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.

Quasi-periodic components in SSA

Quasi-periodic time series $P = (x_0, \ldots, x_{N-1})$ of period $T$

$$p_k = \sum_{i=1}^{N} A_i \cos(2 \pi k / T_i + \phi_i), \quad A_i > 0, \quad \phi_i \in [0, 2\pi),$$

$$\omega_k = k / T$$

Proposition:

Let $P = (x_0, \ldots, x_{N-1})$ be an exponentially-damped (e-d) sine time series:

$$x_0 = A e^{\alpha t} \cos(2 \pi f_0 t + \phi), \quad A, \alpha > 0, \quad f_0 \in [0, 2\pi),$$

Under some conditions it generates two eigenvectors, denoted by $U_1$ and $U_2$, $U_j = (u_j(1), u_j(2), \ldots, u_j(L))^T$ with e-d sine elements:

$$u_j(1) = B e^{\alpha t} \cos(2 \pi f_j t + \phi),$$

$$u_j(2) = B e^{\alpha t} \cos(2 \pi f_j t + \phi)$$

Moreover, when $L \to \infty$, $\alpha > 0$, $L_0 \to \gamma \in R$.

Example.

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

Time series $X$ of length $N$ given by:

$$x = a_0 + 0.6 \sin(7t) + z(t), \quad a_0 \sim N(0, 1)$$

Decomposition:

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

$$X = (x_0, \ldots, x_{N-1}) \rightarrow X = (x_1, \ldots, x_K), \quad x_j = (x_{j-L+1}, \ldots, x_j)^T.$$  

(2) Perform SVD of the trajectory matrix (where $\Lambda, U, V$ are $K$th eigenvalue and eigenvector of $XX^T$):

$$X = \sum_{i=1}^{K} \sqrt{\lambda_i} u_i v_i^T, \quad U = V \Lambda^{1/2}.$$  

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

Reconstruction:

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

$$X_F = \sum_{j=1}^{L} \sqrt{\lambda_j} u_j v_j^T, \quad X_F \subset \mathbb{R}^{K \times \mathcal{F}}.$$  

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.

Method

Periodogram of $m$th EOF is defined as:

$$\Pi_m^2(\omega) = \frac{1}{L} \sum_{k=0}^{L-1} \sum_{i=1}^{K} \alpha_k^2 m_{i,k}^2 \omega_k^2, \quad \omega_k = k / L.$$  

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

$$S_j(x, t) = \sum_{k=1}^{K} \alpha_k^2 m_{i,k} \Pi_m^2(\omega_k).$$

Part 2

Problem: Part 1 produces many false-positives.

Solution: consider all pairs of EOFs identified after Part 1 and check whether the maxima values of their periodograms are large enough.

For given verify if $R(x, t) > r_0$ where

$$R(x, t) = \max_{\omega_k} \frac{1}{2} \left( \Pi_m^2(\omega_k) + \Pi_m^2(\omega_k) \right) \geq r_0.$$  

The enhanced version (more better when $L_0 \notin \mathbb{N}$)

$$\max_{\omega_k} \frac{1}{2} \left( \Pi_m^2(\omega_k) + \Pi_m^2(\omega_k) \right) \geq r_0.$$  

Parameters selection

The proposed method has two parameters $(\alpha_0, \beta_0)$.

One can show numerically, that $\alpha_0$ can be selected a priori: $\alpha_0 = 1$.

The strategy for the choice of $\beta_0$:

- estimate the lower bound of amplitude of e-d sine wave: $A > A_0$
- for some $\beta \in [0, 1]$ (by default $\beta = 0.75$) calculate $\tilde{G}_0(A, \beta) = \beta^2 A^2$
- take $r_0 = \tilde{G}_0(A, \beta) = \min_{\alpha_0} \left\{ \left[ \Pi_m^2(\omega_k) \right] \geq \tilde{G}_0 \right\}$, where $\Pi_m^2(\omega_k)$ denotes a group of SVD components identified with $\alpha_0$.
- $\mathcal{F} \subset \{1, \ldots, L\}$ is a time series reconstructed by the group $\mathcal{F}$.

Frequency estimation

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

$$f_{i+0} = \sum_{i} a_i f_{i+1} + a_i.$$  

Firstly, find the roots of the characteristic polynomial:

$$P(\lambda) = \lambda^L - \sum_{i} a_i \lambda^{L-i}.$$  

The frequency is estimated through two conjugate roots $\lambda_{\pm} = e^{\pm \pi i / L}$ of maximum modulus as following:

$$\omega = (2 \pi)^{-1} \arccos(|\Lambda(\alpha_0, \beta)||\Lambda(\alpha_0, \beta)|)$$

Example


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

Parameters:

- window length: $L=84$ (close to $N/2$ and divisible by 12)
- $\alpha_0 = 1$
- $\beta_0 = 0.92$ selects EOFs with numbers 2-14. Limiting only to those of period from $11k/12$ to $2k < k < L$ we obtain:

$$\omega = (2 \pi)^{-1} \arccos(|\Lambda(\alpha_0, \beta)||\Lambda(\alpha_0, \beta)|)$$

Results: The proposed method with $L = 84$, $\alpha_0 = 1$ and $\beta_0 = 0.92$ selects EOFs with numbers 2-14. Limiting only to those of period from $11k/12$ to $2k < k < L$ we obtain: