

Модификации метода анализа сингулярного спектра для анализа временных рядов: Circulant SSA и Generalized SSA

Погребников Н. В., гр. 21.Б04-мм

Санкт-Петербургский государственный университет
Прикладная математика и информатика
Вычислительная стохастика и статистические модели

Научный руководитель: д. ф.-м. н., проф. Голяндина Н. Э.

Санкт-Петербург, 2025

Пусть $X = (x_1, \dots, x_N)$ – временной ряд длины N , $x_i \in \mathbb{R}$ – наблюдение в момент времени i .

$X = X_{\text{Trend}} + X_{\text{Periodics}} + X_{\text{Noise}}$, где:

- X_{Trend} – тренд, медленно меняющаяся компонента;
- $X_{\text{Periodics}}$ – сумма периодических компонент;
- X_{Noise} – шум, случайная составляющая.

Методы: **SSA** – метод, позволяющий раскладывать временной ряда в сумму интерпретируемых компонент (Golyandina, Nekrutkin и Zhigljavsky 2001); **GSSA** – модификация **SSA** на основе добавления весов (Gu и др. 2024); **CiSSA** – модификация **SSA** на основе циркулярной матрицы (Bogalo, Poncela и Senra 2020).

Задача: Описание модификаций в контексте теории **SSA**, сравнение алгоритмов, реализация их на языке R.

Пример

$$X = S + X_{\text{Noise}} = S^{(1)} + S^{(2)} + X_{\text{Noise}} = e^{An} \sin(2\pi\omega_1 n) + \cos(2\pi\omega_2 n) + \varepsilon_n.$$

ω_1, ω_2 – частоты; $\varepsilon_n \sim N(0, \sigma^2)$ – шум;

S – сигнал.

\hat{S} – оценка выделения сигнала методом.

$\hat{S}^{(1)}, \hat{S}^{(2)}$ – оценки разделения компонент $S^{(1)}, S^{(2)}$.

Критерии сравнения методов:

- Выделение сигнала;
- Разделимость;
- Постановка задачи (для CiSSA частоты предполагаются известными).

$X_N = X_N^{(1)} + X_N^{(2)}$. М – метод разделения ряда на компоненты с параметрами Θ . $\hat{X}_N^{(1)}$ – восстановленная методом М компонента, отвечающая за $X_N^{(1)}$.

Определение 1

Ряды $X_N^{(1)}$ и $X_N^{(2)}$ точно разделимы методом М, если существует такое Θ , что $\text{MSE} \left(X_N^{(1)}, \hat{X}_N^{(1)} \right) = 0$.

Определение 2

Ряды $X_N^{(1)}$ и $X_N^{(2)}$ асимптотически разделимы методом М, если существует последовательность $\Theta(N)$, $N \rightarrow \infty$, что $\text{MSE} \left(X_N^{(1)}, \hat{X}_N^{(1)} \right) \rightarrow 0$.

Метод SSA. Алгоритм

$X = (x_1, \dots, x_N)$ — временной ряд. $1 < L < N$ — длина окна.

Алгоритм SSA:

- 1 Построение траекторной матрицы:

$$\mathbf{X} = \mathcal{T}_L(X) = [X_1 : \dots : X_K], \quad X_i = (x_i, \dots, x_{i+L-1})^T, \\ 1 \leq i \leq K, \quad K = N - L + 1.$$

- 2 Сингулярное разложение (SVD) траекторной матрицы:

$$\mathbf{X} = \sum_{i=1}^d \sqrt{\lambda_i} U_i V_i^T = \sum_{i=1}^d \mathbf{X}_i, \quad d = \text{rank}(\mathbf{X}).$$

\mathbf{X}_i — элементарные матрицы ранга 1.

$(\sqrt{\lambda_i}, U_i, V_i^T)$ — i -ая собственная тройка.

