# A method of extraction of quasi-periodic time series components using Singular Spectrum Analysis

**Theodore Alexandrov**, University of Bremen, Center of Industrial Mathematics (ZeTeM), Bremen, Germany

## Introduction

**Problem:** detection and extraction of quasi-periodic additive components of time series. The quasi-periodic component is defined with componentwise decreasing or increasing behavior.

The study investigates the posed problem in the framework of **Singular Spectrum Analysis** (SSA), see Vautard et al. (1992), Golyandina et al. (2001).

**The main contribution:** a parametric method for extraction of quasi-periodic additive components. This method is simple in use and requires very little prior information about time series.

Strategies for the choice of the parameters are proposed.

## Singular Spectrum Analysis

The central idea of SSA is to embed a time series into some high-dimensional euclidean space, then find a subspace corresponding to a sought-for component and, finally, reconstruct the time series component defined by this subspace.

**Given the time series of length *N*:**

$$X=(x_0,\ldots,x_{N-1}),\ x_n\in\mathbb{R}.$$

**Decomposition:**

(1) Set window length *L* and wrap the time series into the trajectory matrix $\mathbf{X}\in\mathbb{R}^{L\times K}$, $K=N-L+1$,

$$X=(x_0,\ldots,x_{N-1})\ \rightarrow\ \mathbf{X}=[C_1:\ldots:C_K],$$
$$C_j=(x_{j-1},\ldots,x_{j+L-2})^{\mathrm{T}}.$$

(2) Perform SVD of the trajectory matrix (where $\lambda_j, U_j$ are *j*'th eigenvalue and eigenvector of $\mathbf{X}\mathbf{X}^{\mathrm{T}}$):

$$\mathbf{X}=\sum_{j=1}^{L}\sqrt{\lambda_j}U_jV_j^{\mathrm{T}},\quad V_j=\mathbf{X}^{\mathrm{T}}U_j/\sqrt{\lambda_j}.$$

*SVD components are numbered in the decreasing order of their eigenvalues*

**Reconstruction:**

(3) select a group $\mathcal{J}\subset\{1,\ldots,L\}$ of SVD components, hankelize the matrix, unwrap a time series from this matrix (like construction of the trajectory matrix, but in reverse direction):

$$\mathbf{X}_{\mathcal{J}}=\sum_{j\in\mathcal{J}}\sqrt{\lambda_j}U_jV_j^{\mathrm{T}}\xrightarrow{\text{hankelization}}\widetilde{\mathbf{X}}_{\mathcal{J}}\in\mathbb{R}^{L\times K}$$
$$\rightarrow\ \widetilde{P}=(\tilde{p}_0,\ldots,\tilde{p}_N).$$

The problem of extraction of an additive component is reduced to (i) choice of window length *L* (Golyandina et al, 2001), (ii) selection of necessary SVD components. We consider the latter problem.

## Quasi-periodic components in SSA

Quasi-periodic time series $P=(p_0,\ldots,p_{N-1})$ of period *T*

$$p_n=\sum_{k=1}^{\lfloor T/2\rfloor}A_ke^{\alpha_kn}\cos(2\pi\omega_k+\phi_k),\quad A_k>0,\ \phi_k\in[0,2\pi),$$
$$\omega_k=k/T.$$

### Proposition:

Let $F=(f_0,\ldots,f_{N-1})$ be an exponentially-dumped (e-d) sine time series:

$$f_n=Ae^{\alpha n}\cos(2\pi\omega n+\phi),\quad A,\alpha\in\mathbb{R},\quad \phi\in[0,2\pi).$$

Under some conditions it generates two eig.vectors, denoted by $U_1$ and $U_2$, $U_j=(u_1^{(j)},u_2^{(j)},\ldots,u_L^{(j)})^T$ with e-d sine elements:

$$u_n^{(1)}=B_1e^{\alpha n}\cos(2\pi\omega n+\widetilde{\phi}_1),$$
$$u_n^{(2)}=B_2e^{\alpha n}\cos(2\pi\omega n+\widetilde{\phi}_2),\quad 1\leqslant n\leqslant L$$

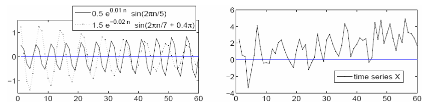
Moreover, when $L\to\infty,\ \alpha\to0,\ L\alpha\to\gamma\in\mathbb{R}$, then

$$|\widetilde{\phi}_1-\widetilde{\phi}_2|\to\pi/2.$$

### Example. Extraction of two e-d sine waves in the presence of noise and trend without prior information

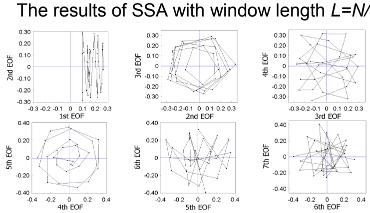
Time series *X* of length 60 given by:

$$x_n=0.05n+0.5e^{0.01n}\sin(2\pi n/5)+$$
$$1.5e^{-0.02n}\sin(2\pi n/7+0.4\pi)+\varepsilon_n,\ \varepsilon_n\sim N(0,1)$$



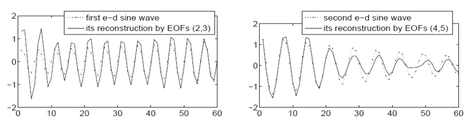
*E-d sine components and the considered time series X.*

