First-order SSA-errors for long time series: model examples of simple noisy signals[1]

Nina Golyandina[2], Ekaterina Vlassieva[3]

Abstract

In this paper we consider the problem of reconstruction of a noisy signal by means of Singular Spectrum Analysis and its extension to systems of several time series. We use the formal expansion of the reconstruction error and investigate the first-order (linear by perturbation value) term in the case of the time-series length tending to infinity. An explicit form of the asymptotic variance of the first-order error is derived for simple model examples with constant signals. Simulations confirm that the obtained conclusions are still valid for a wider class of signals including sine-waves.

1 Introduction

Let us consider the problem of reconstruction of a noisy signal by means of SSA (Singular Spectrum Analysis) [2, 3]. Let $\mathrm{F} = \mathrm{S} + \delta\mathrm{E}$ be a noisy signal of length N. Here E is a random noise, δ is a technical parameter and is used only formally, e.g., it can be equal to 1. The aim of this paper is to obtain the variance of reconstruction errors as a function of numbers of the time-series points, asymptotically by signal length. In the paper, we consider the linear in δ first term errors $\mathrm{S}^{(1)} = (s_0^{(1)}, ..., s_{N-1}^{(1)})$ and demonstrate results on asymptotic variance $\mathbb{D}s_l^{(1)}$ considering simple model examples with constant and sine-wave signals.

Section 2 includes a short description of the SSA algorithm and the basic formulas for perturbations.

Section 3 contains an explicit asymptotic ($N \to \infty$) form of $\mathbb{D}s_l^{(1)}$ for the constant signal. This form allows us to investigate dependence of reconstruction errors on noise variance σ^2, window length L, which is the main parameter of SSA, and time-series length N.

Similar considerations take place for Multi-channel SSA (MSSA), an extension of SSA to analysis of a system of time series (see [2, 4] for the MSSA algorithm and theory). Results analogous to that in the one-dimensional case are considered in Section 4 for a system of two noisy signals .

The simulation results confirm that the obtained conclusions are also valid for application of SSA and MSSA to sine-wave signals. These conclusions are formulated in Section 5.

---

[1]Dear editors, please give us a possibility to insert a reference to a grant here

[2]St.Petersburg University, E-mail: nina@gistatgroup.com

[3]St.Petersburg University, E-mail: kate@vlassiev.info

2 Algorithm of SSA and first-order errors

We start with a short description of the SSA algorithm. In the case of signal extraction, the SSA method can be written as a sequence of mappings applied to the initial time series $\mathrm{F} = (f_0, ..., f_{N-1})$ of length N. Introduce these mappings. Fix the so called window length L, $1 < L < N$, and denote $K = N - L + 1$. Let $\mathcal{M}_{L,K}$ be the space of matrices $L \times K$, $\mathcal{M}_{L,K}^{(H)} \subset \mathcal{M}_{L,K}$ be the space of Hankel matrices, and $\mathcal{M}_{L,K}^{(d)} \subset \mathcal{M}_{L,K}$ be the space of matrices of rank d (all of them are with Frobenious inner product).

Let $\mathcal{T} : \mathbb{R}^N \to \mathcal{M}_{L,K}^{(H)}$ be the embedding operator: $\mathbf{F} = \mathcal{T}\mathrm{F}$ has (i, j)-entries f_{i+j-2}, $i = 1, \ldots, L$, $j = 1, \ldots, K$. Note that $\mathcal{T}$ is a one-to-one correspondence and $\mathcal{T}^{-1}$ exists. For a fixed d, let $U_1, ..., U_d$ be left singular vectors of $\mathbf{F}$ corresponding to its d largest singular values, $\mathcal{L}_{\mathrm{F},d} = \mathrm{span}\{U_1, ..., U_d\} \subset \mathbb{R}^L$ and P_F be the matrix of the orthogonal projector from $\mathbb{R}^L$ to the space $\mathcal{L}_{\mathrm{F},d}$. Denote $\mathcal{H}$ the orthogonal projector from $\mathcal{M}_{L,K}$ to $\mathcal{M}_{L,K}^{(H)}$ (hankelisation operator).

