ISSN 1063-4541, Vestnik St. Petersburg University, Mathematics, 2024, Vol. 57, No. 3, pp. 335–344. © Pleiades Publishing, Ltd., 2024.
Russian Text © The Author(s), 2024, published in Vestnik Sankt-Peterburgskogo Universiteta: Matematika, Mekhanika, Astronomiya, 2024, Vol. 11, No. 3, pp. 495–507.

MATHEMATICS

On the Robustness of Singular Spectrum Analysis for Long Time Series

V. V. Nekrutkin

St. Petersburg State University, St. Petersburg, 199034 Russia

e-mail: vnekr@statmod.ru

Received January 13, 2024; revised February 21, 2024; accepted February 22, 2024

Abstract—This paper is devoted to the theoretical investigation of the robustness of singular spectrum analysis (SSA) if the length N of a time series tends to infinity. The latter condition distinguishes the work from quite a lot of works on the robustness of SSA. Here, we used a version of the SSA method that is intended for extraction of the signal from the sum of the signal and noise. Therefore, taking the series corresponding to the available outliers as noise, we can obtain uniform estimates for the signal-approximation errors at large N. If these estimates tend to zero as $N \to \infty$, then the method is robust. Several examples of this approach for specific signals and outliers are considered; some of them are illustrated using computer experiments.

Keywords: signal processing, singular spectrum analysis, outliers, robustness, asymptotical analysis

DOI: 10.1134/S1063454124700183

1. INTRODUCTION

The paper discusses theoretical issues of robustness of the singular spectrum analysis (SSA) method if the length N of a time series tends to infinity. We note straightaway that precisely the condition $N \to \infty$ distinguishes this paper from quite a lot of works on the robustness of SSA (see, e.g., [1] or [2], § 3.12]).

In this case, we use a version of the SSA method designed to extract the signal from the sum of the signal and noise.

Therefore, considering the series corresponding to the existing outliers as noise and using the technique developed in ([3], § 5.3), we are able to obtain uniform estimates for the signal-approximation errors at large N. If these estimates tend to zero as $N \to \infty$, then the method is robust.

Subsection 2.1 describes the used version of the SSA method, and Subsection 2.2 describes a method for estimating approximation errors with some constraints on the kind of signals.

Section 3 is devoted to the outliers themselves: the required definitions are given there, as well as the necessary theory devoted to the so-called single outliers, see Section 3.1. Section 3.2 applies this theory to several specific signals, and Section 3.3 presents considerations for multiple outliers.

Finally, Section 4 contains computational experiments that confirm and clarify some of the theoretical facts proved.

2. GENERAL CONSIDERATIONS

2.1. SSA Method

Let us first look at the version of the singular-spectrum-analysis method that is used in this paper; a general description of this method can be found in [4] or [5].

A real-valued “signal” $\mathrm{F} = (f_0, \ldots, f_n, \ldots)$ is considered. It is assumed that the series F is governed by a linear recurrent formula (LRF) of order d

$$f_n = \sum_{k=1}^{d} a_k f_{n-k}, \quad n \geq d, \tag{1}$$

which is minimal in the sense that there is no LRF of lower order governing series F.

In addition, we introduce “noise” $\mathrm{E} = (e_0, ..., e_n, ...)$ and assume that the series $\mathrm{X}_N = \mathrm{F}_N + \delta\mathrm{E}_N$ is observed, where F_N and E_N are matched segments of length N of the signal and noise, and δ is the perturbation parameter. In other words,

$$\mathrm{F}_N = (f_0,\ldots,f_{N-1}), \quad \mathrm{E}_N = (e_0,\ldots,e_{N-1}), \quad \text{and} \quad \mathrm{X}_N = (f_0 + \delta e_0,\ldots,f_{N-1} + \delta e_{N-1}).$$

The general problem is to extract (approximately) the signal F_N from the sum X_N. In this case, it is assumed that only the order value d of the LRF (1) is known.

Description of the SSA method. The SSA method looks in this case as follows.

(i) The *window length* $L < N$ is chosen and the Hankel *trajectory* matrix $\mathbf{H}(\delta)$ of dimension $L \times K$, $K = N - L + 1$, with elements $\mathbf{H}(\delta)[ij] = x_{i+j-2}$, $1 \le i \le L$, $1 \le j \le K$, is constructed from the series X_N. Here we assume that $\min(L, K) \ge d$. In [4], this operation is called *embedding*.

If $\mathbf{H}$ and $\mathbf{E}$ denote the Hankel matrices obtained from the series F_N and E_N by the embedding operation with the same window length L, then, of course, $\mathbf{H}(\delta) = \mathbf{H} + \delta\mathbf{E}$.

(ii) The matrix $\mathbf{H}(\delta)$ is subjected to a singular-value decomposition and d principal (i.e., corresponding to the largest singular values) elementary matrices of this decomposition are summed. The result of this operation is denoted by $\tilde{\mathbf{H}}(\delta)$.

(iii) We search for the Hankel matrix $\hat{\mathbf{H}}(\delta)$ that is closest to $\tilde{\mathbf{H}}(\delta)$ in the Frobenius norm. Explicitly, this means that, for each secondary diagonal $i + j = \text{const}$, all elements of the matrix $\tilde{\mathbf{H}}(\delta)$ are replaced by their average values. Therefore, this operation is called *diagonal averaging* in [4]. Denoting it by $\mathcal{S}$, we obtain $\hat{\mathbf{H}}(\delta) = \mathcal{S}\tilde{\mathbf{H}}(\delta)$.

(iv) Finally, applying to $\hat{\mathbf{H}}(\delta)$ the operation inverse to embedding, we arrive at the *reconstructed* series $\mathrm{F}_N(\delta) = (f_0(\delta), ..., f_{N-1}(\delta))$, which is declared as an approximation to the signal F_N.

It is natural to call the series $R_N(\delta) = (r_0(\delta), ..., r_{N-1}(\delta))$ with $r_i(\delta) = f_i(\delta) - f_i$ *a series of reconstruction errors*. In ([3], § 5.3), a general scheme of asymptotic analysis of reconstruction errors is proposed. Let us give a brief description of it.

2.2. Approach to the Analysis of Reconstruction Errors

First of all, we are interested in the uniform convergence of the residuals $r_i(\delta)$ to zero, i.e., the behavior of the norm $\|\mathrm{F}_N(\delta) - \mathrm{F}_N\|_{\max} = \max_{0 \le i < N}|r_i(\delta)|$ as $N \to \infty$.

Further, if $U_0^{\perp}$ is a linear space generated by columns of the matrix $\mathbf{H}$, then it follows from (1) that, for L, $K \ge d$, the dimension of $U_0^{\perp}$ equals d.

