

# О зависимости ошибки восстановления сигнала от ранга при возмущении в виде выброса

Храмов А.П., СПбГУ, Санкт-Петербург alexsandr.khramov@gmail.com,  
Голяндина Н.Э., СПбГУ, Санкт-Петербург n.golyandina@spbu.ru

## Аннотация

Рассматривается временной ряд состоящий из суммы сигнала и возмущения. Исследуется ошибка восстановления сигнала, полученная методом SSA. Для этого применяется теория возмущений, позволяющая исследовать ошибки первого порядка по величине возмущения. При этом численно показано, что результаты, полученные таким образом, асимптотически описывают поведение полной ошибки. Для возмущения в виде шума и выброса показано, что смещение оценки сигнала зависит только от выброса, а дисперсия — только от шума с нулевым математическим ожиданием. Получены явные формулы для первого порядка ошибки в случае экспоненциального сигнала и выброса (до диагонального усреднения). Численное моделирование показало, что MSE оценки сигнала пропорционально рангу сигнала при той же амплитудной модуляции.

## Введение

Исследуется временной ряд  $X = (x_1, \dots, x_N)$ ,  $x_i \in \mathbb{R}$ , длины  $N$ , состоящий из двух компонент: сигнала  $S$  и возмущения  $E$ . Возмущение  $E$  зачастую представляется в виде выброса или шума. Важной задачей является получение оценки сигнала  $\tilde{S}$ . Для этого может применяться метод SSA, представленный, например, в [1], и описанный нами в первой части работы. Интерес представляет исследование ошибки восстановления сигнала  $\tilde{S} - S$ .

В работе применяются теория возмущений [2] и результаты, полученные в [3] для исследования первого порядка ошибки. В [4], применяя этот же подход, были представлены явные формулы для ошибки первого порядка оценки константного сигнала при возмущении в виде выброса. В данной работе получен явный вид матрицы первого порядка ошибки для экспоненциального сигнала  $S = (e^{mb})_{m=1}^N$  и возмущения в виде выброса. Существующие робастные варианты метода SSA [5, 6] являются итеративными, в силу этого трудоемкими и также менее точными при отсутствии выбросов по сравнению с базовым вариантом. Поэтому исследование влияния выбросов на результат восстановления сигнала с помощью базового метода SSA представляет значительный интерес.

В последней части статьи представлены результаты численного моделирования, демонстрирующие поведение полной ошибки и ошибки первого порядка. В статье [4] была показана пропорциональность MSE оценки сигнала рангу для сигнала в виде суммы комплексных экспонент, имеющих постоянный модуль. В данной работе показан аналогичный результат для вещественного экспоненциально-гармонического сигнала.

## Алгоритм SSA

Кратко опишем алгоритм SSA для получения оценки сигнала.

**Входные параметры:** временной ряд  $\mathbf{X} = (x_1, \dots, x_N)$ , длина окна  $L$ , ранг сигнала  $r$ .

**Результат:** восстановленный сигнал  $\tilde{\mathbf{S}}$ .

1. **Вложение.** Построим  $\mathbf{X} \in \mathbb{R}^{L \times K}$  —  $L$ -траекторную матрицу временного ряда  $\mathbf{X}$ :  $\mathbf{X} = \mathcal{T}_L \mathbf{X} = [X_1 : \dots : X_K]$ , где  $K = N - L + 1$ ,  $X_i = (x_i, \dots, x_{i+L-1})^T \in \mathbb{R}^L$ .
2. **Разложение.** Построим сингулярное разложение матрицы  $\mathbf{X}$ :

$$\mathbf{X} = \sum_{k=1}^{\text{rank } \mathbf{X}} \sqrt{\lambda_k} U_k V_k^T = \sum_{k=1}^{\text{rank } \mathbf{X}} \hat{\mathbf{X}}_k,$$

где  $U_k$  и  $V_k$  — левые и правые сингулярные векторы  $\mathbf{X}$ ,  $\sqrt{\lambda_k}$  — сингулярные числа в порядке убывания.

