

# ОШИБКИ ОЦЕНИВАНИЯ СИГНАЛА С ПОМОЩЬЮ КОМПЛЕКСНОГО АНАЛИЗА СИНГУЛЯРНОГО СПЕКТРА <sup>1</sup>

Голяндина Н. Э., доцент кафедры Статистического Моделирования  
СПбГУ, n.golyandina@spbu.ru

Сенов М. А., студент, СПбГУ, senov.mikhail@gmail.com

## Аннотация

Анализ сингулярного спектра (singular spectrum analysis, SSA) — непараметрический метод для разложения временного ряда в сумму интерпретируемых компонент, таких как тренд, периодики и шум. Расширение SSA на комплексный случай называется CSSA. В работе проведено теоретическое сравнение CSSA и SSA, примененного отдельно к вещественной и мнимой части, на основе первого порядка ошибки оценки сигнала, где первый порядок рассматривается по величине возмущения. Проведено численное сравнение полной ошибки и первого порядка ошибки на примерах. Для константного сигнала получен явный вид первого порядка ошибки его оценки в случае наличия в ряде выброса.

## Введение

Анализ сингулярного спектра (singular spectrum analysis, SSA) [2] — мощный метод анализа временных рядов, не требующий предварительного задания параметрической модели ряда. Метод имеет естественное расширение на случай комплексных временных рядов, называемое Complex SSA (CSSA). Есть класс сигналов, а именно временные ряды, управляемые линейными рекуррентными соотношениями, который позволяет получать для него теоретические результаты относительно свойств метода SSA.

Пусть наблюдаемый комплексный временной ряд имеет вид  $X = S + R$ . Для получения оценки сигнала  $\tilde{S}$  будем использовать метод CSSA. Кроме применения CSSA ко всему ряду, будем также применять метод SSA отдельно к вещественной и мнимой части ряда  $X$ .

---

<sup>1</sup>Работа выполнена при поддержке гранта РФФИ 20-01-00067

Для анализа ошибки оценивания сигнала используется теория возмущений [5], которая была применена для случая выделения сигнала методом SSA в ряде работ, см., например, [6].

Хотя теория возмущения Като дает вид полной ошибки, однако ее исследование представляется сложной задачей. Поэтому мы будем рассматривать только первый порядок ошибки в разложении ошибки по величине возмущения.

При этом проведем численное сравнение первого порядка ошибки и полной ошибки для выявления случаев, когда анализ первого порядка ошибки плохо описывает полную ошибку и поэтому его анализ не представляет интереса.

Даже для первого порядка ошибки получение его явного вида — довольно трудоемкая задача. Нам удалось его получить для случая константного сигнала и возмущения в виде выброса. В общем случае результаты касаются сравнения MSE ошибок оценки сигнала методом CSSA и суммарного MSE при применении SSA отдельно к мнимой и вещественной частям.

## Алгоритм CSSA

### Алгоритм 1.

**Вход:** Комплексный временной ряд  $X = (x_1, \dots, x_N)$ , длина окна  $L$ , ранг сигнала  $r$ .

**Выход:** Оценка сигнала  $\tilde{S}$ .

**Алгоритм:**

1. **Вложение.** Построим  $\mathbf{X} \in \mathbb{R}^{L \times K}$ ,  $L$ -траекторную матрицу ряда  $X$ :

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

где  $K = N - L + 1$ , а  $X_i$  — векторы  $L$ -вложения:  $X_i = (x_i, \dots, x_{i+L-1})^T \in \mathbb{R}^L$ .

2. **Разложение.** Построим SVD-разложение матрицы  $\mathbf{X}$ :  $\mathbf{X} = \sum_{k=1}^{\text{rank } \mathbf{X}} \sqrt{\lambda_k} U_k V_k^H = \sum_{k=1}^{\text{rank } \mathbf{X}} \hat{\mathbf{X}}_k$ , где  $U_k$ ,  $V_k$  — правые и левые сингулярные векторы матрицы  $\mathbf{X}$  соответственно,  $\sqrt{\lambda_k}$  — сингулярные числа.

3. **Группировка.** Сгруппируем матрицы компонент сигнала  $\hat{\mathbf{S}}$ :  $\hat{\mathbf{S}} = \sum_{k=1}^r \hat{\mathbf{X}}_k$ .

