the projection operators Π_{row} and Π_{col}, see (6). For double projection, this expansion depends on the order of projections; for definiteness, we will apply the row projector first.

Thus, the matrix $\mathbf{C}$ can be considered as a sum of $p + q$ elementary matrices of the forms $\sigma_i^{(c)} P_i \widetilde{Q}_i^{\mathrm{T}}$, $i = 1, \ldots, p$, and $\sigma_i^{(r)} \widetilde{P}_i Q_i^{\mathrm{T}}$, $i = 1, \ldots, q$ (some of them can be zero), where the triples $(\sigma_i^{(c)}, P_i, \widetilde{Q}_i)$ and $(\sigma_i^{(r)}, \widetilde{P}_i, Q_i)$ have the same meaning as eigentriples. Therefore, the decomposition (7) can be transformed to a decomposition into a sum of $d^\star + p + q$ elementary rank-one matrices, which are orthogonal with respect to the Frobenius norm $\|\cdot\|$, by construction. As a consequence, contribution of the projection term $\mathbf{C}$ into the decomposition is given by $\|\mathbf{C}\|^2 / \|\mathbf{X}\|^2$;

The following lemma describes properties of the decomposition (7).

Lemma 1. *The decomposition (7) satisfies the following properties:*

1. $d^\star \le \operatorname{rank} \mathbf{X}$*;*
2. $d^\star \ge \operatorname{rank} \mathbf{X} - (p + q)$*; the equality holds if* P_i, $i = 1, \ldots, p$, *belong to the column span of* $\mathbf{X}$ *and* Q_i, $i = 1, \ldots, q$, *belong to the row span of* $\mathbf{X}$.

Proof. Due to the definition of orthogonal projection, $\mathbf{X}^\star = \mathbf{X} - \mathbf{C} = (\mathbf{I}_L - \mathbf{P}\mathbf{P}^{\mathrm{T}})\mathbf{X}(\mathbf{I}_K - \mathbf{Q}\mathbf{Q}^{\mathrm{T}})$, where $\mathbf{I}_M$ denotes the $M \times M$ identity matrix. Therefore, $d^\star \le \operatorname{rank} \mathbf{X}$. On the other hand, rank $\mathbf{X}$ does not exceed the number of elementary matrices $d^\star + p + q$; the equality is attained if the rows and columns of the projected matrices belong to the row and column spaces of $\mathbf{X}$ respectively. □

3.1 Algorithm

Let us summarize the steps of SSA with projection in the form of algorithms, splitting the whole algorithm into decomposition and reconstruction.

ALGORITHM 3.1. *SSA with projection: decomposition*

Input: The time series $\mathbb{X}$ of length N, the window length L, an orthonormal basis of the column projection space $(P_i, i = 1, \ldots, p)$ and an orthonormal basis of the row projection space $(Q_i, i = 1, \ldots, q)$. Either p or q can be zero.

Output: Decomposition of the trajectory matrix on elementary matrices $\mathbf{X} = \mathbf{X}_1 + \ldots + \mathbf{X}_d$, where $\mathbf{X}_i = \sigma_i U_i V_i^{\mathrm{T}}$ are either zero or rank-one matrices.

1: Construct the trajectory matrix $\mathbf{X} = \mathcal{T}_{\mathrm{SSA}}(\mathbb{X})$.

2: Subtract the row projection: $\mathbf{X}' = \mathbf{X} - \mathbf{C}_{\mathrm{row}}$, where

$$\mathbf{C}_{\mathrm{row}} = \Pi_{\mathrm{row}}(\mathbf{X}) = \sum_{i=1}^{q} \sigma_i^{(r)} \widetilde{P}_i Q_i^{\mathrm{T}},$$

$\sigma_i^{(r)} = \|\mathbf{X}Q_i\|$, $\widetilde{P}_i = \mathbf{X}Q_i/\sigma_i^{(r)}$ if $\sigma_i^{(r)} > 0$; otherwise, $\widetilde{P}_i$ is the zero vector.

3: Subtract the column projection: $\mathbf{X}^\star = \mathbf{X}' - \mathbf{C}_{\mathrm{col}}$, where

$$\mathbf{C}_{\mathrm{col}} = \Pi_{\mathrm{col}}(\mathbf{X}') = \sum_{i=1}^{p} \sigma_i^{(c)} P_i \widetilde{Q}_i^{\mathrm{T}},$$

$\sigma_i^{(c)} = \|\mathbf{X}'^{\mathrm{T}} P_i\|$, $\widetilde{Q}_i = \mathbf{X}'^{\mathrm{T}} P_i/\sigma_i^{(c)}$ if $\sigma_i^{(c)} > 0$; otherwise, $\widetilde{Q}_i$ is the zero vector.

4: Construct an SVD of the matrix $\mathbf{X}^\star$: $\mathbf{X}^\star = \sum_{i=1}^{d^\star} \mathbf{X}_i^\star$, where $\mathbf{X}_i^\star = \sqrt{\lambda_i^\star} U_i^\star (V_i^\star)^{\mathrm{T}}$.

5: As a result, $\mathbf{X} = \sum_{i=1}^{d} \mathbf{X}_i$, where $d = p + q + d^\star$, $\mathbf{X}_i = \sigma_i^{(r)} \widetilde{P}_i Q_i^{\mathrm{T}}$ for $i = 1, \ldots, q$, $\mathbf{X}_{i+q} = \sigma_i^{(c)} P_i \widetilde{Q}_i^{\mathrm{T}}$ for $i = 1, \ldots, p$, and $\mathbf{X}_{i+p+q} = \sqrt{\lambda_i^\star} U_i^\star (V_i^\star)^{\mathrm{T}}$ for $i = 1, \ldots, d^\star$.

