Iterative signal decomposition and time-frequency representation using singular spectrum analysis

Milton J. Porsani, UFBA
Bjørn Ursin, NTNU/UFBA
Michelângelo G. da Silva, UFBA

April, 2018

CONTENTS

Introduction

- The Singular Spectral Analysis (SSA) method is a particular way to apply the Singular Value Decomposition (SVD) method in a single (or multivariate) time series;
- The SSA method in an iterative and recursive way to estimate individual components of the signal. Applying the short time autoregressive method to obtain a time-frequency representation of the signal;
- For the computing of the instantaneous frequency we provide a new equation which depend on a single autoregressive coefficient;
- The effectiveness of the new approach is demonstrated in a synthetic data example and in the removal of ground-roll noise from land seismic data.

Singular Spectrum Analysis

- Let the vector $\mathbf{d} = [d(0), \ldots, d(N)]^T$ represent single-trace data, and, let $\mathbf{D}$ be the Toeplitz matrix with the data shifted by one time sample in each column. τ represents the variable associated with the time shift, $\tau = 0, \ldots, M$. The matrix $\mathbf{D}$ has dimensions, $(M + N + 1) \times (M + 1)$, and

$$\mathbf{D}^T = \begin{bmatrix} d(0) & \ldots & d(N) & & \mathsf{O}_M \\ & \ddots & \ddots & \ddots & \\ \mathsf{O}_M & & d(0) & \ldots & d(N) \end{bmatrix}.$$

- Where O_M represent a triangle of null coefficients. The matrix $\mathbf{D}^T$ is an extension of the so-called trajectory matrix which is used in SSA.

- We have $\mathbf{D} = [\mathbf{E}_0\mathbf{d} \ \ldots \ \mathbf{E}_M\mathbf{d}] = [\bar{\mathbf{d}}_0 \ \ldots \ \bar{\mathbf{d}}_M]$, where $\mathbf{E}_k$ is a $(N+M+1) \times (M+1)$ shifting matrix:

$$\mathbf{E}_k = \begin{bmatrix} \bar{\mathbf{0}}_k \\ \mathbf{I}_{N+1} \\ \bar{\mathbf{0}}_{M-k} \end{bmatrix},$$

$\mathbf{I}_{N+1}$ is the identity matrix of order $N+1$. $\bar{\mathbf{0}}_k$ and $\bar{\mathbf{0}}_{M-k}$ represent matrices with null coefficients and dimensions $k \times (N+1)$ and $(M-k) \times (N+1)$, respectively, such

$$\bar{\mathbf{d}}_k = \mathbf{E}_k\mathbf{d} = \begin{bmatrix} \mathbf{0}_k \\ \mathbf{d} \\ \mathbf{0}_{M-k} \end{bmatrix}.$$

$\mathbf{0}_k$ and $\mathbf{0}_{M-k}$ represent the vectors with k and $M-k$ null coefficients, respectively. The signal can be expressed as $\mathbf{d} = \frac{1}{M+1}\sum_{k=0}^{M} \mathbf{E}_k^T\bar{\mathbf{d}}_k$.

The reduced SVD

- The reduced SVD of the matrix $\mathbf{D}$ is

$$\mathbf{D} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = \sum_{\tau=0}^{M} \sigma_\tau \mathbf{u}_\tau \mathbf{v}_\tau^T = \sum_{\tau=0}^{M} \widetilde{\mathbf{D}}_\tau \,,$$

where $\widetilde{\mathbf{D}}_\tau = \sigma_\tau \mathbf{u}_\tau \mathbf{v}_\tau^T$ represents the eigenimage of index τ of the data matrix $\mathbf{D}$. Then the eigenvalue decomposition of the data covariance matrix is

$$\mathbf{D}^T\mathbf{D} = \mathbf{V}\boldsymbol{\Sigma}^2\mathbf{V}^T \,.$$

The eigenvalues are σ_τ^2, and the eigenvectors are $\mathbf{v}_\tau$.