4. *Диагональное усреднение.* Применим процедуру диагонального усреднения (проекция в норме Фробениуса на линейное пространство ганкелевых матриц):  $\tilde{\mathbf{S}} = \mathcal{H}\hat{\mathbf{S}}$ , затем сопоставим полученным Ганкелевым матрицам ряды длины  $N$ :  $\tilde{\mathbf{S}} = \mathcal{T}_L^{-1}\tilde{\hat{\mathbf{S}}}$ .

$L$ -Рангом временного ряда называется ранг его траекторной матрицы. Для дальнейших рассуждений потребуется знание рангов конкретных рядов. Из [3] известно, что ранг комплексного сигнала, вещественная и мнимая часть которого являются синусоидами с одинаковой частотой  $\omega$ ,  $0 < \omega < 0.5$ , равен 2, если сдвиг между синусоидами не равен  $\pi/2$ , и равен 1 в случае комплексной экспоненты. Ранг вещественного синусоидального сигнала, равен 2 при тех же ограничениях на частоту. Ранги же комплексной и вещественной констант равны 1 — их можно рассматривать как частный случай с  $\omega = 0$ .

## Применение теории возмущений к SSA и CSSA

Наблюдаем комплексный временной ряд  $\mathbf{X}$  длины  $N$ , данный ряд представляется как  $\mathbf{X} = \mathbf{S} + \mathbf{R}$ , где  $\mathbf{S}$  — сигнал ранга  $r$ ,  $\mathbf{R}$  — возмущение. Возьмем некоторую длину окна  $L$ ,  $L > r$ .

В [6] вводится разложение восстановления сигнала в модели  $\mathbf{S}(\delta) = \mathbf{S} + \delta\mathbf{R}$ , что соответствует  $\mathbf{H}(\delta) = \mathbf{H} + \delta\mathbf{E}$ , где  $\mathbf{H}(\delta) = \mathcal{T}_L\mathbf{S}(\delta)$ ,  $\mathbf{H} = \mathbf{S}(\delta)\mathbf{S}$ ,  $\delta\mathbf{E} = \delta\mathbf{R}$ , и рассматривается линейный по  $\delta$  член ошибки восстановления, называемый первым порядком ошибки восстановления.

Рассмотрим возмущение ряда  $\mathbf{R}$  с  $\delta = 1$ , его траекторная матрица  $\mathbf{E}$ . Первый порядок ошибки восстановления обозначим как  $\mathbf{F}^{(1)} = \mathcal{H}(\mathbf{H}^{(1)})$ .

На основе результатов из [8, стр.12] и теоремы 2.1 из [6] была получена следующая формула для  $\mathbf{H}^{(1)}$  в случае достаточно маленького возмущения.

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

где  $\mathbf{P}_0^\perp$  — проектор на пространство столбцов  $\mathbf{H}$ ,  $\mathbf{Q}_0^\perp$  — проектор на пространство строк  $\mathbf{H}$ ,  $\mathbf{P}_0 = \mathbf{I} - \mathbf{P}_0^\perp$ ,  $\mathbf{I}$  — единичная матрица.

## Сравнение CSSA и SSA в случае совпадающих пространств сигналов

Обозначим за:

$\mathbf{F}^{(1)} = (f_1^{(1)}, \dots, f_N^{(1)})$  первый порядок ошибки восстановления  $\mathbf{S}$  с возмущением  $\mathbf{R}$  метода CSSA,

$F_{\text{Re}}^{(1)} = (f_{\text{Re},1}^{(1)}, \dots, f_{\text{Re},N}^{(1)})$  первый порядок ошибки восстановления  $\text{Re}(S)$  с возмущением  $\text{Re}(R)$  метода SSA,

$F_{\text{Im}}^{(1)} = (f_{\text{Im},1}^{(1)}, \dots, f_{\text{Im},N}^{(1)})$  первый порядок ошибки восстановления  $\text{Im}(S)$  с возмущением  $\text{Im}(R)$  метода SSA.

**Теорема 1.** Пусть пространства столбцов траекторных матриц рядов  $S$ ,  $\text{Re}(S)$  и  $\text{Im}(S)$  совпадают и то же самое верно для пространств строк. Тогда при любом достаточно малым возмущении  $R$

$$F^{(1)} = F_{\text{Re}}^{(1)} + iF_{\text{Im}}^{(1)}.$$

Теорема непосредственно следует из линейности вхождения  $E$  в формулу (1) и линейности диагонального усреднения.

