Singular Spectrum Analysis

Technical report – L. Querella, PhD, MSc, Data Scientist, July 2023

Contents

1 Overview

1.1 Basic SSA algorithm

- Introduction – *Basic* SSA algorithm
  - **SSA**
    - Decomposes a time series into a set of summable components that are grouped together and interpreted as trend, periodicity and noise with no a-priori assumptions about the parametric form of these components
    - Emphasises separability of the underlying components, and can readily separate periodicities that occur on different time scales, even in very noisy time series data. The original time series is recovered by summing together all of its components
    - Can be used to analyse and reconstruct a time series with or without different components as desired
  - Applications:
    - Smoothing & noise reduction – Construct a smoothed version of a time series using a small subset of its components
    - Periodicity detection & seasonal adjustment – Investigate a time series periodic components to understand the underlying processes that generated the time series
    - Trend extraction – Reconstruct the original time series without its periodic components
    - Remove all trend and periodic components from the series, leaving just the 'noise'
    - Outlier detection
    - Forecasting
  - Remark:
    - SSA can be used as a model-free technique so that it can be applied to arbitrary time series including non-stationary time series; unlike the commonly used autoregressive integrated moving average (ARIMA) method, SSA makes no assumptions about the nature of the time series, and has just a single adjustable (and easily interpretable) parameter
    - Furthermore, the base algorithm requires little more than a few lines of linear algebra to implement in Python/R
  - Beyond basic SSA: Multivariate SSA, causality test, business cycles detection, gap-filling, detection of structural changes

1.2 Relation with other methods

Autoregression

Typical model for SSA is $x_n = s_n + e_n$, where $s_n = \sum_{k=1}^{r} a_k s_{n-k}$ (signal satisfying an LRR) and e_n is noise. The model of AR is $x_n = \sum_{k=1}^{r} a_k x_{n-k} + e_n$. Despite these two models look similar they are very different. SSA considers AR as a noise component only. AR(1), which is red noise, is typical model of noise for Monte-Carlo SSA (Allen and Smith, 1996).

Spectral Fourier Analysis

In contrast with Fourier analysis with fixed basis of sine and cosine functions, SSA uses an adaptive basis generated by the time series itself. As a result, the underlying model in SSA is more general and SSA can extract amplitude-modulated sine wave components with frequencies different from k/N. SSA-related methods like ESPRIT can estimate frequencies with higher resolution than spectral Fourier analysis.

Linear Recurrence Relations

Let the signal be modeled by a series, which satisfies a linear recurrence relation $s_n = \sum_{k=1}^{r} a_k s_{n-k}$; that is, a series that can be represented as sums of products of exponential, polynomial and sine wave functions. This includes the *sum of dumped sinusoids model* whose complex-valued form is $s_n = \sum_k C_k \rho_k^n e^{i2\pi\omega_k n}$. SSA-related methods allow estimation of frequencies ω_k and exponential factors ρ_k (Golyandina and Zhigljavsky, 2013, Sect 3.8). Coefficients C_k can be estimated by the least squares method. Extension of the model, where C_k are replaced by polynomials of n, can be also considered within the SSA-related methods (Badeau et al., 2008).

Signal Subspace methods

SSA can be considered as a subspace-based method, since it allows estimation of the signal subspace of dimension r by $\mathrm{span}(U_1, \ldots, U_r)$.

State Space Models

The main model behind SSA is $x_n = s_n + e_n$, where $s_n = \sum_{k=1}^{r} a_k s_{n-k}$ and e_n is noise. Formally, this model belongs to the general class of state space models. The specifics of SSA is in the facts that parameter estimation is a problem of secondary importance in SSA and the data analysis procedures in SSA are nonlinear as they are based on the SVD of either trajectory or lag-covariance matrix.

Independent Component Analysis (ICA)

SSA is used in blind source separation by ICA as a preprocessing step (Pietilä et al., 2006). On the other hand, ICA can be used as a replacement of the SVD step in the SSA algorithm for achieving better separability (Golyandina and Zhigljavsky, 2013, Sect. 2.5.4).

Regression

SSA is able to extract polynomial and exponential trends. However, unlike regression, SSA does not assume any parametric model which may give significant advantage when an exploratory data analysis is performed with no obvious model in hand (Golyandina et al., 2001, Ch.1).