3. **Группировка.** Группируем слагаемые  $\hat{\mathbf{X}}_k$  из разложения, относящиеся к сигналу:  $\hat{\mathbf{S}} = \sum_{k=1}^r \hat{\mathbf{X}}_k$ .
4. **Диагональное усреднение.** Применяем проекцию на пространство ганкелевых матриц:  $\tilde{\mathbf{S}} = \Pi_{\mathcal{H}} \hat{\mathbf{S}}$ , и переходим к форме ряда:  $\tilde{\mathbf{S}} = \mathcal{T}_L^{-1} \hat{\mathbf{S}}$ .

Ошибкой восстановления будем называть ряд  $\mathbf{F} = \tilde{\mathbf{S}} - \mathbf{S}$ .

$L$ -рангом временного ряда называют ранг его  $L$ -траекторной матрицы. Интерес представляют сигналы конечного ранга, т.е. сигналы,  $L$ -ранг которых не зависит от длины окна. В [1] подробнее описаны свойства рядов конечного ранга. Там же приведены примеры сигналов конечного ранга, в частности, экспоненциально-гармонический сигнал  $\mathbf{S} = (s_m)_{m=1}^N$  с

$$s_m = \sum_{j=1}^p A_j e^{\alpha_j m} \cos(2\pi\omega_j m + \varphi_j),$$

где  $A_j, \alpha_j \in \mathbb{R}$ ,  $\omega_j \in [0, 1/2]$ ,  $\varphi_j \in [0, 2\pi)$ . Для суммы из одного слагаемого ранг равен 1, если  $\omega_1 = 0$  или  $\omega_1 = 0.5$ , и равен 2 иначе.

## Исследование ошибки первого порядка

Рассмотрим ряд  $\mathbf{X} = \mathbf{S}(\delta)$ , где  $\mathbf{S}(\delta) = \mathbf{S} + \delta\mathbf{E}$ ; обозначим  $\tilde{\mathbf{S}}$  оценку сигнала методом SSA. Из [3] известно следующее представление (полной) ошибки восстановления:

$$\mathbf{F} = \mathbf{F}(\mathbf{S}, \delta\mathbf{E}) = \tilde{\mathbf{S}} - \mathbf{S} = \mathcal{T}^{-1}\Pi_{\mathcal{H}}(\delta\mathbf{S}^{(1)} + \delta^2\mathbf{S}^{(2)}(\delta)).$$

Назовём ошибкой восстановления первого порядка часть полной ошибки, линейно зависящую от  $\delta$ :  $\delta\mathbf{F}^{(1)} = \mathcal{T}_L^{-1}\Pi_{\mathcal{H}}(\delta\mathbf{S}^{(1)})$ . Примем  $\delta = 1$ . Теорема 2.1 из [3] даёт следующую формулу для  $\mathbf{S}^{(1)} \in \mathbb{R}^{L \times K}$ :

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

где  $\mathbf{P}_0^\perp$  — проектор на пространство столбцов  $\mathbf{S}$ ,  $\mathbf{Q}_0^\perp$  — проектор на пространство строк  $\mathbf{S}$ .

Для сигналов ранга 1 формула (1) принимает вид:

$$\mathbf{S}^{(1)} = \mathbf{S}^{(1)}(\mathbf{E}, \mathbf{S}) = -(\mathbf{U}^\mathbf{T} \mathbf{E} \mathbf{V}) \mathbf{U} \mathbf{V}^\mathbf{T} + \mathbf{U} \mathbf{U}^\mathbf{T} \mathbf{E} + \mathbf{E} \mathbf{V} \mathbf{V}^\mathbf{T}. \quad (2)$$

## Смещение и дисперсия оценки сигнала при возмущении в виде шума и выброса

Пусть  $\mathbf{E} = \mathbf{E}_n + \mathbf{E}_o$ , где  $\mathbf{E}_n$  — случайный стационарный процесс с нулевым математическим ожиданием и достаточно малой дисперсией,  $\mathbf{E}_o$  состоит из нулей за исключением значения  $a \in \mathbb{R}$  на позиции  $k$ . Ряд  $\mathbf{E}_n$  будем называть шумом,  $\mathbf{E}_o$  — выбросом.