To complete the algorithm of SSA with projection, let us describe the algorithm of Reconstruction stage.

ALGORITHM 3.2. *SSA with projection: reconstruction*

Input: Decomposition $\mathbf{X} = \mathbf{X}_1 + \ldots + \mathbf{X}_d$ and grouping $\{1, \ldots, d\} = \bigsqcup_{j=1}^{m} I_j$, which does not split the first $p + q$ projection components, where q and p are the numbers of row and column projection components.

Output: Decomposition of time series on identifiable components $\mathbb{X} = \mathbb{X}_1 + \ldots + \mathbb{X}_m$.

1: Construct the grouped matrix decomposition $\mathbf{X} = \mathbf{X}_{I_1} + \ldots + \mathbf{X}_{I_m}$, where $\mathbf{X}_I = \sum_{i \in I} \mathbf{X}_i$.

2: Compute $\mathbb{X} = \mathbb{X}_1 + \ldots + \mathbb{X}_m$, where $\mathbb{X}_i = \mathcal{T}^{-1}\mathscr{H}(\mathbf{X}_{I_i})$.

The only essential difference with the reconstruction by Basic SSA is that the set of the matrices $\mathbf{X}_i$, $i = 1, \ldots, p + q$, which is produced by projections, should be included in the same group. Then the resultant series decomposition does not depend on the selected bases $\{P_i, i = 1, \ldots, p\}$ and $\{Q_i, i = 1, \ldots, q\}$.

Remark 2. *Algorithms 3.1 and 3.2 can be applied to multidimensional objects $\mathbb{X}$ without changes. The only difference is in the definition of operators $\mathcal{T}$ and $\mathscr{H}$; see [12, 20] for description of multidimensional and shaped SSA extensions.*

3.2 Appropriate class of time series

For SSA with projection, the natural question is what series are preserved with projection; that is, for what kinds of series with a trajectory matrix $\mathbf{X}$ we have $\Pi_{\mathrm{col}}(\mathbf{X}) = \mathbf{X}$ for column projection, $\Pi_{\mathrm{row}}(\mathbf{X}) = \mathbf{X}$ for row projection and $\Pi_{\mathrm{both}}(\mathbf{X}) = \mathbf{X}$ for double projection.

The following lemma is a direct consequence of the definition of projection.

Lemma 2. *Let $\mathcal{L}_{\mathrm{row}}$ contain the row space of a matrix $\mathbf{X}$ and $\mathcal{L}_{\mathrm{col}}$ contain the column space of a matrix $\mathbf{X}$. Then $\Pi_{\mathrm{row}}(\mathbf{X}) = \mathbf{X}$ and $\Pi_{\mathrm{col}}(\mathbf{X}) = \mathbf{X}$.*

For example, it follows from Lemma 2 that to preserve an exponential series with $s_n = C\mu^n$ by SSA with column projection, the column projection should be performed to a space which contains $\mathrm{span}\big((\mu, \mu^2, \ldots, \mu^L)^{\mathrm{T}}\big)$, while to preserve a linear function with $s_n = an + b$ for any b and nonzero a, the column projection should be performed to a space which contains $\mathrm{span}\big((1, 1, \ldots, 1)^{\mathrm{T}}, (1, 2, \ldots, L)^{\mathrm{T}}\big)$.

Let us derive a condition sufficient for $\Pi_{\mathrm{both}}(\mathbf{X}) = \mathbf{X}$ to hold for the general case of the double projection.

Lemma 3. *Let the columns of a matrix $\mathbf{W} \in \mathcal{M}_{K,q}$ belong to $\mathcal{L}_{\mathrm{row}}$ and the columns of a matrix $\mathbf{S} \in \mathcal{M}_{L,p}$ belong to $\mathcal{L}_{\mathrm{col}}$. Then $\Pi_{\mathrm{both}}(\mathbf{X}) = \mathbf{X}$ for*

$$\mathbf{X} = \widetilde{\mathbf{S}}\mathbf{W}^{\mathrm{T}} + \mathbf{S}\widetilde{\mathbf{W}}^{\mathrm{T}} \tag{8}$$

for any $\widetilde{\mathbf{S}} \in \mathcal{M}_{L,q}$ and $\widetilde{\mathbf{W}} \in \mathcal{M}_{K,p}$.

Proof. By the assumption, $\Pi_{\mathrm{row}}(\mathbf{A}\mathbf{W}^{\mathrm{T}}) = \mathbf{A}\mathbf{W}^{\mathrm{T}}$ for any matrix $\mathbf{A} \in \mathcal{M}_{t,q}$, where $t \geq 1$, while $\Pi_{\mathrm{col}}(\mathbf{S}\mathbf{B}^{\mathrm{T}}) = \mathbf{S}\mathbf{B}^{\mathrm{T}}$ for any matrix $\mathbf{B} \in \mathcal{M}_{t,p}$.

$$\Pi_{\mathrm{both}}(\mathbf{X}) = \widetilde{\mathbf{S}}\mathbf{W}^{\mathrm{T}} + \Pi_{\mathrm{row}}(\mathbf{S}\widetilde{\mathbf{W}}^{\mathrm{T}}) + \mathbf{S}\widetilde{\mathbf{W}}^{\mathrm{T}} + \Pi_{\mathrm{col}}(\widetilde{\mathbf{S}}\mathbf{W}^{\mathrm{T}}) - \Pi_{\mathrm{col}}(\Pi_{\mathrm{row}}(\widetilde{\mathbf{S}}\mathbf{W}^{\mathrm{T}} + \mathbf{S}\widetilde{\mathbf{W}}^{\mathrm{T}})) = \mathbf{X},$$

since $\Pi_{\mathrm{col}} \circ \Pi_{\mathrm{row}} \equiv \Pi_{\mathrm{row}} \circ \Pi_{\mathrm{col}}$ by the associativity of matrix multiplication. □

It is easy to check that the trajectory matrix of a linear series satisfies the conditions of Lemma 3 for the case of double centering. However, for a general case an approach based on characteristic roots is more convenient. We start with a technical lemma.