Let us denote by $\mathbf{P}_0^{\perp}$ an orthogonal projector onto the linear space $U_0^{\perp}$ and by $\mathbf{P}_0^{\perp}(\delta)$, an orthogonal projector onto a linear space generated by columns of the matrix $\tilde{\mathbf{H}}(\delta)$. Then, as shown in ([3], § 5.3),

$$\tilde{\mathbf{H}}(\delta) - \mathbf{H} = \left(\mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp}\right)\mathbf{H}(\delta) + \delta\mathbf{P}_0^{\perp}\mathbf{E}. \tag{2}$$

In this paper, following [3], we use two matrix norms. For a matrix $\mathbf{A}$ of size $L \times K$, the spectral norm $\|\mathbf{A}\|$ is the maximum singular number of this matrix and the uniform norm $\|\mathbf{A}\|_{\max}$ is the maximum of the moduli of the elements of $\mathbf{A}$. The relation between these norms is well-known, according to [6] (see § 2.3.2), as

$$\|\mathbf{A}\|_{\max} \le \|\mathbf{A}\| \le \sqrt{LK}\,\|\mathbf{A}\|_{\max}. \tag{3}$$

Since $\|\mathcal{S}\mathbf{A}\|_{\max} \le \|\mathbf{A}\|_{\max}$, we obtain

$$\max_{0 \le i < N}|r_i(\delta)| \le \left\|(\mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp})\mathbf{H}(\delta) + \delta\mathbf{P}_0^{\perp}\mathbf{E}\right\|_{\max}. \tag{4}$$

Inequality (4) is used as follows. Some operator $\mathbf{N} : \mathbb{R}^L \mapsto \mathbb{R}^L$ is chosen and equality (2) is rewritten in the form

$$\tilde{\mathbf{H}}(\delta) - \mathbf{H} = \left(\mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp} - \mathbf{N}\right)\mathbf{H}(\delta) + \delta\mathbf{P}_0^{\perp}\mathbf{E} + \mathbf{N}\mathbf{H}(\delta).$$

If it turns out in this case that, as $N \to \infty$,

$$\left\|\left(\mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp} - \mathbf{N}\right)\mathbf{H}(\delta)\right\| \le \left\|\left(\mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp} - \mathbf{N}\right)\right\|\|\mathbf{H}(\delta)\| \to 0,$$

then (see (3)) checking the asymptotic behavior of the elements of a specific (even perhaps complex) residual matrix $\delta\mathbf{P}_0^{\perp}\mathbf{E} + \mathbf{N}\mathbf{H}(\delta)$ remains.

In this paper, we use the following option for choosing the operator $\mathbf{N}$.

Let us denote the maximum and minimum positive eigenvalues of the matrix $\mathbf{H}\mathbf{H}^{\mathrm{T}}$ by $\mu_{\max} = \|\mathbf{H}\|^2$ and $\mu_{\min}$, respectively. In addition, let $\mathbf{S}_0$ be the pseudo-inverse Moore–Penrose matrix to the matrix $\mathbf{H}\mathbf{H}^{\mathrm{T}}$ with $\|\mathbf{S}_0\| = 1/\mu_{\min}$. Further, we put

$$\mathbf{B}(\delta) = \mathbf{H}(\delta)(\mathbf{H}(\delta))^{\mathrm{T}} - \mathbf{H}\mathbf{H}^{\mathrm{T}} = \delta(\mathbf{H}\mathbf{E}^{\mathrm{T}} + \mathbf{E}\mathbf{H}^{\mathrm{T}}) + \delta^2\mathbf{E}\mathbf{E}^{\mathrm{T}} \quad \text{and}$$

$$\mathbf{W}_1(\delta) = \mathbf{P}_0\mathbf{B}(\delta)\mathbf{S}_0 + \mathbf{S}_0\mathbf{B}(\delta)\mathbf{P}_0$$
$$= \mathbf{P}_0\mathbf{B}(\delta)\mathbf{S}_0 + \mathbf{S}_0\mathbf{B}(\delta)\mathbf{P}_0 = \delta\mathbf{V}_0^{(1)} + \delta^2\left(\mathbf{P}_0\mathbf{E}\mathbf{E}^{\mathrm{T}}\mathbf{S}_0 + \mathbf{S}_0\mathbf{E}\mathbf{E}^{\mathrm{T}}\mathbf{P}_0\right),$$

where $\mathbf{V}_0^{(1)} = \mathbf{P}_0\mathbf{E}\mathbf{H}^{\mathrm{T}}\mathbf{S}_0 + \mathbf{S}_0\mathbf{H}\mathbf{E}^{\mathrm{T}}\mathbf{P}_0$, $\mathbf{P}_0 = \mathbf{I} - \mathbf{P}_0^{\perp}$, and $\mathbf{I}$ is the identity $L \times L$ matrix.

Then the following assertion holds (see [3], Theorem 2.4), which is derived using the classical results of Kato ([7], Ch. 2, § 3).

Theorem 1. *Let* $\delta_0 > 0$ *and* $\|\mathbf{B}(\delta)\|/\mu_{\min} < 1/4$ *for all* $\delta \in [-\delta_0; \delta_0]$. *Then there exists an absolute constant* C *such that*

$$\left\|\mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp} - \mathbf{W}_1(\delta)\right\| \le C\left(\frac{\|\mathbf{B}(\delta)\|}{\mu_{\min}}\right)^2 \frac{1}{1 - 4\|\mathbf{B}(\delta)\|/\mu_{\min}}.$$

Thus, we arrive at the following inequality, which holds under the conditions of Theorem 1:

$$\max_{0\le i<N} |r_i(\delta)| \le C\left(\frac{\|\mathbf{B}(\delta)\|}{\mu_{\min}}\right)^2 \frac{\|\mathbf{H}(\delta)\|}{1 - 4\|\mathbf{B}(\delta)\|/\mu_{\min}} + \left\|\delta\mathbf{P}_0^{\perp}\mathbf{E} + \mathbf{W}_1(\delta)\mathbf{H}(\delta)\right\|_{\max}.$$