- For a given eigenimage, $\widetilde{\mathbf{D}}_\tau$, we can apply the matrix $\mathbf{E}_k$ and restore a transformed data component

$$\widetilde{\mathbf{d}}_\tau = \sigma_\tau \sum_{k=0}^{M} v_\tau(k) \mathbf{E}_k^T \mathbf{u}_\tau \,.$$

That is, the left singular vector $\mathbf{u}_\tau$ is cut, shifted and added to the result with a weight $\sigma_\tau v_\tau(k)$. The previous equation can be expressed by

$$\widetilde{\mathbf{d}}_\tau = \sigma_\tau \mathbf{V}_\tau^T \mathbf{u}_\tau$$

where $\mathbf{V}_\tau^T$ is a $(N+1) \times (M+N+1)$ banded Toeplitz matrix.

- The banded Toeplitz matrix given by

$$\mathbf{V}_\tau^T = \begin{bmatrix} v_\tau(0) & \dots & v_\tau(M) & & \mathsf{O}_N \\ & \ddots & \ddots & \ddots & \\ \mathsf{O}_N & & v_\tau(0) & \dots & v_\tau(M) \end{bmatrix}.$$

It may be shown that: $\mathbf{I} = \frac{1}{M+1}\sum_{\tau=0}^{M}(\mathbf{V}_\tau^T\mathbf{V}_\tau)\,.$

- Where $\mathbf{V}_\tau^T\mathbf{V}_\tau$ is a symmetric Toeplitz matrix of dimension $(N+1)\times(N+1)$, formed by the autocorrelations coefficients, of the eigenvector $\mathbf{v}_\tau$. By multiplying $\mathbf{I}$ by $\mathbf{d}$ we obtain:

$$\mathbf{d} = \frac{1}{M+1}\sum_{\tau=0}^{M}(\mathbf{V}_\tau^T\mathbf{V}_\tau)\mathbf{d} = \frac{1}{M+1}\sum_{\tau=0}^{M}\widetilde{\mathbf{d}}_\tau \quad \text{where} \quad \widetilde{\mathbf{d}}_\tau = \mathbf{V}_\tau^T\mathbf{V}_\tau\,\mathbf{d}\,.$$

- The data vector **d** may be decomposed in terms of the eigentraces $\widetilde{\mathbf{d}}_\tau$.
- From equation:

$$\widetilde{\mathbf{d}}_\tau = \mathbf{V}_\tau^T \mathbf{V}_\tau \, \mathbf{d}$$

- We see that the output trace of the eigenimage number τ is the convolution of the data vector **d** with the autocorrelation of the eigenvector $\mathbf{v}_\tau$.
- Since the autocorrelation is zero phase, the phase of the output trace is equal to the phase of the data trace.

Iterative and recursive signal decomposition via SSA

- We use the SSA method to iteratively decompose a signal into high and low energy components using a three-loop algorithm.
- In the inner loop we compute a high-energy component of the signal by recursions in the number of rows in the trajectory matrix using only the first right singular vector, corresponding to the component with highest energy.
- The result is subtracted from the input signal in the second loop and the process is repeated. This gives an estimate of the low-energy part of the signal.
- In the outer loop this low-energy signal is subtracted from the input signal and the result is output as one signal component.
- The whole procedure is then repeated with the low-energy component as the new input signal.

Algorithm to iterative and recursive signal decomposition via SSA

Initial vector
$\mathbf{d}_1 = \mathbf{d}$
DO $k = 1, \ldots, K$ (Components)
$M = \max\{1, K - k + 1\}$
$\widehat{\mathbf{d}}_0 = \mathbf{d}_k$
DO $j = 1, \ldots, J$ (Iterations)
$\widetilde{\mathbf{d}}_0 = \widehat{\mathbf{d}}_{j-1}$
DO $\tau = 1, \ldots, M$ (Recursion in order M)

- Form the matrix $\mathbf{D}_\tau = [\ \overline{\mathbf{d}}_0 \ldots \overline{\mathbf{d}}_\tau\ ]$ from $\widetilde{\mathbf{d}}_{\tau-1}$
- Compute the first right singular vector $\mathbf{v}_0$ (of dimension $(\tau + 1) \times 1$)
- Compute the auto-correlation of $\mathbf{v}_0$, $\mathbf{r}_\tau = (1, r_\tau(1), \ldots, r_\tau(\tau))^T$
- Compute the update $\widetilde{\mathbf{d}}_\tau = \mathbf{V}_\tau^T \mathbf{V}_\tau \widetilde{\mathbf{d}}_{\tau-1}$ (equation (11))