Lemma 4. *For any polynomial P_d of degree d and for any l and k such that $l+k=d-1$ the following expansion can be constructed:*

$$P_d(i+j) = P_{l,d}(i,j) + P_{d,k}(i,j),$$

where $P_{\tau,\varkappa}(i,j)$ denotes a polynomial of i and j of bidegree $(\tau,\varkappa)$.

Proof. This lemma is proved by an appropriate grouping of the monomials $C_{u,v} i^u j^v$, $u+v \le d$, of $P_d(i+j)$. □

Recall that a series governed by an LRR, whose characteristic polynomial has the given set of roots called characteristic roots, is of the form (5).

Theorem 1. *Let series $\mathbb{Y}^{(m)}$, $m=1,2$, be governed by minimal LRRs of orders r_m, $\mathbf{Y}^{(m)}$ be their trajectory matrices. Let $\{\mu_j;\ j=1,\ldots,s\}$ be the set containing the characteristic roots of both series. Assume that $\mathbb{Y}^{(m)}$, $m=1,2$, have the characteristic roots μ_j, $j=1,\ldots,s$, with multiplicities $d_j^{(m)} \ge 0$, $\sum_{j=1}^s d_j^{(m)} = r_m$. Let Π_{col} be the projector on the column space $\mathscr{C}$ of $\mathbf{Y}^{(1)}$, Π_{row} be the projector on the row space $\mathscr{R}$ of $\mathbf{Y}^{(2)}$, Π_{both} be given in (6). Then $\Pi_{\mathrm{both}}(\mathbf{X}) = \mathbf{X}$ if and only if the set of characteristic roots of the series $\mathbb{X}$ consists of the roots μ_j, $j=1,\ldots,s$, of multiplicities $d_j \le d_j^{(1)} + d_j^{(2)}$.*

Proof. Due to linearity of projectors and linear dependence of Π_{both} on Π_{row} and Π_{col}, it is sufficient to prove the theorem for the case of one root μ. Let $\mathbb{Y}^{(1)}$ have the characteristic root μ of multiplicity p, $\mathbb{Y}^{(2)}$ have the characteristic root μ of multiplicity q.

Thus, we should prove that $\Pi_{\mathrm{both}}(\mathbf{X}) = \mathbf{X}$ if and only if the series $\mathbb{X}$ has the form $x_k = P_t(k)\mu^k$, where $t \le p+q-1$. It is sufficient to consider $t = p+q-1$.

By Lemma 4

$$\begin{aligned} &P_{p+q-1}(i+j)\mu^{i+j} \\ &\qquad = P_{p-1,p+q-1}(i,j)\mu^i\mu^j + P_{p+q-1,q-1}(i,j)\mu^i\mu^j. \end{aligned}$$

This means that (8) holds for $\mathbf{W} \in \mathcal{M}_{K,q}$ and $\mathbf{S} \in \mathcal{M}_{L,p}$ such that the column space of $\mathbf{W}$ coincides with $\mathscr{R}$ and the column space of $\mathbf{S}$ coincides with $\mathscr{C}$.

Since the dimension of the space of trajectory matrices that are preserved by the projector Π_{both} is equal to $r = r_1 + r_2$, we found all such matrices. This completes the proof. □

Corollary 1. *Let $\mathbb{Y}$ be a series of dimension r, $\mathbf{Y}$ be its trajectory matrix, Π_{row} be the projector on its row trajectory space, Π_{col} be the projector on its column trajectory space. Consider the series $\mathbb{X}$ with $x_n = (an+b)y_n$. Then $\Pi_{\mathrm{both}}(\mathbf{X}) = \mathbf{X}$.*

Remark 3. *Note that multiplication of a series by $an+b$, where $a \neq 0$, means that the multiplicities of the series characteristic roots increase by 1.*

Note that formally the sets $\{P_i, i=1,\ldots,p\}$ and $\{Q_i, i=1,\ldots,q\}$ can be arbitrary. However, if the model of the series is partly known, then in the context of SSA this means that a time series component satisfies an LRR and we know its characteristic roots (see Section 2.3). Therefore, to extract, for example, a sine wave using projections, we should know its period, and to extract an exponential trend, we should know its rate. Such conditions are often too restrictive. A clear exception is extraction of polynomial trends of a degree k, when there is the unique characteristic root equal to 1 of multiplicity $k+1$ and we should assume only the degree of the polynomial trend to obtain its trajectory space.

Corollary 2. *Let Π_{row} be the projector on the row trajectory space of a polynomial of degree l, Π_{col} be the projection on the column trajectory space of a polynomial of degree k. Then for any polynomial $\mathbb{X} = P_{l+k+1}$ of degree $l+k+1$ we have $\Pi_{\mathrm{both}}(\mathbf{X}) = \mathbf{X}$.*

Remark 4. *It immediately follows from Lemma 2 that in the conditions of Corollary 2, for any polynomial $\mathbb{X} = P_l$ of degree l we have $\Pi_{\mathrm{row}}(\mathbf{X}) = \mathbf{X}$ and for any polynomial $\mathbb{X} = P_k$ of degree k we have $\Pi_{\mathrm{col}}(\mathbf{X}) = \mathbf{X}$.*

3.3 Separability

We expect that if a time series component is governed by a minimal LRR and this LRR is known, then the series component can be separated by a suitable version of SSA with projection better than it can be done by Basic SSA.