Considering that $\mathbf{H}^{\mathrm{T}}\mathbf{S}_0\mathbf{H} = \mathbf{Q}_0^{\perp}$, where $\mathbf{Q}_0^{\perp}$ is the matrix of orthogonal projection onto the space of rows of the matrix $\mathbf{H}$, we write

$$\begin{aligned}\delta\mathbf{P}_0^{\perp}\mathbf{E} + \mathbf{W}_1(\delta)\mathbf{H}(\delta) &= \delta\left(\mathbf{P}_0^{\perp}\mathbf{E} + \mathbf{P}_0\mathbf{E}\mathbf{H}^{\mathrm{T}}\mathbf{S}_0\mathbf{H}(\delta) + \mathbf{S}_0\mathbf{H}\mathbf{E}^{\mathrm{T}}\mathbf{P}_0\mathbf{H}(\delta)\right)\\ &+ \delta^2\left(\mathbf{P}_0\mathbf{E}\mathbf{E}^{\mathrm{T}}\mathbf{S}_0 + \mathbf{S}_0\mathbf{E}\mathbf{E}^{\mathrm{T}}\mathbf{P}_0\right)\mathbf{H}(\delta) = \delta\left(\mathbf{P}_0^{\perp}\mathbf{E} + \mathbf{E}\mathbf{Q}_0^{\perp} - \mathbf{P}_0^{\perp}\mathbf{E}\mathbf{Q}_0^{\perp}\right)\\ &+ \delta^2\left(\mathbf{P}_0\mathbf{E}\mathbf{H}^{\mathrm{T}}\mathbf{S}_0\mathbf{E} + \mathbf{S}_0\mathbf{H}\mathbf{E}^{\mathrm{T}}\mathbf{P}_0\mathbf{E} + \mathbf{P}_0\mathbf{E}\mathbf{E}^{\mathrm{T}}\mathbf{S}_0\mathbf{H}\right) + \delta^3\left(\mathbf{S}_0\mathbf{E}\mathbf{E}^{\mathrm{T}}\mathbf{P}_0\mathbf{E} + \mathbf{P}_0\mathbf{E}\mathbf{E}^{\mathrm{T}}\mathbf{S}_0\mathbf{E}\right).\end{aligned} \tag{5}$$

Thus, we obtain a natural estimate:

$$\max_{0\le i<N} |r_i(\delta)| \le J_1 + J_2 + J_3, \quad \text{where} \tag{6}$$

$$J_1 = C\left(\frac{\|\mathbf{B}(\delta)\|}{\mu_{\min}}\right)^2 \frac{\|\mathbf{H}(\delta)\|}{1 - 4\|\mathbf{B}(\delta)\|/\mu_{\min}}, \tag{7}$$

$$J_2 = |\delta|\left(\left\|\mathbf{P}_0^{\perp}\mathbf{E} + \mathbf{E}\mathbf{Q}_0^{\perp} - \mathbf{P}_0^{\perp}\mathbf{E}\mathbf{Q}_0^{\perp}\right\|_{\max}\right), \quad \text{and} \tag{8}$$

$$\begin{aligned}J_3 = \delta^2&\left(\left\|\mathbf{P}_0\mathbf{E}\mathbf{H}^{\mathrm{T}}\mathbf{S}_0\mathbf{E}\right\|_{\max} + \left\|\mathbf{S}_0\mathbf{H}\mathbf{E}^{\mathrm{T}}\mathbf{P}_0\mathbf{E}\right\|_{\max} + \left\|\mathbf{P}_0\mathbf{E}\mathbf{E}^{\mathrm{T}}\mathbf{S}_0\mathbf{H}\right\|_{\max}\right)\\ &+ |\delta|^3\left(\mathbf{S}_0\mathbf{E}\mathbf{E}^{\mathrm{T}}\mathbf{P}_0\mathbf{E}\|_{\max} + \left\|\mathbf{P}_0\mathbf{E}\mathbf{E}^{\mathrm{T}}\mathbf{S}_0\mathbf{E}\right\|_{\max}\right).\end{aligned} \tag{9}$$

We now introduce additional constraints.

Lemma 1. *Let* $N \to \infty$ *and* $\|\mathbf{E}\| \ge c = \mathrm{const}$. *If* $\mu_{\max}/\mu_{\min} \le æ = \mathrm{const}$ *and* $|\delta|\,\|\mathbf{E}\|/\|\mathbf{H}\| \to 0$, *then the term* $J_1 + J_3$ *on the right side of* (6) *has the form* $O(\delta^2\|\mathbf{E}\|^2/\|\mathbf{H}\|)$.

Proof. First of all, $\|\mathbf{B}(\delta)\|/\mu_{\min} = O(|\delta|\,\|\mathbf{E}\|/\|\mathbf{H}\|)$ as $N \to \infty$. Indeed, since $\|\mathbf{B}(\delta)\| \le 2|\delta|\,\|\mathbf{H}\|\,\|\mathbf{E}\| + \delta^2\|\mathbf{E}\|^2$ and $\mu_{\max} = \|\mathbf{H}\|^2$, then

$$\begin{aligned}\|\mathbf{B}(\delta)\|/\mu_{\min} &\le 2|\delta|\|\mathbf{E}\|\sqrt{\mu_{\max}}/\mu_{\min} + \delta^2\|\mathbf{E}\|/\mu_{\min}\\ &\le 2|\delta|\|\mathbf{E}\|/\sqrt{\mu_{\max}} + \delta^2\|\mathbf{E}\|^2 æ/\mu_{\max} = O(|\delta|\|\mathbf{E}\|/\|\mathbf{H}\|).\end{aligned} \tag{10}$$

Thus, the condition of Theorem 1 is satisfied, $1 - 4\|\mathbf{B}(\delta)\|/\mu_{\min} \to 1$, $(\|\mathbf{B}(\delta)\|/\mu_{\min})^2 = O(|\delta|^2 \|\mathbf{E}\|^2 \|\mathbf{H}\|^{-2})$, and $\|\mathbf{H}(\delta)\| \sim \|\mathbf{H}\|$. The required assertion about J_1 now follows from (7).

Since $\|\mathbf{S}_0\mathbf{E}\| \le \|\mathbf{S}_0\|\,\|\mathbf{E}\| = \|\mathbf{E}\|/\|\mathbf{H}\|^2$ and $\|\mathbf{S}_0\mathbf{H}\| \le \|\mathbf{S}_0\|\,\|\mathbf{H}\| = \|\mathbf{H}\|^{-1}$, then, according to (3),

- $\|\mathbf{P}_0\mathbf{E}\mathbf{H}^{\mathrm{T}}\mathbf{S}_0\mathbf{E}\|_{\max} \le \|\mathbf{P}_0\mathbf{E}\mathbf{H}^{\mathrm{T}}\mathbf{S}_0\mathbf{E}\| \le \|\mathbf{P}_0\|\,\|\mathbf{E}\|\,\|\mathbf{H}\|\,\|\mathbf{S}_0\mathbf{E}\| = \|\mathbf{E}\|^2/\|\mathbf{H}\|$,
- similarly, $\|\mathbf{S}_0\mathbf{H}\mathbf{E}^{\mathrm{T}}\mathbf{P}_0\mathbf{E}\|_{\max} \le \|\mathbf{E}\|^2/\|\mathbf{H}\|$ and $\|\mathbf{P}_0\mathbf{E}\mathbf{E}^{\mathrm{T}}\mathbf{S}_0\mathbf{H}\|_{\max} \le \|\mathbf{E}\|^2/\|\mathbf{H}\|$,
- $\|\mathbf{S}_0\mathbf{E}\mathbf{E}^{\mathrm{T}}\mathbf{P}_0\mathbf{E}\|_{\max} \le \|\mathbf{S}_0\mathbf{E}\mathbf{E}^{\mathrm{T}}\mathbf{P}_0\mathbf{E}\| \le \|\mathbf{E}\|^2/\|\mathbf{H}\|^2$ and, similarly,
- $\|\mathbf{P}_0\mathbf{E}\mathbf{E}^{\mathrm{T}}\mathbf{S}_0\mathbf{E}\| \le \|\mathbf{E}\|^2/\|\mathbf{H}\|^2$.