The results of SSA with window length *L=N/2=30:*



*Scatterplots for the pairs of the first neighbor EOFs*

*Empirical Orthogonal Function (EOF) = sequence of elements of an eigenvector*



*Original e-d sine waves (dotted) and their estimations reconstructed by SVD components 2,3 (left) and 4,5 (right) (solid)*

## Method

Periodogram of *m*'th EOF is defined as

$$\Pi_{U_m}^L(\omega)=\frac{1}{L}\left|\sum_{n=0}^{L-1}e^{-i2\pi\omega n}u_n^{(m)}\right|^2,\quad \omega\in\{k/L\}_{k=0}^{\lfloor L/2\rfloor}.$$

### Part 1

For all pairs of SVD components check if the periodograms of EOFs from a pair have maxima at near points, i.e. if

$$\mathcal{S}_L(i,j)=L|\theta_i-\theta_j|\leqslant s_0,\quad \theta_i=\arg\max_{0\leqslant k\leqslant L/2}\{\Pi_{U_i}^L(k/L)\}.$$

### Part 2

**Problem:** Part1 produces many false-positives.

**Solution:** consider all pairs of EOFs identified after Part1 and check whether the maxima values of their periodograms are large enough.

For given verify if $\mathcal{R}(i,j)\geqslant r_0$, where

$$\mathcal{R}(i,j)=\max_{0\leqslant k\leqslant L/2}\frac{1}{2}\left(\Pi_{U_i}^L(k/L)+\Pi_{U_j}^L(k/L)\right)$$

The enhanced version (much better when $L\omega\notin\mathbb{N}$)

$$\max_{0\leqslant k\leqslant L/2}\frac{1}{2}\left(\gamma_{i,j}^L(k/L)+\gamma_{i,j}^L((k+1)/L)\right)\geqslant r_0,$$

$$\text{where}\quad \gamma_{i,j}^L(k/L)=\Pi_{U_i}^L(k/L)+\Pi_{U_j}^L(k/L).$$

## Parameters selection

The proposed method has two parameters ($s_0$ $r_0$).

One can show numerically, that $s_0$ can be selected a priori: $s_0=1$.

The strategy for the choice of $r_0$:

- estimate the lower bound of amplitude of e-d sine wave: $A>A_0$
- for some $\beta\in[0,1]$ (by default $\beta\cong0.75$) calculate $\mathcal{G}_0(A,\beta)=\beta A^2/2$
- take $r_0(\mathcal{G}_0,\Delta r)=\min\{r_0:\ \|X(r_0,\Delta r)\|^2\geqslant\mathcal{G}_0\}$, where
  - $\mathcal{I}(r_0)$ denotes a group of SVD components identified with $r_0$;
  - $\mathcal{J}(r_0,\Delta r)=\mathcal{I}(r_0)\backslash\mathcal{I}(r_0+\Delta r)$;
  - $X(r_0,\Delta r)$ is a time series reconstructed by the group $\mathcal{J}(r_0)$.

## Frequency estimation

SSA finds the recurrent representation for any e-d sine wave:

$$f_{i+L-1}=\sum_{k=1}^{L-1}a_kf_{i+L-1-k}$$

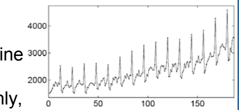
Firstly, find the roots of the characteristic polynomial:

$$P(\lambda)=\lambda^d-\sum_{k=1}^{d}a_k\lambda^{d-k}$$

The frequency is estimated through two conjugate roots $\lambda_{1,2}=e^{\alpha\pm i2\pi\omega}$ of maximum modulus as following:

$$\omega=(2\pi)^{-1}\arccos(\Re(\lambda_1)/|\lambda_1|)$$

## Example



**Data:** US retail sails at beer, wine and liquor stores 1992/01-2007/01 in millions US$. Monthly, not seasonally adjusted data. Length *N=187*.

**Task:** seasonal adjustment (removal of quasi-periodical component of period *12*).

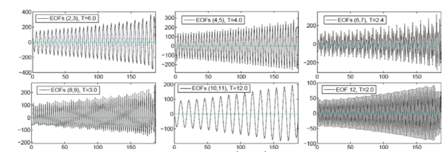
**Parameters:**

- window length: *L=84* (close to *N/2* and divisible by *12*)
- Part1: $s_0=1$
- Part2: let the lower bound of amplitudes be $A_0=100$ Then $\mathcal{G}_0=0.75\cdot100^2/2=3750$ and the strategy (see section "Param. select.") with $\Delta r=0.01$ and $\mathcal{G}_0=3750$ gives $r_0=0.92$.

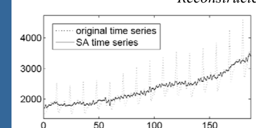
*Remark: The same* $r_0$ *is resulted for a wide range of* $A_0\in[42,181]$

**Results:** The proposed method with $L=84$, $s_0=1$, and $r_0=0.92$ selects EOFs with numbers 2-14. Limiting only to those of period from $[11/k,13/k]$, $2\leqslant k\leqslant T/2$, we obtain:

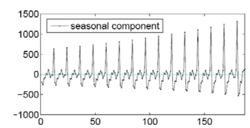
| Numbers of EOFs | (2,3) | (4,5) | (6,7) | (8,9) | (10,11) | 12 |
|---|---|---|---|---|---|---|
| Estimated periods | 6.0 | 4.0 | 2.4 | 3.0 | 12.0 | 2.0 |



*Reconstructed e-d sine waves*



*The original data (dotted) and seasonally adjusted data (solid)*



*Extracted seasonal component*