Suppose that $d = \mathrm{rank}\, \mathcal{T}\mathrm{S} < \min(L, K)$, that is, we choose d equal to the SSA-rank of the signal S. Then we have the following sequence of objects: the initial time series F, its trajectory matrix $\mathbf{F} = \mathcal{T}\mathrm{F}$, the reconstructed matrix $\widetilde{\mathbf{S}} = P_\mathrm{F}\mathbf{F}$, and the reconstructed signal $\widetilde{\mathrm{S}} = \mathcal{T}^{-1}\mathcal{H}\,\widetilde{\mathbf{S}}$.

Thus, the signal reconstructed by SSA is defined as

$$\widetilde{\mathrm{S}} = \mathcal{T}^{-1}\mathcal{H}P_\mathrm{F}\mathcal{T}\,\mathrm{F}.$$

Denote the trajectory matrices $\mathbf{E} = \mathcal{T}\mathrm{E}$, $\mathbf{S} = \mathcal{T}\mathrm{S}$. Note that $\mathbf{S} \in \mathcal{M}_{L,K}^{(d)}$ and $P_\mathrm{S}\mathbf{S} = \mathbf{S} \in \mathcal{M}_{L,K}^{(H)}$. The reconstruction can be represented in the form $\widetilde{\mathrm{S}} = \mathrm{S} + \mathrm{S}(\delta)$, where the reconstruction error is

$$\mathrm{S}(\delta) = \mathcal{T}^{-1}\mathcal{H}\big((P_{\mathrm{S}+\delta\mathrm{E}} - P_\mathrm{S})\mathbf{S} + \delta P_{\mathrm{S}+\delta\mathrm{E}}\mathbf{E}\big).$$

Conditions for convergence in norm of $P_{\mathrm{S}+\delta\mathrm{E}} - P_\mathrm{S}$ to 0 as $N \to \infty$ are formulated in [5]. The models of noisy signals considered in this paper satisfy these conditions. Since the results of [5] do not guarantee the convergence of $\mathrm{S}(\delta)$ to 0, the convergence was checked by computer simulations.

Considering a formal expansion of $P_{\mathrm{S}+\delta\mathrm{E}}$ in δ, $P_{\mathrm{S}+\delta\mathrm{E}} - P_\mathrm{S} = \delta P^{(1)} + \delta^2 P^{(2)} + \ldots$, we obtain a formal expansion $\widetilde{\mathbf{S}} - \mathbf{S} = \delta\mathbf{S}^{(1)} + \delta^2\mathbf{S}^{(2)} + \ldots$ and therefore, $\mathrm{S}(\delta) = \delta\mathrm{S}^{(1)} + \delta^2\mathrm{S}^{(2)} + \ldots$. Taking into consideration the simulations for noisy signals, we confirm that $\delta\mathrm{S}^{(1)}$ is the main term of the reconstruction error for a wide range of values of δ, as N tends to infinity. That is why we call the linear in δ term of $\mathrm{S}(\delta)$ the first-order reconstruction error. It is easy to see that $\mathbf{S}^{(1)} = P^{(1)}\mathbf{S} + P_\mathrm{S}\mathbf{E}$, where the first summand appears due to errors in the projector operator while the second term is caused by the perturbations of the signal by noise.

The perturbation technique [1] gives the possibility to obtain the expansion of $P_{\mathrm{S}+\delta\mathrm{E}}$. In particular, we can calculate $\mathrm{S}^{(1)} = \mathcal{T}^{-1}\mathcal{H}\,\mathbf{S}^{(1)}$, where for $d = 1$

$$\mathbf{S}^{(1)} = U_1U_1^\mathrm{T}\mathbf{E} + \mathbf{E}V_1V_1^\mathrm{T} - \alpha_{11}U_1V_1^\mathrm{T} \qquad (1)$$

and for $d = 2$

$$\begin{aligned}\mathbf{S}^{(1)} = (U_1 U_1^{\mathrm{T}} + U_2 U_2^{\mathrm{T}})\mathbf{E} + \mathbf{E}(V_1 V_1^{\mathrm{T}} + V_2 V_2^{\mathrm{T}}) - \\ - \left(\alpha_{11} U_1 V_1^{\mathrm{T}} + \alpha_{12} U_1 V_2^{\mathrm{T}} + \alpha_{21} U_2 V_1^{\mathrm{T}} + \alpha_{22} U_2 V_2^{\mathrm{T}}\right).\end{aligned} \tag{2}$$

Here U_i and V_i are left and right singular vectors of $\mathbf{S}$, $\alpha_{ij} = U_i^{\mathrm{T}} \mathbf{E} V_j$. These formulas help us to calculate first-order reconstruction errors for constant (rank equals 1) and sine-wave (rank equals 2) signals.