Therefore, $J_3 = O(\delta^2\|\mathbf{E}\|^2/\|\mathbf{H}\|) + O(|\delta|^3\,\|\mathbf{E}\|^2/\|\mathbf{H}\|^2)$. Taking into account the fact that $\|\mathbf{E}\|$ is separated from zero and $|\delta|\,\|\mathbf{E}\|/\|\mathbf{H}\| \to 0$, we obtain the desired assertion regarding J_3.

□

Remark 1. The boundedness of the ratio $\mu_{\max}/\mu_{\min}$ holds, in particular, for $d = 1$ and also (under natural conditions for the dimension of the matrix $\mathbf{H}$) for polynomial signals and signals that are a linear combination of different harmonics (see, for example, [3], Lemma 3.1).

Remark 2. It is easy to see that, if we replace the condition $|\delta|\,\|\mathbf{E}\|/\|\mathbf{H}\| \to 0$ in Lemma 1 by $|\delta|\Delta/\|\mathbf{H}\| \to 0$, where $\|\mathbf{E}\| \le \Delta = \Delta_N$, then it turns out that $J_1 + J_3 = O(\delta^2\Delta^2/\|\mathbf{H}\|)$.

3. ROBUSTNESS TO OUTLIERS

In our case, the problem lies in examining the robustness of the SSA method to outliers. Let us give the corresponding definitions.

Definition 1. Consider the sequence of series as $N \to \infty$:

$$\mathrm{E}_N = (e_0^{(N)}, \ldots, e_{N-1}^{(N)})$$

and denote by $\mathrm{Re}(N)$ the set of nonzero elements of the series $\mathbf{E}_N$. Let $\mathfrak{n}(N) = \mathrm{card}(\mathrm{Re}(N))$.

For any signal F, we consider the sequence of series $\mathrm{X}_N = \mathrm{F}_N + \delta\mathrm{E}_N$. If $\mathfrak{n}(N)/N \to 0$ as $N \to \infty$, then the sequence of series $\mathbf{E}_N$ will be called an *asymptotic additive outlier with respect to the series* F, and the quantity $|\delta|\,|e_i|$ will be called the *power of the ith element of* E_N for $i \in \mathrm{Re}(N)$.

If $\mathfrak{n}(N) = 1$ for any N, then the series E_N will be called a *single outlier*.

Let us denote by $r_i(\mathrm{E}_N, \delta)$ the error of reconstruction of the series F_N for an additive outlier E_N. In the case of a single outlier with position M, we write $r_i(M, \delta)$ instead of $r_i(\mathrm{E}_N, \delta)$.

Definition 2. The SSA method is called *asymptotically robust with respect to outliers* E_N if $\max_{0 \le i < N}|r_i(\mathrm{E}_N, \delta)| \to 0$ as $N \to \infty$.

Remark 3. (i) Thus, the problem of checking the robustness of the SSA method with respect to outliers can be solved in the way described in Subsection 2.1.

(ii) Here, the formal parameter δ is not assumed to be constant. That is, we believe it possible that $|\delta| = |\delta(N)| \to \infty$ or $|\delta| \to 0$ as $N \to \infty$. In particular, the condition $|\delta|\,\|\mathbf{E}\|/\|\mathbf{H}\| \to 0$ of Lemma 1 can be satisfied not only at $\|\mathbf{H}\| \to \infty$ but also under a bounded norm $\|\mathbf{H}\|$ due to $\delta \to 0$.

(iii) In this formulation, the problem of filling gaps can be formally reduced to the problem of robustness to outliers. Indeed, if the gap numbers make up the set $\mathrm{Im}(N)$, and the gaps themselves are coded as zeros, then for this it suffices to take $\mathrm{Re}(N) = \mathrm{Im}(N)$ and put $\delta e_i^{(N)} = -f_i(N)$ for $i \in \mathrm{Re}(N)$.

3.1. Single Outlier

If $\mathrm{card}(\mathrm{Re}(N)) = 1$, then we can assume that $e_M = 1$ for $M \in \mathrm{Re}(N)$ and the magnitude of the outlier is equal to δ.

Let $\mathbf{E}_{k,\ell}$ be an $L \times K$ matrix with $e_{k,\ell} = 1$ and the remaining zero terms, $1 \le k \le L$, $1 \le \ell \le K$. Then $\mathbf{E}_{k,\ell} = \mathfrak{o}_k(L)\mathfrak{o}_\ell^{\mathrm{T}}(K)$, where $\mathfrak{o}_j(k) \in \mathbb{R}^k$ is the jth unit vector in $\mathbb{R}^k$. Therefore, $\|\mathbf{E}_{k,\ell}\| = 1$.

We consider noise of the form $\mathrm{E}_N = (e_0, \ldots, e_{N-1})$, where $e_M = 1$ for $0 \le M \le N - 1$ and the rest are zeros: $e_j = 0$. Thus, M is the position of a single outlier and $M = M(N)$.

Let $\mathbf{E}$ be an $L \times K$-trajectory matrix of the series E_N, $L \le K$. Then

$$\mathbf{E} = \sum_{k+\ell=M+2} \mathbf{E}_{k,\ell} = \sum_{k+\ell=M+2} \mathfrak{o}_k(L)\mathfrak{o}_\ell^{\mathrm{T}}(K) = \sum_k \mathfrak{o}_k(L)\mathfrak{o}_{M+2-k}^{\mathrm{T}}(K),$$

where

$$\sum_k = \begin{cases} \sum_{k=1}^{M+1} & \text{at} \quad 0 \le M < L-1, \\ \sum_{k=1}^{L} & \text{at} \quad L-1 \le M < K, \\ \sum_{k=M-K+2}^{N-K+1} & \text{at} \quad K \le M < N. \end{cases}$$

Nevertheless, $\|\mathbf{E}\|_{\max} = \|\mathbf{E}\| = 1$ for any behavior of M.