ENDDO
$\widehat{\mathbf{d}}_j = \widehat{\mathbf{d}}_{j-1} - \widetilde{\mathbf{d}}_M$
ENDDO
Output $\mathbf{x}_k = \mathbf{d}_k - \widehat{\mathbf{d}}_J$
$\mathbf{d}_{k+1} = \widehat{\mathbf{d}}_J$
ENDDO

Time-frequency representation using AR coefficients

- From each estimated signal component $x(t)$ we form the analytic signal

$$z(t) = x(t) + iy(t) \tag{1}$$

- where $y(t)$ is the Hilbert transform of $x(t)$. For a data window $\{z(t - L\Delta t), \ldots, z(t), \ldots, z(t + L\Delta t)\}$ (Δt is the sample interval) we define the instantaneous auto-correlation,

$$R_t(\tau) = w_t(\tau) \oplus w_t^*(\tau) = A_t(\tau)e^{i\phi_t(\tau)} \tag{2}$$

- where $w_t(\tau) = \{1, c(t), c(t)^2, \ldots, c(t)^\infty\}$ is the minimum-phase wavelet corresponding to the inverse of the prediction error operator of order 1, $\{1, -c(t)\}$, $\oplus$ represents correlation and $*$ represents complex conjugate.

Time-frequency representation using AR coefficients

- The coefficient $c(t)$ may be computed by using the Burg algorithm (Burg, 1975). We remark that $|c(t)| < 1$. It may be shown that,

$$R_t(\tau) = R_t(0)w_t(\tau),\ \tau \geq 0 \tag{3}$$

where $R_t(0) = 1/(1 - c(t)c^*(t))$. The normalized derivative of eq. (2) gives,

$$\frac{R_t'(\tau)}{R_t(\tau)} = \frac{A_t'(\tau)}{A_t(\tau)} + i\phi_t'(\tau) \tag{4}$$

where,

$$\phi_t'(0) = \left.\frac{d\phi_t(\tau)}{d\tau}\right|_{\tau=0} = 2\pi f(t) = \frac{1}{R_t(0)} Imag\{R_t'(\tau)\}\Big|_{\tau=0} \tag{5}$$

Time-frequency representation using AR coefficients

- Similar equation was presented by Zoukaneri and Porsani (2015). Taking into considerations the characteristics of the minimum-phase wavelet, $w_t(\tau)$, and both the anti-symmetries of the derivative operator, and the imaginary part of the auto-correlation function $R_t(\tau)$ one obtains the equation for the instantaneous frequency,

$$f(t) = \frac{1}{\pi \Delta t} Imag\{\sum_{n=1}^{\infty} \frac{(-1)^{n-1} c(t)^n}{n}\}\,. \tag{6}$$

Additionally, it can be shown that equation (6) may be written as,

$$f(t) = \frac{1}{\pi \Delta t} Imag\{log(1 + c(t))\} = \frac{1}{\pi \Delta t} arg(1 + c(t)) \\ = \frac{1}{\pi \Delta t} \arctan\{\frac{Imag\{c(t)\}}{1 + real\{c(t)\}}\}\,. \tag{7}$$

Time-frequency representation using AR coefficients

- Then a time-frequency representation of the signal component $x_k(t)$ is given by

$$D_k(t,f) = \sqrt{z_k(t) z_k^*(t)}\; \delta(f - f_k(t))\,. \tag{8}$$

- The following pseudo-code illustrates the process:

DO $k = 1, \ldots, K$ (Components)
- compute the complex trace $z_k(t) = x_k(t) + iy_k(t)$
- compute the coefficients $c_k(t)$
- compute the instantaneous frequency $f_k(t)$ (eq. (7))
- obtain the time-frequency representation $D_k(t,f)$ (eq. (8))

ENDDO

Numerical example

▷ Real seismic data:

- Illustrates the decomposition of a seismic trace into five eigentraces;
- The single-channel SSA method was applied to each trace of the split-spread shot gather to test effect of the recursion and iterations;
- Generate of the average amplitude spectra of the shot gathers obtained after applications of single-channel SSA method.