3 First-order SSA-errors for a constant signal

Consider $f_n = s_n + \varepsilon_n$, where $s_n \equiv c$ and ε_n, $n = 0, \ldots, N-1$, is Gaussian white noise with $\mathbb{D}\varepsilon_n = \sigma^2$ (formally, we take $\delta = 1$). Let σ^2 be not too big to provide separability of the signal from noise. Relation (1) implies that the first-order reconstruction error does not depend on c. Therefore, we set $c = 1$.

Asymptotic formulas Let the window length $L \sim \alpha N$, $0 \le \alpha \le 1/2$, and the number of the considered series point $l \sim \lambda N/2$, $0 \le \lambda \le 1$, as $N \to \infty$. The point with proportion $\lambda = 1$ lies at the middle of the time series, that is, we start with consideration of errors for the first half of the time series and of window lengths less than half of the time-series length.

By direct calculations, we obtain the following asymptotic form of variance of the first-order error as $N \to \infty$:

$$\mathbb{D}s_l^{(1)} \sim F(\alpha, \lambda, N) = \frac{\sigma^2}{N}\psi(\alpha, \lambda) = \frac{\sigma^2}{N}\begin{cases} D_1(\alpha, \lambda), & 0 \le \lambda \le 2(1-2\alpha), \\ D_2(\alpha, \lambda), & 2(1-2\alpha) < \lambda < 2\alpha, \\ D_3(\alpha, \lambda), & 2\alpha \le \lambda \le 1, \end{cases} \tag{3}$$

where

$$\begin{aligned}
D_1(\alpha, \lambda) &= \frac{1}{12\alpha^2(1-\alpha)^2}\left(\lambda^2(1+\alpha) - 2\lambda\alpha(1+\alpha)^2 + 4\alpha\left(3 - 3\alpha + 2\alpha^2\right)\right), \\
D_2(\alpha, \lambda) &= \frac{1}{6\alpha^2(1-\alpha)^2\lambda^2}\left(\lambda^4 + 2\lambda^3\left(-2 + 3\alpha - 3\alpha^2\right) + \right. \\
&+ 2\lambda^2\left(3 - 9\alpha + 12\alpha^2 - 4\alpha^3\right) + 4\lambda\left(-1 + 4\alpha - 3\alpha^2 - 4\alpha^3 + 4\alpha^4\right) + \\
&+ \left. 8\,\alpha - 56\,\alpha^2 + 144\,\alpha^3 - 160\,\alpha^4 + 64\,\alpha^5\right), \\
D_3(\alpha, \lambda) &= \frac{2}{3\alpha}.
\end{aligned}$$

The change points in the conditions of (3) correspond to $l = K - L$ (i.e., with proportion $2(1-2\alpha)$) and $l = L$ (2α). The former change point exists if $K < 2L$ ($\alpha > 1/3$). Note that these formulas can be extended to window lengths $2 < L < N - 1$ ($0 \le \alpha \le 1$) and to numbers of time-series points $0 \le l \le N - 1$ ($0 \le \lambda \le 2$) by the symmetry of error with respect to the middle of the time series and by the equality of results for change $L \leftrightarrow K$ ($\alpha \leftrightarrow 1 - \alpha$).