Using the notion of separability, we can formulate this improvement as follows. Let $\mathbb{X} = \mathbb{X}^{(1)} + \mathbb{X}^{(2)}$. We will say that a time series component $\mathbb{X}^{(1)}$ is separated by SSA with projection if $\mathbf{X}^{(1)} = \mathbf{C}$, where $\mathbf{C}$ is equal to $\Pi_{\mathrm{row}}(\mathbf{X})$, $\Pi_{\mathrm{col}}(\mathbf{X})$ or $\Pi_{\mathrm{both}}(\mathbf{X})$, in dependence on the type of projection.

Let $\mathbb{X}^{(1)}$ be a series of finite rank, $\mathbb{X} = \mathbb{X}^{(1)} + \mathbb{X}^{(2)}$. Similarly to [9, Section 6.3], where conditions for separability by SSA with centering are considered, the following conditions of separability can be obtained.

1. Basic SSA:
   $\mathbb{X}^{(1)}$ and $\mathbb{X}^{(2)}$ are separable if (if and only if, by definition) their row and column spaces are orthogonal.
2. SSA with row projection on the row space of $\mathbb{X}^{(1)}$:
   $\mathbb{X}^{(1)}$ and $\mathbb{X}^{(2)}$ are separable if their row spaces are orthogonal.
3. SSA with column projection on the column space of $\mathbb{X}^{(1)}$:
   $\mathbb{X}^{(1)}$ and $\mathbb{X}^{(2)}$ are separable if their column spaces are orthogonal.
4. SSA with double projection on the row and column space of $\mathbb{Y}$, where $\mathbb{X}^{(1)}$ is expressed through $\mathbb{Y}$ as $x_n^{(1)} = (an+b)y_n$, $a \neq 0$:
   $\mathbb{X}^{(1)}$ and $\mathbb{X}^{(2)}$ are separable by SSA with double projection if $\mathbb{Y}$ and $\mathbb{X}^{(2)}$ are separable by Basic SSA.

Note that the separability by SSA with projection is always strong, since projections on linear spaces are uniquely defined.

For the approximate separability, where $\mathbf{X}^{(1)} \approx \mathbf{C}$, the approximate orthogonality is necessary. Also, the asymptotic separability can be considered by analogy with the conventional separability for Basic SSA and SSA with centering.

Recall that the usual double centering in SSA corresponds to a constant series $\mathbb{Y}$ and therefore to a linear series $\mathbb{X}^{(1)}$. Orthogonality to a constant series is a much weaker condition than that to a linear series (moreover, the condition of orthogonality to a linear series can never be exactly satisfied). In particular, any sinusoid with frequency ω is asymptotically separable from the linear trend and the exact separability by SSA with projection takes place if $L\omega$ and $K\omega$ are integers, that is, if L and K are divisible by the period of the sinusoid. Therefore, for extraction of linear trends, the double centering is recommended.

In the case of a polynomial trend of degree larger than 1, the conditions of exact separability cannot be satisfied at all, even for SSA with double projection. However, we still can expect that in the case of polynomial trends, SSA with double projection also will work better than SSA with only row or column projection and also better than Basic SSA. This will be checked in the next section.

4. EXAMPLES

The presented examples are related to finding polynomial trends. For convenience, if the row and column projections are performed on the subspace generated by polynomials of degree $q-1$ and $p-1$ respectively, then we denote the method as ProjSSA(q,p). Recall (see Corollary 2 and Remark 4) that the choice ProjSSA(q,p) corresponds to extraction of a polynomial trend of degree $q+p-1$. The zero value for p or q means that the corresponding projection is not performed. For example, both ProjSSA(2,0) and ProjSSA(1,1) can be used for extraction of a linear trend. In ProjSSA(q,p), the projection part of the decomposition, i.e., the decomposition of the matrix $\mathbf{C} = \Pi_{\text{both}}(\mathbf{X})$, consists of $p+q$ rank-one matrices.

All the examples are implemented in R [18] with the help of the Rssa package [16]. For example, to perform ProjSSA(q,p) for a time series taken from the variable `x` with a window length `L`, the following code should be called:

```
s <- ssa(x, L = L,
         row.projector = q,
         column.projector = p)
r <- reconstruct(s,
       groups = list(trend = 1:nspecial(s)))
plot(r, add.residuals = FALSE,
     plot.method = "xyplot", superpose = TRUE)
```

For more details on Rssa, see the help files in [16].

Note that the implementation of SSA with projection in Rssa is efficient, since it uses the approach described in [15, 12]. The computational cost of Decomposition stage for the common case is $\mathcal{O}((q+p)N\log N + qpN)$ for the projection computation using fast convolution and $\mathcal{O}((q+p+r+\log N)Nr)$ for the SVD decomposition itself. Here r is the number of required leading SVD eigentriples, which is small in practice. The computational cost of Reconstruction stage is exactly the same as that for Basic SSA.

4.1 SSA with projection and regression

Let us demonstrate that the conventional linear regression and SSA with double centering, i.e., ProjSSA(1,1), use different statements of the solved problem and therefore can yield different results. It is clearly seen in short time series. For long time series the results are very close. Also, in the model of linear regression with white noise, the least-squares regression solution is the best linear unbiased estimate, see the Gauss-Markov theorem. Therefore, to demonstrate the difference, we consider a time series, which contains a seasonal component.