**Утверждение 1.** Рассмотрим сигнал  $\mathbf{S}$ , возмущение  $\mathbf{E} = \mathbf{E}_n + \mathbf{E}_o$ , где  $\mathbf{E}_n$  — шум,  $\mathbf{E}_o$  — выброс. Тогда

$$\begin{aligned} \mathbb{E}\mathbf{F}^{(1)}(\mathbf{S}, \mathbf{E}) &= \mathbf{F}^{(1)}(\mathbf{S}, \mathbf{E}_o), \\ \mathbb{D}\mathbf{F}^{(1)}(\mathbf{S}, \mathbf{E}) &= \mathbb{D}\mathbf{F}^{(1)}(\mathbf{S}, \mathbf{E}_n), \end{aligned}$$

где  $\mathbf{E}_n = \mathcal{T}_L \mathbf{E}_n$ ,  $\mathbf{E}_o = \mathcal{T}_L \mathbf{E}_o$ .

*Доказательство.* Из линейности  $\mathcal{T}_L$  следует, что  $\mathbf{E} = \mathbf{E}_n + \mathbf{E}_o$ . Из линейности формулы (1) по  $\mathbf{E}$  получаем

$$\mathbf{S}^{(1)}(\mathbf{S}, \mathbf{E}) = \mathbf{S}^{(1)}(\mathbf{S}, \mathbf{E}_n) + \mathbf{S}^{(1)}(\mathbf{S}, \mathbf{E}_o),$$

а из линейности  $\Pi_{\mathcal{H}}$  и  $\mathcal{T}_L^{-1}$  следует, что

$$\mathbf{F}^{(1)}(\mathbf{S}, \mathbf{E}) = \mathbf{F}^{(1)}(\mathbf{S}, \mathbf{E}_n) + \mathbf{F}^{(1)}(\mathbf{S}, \mathbf{E}_o).$$

Так как слагаемое  $\mathbf{F}^{(1)}(\mathbf{S}, \mathbf{E}_o)$  не случайно,

$$\mathbb{D}\mathbf{F}^{(1)}(\mathbf{S}, \mathbf{E}) = \mathbb{D}\mathbf{F}^{(1)}(\mathbf{S}, \mathbf{E}_n),$$

а пользуясь линейностью математического ожидания и тем, что  $\mathbb{E}\mathbf{E}_n = \mathbf{0}$ , получаем, что

$$\mathbb{E}\mathbf{F}^{(1)}(\mathbf{S}, \mathbf{E}) = \mathbf{F}^{(1)}(\mathbf{S}, \mathbf{E}_o).$$

□

Далее будут рассматриваться сигналы, для которых первый порядок ошибки адекватно описывает полную ошибку при росте  $N$ . Для таких сигналов утверждение 1 позволяет сделать следующий вывод: при возмущении в виде шума и выброса, шум отвечает за дисперсию оценки сигнала, а выброс — за смещение.

**Утверждение 2.** Пусть  $E$  — возмущение в виде выброса  $a$  на позиции  $k$ . Тогда ошибка первого порядка  $\mathbf{F}^{(1)}$  пропорциональна  $a$ .