Since $\|\mathbf{E}\| = 1$ for a single outlier, we have $J_1 + J_3 = O(|\delta^2|/\|\mathbf{H}\|)$ under the conditions of Lemma 1. Thus, if $|\delta|^2/\|\mathbf{H}\| \to 0$ as $N \to \infty$, then, under the conditions of Lemma 1, we still have to take a look at the term J_2 on the right side of (6).

The next estimation will be useful here. Let $\mathbf{G}^{(1)}$ be some matrix of size $L \times L$, and let $\mathbf{G}^{(2)}$ be a $K \times K$ matrix.

Lemma 2. *The inequalities* $\|\mathbf{G}^{(1)}\mathbf{E}\|_{\max} \le \|\mathbf{G}^{(1)}\|_{\max}$ *and* $\|\mathbf{E}\mathbf{G}^{(2)}\|_{\max} \le \|\mathbf{G}^{(2)}\|_{\max}$ *hold for any position M of a single outlier.*

Proof. The assertion for $\|\mathbf{G}^{(1)}\mathbf{E}\|_{\max}$ follows from the fact that every nonzero element of the matrix $\mathbf{G}^{(1)}\mathbf{E}$ coincides with some element of the matrix $\mathbf{G}^{(1)}$. More precisely, let $G_i \in \mathbb{R}^L$ denote the ith column of the matrix $\mathbf{G}^{(1)}$. Next, let $\mathbf{G}_{ij} = [G_i : \ldots : G_j]$ for $1 \le i, j \le K$. Then, for $L \le K$,

$$\mathbf{G}_1\mathbf{E} = \begin{cases} \left[\mathbf{G}_{M+1,1} : \mathbf{0}^{(L,K-M+1)}\right] & \text{at} \quad 0 \le M \le L-1, \\ \left[\mathbf{0}^{(L,M-2)} : \mathbf{G}_{L,1} : \mathbf{0}^{(L,K-L-M+2)}\right] & \text{at} \quad L-1 < M \le K-1, \\ \left[\mathbf{0}^{(L,M-2)} : \mathbf{G}_{L,K-M+2}\right] & \text{at} \quad K-1 < M \le N-1, \end{cases}$$

where $\mathbf{0}^{(i,j)}$ is the zero matrix of size $i \times j$. Therefore,

$$\|\mathbf{G}_1\mathbf{E}\|_{\max} = \begin{cases} \|\mathbf{G}_{M+1,1}\|_{\max} & \text{at} \quad 0 \le M < L-1, \\ \|\mathbf{G}_{L,1}\|_{\max} & \text{at} \quad L-1 \le M \le K-1, \\ \|\mathbf{G}_{L,K-M+2}\|_{\max} & \text{at} \quad K-1 < M \le N-1. \end{cases} \tag{11}$$

This immediately implies the assertion of the lemma for $\mathbf{G}_1\mathbf{E}$. For $\mathbf{E}\mathbf{G}_2$, everything is the same.

□

3.2. Examples. A Single Outlier

We consider a real-valued signal $\mathrm{F} = (f_0, \ldots, f_n, \ldots)$ and assume that the series F is governed by the minimal linear recurrent formula of order d.

Noise is a sequence of single outliers $\mathrm{E}_N = (e_0, \ldots, e_{N-1})$, where $e_M = 1$ for $0 \le M \le N-1$, and the rest are zeros: $e_j = 0$; so, $M = M(N)$ is the position of a single outlier in the noise series E_N. As usual, we assume that $L/N \to \alpha \in (0, 0.5]$.

Let $N \to \infty$. We apply the SSA method to the series $\mathrm{X}_N = \mathrm{F}_N + \delta\mathrm{E}_N$ with the choice of the d principal components. In addition, let $r_i(M, \delta)$ be the error of reconstruction of the term f_i in the series F_N when the noise E_N is used.

Theorem 2. 1. *If* $|\delta| = o(a^{N/2})$, *then, as* $M/N \to \beta$, *for a growing exponential signal* $f_n = a^n$ *with* $a > 1$,

$$\limsup_N \max_i r_i(M,\delta) \le |\delta|(1-a^{-2})\begin{cases} a^{-(\alpha-\beta)N} + a^{-(1-\alpha-\beta)N} & \beta < \alpha, \\ 1 & \alpha \le \beta < 1-\alpha, \\ 2 & \beta \ge 1-\alpha. \end{cases}$$

2. *If* $|\delta|/N \to 0$, *then, for a linear signal* $f_n = \theta_1 n + \theta_0$ *with* $\theta_1 \ne 0$,

$$\max_M \max_i |r_i(M,\delta)| = O(|\delta|/N) \to 0 \quad at \quad N \to \infty. \tag{12}$$

3. *We consider a constant signal* F *with* $f_n = 1$. *Then*

- *if* $|\delta| \le \delta_0$, *then* $\max_M \max_i |r_i(M, \delta)| = O(|\delta|/N)$;
- *if* $|\delta| \to \infty$ *and* $\delta^2/N \to 0$, *then* $\max_M \max_i |r_i(M, \delta)| = O(\delta^2/N)$.

4. *For the signal*

$$f_n = \sum_{i=1}^{r} f_{i,n}, \quad where \quad f_{i,n} = b_k \cos(\omega_i n + \gamma_i) \tag{13}$$

with pairwise different frequencies $\omega_i \in (0, 1/2)$ *and nonzero amplitudes* $|b_i|$

- *if* $|\delta| \le \delta_0$, *then* $\max_M \max_i |r_i(M, \delta)| = O(|\delta|/N)$;
- *if* $|\delta| \to \infty$ *and* $\delta^2/N \to 0$, *then* $\max_M \max_i |r_i(M, \delta)| = O(\delta^2/N)$.

Proof. In all the cases, we use inequality (6). To estimate the quantities J_1, J_2, and J_3, we use the results of Lemmas 1 and 2; moreover, representation (11) may be useful to us in the last case.

In other words, as follows from Subsection 3.1, if $|\delta|/\|\mathbf{H}\| \to 0$, then

$$\max_i |r_i(M,\delta)| = O\left(\delta^2/\|\mathbf{H}\| + |\delta| \left\|\mathbf{P}_0^{\perp}\mathbf{E} + \mathbf{E}\mathbf{Q}_0^{\perp} - \mathbf{P}_0^{\perp}\mathbf{E}\mathbf{Q}_0^{\perp}\right\|_{\max}\right).$$

1. For a growing exponential signal, $d = 1$ with $\|\mathbf{H}\| \asymp a^N$ and therefore, since $|\delta|/a^N \to 0$, we have $J_1 + J_3 = O(\delta^2 a^{-N})$.

