On signal estimation errors using complex singular spectrum analysis

Nina Golyandina [1,*] **and Mikhail Senov** [1]

1   Faculty of Mathematics and Mechanics, St.Petersburg State University, Universitetskaya nab. 7/9, St.Petersburg, 199034, Russia

*   Correspondence: n.golyandina@spbu.ru

Abstract: Singular spectrum analysis (SSA) is a nonparametric method that can be applied to signal estimation. The extension of SSA to the complex case is called CSSA and is applied to complex-valued time series. The accuracy of signal estimation using CSSA is investigated. An explicit form of the first-order errors is obtained for the case of a constant signal and two forms of perturbations, by an outlier and by independent Gaussian noise. It is shown numerically that (1) the first-order errors describe the full errors well, and (2) the formulas for accuracy for estimating a constant signal can be applied to a sum of complex exponentials using multiplication by the number of exponentials.

Keywords: time series; singular spectrum analysis; perturbation analysis; signal estimation.

1. Introduction

The analysis of time series with complex values is an important area in signal processing. Although complex time series do not occur directly in real data, they are constructed from two related features [1,2]) and can also arise as a result of applying the Fourier transform to real data [3,4].

Singular spectrum analysis (SSA) [5] is a powerful method of time series analysis that does not require a prior parametric model of the time series. The method has a natural extension to the case of complex time series, called Complex SSA (CSSA) [6,7]. In fact, one could say that SSA is a special case of CSSA. There is a class of signals, namely time series governed by linear recurrence relations, which allows us to obtain theoretical results regarding the SSA method.

Let the observed complex time series be of the form $X = S + E$. To obtain an estimate $\widetilde{S}$ of the signal S, we will use the CSSA method.

To analyse the signal estimation error we use the perturbation theory [8], which has been applied to the case of signal extraction by the SSA method in a number of papers, see for example [9]. Although Kato's perturbation theory gives a form of the full error, its examination is a difficult task. Therefore, we will only consider the first order error, which is the part that is linear in the perturbation size.

Even for the first order error, obtaining its explicit form is a rather complicated task. Theoretically, we consider only the case of a constant signal. Two types of perturbations are considered, in the form of an outlier and a random noise. The theoretical results obtained allow us to study how CSSA deals with signal extraction.

Certainly, constant signals are not a common case in time series analysis. Therefore, special attention is given to a possible generalisation of the theoretical results, which are demonstrated numerically.

2. Algorithm CSSA

Let us briefly describe singular spectrum analysis and its features. Algorithm 1 presents the steps of CSSA for estimating the signal.

Algorithm 1.
Input: *Complex time series $X = (x_1, \ldots, x_N)$, window length L, signal rank r.*
Output: *Signal estimate $\widetilde{S}$.*
Algorithm:

1. **Embedding.** *Construct $\mathbf{X} \in \mathsf{R}^{L \times K}$, the L-trajectory matrix of the time series X:*

$$\mathbf{X} = \mathcal{T}_L X = [X_1 : \ldots X_K],$$

   where $K = N - L + 1$ and X_i are the L-embedding vectors: $X_i = (x_i, \ldots, x_{i+L-1})^{\mathrm{T}} \in \mathsf{R}^L$.

2. **Decomposition.** *Construct the singular value decomposition (SVD) of the matrix $\mathbf{X}$: $\mathbf{X} = \sum_{k=1}^{\mathrm{rank}\,\mathbf{X}} \sqrt{\lambda_k} U_k V_k^{\mathrm{H}} = \sum_{k=1}^{\mathrm{rank}\,\mathbf{X}} \widehat{\mathbf{X}}_k$, where U_k and V_k are the left and right singular vectors of the matrix $\mathbf{X}$, $\sqrt{\lambda_k}$ are the singular values in decreasing order.*

3. **Grouping.** *Group the matrix components $\widehat{\mathbf{X}}_k$ associated with the signal: $\widehat{\mathbf{S}} = \sum_{k=1}^{r} \widehat{\mathbf{X}}_k$.*

4. **Diagonal averaging.** *Apply the hankelization (which is the projection into the linear space of Hankel matrices): $\widetilde{\mathbf{S}} = \mathcal{H}\widehat{\mathbf{S}}$, and return back to the series form: $\widetilde{S} = \mathcal{T}_L^{-1}\widetilde{\mathbf{S}}$.*

Recommendations for the choice of the window length L can be found in [5,10]. The choice of r is a much more sophisticated problem; information criteria for the choice of r are proposed in [11].