*Доказательство.* Определим матрицу  $\mathbf{E}_1 \in \mathbb{R}^{L \times K}$  следующим образом:

$$\mathbf{E}_1[p, q] = \begin{cases} 1, & p + q - 1 = k; \\ 0, & \text{иначе.} \end{cases} \quad (3)$$

Тогда  $a\mathbf{E}_1$  — траекторная матрица возмущения в виде рассматриваемого выброса. Подставляя её в формулу (1), получаем

$$\mathbf{S}^{(1)}(\mathbf{S}, a\mathbf{E}_1) = a\mathbf{S}^{(1)}(\mathbf{S}, \mathbf{E}_1),$$

откуда

$$\mathbf{F}^{(1)}(\mathbf{S}, a\mathbf{E}_1) = a\mathbf{F}^{(1)}(\mathbf{S}, \mathbf{E}_1).$$

□

В случае сигнала, для которого первый порядок ошибки адекватно описывает полную, и возмущения в виде шума и выброса, утверждение 2 даёт следующий результат: смещение оценки сигнала пропорционально значению выброса  $a$ . Значит достаточно рассматривать смещение в случае  $a = 1$ .

### Вещественный экспоненциальный сигнал

Рассмотрим сигнал в виде вещественной экспоненты  $\mathbf{S} = (e^{mb})_{m=1}^N$ ,  $b \in \mathbb{R}$  и возмущение  $\mathbf{E}$  в виде выброса  $a = 1$ , на позиции  $k$ . Получим явный вид матрицы первого порядка ошибки восстановления сигнала  $\mathbf{S}$ .

Сингулярные векторы этого сигнала имеют следующий вид

$$U = \frac{1}{v_L} (e^b, e^{2b}, \dots, e^{Lb})^T, \quad V = \frac{1}{v_K} (e^b, e^{2b}, \dots, e^{Kb})^T, \quad (4)$$

где

$$v_j^2 = e^{2b} + e^{4b} + \dots + e^{2jb} = e^{2b} \frac{e^{2jb} - 1}{e^{2b} - 1}, \quad j \in \mathbb{N}.$$

Траекторная матрица возмущения имеет вид (3).

В формуле (2) обозначим слагаемые следующим образом

$$\mathbf{A} = U^T \mathbf{E} V U V^T, \quad \mathbf{B} = U U^T \mathbf{E}, \quad \mathbf{C} = \mathbf{E} V V^T$$

и подставим в них векторы  $U$  и  $V$  из (4) и матрицу  $\mathbf{E} = \mathbf{E}_1$  из (3). Найдём элементы этих матриц. Имеем:

$$\begin{aligned} \mathbf{A}[p, q] &= U^T \mathbf{E} V (U V^T)[p, q] = \left( \frac{1}{v_L v_K} \sum_{m_1, m_2} e^{(m_1+m_2)b} \mathbf{E}[p, q] \right) \cdot \frac{e^{(p+q)b}}{v_L v_K} = \\ &= \frac{e^{(p+q)b}}{v_L^2 v_K^2} \sum_{m_1+m_2-1=k} e^{(m_1+m_2)b} = \frac{w_k}{v_L^2 v_K^2} e^{(p+q+k+1)b}, \end{aligned}$$

где  $w_k$  — число элементов на  $k$ -ой диагонали, перпендикулярной главной, значит

$$w_k = \begin{cases} k, & k \leq L; \\ L, & L < k < K; \\ N - k + 1, & k \geq K. \end{cases}$$

Далее,

$$\mathbf{B}[p, q] = \sum_{m=1}^L (U U^T)[p, m] \mathbf{E}[m, q] = \begin{cases} \frac{1}{v_L^2} e^{(p-q+k+1)b}, & k+1-L \leq q \leq k; \\ 0, & \text{иначе;} \end{cases}$$

и

$$\mathbf{C}[p, q] = \sum_{m=1}^K \mathbf{E}[p, m] (V V^T)[m, q] = \begin{cases} \frac{1}{v_K^2} e^{(-p+q+k+1)b}, & k+1-K \leq p \leq k; \\ 0, & \text{иначе.} \end{cases}$$

Обозначим  $A = e^{(p+q+k+1)b} / (v_L^2 v_K^2)$ . Итого получаем: если  $k \leq L$ ,

$$\mathbf{S}^{(1)}[p, q] = A \left( -k + v_L^2 e^{-2pb} \mathbb{1}_{[1, k]}(p) + v_K^2 e^{-2qb} \mathbb{1}_{[1, k]}(q) \right); \quad (5)$$