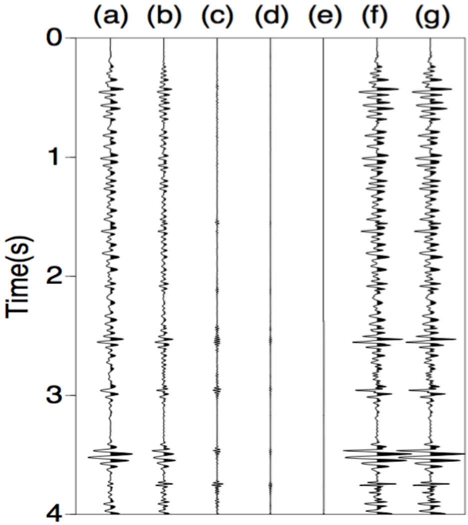

Figure 1: Decomposition of a seismic trace. Signal components 0 to 4, from high energy to low energy corresponding to (a), (b), (c), (d), (e), respectively. The sum of the components in (f) and the original seismic trace in (g).

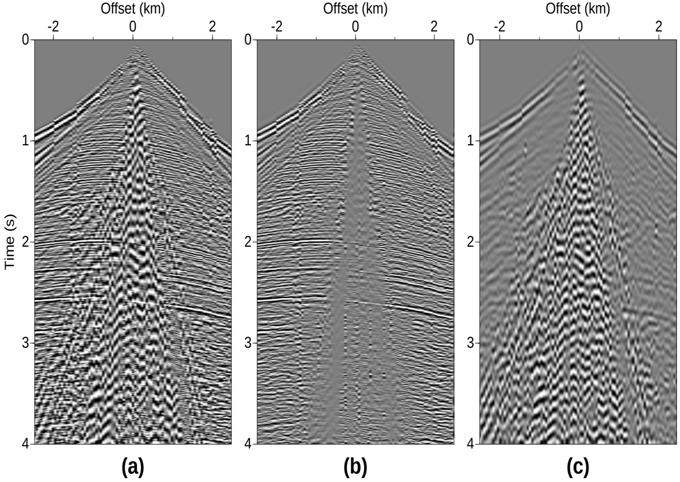

Figure 2: The result of SSA with $M = 11$ recursions in matrix dimension and $J = 20$ iterations in frequency content. The original shot gather in (a), the high-frequency part in (b), and the low-frequency part in (c).

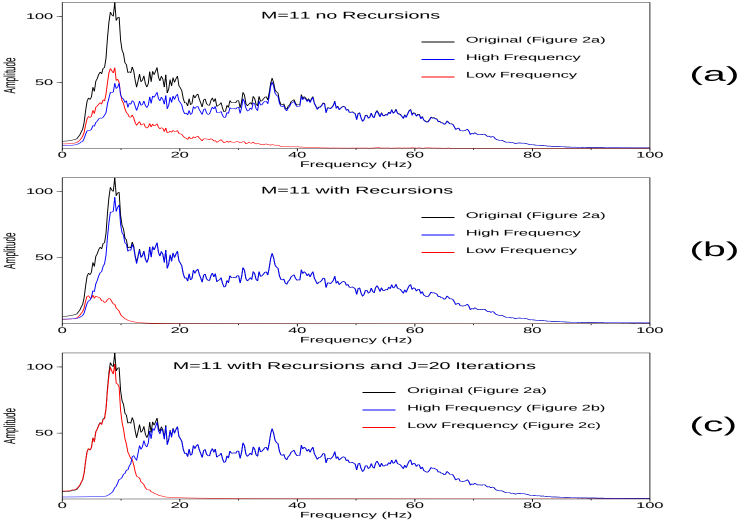

Figure 3: Average amplitude spectra of the data in Fig. 2a with $M = 11$ and no recursions (a), with $M = 11$ recursions in matrix dimension (b), and with $M = 11$ recursions in matrix dimension and $J = 20$ iterations in frequency content (c).