The SSA/CSSA estimation of the signal can be presented in a short form

$$\widetilde{S} = \mathcal{T}_L^{-1} \Pi_{\mathcal{H}} \Pi_r \mathcal{T}_L(X),$$

where $\Pi_{\mathcal{H}}$ is the projector on the space of Hankel matrices (in Frobenius norm), Π_r is the projector on the set of matrices of rank not larger than r.

The L-rank of a time series is the rank of its trajectory matrix. For further considerations we need to know the ranks of certain series. It is known from [12] that the rank of a complex signal whose real and imaginary parts are sinusoids with the same frequency ω, $0 < \omega < 0.5$, is equal to 2 if the absolute phase shift of the sinusoids, given in the interval from $-\pi$ to π, is not equal to $\pi/2$, and is equal to 1 in the case of a complex exponential. The rank of a real sine-wave is 2 under the same frequency constraints. The ranks of complex and real constants are equal to 1 as they can be considered as a special case with $\omega = 0$.

3. Application of perturbation theory to CSSA

Let us observe $X = S(\delta)$, where $S(\delta) = S + \delta E$, and estimate the signal S. It is known from [9] that the reconstruction error $F = \widetilde{S} - S = \mathcal{T}_L^{-1} \Pi_{\mathcal{H}}(\delta \mathbf{S}^{(1)} + \delta^2 \mathbf{S}^{(2)})$.

For simplicity, let $\delta = 1$. We call the first-order reconstruction error the part of the error that is linear in δ: $F^{(1)} = \mathcal{T}_L^{-1} \Pi_{\mathcal{H}}(\delta \mathbf{S}^{(1)})$. Theorem 2.1 from [9] gives the following formula for $\mathbf{S}^{(1)}$ in the case of sufficiently small perturbations:

$$\mathbf{S}^{(1)} = \mathbf{P}_0 \mathbf{E} \mathbf{Q}_0^{\perp} + \mathbf{P}_0^{\perp} \mathbf{E}, \tag{1}$$

where $\mathbf{P}_0^{\perp}$ is the projector onto the column space of $\mathbf{S}$, $\mathbf{Q}_0^{\perp}$ is the projector onto the row space of $\mathbf{S}$, $\mathbf{P}_0 = \mathbf{I} - \mathbf{P}_0^{\perp}$, $\mathbf{I}$ is the unit matrix.

3.1. Relation between complex and real first-order errors for rank 1

In [9] formulas were derived for $\mathbf{S}^{(1)}$ for real series of ranks 1 and 2. Below we give the complex form for the case of rank 1. For rank $\mathsf{S} = 1$ and the leading left and right singular vectors U and V of the trajectory matrix of the series S,

$$\mathbf{S}^{(1)}(\mathbf{E}, \mathsf{S}) = -U^{\mathrm{H}}\mathbf{E}VUV^{\mathrm{H}} + UU^{\mathrm{H}}\mathbf{E} + \mathbf{E}VV^{\mathrm{H}}. \tag{2}$$

Proposition 1. *Let* $\operatorname{rank} \mathsf{S} = \operatorname{rank} \Re(\mathsf{S}) = \operatorname{rank} \Im(\mathsf{S}) = 1$. *Then*

$$F^{(1)} = F_{\Re(S)}^{(1)*} + \mathrm{i} F_{\Im(S)}^{(1)*},$$

where $F^{(1)}$ *is the first-order error of the CSSA estimate of* S, $F_{\Re(S)}^{(1)}$ *is the first-order error of the SSA estimate of* $\Re(S)$, $F_{\Im(S)}^{(1)}$ *is the first-order error of the SSA estimate of* $\Im(S)$.