Заметим, что хотя в утверждении теоремы возмущение  $R$  может быть любым по форме, однако теорема имеет практическое применение только если первый порядок ошибки адекватно описывает полную ошибку.

### Случайное возмущение

Рассмотрим случайное возмущение  $R$ .

Для дальнейших рассуждений приведём известный результат.

**Лемма 1.** Пусть  $\zeta = \xi + i\eta$ . Тогда  $\mathbb{D}\zeta = \mathbb{D}\xi + \mathbb{D}\eta$ .

**Следствие 1** (из теоремы 1). Пусть выполнены условия теоремы 1. Тогда для любого  $l$ ,  $1 \leq l \leq N$ ,

$$\mathbb{D}f_l^{(1)} = \mathbb{D}f_{\text{Re},l}^{(1)} + \mathbb{D}f_{\text{Im},l}^{(1)}. \quad (2)$$

Утверждение получается автоматически из теоремы 1 и леммы 1.

### Случай двух зашумленных синусоид

Пусть сигнал  $S$  имеет вид

$$s_l = A \cos(2\pi\omega l + \phi_1) + iB \cos(2\pi\omega l + \phi_2), \quad (3)$$

где  $0 < \omega \leq 0.5$  и  $0 \leq \phi_i < 2\pi$ . Заметим, что случай  $|\psi_2 - \psi_1| = \pi/2$  и  $A = B$  соответствует комплексной экспоненте.

Пусть возмущение  $R$  — шум, т.е. случайный вектор с нулевым математическим ожиданием и достаточно малой дисперсией.

**Следствие 2** (из теоремы 1). Для комплексного ряда вида (3), кроме случая  $|\psi_2 - \psi_1| = \pi/2$  и  $A = B$ , выполняется формула (2).

Выполнение условий теоремы 1 (совпадение столбцовых и строковых траекторных пространств сигналов) для ряда вида (3) следует из результатов работы [3].

**Замечание 1.** Численные эксперименты, проведённые в [1], показывают, что для сигнала в виде комплексной экспоненты суммарная MSE CSSA-оценки сигнала равна полусумме суммарных MSE SSA-оценок сигнала его вещественной и мнимой частей. Численные эксперименты, проведённые нами (не приведены), показывают, что для сигнала в виде комплексной экспоненты MSE CSSA-оценки сигнала равны полусумме MSE SSA-оценок сигнала его вещественной и мнимой частей для каждой точки ряда.

### Случай константных сигналов с выбросом

Рассматриваем сигнал  $\mathbf{S} = (c_1 + ic_2, \dots, c_1 + ic_2)$ , возмущённый выбросом  $a_1 + ia_2$  на позиции  $k$ , т.е. ряд  $\mathbf{R}$  состоит из нулей кроме значения  $a_1 + ia_2$  на  $k$ -м месте. Исходя из теоремы 1, достаточно уметь вычислять первый порядок ошибки восстановления сигнала  $\mathbf{S} = (c, \dots, c)$ , возмущённого выбросом  $a$  на позиции  $k$ .

В работе [7] была получена частный случай формулы (1) для вещественных сигналов единичного ранга:

$$\mathbf{H}^{(1)} = -U^T \mathbf{E} V U V^T + U U^T \mathbf{E} + \mathbf{E} V V^T,$$

где  $U, V$  — сингулярные вектора матрицы  $\mathbf{H}$ .

Подстановкой  $U = \{1/\sqrt{L}\}_{i=1}^L$ ,  $V = \{1/\sqrt{K}\}_{i=1}^K$ ,  $K = N - L + 1$  и последующим диагональным усреднением матрицы  $\mathbf{H}^{(1)}$  был получен явный вид первого порядка ошибки восстановления.

Приведем результат для случая  $k \leq \min(L/2, K - L)$  и  $L < K$

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

**Замечание 2.** Из данной формулы видно, что при фиксированном  $L$  первый порядок ошибки не стремится к 0 с ростом  $N$ , тогда как численные эксперименты показывают, что полная ошибка восстановления стремится к 0 с ростом  $N$ . Как показано в следующем разделе, это следствие того, что полная ошибка не описывается ее первым порядком. Если же  $L$  и  $K$  пропорциональны  $N$ , то первый порядок ошибки стремится к нулю.