Here we examine the time series 'Gasoline' taken from [1] and containing the data 'Gasoline', which contains gasoline demand, monthly, Jan 1960 – Jun 1967, Ontario, gallon millions.

Let us consider the first two years and apply the linear regression and ProjSSA(1,1) with $L = 12$. To show the difference, we continue the linear regression line with the help of the estimated coefficients. In the Rssa, a method of forecasting for SSA with projection is implemented. However, since this forecasting method is not thoroughly tested and proved, we do not use it. We will construct the forecast by a linear regression applied to the reconstruction, which is performed by ProjSSA(1,1). Note that the forecasting procedure from Rssa provides a similar prediction. As a benchmark, the linear regression constructed by the whole series is considered.

One can see in Figure 1 that the ProjSSA(1,1) linear trend (the black thin line) is very close to a linear trend constructed by the whole long time series (the grey dash line). The linear regression line (a line with circle points) gives a much worse approximation of the trend. This is explained by the following reasons. The least-squares approach to the linear regression estimation minimizes the prediction error and therefore the seasonal component can shift the linear regression trend. For ProjSSA(1,1), the seasonal component is well separated from the linear trend, since for the chosen parameters $L = K = 12$ are divisible by the seasonal period 12.

4.2 SSA with projection and Basic SSA

The example introduced in this section demonstrates that both SSA with projection and Basic SSA can extract trends in a similar manner. Let us consider the example 'co2'

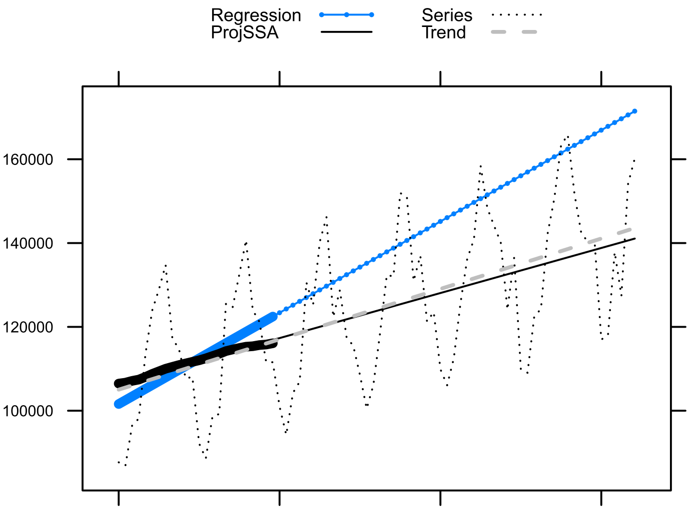

Figure 1. 'Gasoline': SSA with projection, linear trend detection.

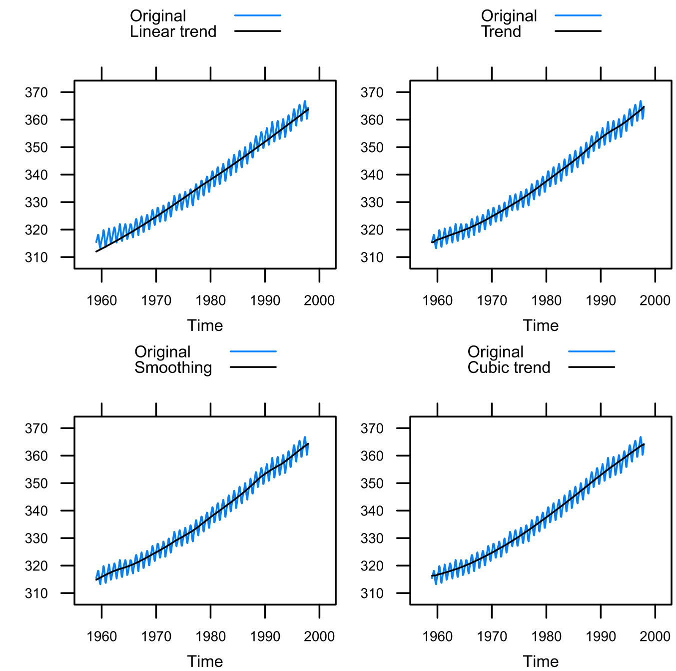

Figure 2. 'co2': Reconstructions of the trend. Left-top: ProjSSA(1,1), $L = 228$; right-top: ProjSSA(1,1), $L = 228$, complemented by the ET 5 and 8; left-bottom: ProjSSA(1,1), $L = 36$; right-bottom: ProjSSA(2,2), $L = 228$.

(Mauna Loa Atmospheric CO_2 Concentration, 468 observations, monthly from 1959 to 1997 [14]).

We start with extraction of the linear trend and therefore choose ProjSSA(1,1) to perform SSA with double centering.

By analogy with SSA, large window lengths help to extract separable series components, while small window lengths correspond to smoothing. Therefore, we take $L = 228$, which is divisible by 12 and is close to half of the time series length to obtain better separability, and a small value $L = 36$ to smooth the series. Three of four versions of the extracted trends presented in Figure 2 almost coincide.

For the choice $L = 228$, the extracted trend is close to linear, see Figure 2 (left-top). Certainly, the accurate trend of 'co2' series is not linear. However, the projection components can be supplemented by the third and sixth SVD components (ET5,8) to improve the trend (Figure 2 (right-top)). Figure 2 (left-bottom) shows the result of smoothing with $L = 36$. Finally, the result of ProjSSA(2,2) with $L = 228$, which is designed for extraction of a cubic trend, is depicted in Figure 2 (right-bottom). The extracted trend is very similar to that in [8], which was extracted by Basic SSA (not depicted).