Proof. By the linearity of (2) in $\mathbf{E}$,

$$\mathbf{S}^{(1)}(\mathbf{E}, \mathsf{S}) = \mathbf{S}^{(1)}(\Re(\mathbf{E}), \mathsf{S}) + \mathrm{i}\mathbf{S}^{(1)}(\Im(\mathbf{E}), \mathsf{S}). \tag{3}$$

Since the rank is equal to 1, the singular vectors of the trajectory matrices of $\Re(\mathsf{S})$, $\Im(\mathsf{S})$ and S coincide. From (4), the linearity of the diagonal averaging operation and the equality of the singular vectors, we get

$$\mathbf{S}^{(1)}(\mathbf{E}, \mathsf{S}) = \mathbf{S}^{(1)}(\Re(\mathbf{E}), \Re(\mathsf{S})) + \mathrm{i}\mathbf{S}^{(1)}(\Im(\mathbf{E}), \Im(\mathsf{S})). \tag{4}$$

□

Note that although the perturbation E can take any form in the statement of the proposition, the proposition has practical application only if the first-order error adequately describes the full error.

An example where the conditions of Proposition 1 are fulfilled is the constant signal. In Section 3.2, we consider the perturbation in the form of an outlier; in Section 3.2, the perturbation is white noise.

Corollary 1. *Let* $\Re(E)$ *and* $\Im(E)$ *be independent. Then in the conditions of Proposition 1*

$$\mathbb{D}F^{(1)} = \mathbb{D}F_{\Re(S)}^{(1)*} + \mathrm{i}\mathbb{D}F_{\Im(S)}^{(1)*}.$$

3.2. Case of constant signals with outliers

In this section we consider the signal $\mathsf{S} = (c_1 + \mathrm{i}c_2, \ldots, c_1 + \mathrm{i}c_2)$ perturbed by the outlier $a_1 + \mathrm{i}a_2$ at position k, i.e. the series E consists of zeros except for the value $a_1 + \mathrm{i}a_2$ at the k-th position. Based on Proposition 1, it is sufficient to compute the first-order reconstruction error for the signal $\mathsf{S} = (c, \ldots, c)$ perturbed by the outlier a at position k.

By substituting $U = \{1/\sqrt{L}\}_{i=1}^{L}$, $V = \{1/\sqrt{K}\}_{i=1}^{K}$, $K = N - L + 1$ into (2) and then diagonal averaging the matrix $\mathbf{S}^{(1)}$, an explicit first-order form of the reconstruction error was obtained.

$1 \leq k < L$

- $k \leq L/2$
  $k \leq K - L$

$$f_l^{(1)} = \frac{a}{LK} \begin{cases} (L + K - k), & 1 \le l \le k \\ \frac{1}{l}(L + K - l)k, & k < l \le L \\ \frac{1}{L}K(L + k - l), & L < l < L + k \\ 0, & L + k \le l \le K \\ \frac{1}{N-l+1}(K - l)(L - k), & K < l < K + k \\ -k, & K + k \le l \le N \end{cases}.$$

- $k \le L/2$
  $k > K - L$

$$f_l^{(1)} = \frac{a}{LK} \begin{cases} (L + K - k), & 1 \le l \le k \\ \frac{1}{l}(L + K - l)k, & k < l \le L \\ \frac{1}{L}K(L + k - l), & L < l < K \\ \frac{1}{N-l+1}(2KL - l(L + K - k)), & K \le l \le L + k \\ \frac{1}{N-l+1}(K - l)(L - k), & L + k < l < K + k \\ -k, & K + k \le l \le N \end{cases}.$$

- $k > L/2$
  $k \le K - L$

$$f_l^{(1)} = \frac{a}{LK} \begin{cases} (L + K - k), & 1 \le l \le k \\ \frac{1}{l}(L + K - l)k, & k < l < L \\ \frac{1}{L}K(L + k - l), & L \le l < L + k \\ 0, & L + k \le l \le K \\ \frac{1}{N-l+1}(L - K)(L - k), & K < l < K + k \\ -k, & K + k \le l \le N \end{cases}.$$