Since, for $1 \le i, j \le L$,

$$\mathbf{P}_0^{\perp}[i,j] = (a^2-1)\frac{a^{i+j-2}}{a^{2L}-1}$$

it follows that $\left\|\mathbf{P}_0^{\perp}\right\|_{\max} \to \text{const} = 1 - a^{-2}$. Therefore, for the “worst” location of the place M of a single outlier, $\left\|\mathbf{P}_0^{\perp}\mathbf{E}\right\|_{\max}$ does not tend to zero. Similarly, for $1 \le i, j \le K$,

$$\mathbf{Q}_0^{\perp}[i,j] = (a^2-1)\frac{a^{i+j-2}}{a^{2K}-1}$$

and $\max_M \left\|\mathbf{E}\mathbf{Q}_0^{\perp}\right\|_{\max}$ does not tend to zero. At the same time, this may not be the case for specific locations of the outlier. Let $L \le K$, $L/N \to \alpha \in (0, 1/2]$. Then

$$\max_i \mathbf{P}_0^{\perp} E[i,j] = \mathbf{P}_0^{\perp}\mathrm{E}[L,j] = (a^2-1)\frac{a^{L+j-2}}{a^{2L}-1} = (1-a^{-2})\frac{a^j}{a^L - a^{-L}}$$

and therefore, it follows from (11) that $\left\|\mathbf{P}_0^{\perp}\mathbf{E}\right\|_{\max} \sim (1-a^{-2})a^{-(\alpha-\beta)N}$ for $M/N \to \beta < \alpha$.

But if $M/N \to \beta \ge \alpha$, then $\left\|\mathbf{P}_0^{\perp}\mathbf{E}\right\|_{\max} \to 1 - a^{-2}$.

Similarly, if $M/N \to \beta < 1-\alpha$, then $\left\|\mathbf{E}\mathbf{Q}_0^{\perp}\right\|_{\max} \sim (1-a^{-2})a^{-(1-\alpha-\beta)N}$, and if $M/N \to \beta \ge 1-\alpha$, then $\left\|\mathbf{E}\mathbf{Q}_0^{\perp}\right\|_{\max} \to 1-a^{-2}$. In addition,

$$\min_M \left\|\mathbf{P}_0^{\perp}\mathbf{E}\mathbf{Q}_0^{\perp}\right\|_{\max} \ge \left\|\mathbf{P}_0^{\perp}\mathfrak{o}_1(L)\mathfrak{o}_1(K)^{\mathrm{T}}\mathbf{Q}_0^{\perp}\right\|_{\max} = \left\|\mathbf{P}_0^{\perp}(1)\mathbf{Q}_0^{\perp}(1)\right\|_{\max}$$
$$= (1-a^{-2})^2\frac{(a^{L-1}-1)(a^{K-1}-1)}{(a^{2L}-1)(a^{2K}-1)} \asymp O(a^{-N}).$$

It immediately follows that, for $L/N \to \alpha \in (0, 1/2]$ and $M/N \to \beta$,

$$J_2 \le |\delta|\left(\left\|\mathbf{P}_0^{\perp}\mathbf{E}\right\|_{\max} + \left\|\mathbf{E}\mathbf{Q}_0^{\perp}\right\|_{\max} - \min_M\left\|\mathbf{P}_0^{\perp}\mathbf{E}\mathbf{Q}_0^{\perp}\right\|_{\max}\right)$$
$$\sim |\delta|(1-a^{-2})\begin{cases} a^{-(\alpha-\beta)N} + a^{-(1-\alpha-\beta)N} & \beta < \alpha, \\ 1 & \alpha \le \beta < 1-\alpha, \\ 2 & \beta \ge 1-\alpha. \end{cases}$$

The statement is proved.

2. First of all, we can restrict ourselves to the case $\theta_1 = 1$.

Then, according to Lemma 3.1 in [3], if $L/N \to \alpha \in (0, 1)$, then $\mu_{\max}/N^4 \to \Theta_{\max}$ and $\mu_{\min}/N^4 \to \Theta_{\min}$ for some $\Theta_{\max} \ge \Theta_{\min} > 0$.

Thus, since $\|\mathbf{E}\| = 1$ and $\|\mathbf{H}\| \asymp N^2$ for $L/N \to \alpha \in (0, 1)$, it follows from Lemma 1 that the term $J_1 + J_3$ on the right side of (6) has the order $O(\delta^2 N^{-2}) = o(|\delta|/N)$.

Let us now prove that the equality $\max_M \left\|\mathbf{P}_0^{\perp}\mathbf{E}\right\|_{\max} = O(N^{-1})$ is true.

Indeed, elementary calculations show that, for $1 \le i, j \le L$,

$$\begin{aligned} L\mathbf{P}_0^{\perp}[i,j] &= 1 + \frac{3(L-1)}{L+1} + \frac{12}{L^2-1}(i-1)(j-1) - \frac{6}{L+1}(i+j-2) \\ &= 1 + \frac{3(L-1)}{L+1} - \frac{6}{L+1}\left(\left(1-\frac{j-1}{L-1}\right)(i-1) + \left(1-\frac{i-1}{L-1}\right)(j-1)\right). \end{aligned} \tag{14}$$

It immediately follows that $\left\|\mathbf{P}_0^{\perp}\right\|_{\max} \le 4/L$. A reference to Lemma 2 and to the relation $L/N \to \alpha$ completes the proof.

In complete analogy, $\max_M \left\|\mathbf{Q}_0^{\perp}\mathbf{E}\right\|_{\max} = O(N^{-1})$.

Further, since $\left\|\mathbf{P}_0^{\perp}\mathbf{E}\mathbf{Q}_0^{\perp}\right\|_{\max} \le \left\|\mathbf{P}_0^{\perp}\right\|_{\max} L\left\|\mathbf{E}\mathbf{Q}_0^{\perp}\right\|_{\max} = O(N^{-1})$, then $\left\|\mathbf{P}_0\mathbf{E}\mathbf{Q}_0^{\perp}\right\|_{\max} \le \left\|\mathbf{E}\mathbf{Q}_0^{\perp}\right\|_{\max} + \left\|\mathbf{P}_0^{\perp}\mathbf{E}\mathbf{Q}_0^{\perp}\right\|_{\max} = O(N^{-1})$ uniformly in M. Therefore, $J_2 = O(|\delta|N^{-1})$ and the statement is proved.

3. In the case of a constant signal $\mathbf{H} = \sqrt{LK}UV^{\mathrm{T}}$, where $U \in \mathbb{R}^L$ and $V \in \mathbb{R}^K$ are vectors with all the same coordinates equal to $1/\sqrt{L}$ and $1/\sqrt{K}$, respectively. Therefore, $\|\mathbf{H}\| \sim \alpha(1-\alpha)N$.