The case of the known projector Suppose that the projector P_{S} can be found exactly. Then $P^{(1)} = 0$ and $\mathbf{S}^{(1)} = P_{\mathrm{S}}\mathbf{E}$. The following formula for the asymptotic first-order reconstruction error takes place:

$$\mathbb{D}s_l^{(1)} \sim F_0(\alpha, \lambda, N) = \frac{\sigma^2}{3N(1-\alpha)^2} \begin{cases} (3 - 3\alpha - \lambda/2), \ 0 \le \lambda \le 2\alpha, \\ (3 - 4\alpha), \ 2\alpha \le \lambda \le 1. \end{cases} \tag{4}$$

It is interesting that the full first-order error $F(\alpha, \lambda, N)$ generally increases as window length (α) decreases, while $F_0(\alpha, \lambda, N)$ decreases together with α. The equality $F(0.5, 1, N) = F_0(0.5, 1, N)$ holds for the middle point and $L \sim N/2$. Note that for $\alpha = 0$ we get the value $F(0, \lambda, N) = \sigma^2/N$, which is equal to the least square error if we estimate the constant in the parametric model.

Optimal window length Theory of SSA [3] recommends the choice $L = N/2$ to decrease the whole error of reconstruction. The obtained formula (3) allows us to find the window length providing minimal reconstruction errors for the given time-series points. That is, we are interested in $\alpha_{\mathrm{opt}}(\lambda) = \arg \min\limits_{\alpha \in [0,1]} \psi(\alpha, \lambda)$. The graph of α_{opt} is depicted in Fig. 1. Fig. 2 contains the rate of improving the error with respect to the choice $L \sim N/2$:

$$q(\lambda) = \left(\frac{\psi(0.5, \lambda)}{\psi(\alpha_{\mathrm{opt}}, \lambda)} - 1 \right) 100\%.$$

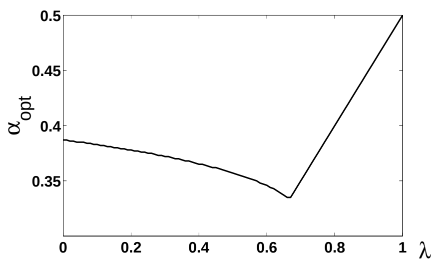

Figure 1: Optimal windows, $\alpha_{\mathrm{opt}}(\lambda)$

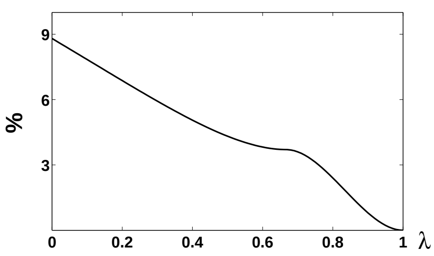

Figure 2: Percentage of improving, $q(\lambda)$

One can see that the choice $L < N/2$ can improve reconstruction of edge time-series points up to 9%.

4 First-order MSSA-errors for constant signals

Consider the system of two time series $(\mathrm{F}, \hat{\mathrm{F}}) = (\mathrm{S}, \hat{\mathrm{S}}) + (\mathrm{E}, \hat{\mathrm{E}})$, with signal terms $s_n \equiv c$, $\hat{s}_n \equiv \hat{c}$ and noise terms ε_n, $\hat{\varepsilon}_n$ with variances σ^2 and $\hat{\sigma}^2$ correspondingly ($n = 0, \ldots, N-1$). By the similar to the one-dimensional case way, we obtain the asymptotic form of variance of the first-order error $\mathrm{S}^{(1)}$. In view of much more complicated calculations, we get the formulas for $L \sim N/2$ ($\alpha = 0.5$) only.

Asymptotic formulas Let $L \sim 0.5N$ and $l \sim \lambda N/2$, $0 \le \lambda \le 1$, as $N \to \infty$. Then for the first time series

$$\mathbb{D}s_l^{(1)} \sim G(\lambda, N) = \frac{2}{3N(c^2+\hat{c}^2)^2}\Big(\lambda^2\big((4c^4+2c^2\hat{c}^2)\sigma^2+2c^2\hat{c}^2\hat{\sigma}^2\big)-$$
$$-\lambda\big((10c^4+\hat{c}^4+7c^2\hat{c}^2)\sigma^2+4c^2\hat{c}^2\hat{\sigma}^2\big)+\big((8c^4+3\hat{c}^4+9c^2\hat{c}^2)\sigma^2+2c^2\hat{c}^2\hat{\sigma}^2\big)\Big).$$

$G(\lambda, N)$ is a quadratic polynomial in λ which decreases from the edges ($\lambda = 0$) of the time series to its middle ($\lambda = 1$).