- $k > \max(L/2, K - L)$
  $k \le K/2$

$$f_l^{(1)} = \frac{a}{LK} \begin{cases} (L + K - k), & 1 \le l \le k \\ \frac{1}{l}(L + K - l)k, & k < l < L \\ \frac{1}{L}K(L + k - l), & L \le l < K \\ \frac{1}{N-l+1}(2KL - l(L + K - k)), & K \le l \le L + k \\ \frac{1}{N-l+1}(L - K)(L - k), & L + k < l < K + k \\ -k, & K + k \le l \le N \end{cases}.$$

- $k > K/2$

$$f_l^{(1)} = \frac{a}{LK} \begin{cases} (L + K - k), & 1 \le l \le k \\ \frac{1}{l}(L + K - l)k, & k < l < L \\ \frac{1}{L}K(L + k - l), & L \le l < K \\ \frac{1}{N-l+1}(2KL - l(L + K - k)), & K \le l \le L + k \\ \frac{1}{N-l+1}(K - l)(L - k), & L + k < l < K + k \\ -k, & K + k \le l \le N \end{cases}.$$

$L \leq k \leq K$ 105

$$f_l^{(1)} = \frac{a}{L} \begin{cases} \frac{1}{\min(L,l)}(l-k+L), & k-L \leq l \leq k \\ \frac{1}{\min(L,N-l+1)}(L+k-l), & k < l < L+k \\ 0, & \text{otherwise} \end{cases}$$

$K < k \leq N$ 106

This case is completely analogous to the first case if we invert the time series and take 107
the outlier at position $N - k + 1$. 108

Comments 109

Note that the first-order error tends to 0 only if L and K are proportional to N. Then 110
the first-order error is proportional to $\sqrt{1/\min(L,K)}$, i.e. $1/N$. Therefore, the rate of 111
convergence of the first-order error to zero can be estimated as $\mathcal{O}(1/N)$. 112

Remark 1. *From the above formulas we can see that for fixed L the first order error does not tend to* 113
0 as N increases. As will be shown in the next section, the first order still describes the full error 114
quite accurately in this case. Therefore, the full error does not tend to 0 as N increases. If L and K 115
are proportional to N, the first order error tends to zero. The theoretical basis for the real case and 116
the maximum absolute deviation error can be found in [13]. 117

Note that the first-order error does not depend on the signal and is therefore not 118
determined by the signal-to-noise ratio. 119

3.3. Case of noisy constant signals 120

Consider the time series S with common terms $s_n = c_1 + ic_2$ and the noise matrix $\mathbf{E}$ 121
with variances of real and imaginary parts σ_1^2 and σ_2^2 respectively. Singular vectors of **S**. 122
The conditions of Proposition 1 are obviously satisfied, and therefore the first-order CSSA 123
error is expressed by the first-order errors of the SSA estimates of the real and imaginary 124
parts. 125

In [14] the analytic formula was obtained for the variances of a real constant series, i.e. 126
for $F_{\Re}^{(1)*}$ and $F_{\Im}^{(1)*}$ separately. 127

Let $L = \alpha N$, $\alpha \leq \frac{1}{2}$, $\lambda = \lim_{N\to\infty} 2l/N$. Following [14] and correcting a typo, we get 128

$$\mathbb{D}f_l^{(1)} = \mathbb{D}f_{\Re,l}^{(1)*} + \mathbb{D}f_{\Re,l}^{(1)*} \sim \frac{\sigma_1^2 + \sigma_2^2}{N} \begin{cases} D_1(\alpha,\lambda), & 0 \leq \lambda \leq 2(1-2\alpha) \\ D_2(\alpha,\lambda), & 2(1-2\alpha) < \lambda < 2\alpha, \\ D_3(\alpha,\lambda), & 2\alpha \leq \lambda \leq 1 \end{cases} \quad (5)$$

where

$$D_1(\alpha,\lambda) = \frac{1}{3\alpha^2(1-\alpha)^2}(\lambda^2(\alpha+1) - 2\lambda(1+\alpha)^2 + 4\alpha(-3\alpha+3+2\alpha^2))$$

$$D_2(\alpha,\lambda) = \frac{1}{6\alpha^2\lambda^2(\alpha-1)}(\lambda^4 + 2\lambda^3(3\alpha - 2 - 3\alpha^2) +$$

$$+2\lambda^2(3 - 9\alpha + 12\alpha^2 - 4\alpha^3) + 4\lambda(4\alpha^4 - 4\alpha^3 - 3\alpha^2 + 4\alpha - 1) +$$

$$+8\alpha - 56\alpha^2 + 144\alpha^3 - 160\alpha^4 + 64\alpha^5$$

$$D_3(\alpha,\lambda) = \frac{2}{3\alpha}.$$

Formulas are written down for $l \leq N/2$. For $l \geq N/2$, the error at l is equal to the error at $N - l$.