Linear Filters

The reconstruction of the series by SSA can be considered as adaptive linear filtration. If the window length L is small, then each eigenvector $U_i = (u_1, \ldots, u_L)^{\mathrm{T}}$ generates a linear filter of width $2L-1$ for reconstruction of the middle of the series $\tilde{x}_s$, $L \le s \le K$. The filtration is non-causal. However, the so-called Last-point SSA can be used as a causal filter (Golyandina and Zhigljavsky 2013, Sect. 3.9).

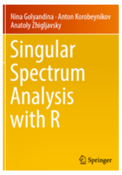
Density Estimation

Since SSA can be used as a method of data smoothing it can be used as a method of non-parametric density estimation (Golyandina et al., 2012).

https://en.wikipedia.org/wiki/Singular_spectrum_analysis

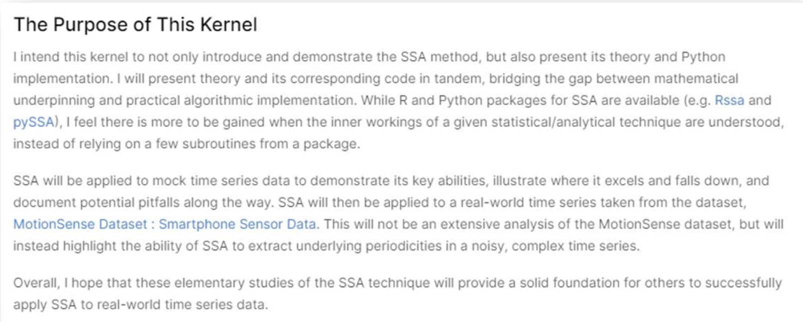
1.3 References

- References
  - Tutorial
    - https://www.kaggle.com/jdarcy/introducing-ssa-for-time-series-decomposition
  - Python packages
    - PySSA: https://github.com/aj-cloete/pySSA
    - Pyts: pyts.decomposition.SingularSpectrumAnalysis
    - Multivariate SSA: pymssa
    - Example (outlier detection): http://www.xavierdupre.fr/app/ensae_teaching_cs/helpsphinx/notebooks/timeseries_ssa.html
  - R:
    - Very comprehensive literature (reference books/articles/examples)
    - http://karthur.org/2017/learning-for-time-series-ssa-vs-pca.html



1.4 Tutorial

- Tutorial (Python)
  - https://www.kaggle.com/jdarcy/introducing-ssa-for-time-series-decomposition

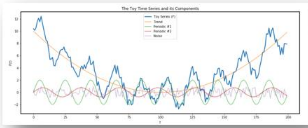


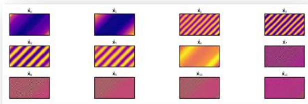
The Purpose of This Kernel

I intend this kernel to not only introduce and demonstrate the SSA method, but also present its theory and Python implementation. I will present theory and its corresponding code in tandem, bridging the gap between mathematical underpinning and practical algorithmic implementation. While R and Python packages for SSA are available (e.g. Rssa and pySSA), I feel there is more to be gained when the inner workings of a given statistical/analytical technique are understood, instead of relying on a few subroutines from a package.

SSA will be applied to mock time series data to demonstrate its key abilities, illustrate where it excels and falls down, and document potential pitfalls along the way. SSA will then be applied to a real-world time series taken from the dataset, MotionSense Dataset : Smartphone Sensor Data. This will not be an extensive analysis of the MotionSense dataset, but will instead highlight the ability of SSA to extract underlying periodicities in a noisy, complex time series.

Overall, I hope that these elementary studies of the SSA technique will provide a solid foundation for others to successfully apply SSA to real-world time series data.





2 Theory – Basic SSA

Cf. written notes and schematic view (excerpt from [1])