Identification of the components in the decomposition produced by SSA with projection is exactly the same as it is performed in Basic SSA.

4.3 Numerical comparison

The real-life examples presented in Sections 4.1 and 4.2 show that the results of Basic SSA, SSA with projection and linear regression can be either different or similar. To understand, which method is better, let us perform a numerical study.

We consider a time series of length $N = 199$ with the common term

$$x_n = t_n + s_n + \varepsilon_n, \tag{9}$$

where t_n is a trend, $s_n = A\sin(2\pi n\omega + \phi)$, ε_n is a Gaussian white noise with standard deviation σ.

For obtained estimates $\hat{t}_n^{(i)}$, where i is the number of series with ith realization of noise $\varepsilon_n^{(i)}$, $i = 1, \ldots, M$, we will calculate the root-mean-square error (RMSE) as $\sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{n=1}^{N}(\hat{t}_n^{(i)} - t_n)^2}$.

Linear trend and sine wave. Let us start with the noiseless case ($\sigma = 0$) and therefore take $M = 1$. Let $t_n = an + b$. We fix $a = 1$, $b = -100$, $A = 1$ and change ω from 0.02 to 0.1 (that is, the period is changed from 50 to 10).

Since the result of the least-squares method strongly depends on the form of the residual, we consider the values of the phase, $\phi = 0$ and $\phi = \pi/2$.

Figure 3 (left) contains the RMSE values in the case $\phi = 0$ for Basic SSA with reconstruction by ET1–2, ProjSSA(2,0), ProjSSA(1,1) with $L = 100$, and for the linear regression. One can see that the worse cases for ProjSSA(1,1) are approximately equal to the best cases for the linear regression.

In Section 4.1, we performed forecasting by the linear regression applied to the trend reconstruction. The same linear regression, which is applied to the trend reconstruction, can be considered as a different trend estimate. Figure 3 (right) contains the RMSE for the linear regression lines constructed in this way; 'regr' is added to the legend. Note

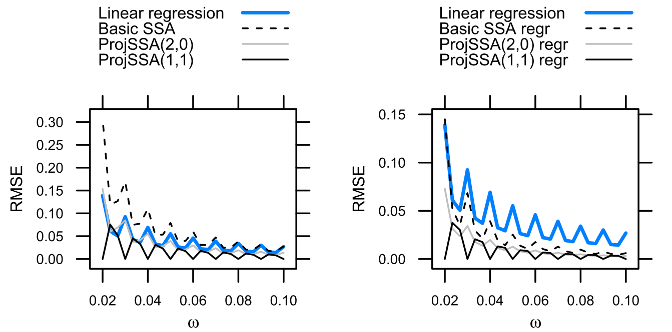

Figure 3. Dependence of the RMSE of linear-trend estimates on frequency of the periodic component, $\phi = 0$.

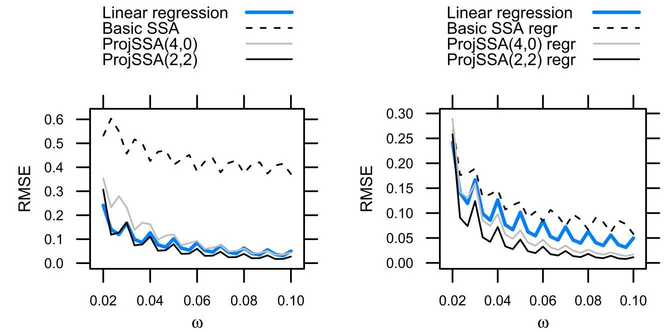

Figure 5. Dependence of the RMSE of cubic-trend estimates on frequency of the periodic component, $\phi = 0$.

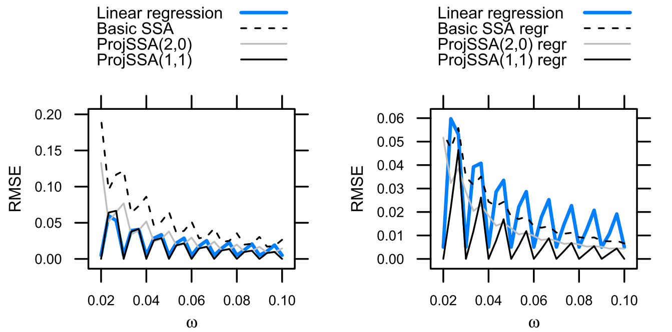

Figure 4. Dependence of the RMSE of linear-trend estimates on frequency of the periodic component, $\phi = \pi/2$.

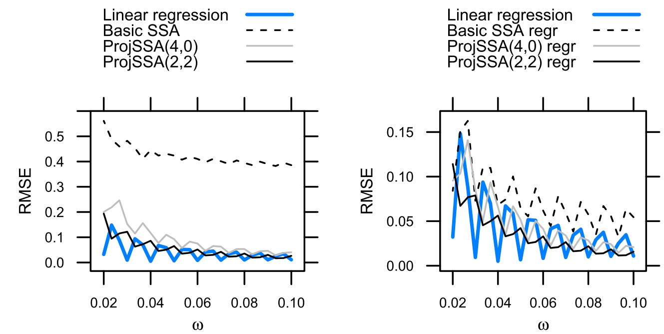

Figure 6. Dependence of the RMSE of cubic-trend estimates on frequency of the periodic component, $\phi = \pi/2$.

that the ‘linear regression’ thick line is the same on the left and right plots, which are depicted in different scales. The ordering of the SSA methods is generally the same, while the SSA methods become better than the linear regression. Probably, 0 is one of the worst values of ϕ for linear regression.