## Численное сравнение первого порядка ошибки и полной ошибки оценивания сигнала

Для случая зашумленных гармоник рассмотрен пример с сигналом  $s_l = \cos(2\pi l/10) + i \cos(2\pi l/10 + \pi/4)$ ,  $\sigma^2 = 0.1$ ,  $N = 49$ ,  $L = 5$ . Результат для одной из реализаций шума представлен на рис. 1.

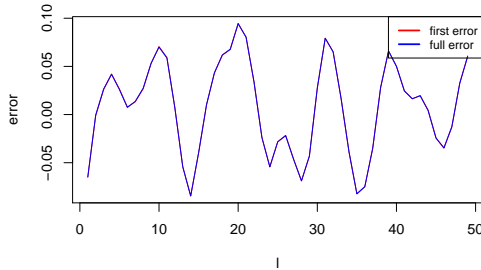


Рис. 1: Вещественные части первого порядка и полной ошибок.

Из графика видно, что ошибки совпадают даже при маленьком  $L$ . Аналогичные численные эксперименты подтверждают, что для комплексной экспоненты также есть совпадение.

Для случая возмущения в виде выброса был рассмотрен пример с сигналом  $s_l = 1 + il$ , с возмущением в виде выброса  $a_1 + ia_2 = 10 + i10$  на позиции  $k = L - 1$ . Результаты представлены в таблице 1.

Таблица 1: Максимальное различие первого порядка и полной ошибок.

$N$	50	100	400	1600
$L = N/2$	0.1313	0.0419	0.0033	0.0002
$L = 20$	0.3074	0.1965	0.5655	0.6720

Аналогичные численные эксперименты показывают, что при расположении выброса в середине ряда результаты качественно совпадают, при  $L = N/2$  различие стремится к 0, при  $L = 20$  не стремится к 0.

Численные результаты показывают, что для случая зашумленных гармоник первый порядок адекватно оценивает полную ошибку восстановления сигнала в каждой точке при любых рассматриваемых параметрах сигналов.

Однако для случая возмущения в виде выброса это верно, только когда  $L$  и  $K$  пропорциональны  $N$ .

Все численные результаты были получены при помощи пакета [4].

## Заключение

В работе удалось подвести теоретическую базу под имеющиеся ранее численные результаты ([1]) по сравнению CSSA и SSA для двух зашумленных гармоник с одинаковой частотой и сдвигом, не равным  $\pi/2$ . Для зашумленной комплексной экспоненты был получен более общий, нежели имеющиеся ранее, численный результат. Результаты показывают, что только в случае сигнала в виде комплексной экспоненты применение CSSA имеет смысл с точки зрения уменьшения ошибки восстановления сигнала.

Для константного ряда с выбросами был получен явный вид первого порядка ошибок оценки сигнала в каждой точке.

Для обоих случаев было численно исследовано соотношение между первым порядком ошибки и полной ошибкой. В случае случайного возмущения оказалось, что первый порядок ошибки практически совпадает с полной ошибкой. Однако в случае неслучайного возмущения выбросом это не так и требуются дополнительные условия на пропорциональность длины окна  $L$  длине ряда  $N$ .

## Литература

- [1] N. Golyandina, A. Korobeynikov, A. Shlemov, and K. Usevich. Multivariate and 2D extensions of singular spectrum analysis with the Rssa package. *Journal of Statistical Software*, 67(2):1–78, 2015.
- [2] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky. *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman&Hall/CRC, 2001.
- [3] Д. Степанов, Н. Голяндина. Варианты метода "Гусеница"-SSA для прогноза многомерных временных рядов. *Труды IV Международной конференции "Идентификация систем и задачи управления" SICPRO'05*. Москва, 2005, с. 1831-1848.
- [4] A. Korobeynikov, A. Shlemov, K. Usevich, and N. Golyandina. *Rssa: A collection of methods for singular spectrum analysis* <http://CRAN.R-project.org/package=Rssa>, 2021. R package version 1.04.
- [5] Т. Като. Perturbation theory for linear operators. *Springer-Verlag*, 1966.
- [6] V. Nekrutkin. Perturbation expansions of signal subspaces for long signals. *Statistics and Its Interface.*, Vol.3, P. 297-319, 2010.
- [7] V. Nekrutkin. Perturbations in SSA. *Manuscript*, 2008.
- [8] А. Константинов. Некоторые задачи анализа временных рядов (теория методов "Singal Subspace"). *Курсовая работа, науч. рук. В. Некруткин*, 2018.