1.1 Generic Scheme of the SSA Family and the Main Concepts

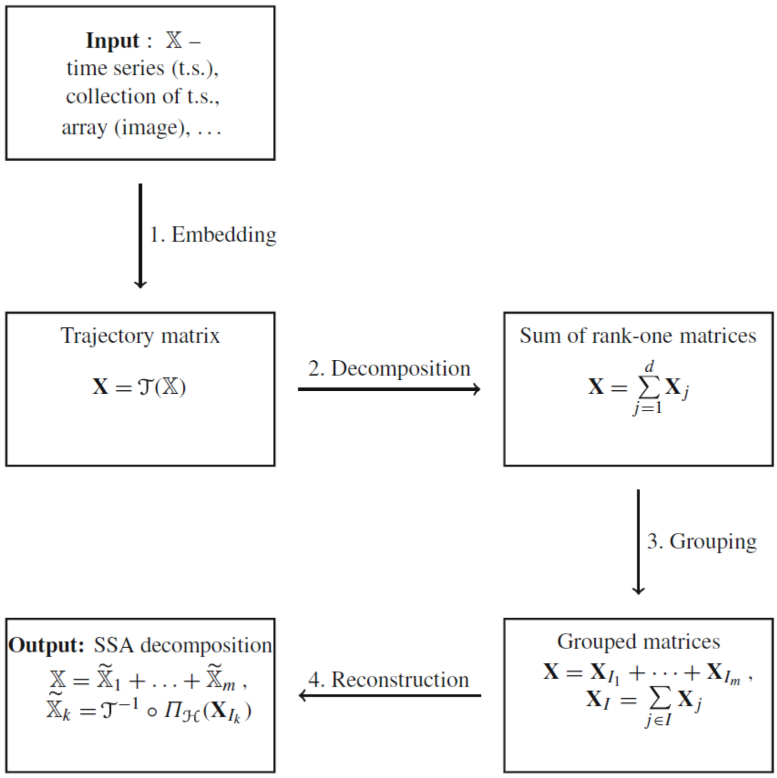


Fig. 1.1 SSA family: Generic scheme

3.1 Decomposition

The first step of SSA is the decomposition of the original time series into a set of elementary matrices required during the reconstruction phase. This process is based on the *embedding* and *singular value decomposition* that we describe next.

Embedding. Given an N-time series $\boldsymbol{l} = (l^1, l^2, \ldots, l^N)$ and the window length W such that $2 < W \leq N/2$, we define $D = N - W + 1$ delayed vectors:

$$\mathbf{l}_i = (l^i, l^{i+1}, \ldots, l^{i+W-1})^\top \text{ for } 1 \leq i \leq D,$$

and the trajectory matrix:

$$\mathbf{L} = (\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_D) = \begin{pmatrix} l^1 & l^2 & \cdots & l^D \\ l^2 & l^3 & \cdots & l^{D+1} \\ \vdots & \vdots & \ddots & \vdots \\ l^W & l^{W+1} & \cdots & l^N \end{pmatrix},$$

which is a Hankel matrix, i.e., a matrix with constant skew diagonals. The window length W plays a key role in the performance and accuracy of SSA, hence the importance of choosing its optimal value. Too large or too small values can lead to decompositions where the components are mixed-up between them, making the reconstruction step difficult. In the context of noise reduction, conditions are relaxed allowing practitioners to consider other aspects, e.g., performance, while keeping the accuracy of the tool high. In this work, we will use the following rule-of-thumb:

$$W = \lfloor \log(N)^c \rfloor \text{ with } c \in [1.5, 3], \tag{1}$$

which has been shown near optimal for signal vs. noise separation [35].

Singular Value Decomposition. Given the trajectory matrix $\mathbf{L}$ from the previous step, its *singular value decomposition* (SVD) [18] is computed. First, the eigenvalues of $\mathbf{L}\mathbf{L}^\top$ in decreasing order of magnitude $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$, i.e., the so called *singular spectrum* which gives name to SSA, and the corresponding

eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_d$ (with $d = W$ if none of the eigenvalues is zero) are obtained. The SVD decomposition of the trajectory matrix can be written as:

$$\mathbf{L} = \mathbf{L}_1 + \mathbf{L}_2 + \cdots + \mathbf{L}_d, \tag{2}$$

where the $W \times D$ elementary matrix $\mathbf{L}_i = \sqrt{\lambda_i}\mathbf{u}_i\mathbf{v}_i^\top$ and $\mathbf{v}_i = \mathbf{L}^\top\mathbf{u}_i/\sqrt{\lambda_i}$, for $1 \leq i \leq d$. Computing (2) is the most time-consuming step in SSA, however only the leading components are required during the reconstruction phase. In order to alleviate this complexity burden the *partial SVD* (PSVD), which only calculates a subset of the $\mathbf{L}_i$ matrices in (2), is considered in the remaining sections.

3.2 Reconstruction

After having obtained the SVD decomposition of the original time series in the previous phase, the reconstruction step aims for the extraction of its underlying components. It is based on the *diagonal averaging* and *grouping* described next.