Now consider $\phi = \pi/2$ as one of the best cases for the linear regression. The behavior of the errors is quite different (Figure 4 (left)). However, the accuracy of ProjSSA(1,1) is still better than that of the linear regression. Linear least-squares approximation of the SSA reconstructions considerably improves the accuracy of the SSA methods (Figure 4 (right)).

Note that zero values of the RMSE for ProjSSA(1,1) for frequencies $\omega = 0.01k$ are explained by the theory, since then $L\omega$ and $K\omega$ are integers. The errors for ProjSSA(2,0) lie between that for Basic SSA and ProjSSA(1,1). It is interesting that the minimal errors for Basic SSA are achieved for the middle points, when $L\omega + 0.5$ and $K\omega + 0.5$ are integers.

Remark 5. *The accuracy of separation of a periodic component from a trend depends on (a) is the window length L small or close to half of the time series length and (b) is L divisible by the period of the periodic component [9, Section 6.1], [7]. The accuracy is more sensitive to the divisibility. Therefore, taking different periods, we check the stability of the comparison with respect to the choice of the window length.*

Cubic trend and sine wave. Let us consider a more complex case of the cubic trend $t_n = 0.0001n^3$, $\sigma = 0$. Since there is no exact separability for any choice of parameters, the results are unpredictable. Figures 5 (left) and 6 (left) contain the RMSE values for Basic SSA with reconstruction by ET1–4, ProjSSA(4,0), ProjSSA(2,2) with $L = 100$ and for the cubic regression. One can see that ProjSSA(2,2) is the best method for $\phi = 0$, while it is just comparable with the linear regression for $\phi = \pi/2$. Note that here the best parameters for ProjSSA(2,2) do not correspond to the case when $L\omega$ and $K\omega$ are integers. The cubic least-squares approximation of the reconstructed trend again improves the estimates (Figures 5 (right) and 6 (right)).

Basic SSA fails for the chosen parameters because of lack of strong separability: the fourth trend component has a contribution comparable with the contribution of the periodic components that causes their mixture.

Note that one of the modifications described in [10], Iterative O-SSA, can be used to get strong exact separability for the considered noiseless examples. However, we do not involve this modification into the comparison, since Iterative O-SSA is not able to remove noise and should be applied after denoising in the nested manner, while the compared methods are able to extract the trend without denoising.

Linear trend and noise. For the data which satisfy the model of the linear regression with white Gaussian noise, that is, for the amplitude A equal to zero, we take $\sigma = 1$ and use $M = 1000$. As expected, the smallest error 0.10 is achieved for the regression estimate. However, the RMSE of the ProjSSA(1,1) estimate equal to 0.12 is very close to 0.10. The error of the Basic SSA is equal to 0.17. Application of linear regression to the results of SSA reconstruction improves the SSA estimates. The RMSE for ProjSSA(1,1) and Basic SSA become equal to 0.115 and 0.104 respectively.

We do not show the errors of trend estimates when the series has both periodic component and noise, since the comparison result (ordering of error values) is intermediate between the cases of a noisy trend with no periodic component and a noiseless trend with added periodicity. The general conclusion is that to keep the advantage of SSA with projection, the noise standard deviation σ should be considerably smaller than the amplitude A of the periodic component.

5. CONCLUSION

The considered combination of singular spectrum analysis, which does not need a series model given in advance, and of a subspace-based parametric approach, which is incorporated by means of projections to subspaces given in advance, proves successful for extraction of polynomial (especially, linear) trends, when the residual has unknown structure and can include deterministic oscillations, e.g., the seasonality.

The general form of projections of columns and rows of the trajectory matrix, which keeps this trajectory matrix, was obtained. It was proved that projections to the row and column subspaces (so-called double projection) of the trajectory matrix of a series $\mathbb{Y}$ are related to extraction of the series $(an + b)\mathbb{Y}$. In particular, the linear trend can be obtained by double projection to the column and row subspaces of a constant series. The formulated conditions of separability of a series component, which is kept by projections, show that if a series component can be represented in the form $(an+b)\mathbb{Y}$, then the double projection is preferable.

Thus, the theory provides an additional theoretical support to SSA with double centering (ProjSSA(1,1)), which was known before, and also enlarges the range of applications of semi-nonparametric modifications of Basic SSA.

Applications of SSA with projection considered in the paper were related to the extraction of a polynomial trend, since its trajectory space is determined by the polynomial degree only.

We showed on the example ‘Gasoline’ that the linear regression approach can be inadequate for short series and large oscillations, in comparison with ProjSSA(1,1). Comparison of different SSA versions applied to the ‘co2’ data demonstrates that even if the model of a series component used for projection is wrong, the non-parametric part of SSA with projection can correct the bias.

A numerical study was performed for a better understanding of the difference between SSA with projection and the linear regression approach. First, it appears that if we extract a polynomial trend by SSA with projection, then the polynomial least-squares approximation of the trend reconstruction can considerably improve the accuracy.

The second found effect is related to the influence of the residual geometry on the estimate accuracy. In the considered example, we changed the phase of a sinusoid. The SSA estimates slightly depend on the phase, while the regression estimates demonstrate a considerable dependence.

Numerical experiments confirm that for a linear trend and a sine wave residual, ProjSSA(1,1) is more accurate than the linear regression estimate. For a noisy linear trend, when the model of the linear regression if fulfilled, the linear regression estimate is slightly more accurate than SSA. Thus, we can formulate conditions, when SSA with double projection can be recommended for use: series has a linear or polynomial trend (the polynomial degree is not large) and the regular oscillations are considerably larger than the noise level.