▷ The synthetic signal we analyze is a sum of five elements:

$s_1(t) = 0.8\cos(30\pi t)$	$0\,s \leq t \leq 6\,s$
$s_2(t) = 0.6\cos(70\pi t)$	$0\,s \leq t \leq 6\,s$
$s_3(t) = 0.7\cos(130\pi t) + 5\sin(2\pi t)$	$4\,s \leq t \leq 8\,s$
$s_4(t) = \sin\{\frac{8\pi 100^{t/8}}{log(100)}\}$	$6\,s \leq t \leq 10\,s$
$s_5(t) = 3e^{-1250(t-2)^2}\cos(710(t-2))$	$0\,s \leq t \leq 10\,s$

- It is composed of two harmonic components with frequency of 15 and 35 Hz, a frequency-modulated harmonic around 65 Hz, a sliding harmonic from 35 to 158 Hz, and a Morlet wavelet with central frequency of approximated 113 Hz.
- We generate a composite signal and decomposed using the pseudo-code with K=15 components and J=200 iterations.
- We sum the 15 different signal components into four new signal components with their time-frequency representation. The sum of these gives the composite time-frequency representation is shown.

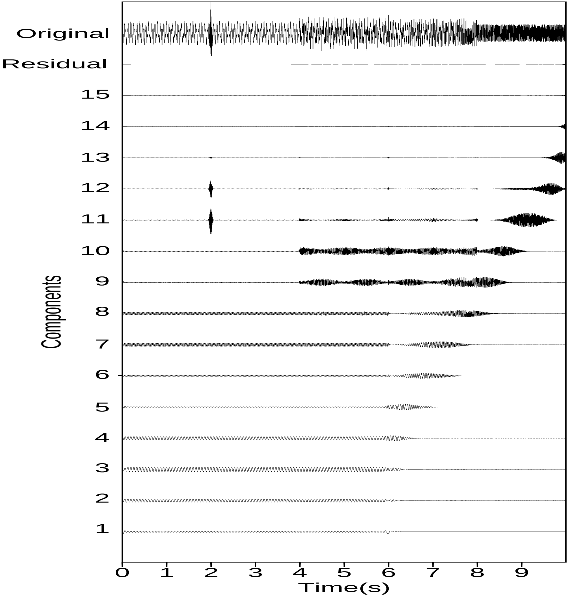

Figure 4: Decomposition of the signal (Original) in top of the figure and 15 signal components.

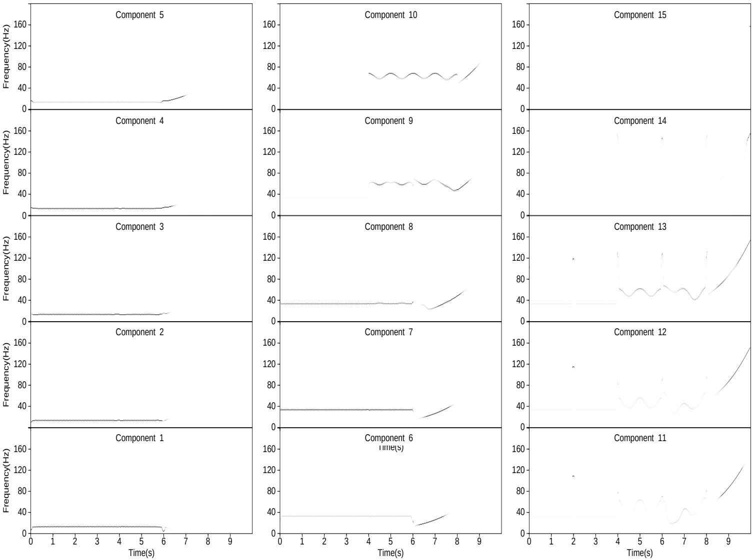

Figure 5: Time-frequency representation of the signal components shown in Figure 4. A window length equal to $2L + 1 = 23$ data samples was used for computation of the AR coefficients.