4. Numerical investigation

4.1. Numerical comparison of first-order error and full signal estimation error

For the case of outlier perturbation, an example with signal $s_l = 1 + i1$, with outlier perturbation $a_1 + ia_2 = 10 + i10$ at position $k = L - 1$ was considered. The results are shown in Table 1.

Table 1. Maximum difference between first-order error and full error for the outlier at $k = L - 1$.

N	50	100	400	1600
$L = N/2$	0.7490	0.1880	0.0109	0.0007
$L = 20$	0.781	0.304	0.063	0.015

Additional numerical experiments allow us to state that when the outlier is in the middle of the series the results are qualitatively the same, for any L the difference tends to 0. However, in the case of a perturbation in the form of an outlier at a fixed point in the series and a fixed L, there is no convergence to zero of the maximum error difference.

In the case of noisy constant series, the numerical results show that the first-order error adequately estimates the full error of the signal reconstruction at each point for all signal parameters considered, even if the errors do not tend to zero. This is demonstrated in the next section.

The result that the first-order error adequately describes the full error was tested for signals of larger ranks (not shown).

4.2. Numerical validation of generalisation to larger ranks

The theoretical results for constant time series appear to be very limited. In this section, we numerically demonstrate that the shape of the errors for signals of larger ranks is similar to the error for signals of rank 1, and their size differs by a multiplier.

Figure 1 shows the behaviour of the absolute errors as a function of the series point. Figure 1 (left) corresponds to the case $r = 1$ (constant signal) and is given by the theoretical formulas presented in Section 3.2. The outlier $10 + i10$ is located at the point $k = 2000$ for a symmetrical figure. Figure 1 (right) corresponds to the signal $s_n = \cos(2\pi n/12) + i\cos(2\pi n/15 + \pi/4)$ of rank 4. One can see that the shapes of the errors are similar. The series lengths are equal to 3999. The ratio of the summed squared errors is 4 for each of the considered window lengths 500, 1000 and 2000.

Thus, the theoretical formulas for an elementary case help to understand the error behaviour for a general case.
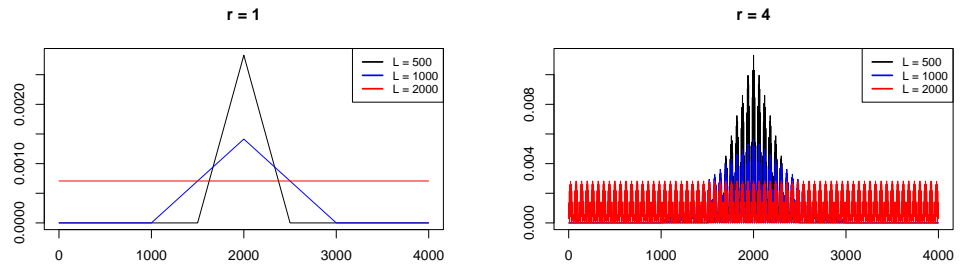

Figure 1. Outlier: Similar squared error behaviour for different ranks r.

The similar relationship of error variances applies to random noise perturbations. We consider the same signals and the complex Gaussian independent noise with variance 1 for real and imaginary parts. The MSE for $L = 1000$ estimated by 500 realisations of the noise is shown in Figure 2. The ratio of the MSE averaging along the series is about 4; this is confirmed by the proximity of the black and blue lines in Figure 2. We also add the line corresponding to the first order error calculated by (5). One can see a good agreement between the first-order error and the full error.
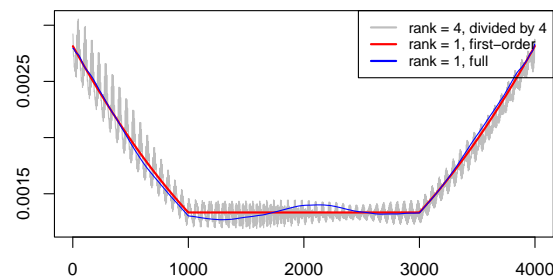

Figure 2. Random noise: Similar error (MSE) behaviour for different ranks *r*.