The further investigation can be performed in two directions. First, the forecasting algorithm for ProjSSA(m,k) implemented in Rssa should be proved. Then, the idea to use projection to involve the structure of a supporting series looks promising.

ACKNOWLEDGEMENTS

The authors are thankful to the anonymous reviewers for their useful comments and suggestions, which helped to considerably improve the paper’s presentation.

Received 16 July 2015

REFERENCES

[1] Abraham, B. and Ledolter, J. (1983). *Statistical Methods for Forecasting*. Wiley, Toronto. MR0719535

[2] Alexandrov, T., Bianconcini, S., Dagum, E. B., Maass, P. and McElroy, T. S. (2012). A Review of Some Modern Approaches to the Problem of Trend Extraction. *Econometric Reviews* **31** 593–624. MR2903313

[3] Barba, L. and Rodríguez, N. (2015). Fishery Forecasting Based on Singular Spectrum Analysis Combined with Bivariate Regression. In *Advances in Artificial Intelligence and Its Applications*, (Lagunas O., Alcántara O. and Figueroa G., eds.). *Lecture Notes in Computer Science* **9414** 487–497. Springer International Publishing.

[4] Chen, X. (2007). Chapter 76 Large Sample Sieve Estimation of Semi-Nonparametric Models. (Heckman J. J. and Leamer E. E., eds.). *Handbook of Econometrics* **6, Part B** 5549–5632. Elsevier.

[5] Elsner, J. B. and Tsonis, A. A. (1996). *Singular Spectrum Analysis: A New Tool in Time Series Analysis*. Plenum.

[6] Florinsky, I. V., Eilers, R. G., Wiebe, B. H. and Fitzgerald, M. M. (2009). Dynamics of Soil Salinity in the Canadian Prairies: Application of Singular Spectrum Analysis. *Environmental Modelling & Software* **24** 1182–1195.

[7] Golyandina, N. (2010). On the Choice of Parameters in Singular Spectrum Analysis and Related Subspace-Based Methods. *Statistics and Its Interface* **3** 259–279. MR2720132

[8] Golyandina, N. and Korobeynikov, A. (2014). Basic Singular Spectrum Analysis and Forecasting with R. *Computational Statistics & Data Analysis* **71** 934–954. MR3132018

[9] Golyandina, N., Nekrutkin, V. and Zhigljavsky, A. (2001). *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman&Hall/CRC. MR1823012

[10] Golyandina, N. and Shlemov, A. (2015). Variations of Singular Spectrum Analysis for Separability Improvement: Non-Orthogonal Decompositions of Time Series. *Statistics and Its Interface* **8** 277–294. MR3341327

[11] Golyandina, N. and Zhigljavsky, A. (2013). *Singular Spectrum Analysis for Time Series. Springer Briefs in Statistics*. Springer. MR3024734

[12] Golyandina, N., Korobeynikov, A., Shlemov, A. and Usevich, K. (2015). Multivariate and 2D Extensions of Singular Spectrum Analysis with the Rssa Package. *Journal of Statistical Software* **67** 1–78.

[13] Ichimura, H. and Todd, P. E. (2007). Chapter 74 Implementing Nonparametric and Semiparametric Estimators. (Heckman J. J. and Leamer E. E., eds.). *Handbook of Econometrics* **6, Part B** 5369–5468. Elsevier.

[14] Keeling, C. D. and Whorf, T. P. (1997). Atmospheric CO2 concentrations — Mauna Loa Observatory, Hawaii, 1959–1997. Scripps Institution of Oceanography (SIO), University of California, La Jolla, California USA 92093-0220.

[15] Korobeynikov, A. (2010). Computation- and Space-Efficient Implementation of SSA. *Statistics and Its Interface* **3** 357–368. MR2720139

[16] Korobeynikov, A., Shlemov, A., Usevich, K. and Golyandina, N. (2015). Rssa: A Collection of Methods for Singular Spectrum Analysis R package version 0.13.

[17] Lu, Y. and Saniie, J. (2015). Singular Spectrum Analysis for Trend Extraction in Ultrasonic Backscattered Echoes. In *Ultrasonics Symposium (IUS), 2015 IEEE International* 1–4.

[18] R Core Team (2015). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.

[19] Shaharudin, S. M., Ahmad, N. and Yusof, F. (2015). Effect of Window Length with Singular Spectrum Analysis in Extracting the Trend Signal on Rainfall Data. *AIP Conference Proceedings* **1643** 321–326.

[20] Shlemov, A. and Golyandina, N. (2014). Shaped Extensions of Singular Spectrum Analysis. In *21st International Symposium on Mathematical Theory of Networks and Systems, July 7–11, 2014. Groningen, The Netherlands* 1813–1820.

[21] Unnikrishnan, P. and Jothiprakash, V. (2015). Extraction of Nonlinear Rainfall Trends Using Singular Spectrum Analysis. *Journal of Hydrologic Engineering* 05015007.

[22] Usevich, K. (2010). On Signal and Extraneous Roots in Singular Spectrum Analysis. *Statistics and Its Interface* **3** 281–295. MR2720133

[23] Vautard, R., Yiou, P. and Ghil, M. (1992). Singular-Spectrum Analysis: A toolkit for Short, Noisy Chaotic Signals. *Physica D* **58** 95–126.

[24] Vityazev, V. V., Miller, N. O. and Prudnikova, E. J. (2010). Singular Spectrum Analysis in Astrometry and Geodynamics. *AIP Conference Proceedings* **1283** 319–328.

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