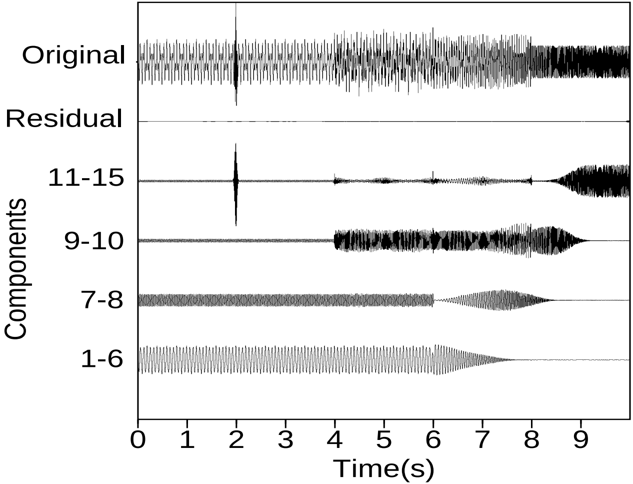

Figure 6: Decomposition of the signal (Original) in top of the figure by combining the signal components in common signal groups.

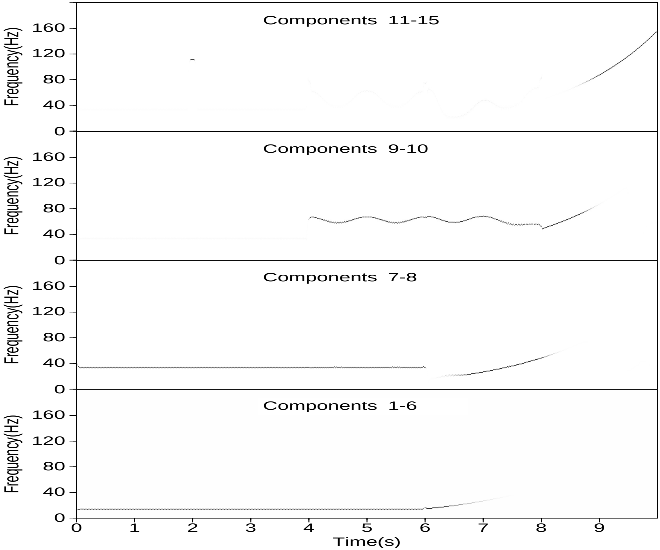

Figure 7: Time-frequency representation of the signal components shown in Figure 6. A window length equal to $2L+1=23$ data samples was used for computation of the AR coefficients.

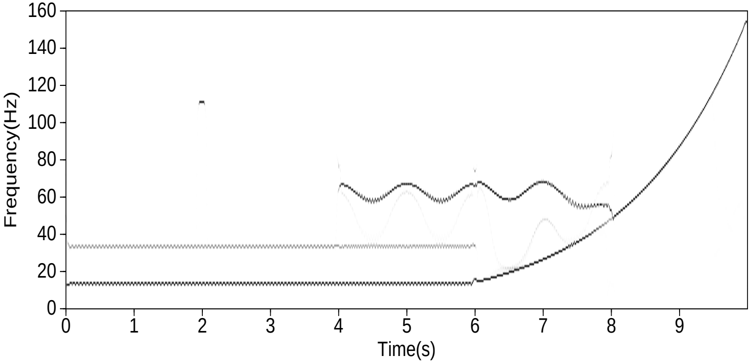

Figure 8: Composite time-frequency representation of the signal shown in Figure 6.

Conclusions

- We proposed an iterative and recursive signal decomposition algorithm based on the SSA method. We demonstrated that the output corresponds to filtering the time series with a zero-phase filter, which is the auto-correlation of the first eigenvector of the covariance matrix of the input signal;
- From the analytic signal and AR modeling we derived a new equation to compute the instantaneous frequency which depend on a single AR coefficient. From each individual component a time-frequency representation is obtained and the sum of these gives a time-frequency distribution of the input signal;
- Application to a synthetic data example shows that the method gives good results compared with other published algorithms.

Acknowledgements

- INCT-GP/CNPq/MCT, ANP, PETROBRAS, FINEP, FAPESB Brazil for financial support;
- Bjorn Ursin has received financial support from Statoil Brazil, from CNPq, Brazil through the project no 400757/2014-1, and from the Norwegian Research Council through the ROSE project.

THANK YOU