если  $L < k < K$ ,

$$\mathbf{S}^{(1)}[p, q] = A \left( -L + v_L^2 e^{-2pb} + v_K^2 e^{-2qb} \mathbb{1}_{[k+1-L, k]}(q) \right); \quad (6)$$

если  $k \geq K$ , пусть  $l = N - k + 1$ , тогда

$$\mathbf{S}^{(1)}[p, q] = A \left( -l + v_L^2 e^{-2pb} \mathbb{1}_{[k+1-K, L]}(p) + v_K^2 e^{-2qb} \mathbb{1}_{[k+1-L, K]}(q) \right). \quad (7)$$

Сформулируем этот результат в виде теоремы.

**Теорема 1.** Пусть сигнал  $\mathbf{S} = (e^{mb})_{m=1}^N$ ,  $b \in \mathbb{R}$ , возмущение  $\mathbf{E}$  состоит из выброса  $a$  на позиции  $k$ ,  $\mathbf{S}^{(1)}$  — матрица первого порядка ошибки. Тогда  $\mathbf{S}^{(1)}/a$  имеет вид: (5), если  $k \leq L$ ; (6), если  $L < k < K$ ; (7), если  $k \geq K$ .

## Численное моделирование

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

При исследовании вещественного экспоненциального сигнала с длиной ряда  $N \rightarrow +\infty$ , рассматривается последовательность сигналов  $\mathbf{S}_N = (e^{mb_N})_{m=1}^N$ , где  $b_N = C/N$ ,  $C \in \mathbb{R}$ . Для такой последовательности сигналов, последний элемент ряда  $e^{Nb_N} = e^C = \text{const}$ , а первый элемент ряда  $e^{b_N} = e^{C/N} \rightarrow 1$ , при  $N \rightarrow +\infty$ .

В Таблице 1 продемонстрированы результаты для сигналов  $\mathbf{S}_N = (e^{mb_N})_{m=1}^N$ ,  $b_N = \ln(2)/N$ , и выброса  $a = 15$  на позиции  $k$ . Столбцы таблицы отвечают за разные длины  $N$ ; строки — за различные позиции выброса  $k$  и длины окна  $L$ . Для каждой пары  $k, L$  указан максимальный модуль полной ошибки восстановления и модуль максимальной разницы между полной ошибкой и первым порядком ошибки. Как можно заметить, разность стремится к нулю при любых значениях  $L$  и  $k$ , но полная ошибка стремится к нулю только, если  $L$  пропорционально  $N$ .

### Численное сравнение ошибки для больших рангов

Продemonстрируем поведение ошибки восстановления при возмущении в виде выброса в случае сигналов больших рангов.

$N$	100	1000	10000
$k = N/2 - 1, L = N/2$ , полная	$3.63 \times 10^{-1}$	$3.98 \times 10^{-2}$	$4.06 \times 10^{-3}$
$k = N/2 - 1, L = N/2$ , разница	$5.58 \times 10^{-2}$	$7.38 \times 10^{-4}$	$7.61 \times 10^{-6}$
$k = 19, L = 20$ , полная	$7.72 \times 10^{-1}$	$7.51 \times 10^{-1}$	$7.50 \times 10^{-1}$
$k = 19, L = 20$ , разница	$6.29 \times 10^{-2}$	$9.67 \times 10^{-3}$	$9.86 \times 10^{-4}$
$k = 19, L = N/2$ , полная	$4.79 \times 10^{-1}$	$4.28 \times 10^{-2}$	$4.17 \times 10^{-3}$
$k = 19, L = N/2$ , разница	$7.51 \times 10^{-2}$	$1.27 \times 10^{-3}$	$1.29 \times 10^{-5}$
$k = N/2, L = 20$ , полная	$7.51 \times 10^{-1}$	$7.50 \times 10^{-1}$	$7.50 \times 10^{-1}$
$k = N/2, L = 20$ , разница	$5.74 \times 10^{-3}$	$5.12 \times 10^{-6}$	$5.07 \times 10^{-9}$

Таблица 1: Максимальная разница между ошибкой I порядка и полной для выброса.