Proposition 1. *Let $\hat{\sigma} = \sigma$. Then $G(\lambda, N) \le F(1/2, \lambda, N)$ and the equality $G(\lambda, N) = F(1/2, \lambda, N)$ holds if and only if $\hat{c} = 0$.*

Proposition 2. $G(1, N) = 4\sigma^2/(3N)$ *for any c, $\hat{c}$, σ, $\hat{\sigma}$.*

It follows from Proposition 1 that for $\hat{\sigma} = \sigma$ MSSA is better than SSA applied to the first time series separately. Proposition 2 demonstrates the specific effect: the error variance at the middle point of the first time series does not depend on characteristics of the second time series. This means (see the paragraph concerning the case of the known projector) that the error at the middle point is determined by P_{S}E only. For non-middle points, it can be shown that the greater is $\hat{c}$, the smaller is the variance of the reconstruction error.

Using supplementary time series to set a model Consider the problem of extraction of signal for the case of one time series. If the signal structure is known, then we can do the following trick: we can formally involve into consideration the additional time series with the given structure and then apply MSSA to the system of two time series.

In the framework of the considered example with a constant signal, let us take $\hat{s}_n \equiv \hat{c}$ and $\hat{\sigma} = 0$. If $\hat{c} \to +\infty$, then $G(\lambda, N)$ tends to $2\sigma^2(3-\lambda)/3N$. That is, we have the 8/3-times decreasing of variance of the edge errors. The error at the middle point does not effected by the second time series due to a specific choice of window length.

Comparing with (4), one can see that the obtained limit value is exactly the reconstruction error in the case of known projector: $F_0(\alpha, \lambda, N) = 2\sigma^2(3-\lambda)/3N$. It is not surprising as in fact we set the model of the signal. The same result is valid for an arbitrary choice of window length L (it was checked by numerical computations).

As we have mentioned before, behavior of $F_0(\alpha, \lambda, N)$ and $F(\alpha, \lambda, N)$ in α is opposite. In particular, $F_0(\alpha, 1, N)$ strongly decreases as α decreases to zero.

5 Sine-wave signals. Simulation

Let us take the noisy sine-wave signal S with $s_n = \sin(2\pi\omega n)$. The time-series length N and the noise variance σ^2 was chosen to provide an approximate separability of the signal from noise. To apply MSSA, the sine-wave signal $\hat{\mathrm{S}}$ of the second time series was taken with the same frequency.

Estimation of variance of the first-order reconstruction errors was realized by means of simulation on the base of formula (2). Note that it is much more quicker than estimation of variance of the full error following the SSA algorithm.

The simulation gives results analogous to that for constant signals. Namely:
(1) SSA: Error variance decreases from the edges toward the middle of the time series and is approximately constant for time-series points with numbers from $[L, K]$.
(2) SSA: Optimal window length for reconstruction of points close to the edges lies between $0.3N$ and $0.4N$.
(3) MSSA, $L = N/2$: Error variance decreases from the edges to the middle of the time series.
(4) MSSA: assignment of the model with the help of an artificial time series with the same signal frequency as that of S diminishes the variance of the reconstruction error.

Thus, there are reasons to believe that the obtained conclusions are valid for a wider class of time series. In particular, the choice of window length less than a half of the time-series length can decrease forecasting errors as the forecast is based mostly on the last points of the time series.

Acknowledgement

We are grateful to Vladimir Nekrutkin for fruitful discussions of application of perturbation theory to SSA, which inspired us to this investigation.

References

[1] T.Kato (1966) *Perturbation theory for linear operators*. Springer-Verlag, Berlin-Heidelberg-New York.

[2] J.Elsner, A.Tsonis (1996) *Singular Spectrum Analysis. A New Tool in Time Series Analysis*. New York: Plenum Press.

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

[4] N.Golyandina, D.Stepanov (2005) SSA-based approaches to analysis and forecast of multidimensional time series.// In: *Proceedings of the 5th St.Petersburg Workshop on Simulation*. St. Petersburg State University, St. Petersburg, 293–298.

[5] V.Nekrutkin (2009) Perturbation expansions of signal subspaces for long signals // The present volume.