Diagonal averaging. If $\mathbf{X}$ is a $W \times D$ matrix with elements $x_{i,j}$ for $1 \le i \le W$ and $1 \le j \le D$, it can be immediately turned into the series $\tilde{\mathbf{x}} = \{\tilde{x}^t\}_{t=1}^N$ if and only if $\mathbf{X}$ is a Hankel matrix. In that case, each entry $\tilde{x}^t$ is equal to all elements $x_{i,j}$ along the anti-diagonal $i + j = t + 1$ of $\mathbf{X}$. However, if $\mathbf{X}$ does not have constant skew diagonals, an additional step is required. Namely, the averaging of the anti-diagonals $i + j = k + 1$ will transform $\mathbf{X}$ into the series $\tilde{\mathbf{x}} = \{\tilde{x}^t\}_{t=1}^N$ in a process which is also known as Hankelization [38]:

$$\tilde{x}^t = \begin{cases} \dfrac{1}{t} \displaystyle\sum_{m=1}^{t} x_{m,t-m+1} & \text{for } 1 \le t < W^*, \\ \dfrac{1}{W^*} \displaystyle\sum_{m=1}^{W^*} x_{m,t-m+1} & \text{for } W^* \le t \le D^*, \\ \dfrac{1}{N-t+1} \displaystyle\sum_{m=t-D^*+1}^{N-D^*} x_{m,t-m+1} & \text{for } D^* < t \le N, \end{cases}$$

where $W^* = \mathsf{min}(W, D)$ and $D^* = \mathsf{max}(W, D)$. Because matrices $\mathbf{L}_i$ in (2) are not Hankel matrices, by applying this procedure each matrix $\mathbf{L}_i$ in (2) is transformed into the i^{th} (so-called) principal component $\mathbf{g}_i$ of length N.

Grouping. Under the assumption of weak separability [19], the original N-time series $\boldsymbol{l}$ can then be reconstructed by:

$$\boldsymbol{l} = \mathbf{g}_1 + \mathbf{g}_2 + \cdots + \mathbf{g}_d.$$

At this stage, the set of indices $I = \{1, \ldots, d\}$ is partitioned into m disjoint subsets $I_1, \ldots, I_m$. Since in the context of this work, SSA aims for signal vs. noise decomposition, we are typically looking for a partitioning such that $m = 2$ and $I = \{I_{\text{signal}}, I_{\text{noise}}\}$. The analysis of the eigenvalues λ_i in the SVD step is the most common method for splitting I according to some criteria that depend on the application area. In our case, indices whose respective eigenvalues are small, usually producing a slowly decreasing sequence, are included in the group of noisy components. The remaining components can be exhaustively tested in order to find the combination bringing a better reconstruction of the noise-free signal. In the following, we will determine the best grouping based on the SNR.

Summarizing, this section presented the main steps involved in SSA, that can be recalled with the following equation:

$$\overbrace{\boldsymbol{l} \longrightarrow \mathbf{L}}^{\text{Embedding}} = \overbrace{\mathbf{L}_1 + \mathbf{L}_2 + \cdots + \mathbf{L}_d}^{\text{SVD}} = \overbrace{\mathbf{g}_1 + \mathbf{g}_2 + \cdots + \mathbf{g}_d}^{\text{Averaging}} = \overbrace{\sum_{i \in I_{\text{signal}}} \mathbf{g}_i + \sum_{i \in I_{\text{noise}}} \mathbf{g}_i}^{\text{Grouping}}.$$

Next, three different case studies will be presented, where raw and preprocessed traces are compared in terms of the corresponding attacks' performance. Additionally, decisions taken for the application of SSA such as the window length W and the grouping of components will be discussed in more details.

[*Blind Source Separation from Single Measurements using Singular Spectrum Analysis*, Santos Merino Del Pozo and Francois-Xavier Standaert, https://perso.uclouvain.be/fstandae/PUBLIS/160.pdf]

3 Applications

Cf. [2]

3.1 Forecasting

Cf. Chapter 3 [1].

3.2 Denoising and Feature extraction

Hybrid methods using SSA as preprocessing step.

Example: In Wang et al. (2016), support vector machine regression (SVR) is applied separately to the trend and fluctuations, which are extracted by SSA. The constructed method is applied to forecast a time series data of failures gathered at the maintenance stage of the Boeing 737 aircraft. It is shown that the suggested hybrid SSA+SVR outperforms Holt-Winters, autoregressive integrated moving average, multiple linear regression, group method of data handling, SSA, and SVR used separately

4 References

[1] *Singular Spectrum Analysis with R*, N. Golyandina et al., Springer, 2018

[2] *Statistics and Its Interface* (2010, v.3, No.3 and 2017, v.10, No.1), https://www.intlpress.com/site/pub/pages/journals/items/sii/content/vols/0003/0003/index.php

[3] http://www.xavierdupre.fr/app/ensae_teaching_cs/helpsphinx/notebooks/timeseries_ssa.html and references therein