Further, in both cases, $|\delta|/\|\mathbf{H}\| \to 0$; therefore, $J_1 + J_3 = O(\delta^2/\|\mathbf{H}\|) = O(\delta^2/N)$.

Moreover, $\mathbf{P}_0^{\perp} = UU^{\mathrm{T}}$ and $\mathbf{Q}_0^{\perp} = VV^{\mathrm{T}}$; and so, $\left\|\mathbf{P}_0^{\perp}\mathbf{E}\right\|_{\max} \asymp \left\|\mathbf{E}\mathbf{Q}_0^{\perp}\right\|_{\max} \asymp N^{-1}$ and $\left\|\mathbf{P}_0^{\perp}\mathbf{E}\mathbf{Q}_0^{\perp}\right\|_{\max} = O(N^{-1})$.

Therefore, $\max_M |r_M(\delta)| = O(\delta^2/N) + O(|\delta|/N)$, which implies the result.

4. As follows from [8], in this case, $d = 2r$, and all the quantities $\|\mathbf{H}\|$, $\left\|\mathbf{P}_0^{\perp}\right\|_{\max}$, and $\left\|\mathbf{Q}^{\perp}\right\|_{\max}$ have the order of decrease $O(N^{-1})$.

Therefore, the result is the same as for a constant signal.

□

Remark 4. It is seen that the order of convergence $O(|\delta|/N)$ in the theorem is completely determined by the first (linear with respect to δ) term on the right side of (5). This means that, in this case, the use of the linear version of perturbation theory (for example, [9]) is correct.

3.3. About Multiple Outliers

As before, we denote the set of nonzero elements of the series E_N by $\mathrm{Re}(N)$ and put $\mathfrak{n}(N) = \mathrm{card}(\mathrm{Re}(N))$. We note that

$$\mathrm{E}_N = \sum_{j\in \mathrm{Re}(N)} e_j^{(N)} \mathrm{E}_j^{(N)} = \sum_{0\le j<N} e_j^{(N)} \mathrm{E}_j^{(N)},$$

where $\mathrm{E}_j^{(N)}$ has a single nonzero element equal to 1 at position j. Then, if $\mathbf{E}$ and $\mathbf{E}_j$ are $L \times K$-trajectory matrices of the series E_N and $\mathrm{E}_j^{(N)}$, we have $\mathbf{E} = \sum_{0\le j<N} e_j^{(N)} \mathbf{E}_j$. Therefore, since $\|\mathbf{E}_i\| = 1$,

$$\|\mathbf{E}\| \le \sum_{j\in \mathrm{Re}(N)} \left|e_j^{(N)}\right| \|\mathbf{E}_i\| = \|\mathrm{E}_N\|_1, \quad \text{where} \quad \|\mathrm{E}_N\|_1 = \sum_{0\le j<N} \left|e_j^{(N)}\right|.$$

In addition, $\|\mathbf{E}\|_{\max} = \max_{j\,\in\,\mathrm{Re}(N)}\left|e_j^{(N)}\right| = \|\mathrm{E}_N\|_{\max}$.

Collecting the results of Lemma 1 and Remark 2 with $\Delta = \|\mathrm{E}_N\|_1$, we find that, under the condition $\|\mathrm{E}_N\|_1 > c = \mathrm{const}$ and $|\delta|\, \|\mathrm{E}_N\|_1/\|\mathbf{H}\| \to 0$, we have $J_1 + J_2 = \delta^2 \|\mathrm{E}_N\|_1^2 / \|\mathbf{H}\|$.

Let us now see what Lemma 2 turns into. Since, for an arbitrary $L \times L$ matrix $\|\mathbf{G}_1\|$,

$$\mathbf{G}_1\mathbf{E} = \sum_j e_j^{(N)} \mathbf{G}_1\mathbf{E}_j \quad \text{and} \quad \|\mathbf{G}_1\mathbf{E}_j\|_{\max} \le \|\mathbf{G}_1\|_{\max}, \quad \text{then}$$

$$\|\mathbf{G}_1\mathbf{E}\|_{\max} \le \sum_j \left|e_j^{(N)}\right| \|\mathbf{G}_1\|_{\max} = \|\mathrm{E}_N\|_1 \|\mathbf{G}_1\|_{\max},$$

and a similar inequality holds for $\|\mathbf{E}\mathbf{G}_2\|_{\max}$.

From here, for example, it immediately follows that, under the condition $\Theta = \sup_N \|\mathrm{E}_N\|_1 < \infty$, all the results of Theorem 2 remain in force with the replacement of $|\delta|$ to $|\delta|\Theta$ on the right-hand side of the corresponding inequalities.

Remark 5. We note that the condition $\Theta < \infty$, generally speaking, applies not only to outliers. For example, as follows from Theorem 2 and previous arguments, for a linear signal and noise $\mathrm{E} = (e_0, \ldots, e_n, \ldots)$ with $\sum_i |e_i| < \infty$, for fixed δ and $L/N \to \alpha \in (0, 1)$, the relation $\max_i |r_i(\delta)| = O(N^{-1})$ holds.

4. COMPUTATIONAL EXPERIMENTS

This section presents several variants of computer experiments aimed at illustrating the theoretical results of Section 3.

We consider three options for the signal f_n:

(i) exponential signal (EXP): $f_n = a^n$ and $a = 1.01$;

(ii) linear signal (LIN): $f_n = an$ and $a = 1/3$;

(iii) harmonic signal (COS): $f_n = \cos(2\pi\omega n)$ with $\omega = \sqrt{2}/4$.

In all experiments, $e_M = 1$ and $L \approx N/2$ were taken.

The first point of Theorem 2 states, in particular, that the SSA method for the EXP signal at a fixed position M of a single outlier is asymptotically robust, while this robustness cannot be obtained for the outlier position at the last point of the series.

Figure 1 confirms both of these estimates: for $M = 0$, it is seen that the maximum reconstruction error quickly tends to zero, while such a tendency is not observed for $M = N - 1$.

It follows from the second point of Theorem 2 that, for a linear signal, the reconstruction errors have the form $O(N^{-1})$ regardless of the position of a single outlier. Figure 2 confirms this theoretical result.

Finally, the fourth point of Theorem 2 and the reasoning of Section 3.3 allow us to state that, for the COS signal in the case of multiple outliers of uniformly limited power with $\mathrm{card}\,\mathfrak{R}(N) <$ const, all the reconstruction errors have the order of decrease $O(N^{-1})$. Figure 3 with $\mathfrak{R}(N) = \{0, 1\}$ and $\delta = 1$ confirms this conclusion.