На Рис. 1 изображено поведение модуля ошибки восстановления сигнала при разных длинах окон  $L$ . На левом рисунке изображён случай вещественного экспоненциального сигнала  $S_N = (e^{mb_N})_{m=1}^N$ ,  $b_N = \ln(2)/N$ ,  $N = 3999$ , имеющего ранг 1, и выброса  $a = 1$ , на позиции  $k = N/3$ . Справа изображена ошибка для сигнала  $s_m = e^{mb_N}(\cos(2\pi m/40) + \cos(2\pi m/20 + \pi/3))$ , имеющего ранг 4. В Таблице 2 показано значение MSE оценки этих сигналов, умноженное на длину ряда, а также отношение MSE для ранга 4 к MSE для ранга 1. Как можно заметить, это отношение близко к 4.

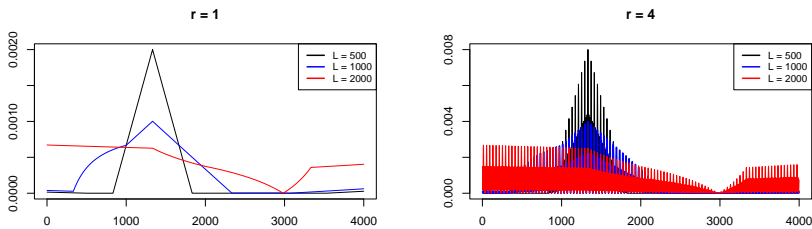


Рис. 1: Поведение ошибки восстановления при разных рангах.

## Заключение

Получен явный вид матрицы первого порядка ошибки для экспоненциального сигнала при возмущении в виде выброса. Численно показано, что для такого сигнала первый порядок по величине выброса адекватно описыва-

$L$	rank = 1	rank = 4	Отношение
500	1.33e-03	5.34e-03	4.002
1000	7.35e-04	2.94e-03	4.001
2000	8.90e-04	3.56e-03	4.004

Таблица 2: MSE оценки сигнала ранга 1 и 4.

ет полную ошибку. Также численно показана зависимость MSE от ранга для сигнала в виде суммы экспоненциально-модулированных гармоник с одинаковым показателем экспонент, что делает теоретический результат потенциально применимым для гораздо более широкого класса сигналов, чем просто вещественные экспоненты.

Так как доказано, что при возмущении в виде суммы шума и выброса смещение первого порядка оценки сигнала зависит только от значения выброса, полученные результаты дают полезную информацию и для зашумленных рядов.

### Список литературы

- [1] Golyandina N., Nekrutkin V., Zhigljavsky A. Analysis of Time Series Structure: SSA and Related Techniques. — New York : Chapman&Hall/CRC, 2001.
- [2] Kato T. Perturbation theory for linear operators. — Springer-Verlag, Berlin-Heidelberg-New York, 1966.
- [3] Nekrutkin V. Perturbation expansions of signal subspaces for long signals // Statistics and Its Interface. — 2010. — Vol. 3. — P. 297–319.
- [4] Golyandina N., Senov M., Khramov A. On errors of signal estimation using complex singular spectrum analysis // Engineering Proceedings. — 2025. — to be published.
- [5] Trickett S., Burroughs L., Milton A. Robust Rank-Reduction Filtering for Erratic Noise // SEG International Exposition and Annual Meeting. — 2012. — Vol. All Days. — P. SEG–2012–0129.
- [6] Kalantari Mahdi, Yarmohammadi Masoud, Hassani Hossein. Singular Spectrum Analysis Based on L1-Norm // Fluctuation and Noise Letters. — 2016. — Vol. 15, no. 01. — P. 1650009.