All numerical results were obtained using the Rssa package http://CRAN.R-project.org/package=Rssa. In the implementation of CSSA, the fast numerical method PRIMME was used to compute the SVDs.

An explicit form of the first-order error of signal estimation at each point was obtained for a constant signal and for both outliers and random noise perturbations. For the random noise, the error was measured by the variance. It appears that the error does not depend on the signal scale. For both examples, the relationship between the first-order error and the full error was investigated numerically.

We have numerically demonstrated how the theoretical results on the accuracy of signal estimation in the simple case of a constant signal help to understand the error behaviour in the general case. These numerical results allow the use of the proposed error formulas in a general case and appear promising for obtaining theoretical results for rank larger than one.

1. Pang, B.; Tang, G.; Tian, T. Complex Singular Spectrum Decomposition and its Application to Rotating Machinery Fault Diagnosis. *IEEE Access (Volume: 7)* **2019**, pp. 143921 – 143934. https://doi.org/10.1109/ACCESS.2019.2945369.
2. Journé, S.; Bihan, N.L.; Chatelain, F.; Flamant, J. Polarized Signal Singular Spectrum Analysis with Complex SSA. In Proceedings of the ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023, pp. 1–5. https://doi.org/10.1109/ICASSP49357.2023.10095954.
3. Trickett, S. F-xy Cadzow noise suppression. *SEG Technical Program Expanded Abstracts* **2008**, pp. 2586–2590.
4. Yuan, S.; Wang, S. A local f-x Cadzow method for noise reduction of seismic data obtained in complex formations. *Pet. Sci. 8* **2011**, p. 269–277. https://doi.org/10.1007/s12182-011-0144-y.
5. Golyandina, N.; Korobeynikov, A.; Zhigljavsky, A. *Singular spectrum analysis with R*; Springer-Verlag Berlin Heidelberg, 2018.
6. Kumaresan, R.; Tufts, D. Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. *IEEE Trans Acoust* **1982**, *30*, 833–840.

7.   Keppenne, C.; Lall, U. Complex singular spectrum analysis and multivariate adaptive regression splines applied to forecasting the southern oscillation. In Proceedings of the Exp. Long-Lead Forcst. Bull, 1996.

8.   Kato, T. *Perturbation theory for linear operators*; Springer-Verlag, Berlin-Heidelberg-New York, 1966.

9.   Nekrutkin, V. Perturbation expansions of signal subspaces for long signals. *Statistics and Its Interface* **2010**, *3*, 297–319.

10. Golyandina, N. On the choice of parameters in Singular Spectrum Analysis and related subspace-based methods. *3*, 259–279.

11. Golyandina, N.; Zvonarev, N. Information Criteria for Signal Extraction Using Singular Spectrum Analysis: White and Red Noise. *Algorithms* **2024**, *17*. https://doi.org/10.3390/a17090395.

12. Golyandina, N.; Korobeynikov, A.; Shlemov, A.; Usevich, K. Multivariate and 2D extensions of singular spectrum analysis with the Rssa package. *Journal of Statistical Software* **2015**, *67(2)*, 1–78.

13. Nekrutkin, V.V. On the Robustness of Singular Spectrum Analysis for Long Time Series. *Vestnik St. Petersburg University, Mathematics* **2024**, *57*, 335–344. https://doi.org/10.1134/S1063454124700183.

14. Golyandina, N.; Vlassieva, E. First-order SSA-errors for long time series: model examples of simple noisy signals. In Proceedings of the Proceedings of the 6th St.Petersburg Workshop on Simulation Vol.1, June 28-July 4, 2009, St. Petersburg. St.-Petersburg State University, 2009, pp. 314–319.