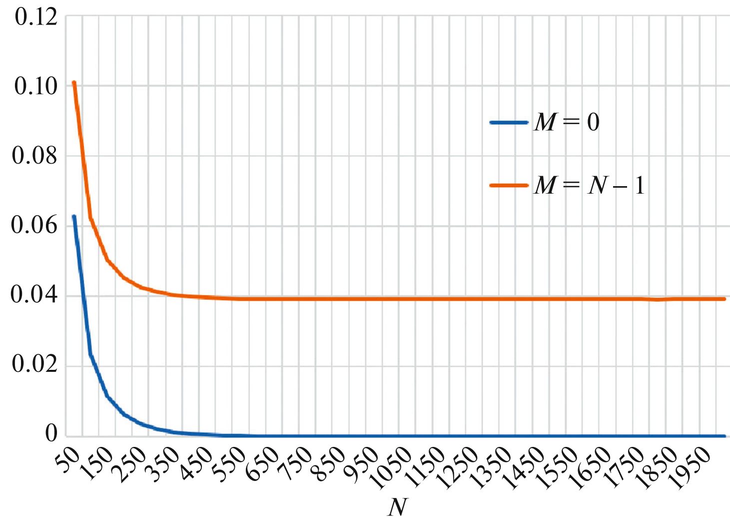

Fig.1. Maximum reconstruction errors for two positions of a single outlier as a function of the series length N for the EXP signal with $\delta = 1$.

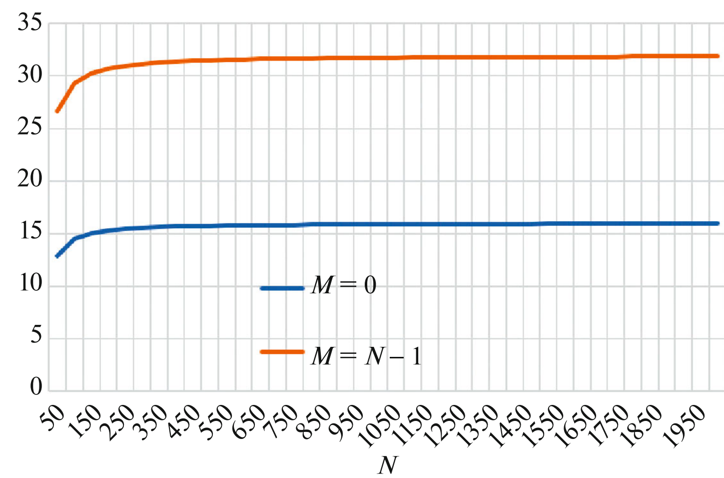

Fig. 2. N multiplied by the maximum reconstruction errors for two positions ($M = 0$ with $\delta = 1$ and $M = N - 1$ with $\delta = 2$) of a single outlier as a function of the series length N for the LIN signal.

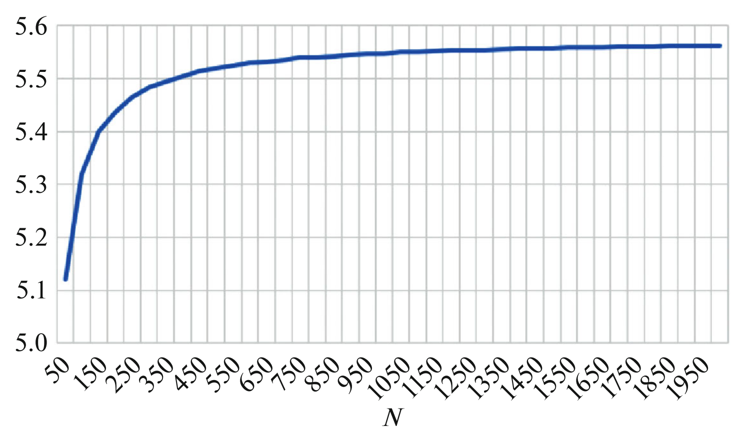

Fig. 3. N multiplied by the maximum reconstruction errors in the case of a double outlier with $e_0 = e_1 = 1$ as a function of the series length N for the COS signal.

5. CONCLUSIONS

For simplicity, we restrict ourselves to the case $\delta = \text{const}$ and $\|\mathbf{H}\| \to \infty$ as $N \to \infty$. Then it follows from the results of the paper that, if the ratio $\mu_{\max}/\mu_{\min}$ is bounded (see Lemma 1), all is determined by the quantity J_2 defined in (8). Namely, if $L/N \to \alpha \in (0, 1)$ and $J_2 \to 0$, then the SSA method is asymptotically robust with respect to any single outlier.

Furthermore, it turns out that J_2 corresponds to the linear term of perturbation theory and, since J_2 is expressed in terms of the projectors $\mathbf{P}_0^{\perp}$ and $\mathbf{Q}_0^{\perp}$, the uniform norms of these projectors play a special role here.

ACKNOWLEDGMENTS

I thank both anonymous reviewers whose comments undoubtedly contributed to significant improvements of the paper.

FUNDING

This work was supported by the Russian Science Foundation grant (project no. 23-21-00222).

CONFLICT OF INTEREST

The author of this work declares that he has no conflicts of interest.

REFERENCES

1. M. Kazemi and P. C. Rodrigues, “Robust singular spectrum analysis: Comparison between classical and robust approaches for model fit and forecasting,” Ann. Appl. Stat. **7**, 1540–1561 (2023).
2. N. Golyandina, “Particularities and commonalities of singular spectrum analysis as a method of time series analysis and signal processing,” Wiley Interdiscip. Rev.: Comput. Stat. **12**, 4 (2020).
3. V. Nekrutkin, “Perturbation expansions of signal subspaces for long signals,” Stat. Its Interface **3**, 297–319 (2010).
4. N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, *Analysis of Time Series Structure. SSA and Related Techniques* (Champan & Hall/CRC, Boca Raton, Fla., 2001).
5. N. Golyandina and A. Zhigljavsky, *Singular Spectrum Analysis for Time Series*, 2nd ed. (Springer-Verlag, Berlin, 2020), in Ser.: *Springer Briefs in Statistics*.
6. G. H. Golub and C. F. van Loan, *Matrix Computations*, 4th ed. (Johns Hopkins Univ. Press, Baltimore, Md., 2013).
7. T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, Berlin, 1966).
8. V. V. Nekrutkin, “Asymptotical separation of harmonics by singular spectrum analysis,” Vestn. St. Petersburg Univ.: Math. **56**, 526–539 (2023).
https://doi.org/10.1134/S1063454123040118
9. M. Senov, “Robust versions of the SSA method” (2022).
https://doi.org/10.5281/zenodo.6563614

Translated by I. Tselishcheva