- 3 Группировка индексов $1, \dots, d$ на m непересекающихся подмножеств I_1, \dots, I_m . $\mathbf{X}_{I_k} = \sum_{i \in I_k} \mathbf{X}_i$.

$$\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_m}.$$

- 4 Восстановление: $\tilde{X}_{I_k} = \mathcal{T}_L^{-1} \circ \mathcal{H}(\mathbf{X}_{I_k})$,
 $X = \tilde{X}_{I_1} + \dots + \tilde{X}_{I_m}.$

$$X = S + X_{\text{Noise}} = S^{(1)} + S^{(2)} + X_{\text{Noise}}$$

Определение 3 (Golyandina, Korobeynikov и Zhigljavsky 2018)

Вложенный вариант SSA — двухэтапный метод:

1. *Задается r . \tilde{S} – сумма первых r слагаемых SVD разложения траекторной матрицы сигнала S с помощью базового **SSA**.*
2. *Применение другого метода к \tilde{S} для улучшения разделимости: $\tilde{S} = \tilde{S}_1 + \tilde{S}_2$.*

SSA EOSSA (Golyandina, Dudnik и Shlemov 2023) является вложенным вариантом **SSA**.

Метод GSSA. Алгоритм

$X = (x_1, \dots, x_N)$ — временной ряд, параметры L и $\alpha \geq 0$.

$$w^{(a)} = (w_1, w_2, \dots, w_L) = \left(\left| \sin \left(\frac{\pi n}{L+1} \right) \right|^\alpha \right), \quad n = 1, 2, \dots, L.$$

Шаг 1 алгоритма GSSA:

$$X^{(\alpha)} = \mathcal{T}_L^{(\alpha)}(X) = [X_1^{(\alpha)} : \dots : X_K^{(\alpha)}],$$

$$X_i^{(\alpha)} = (w_1 x_{i-1}, \dots, w_L x_{i+L-2})^T, \quad 1 \leq i \leq K.$$

Шаги 2-4: аналогичны **SSA**.

Замечание 1

При $\alpha = 0$, **GSSA** — в точности базовый алгоритм **SSA**.

Замечание 2

$w^{(a)}$ называются степенными синусными весами. Они могут иметь другой вид.

Определение 4

Пусть $X = (\dots, x_{-1}, x_0, x_1, \dots)$ — бесконечный временной ряд. **Линейный конечный фильтр** — оператор Φ , преобразующий X в $Y = (\dots, y_{-1}, y_0, y_1, \dots)$ по правилу:

$$y_j = \sum_{i=-r_1}^{r_2} h_i x_{j-i}, \quad j \in \mathbb{Z},$$

где $r_1 + r_2$ — ширина фильтра, $h_i \in \mathbb{R}$ — коэффициенты.

Пример. При применении фильтра Φ к $x_j = \cos 2\pi\omega j$, получается ряд $y_j = A_\Phi(\omega) \cos(2\pi\omega j + \phi_\Phi(\omega))$.

$\phi_\Phi(\omega)$ — фазово-частотная характеристика (ФЧХ).

$A_\Phi(\omega)$ — амплитудно-частотная характеристика (АЧХ).

Сравнение SSA и GSSA. Линейные фильтры 2

$X = (x_1, \dots, x_N)$, $(\sqrt{\lambda}, U, V)$ – собственная тройка **SSA**.

$U = (u_1, \dots, u_L)$. $\tilde{X} = \mathcal{T}_L \circ \mathcal{H}(\sqrt{\lambda}UV^T)$.

Запись SSA через линейный фильтр для средних точек:

$$\tilde{x}_s = \sum_{j=-(L-1)}^{L-1} \left(\sum_{k=1}^{L-|j|} u_k u_{k+|j|} / L \right) x_{s-j}, \quad L \leq s \leq K.$$

Аналогичное представление для GSSA:

$$\tilde{x}_s = \sum_{j=-(L-1)}^{L-1} \left(\sum_{k=1}^{L-|j|} u_k^{(\alpha)} u_{k+|j|}^{(\alpha)} w_k / \sum_{i=1}^L w_i \right) x_{s-j}, \quad L \leq s \leq K.$$

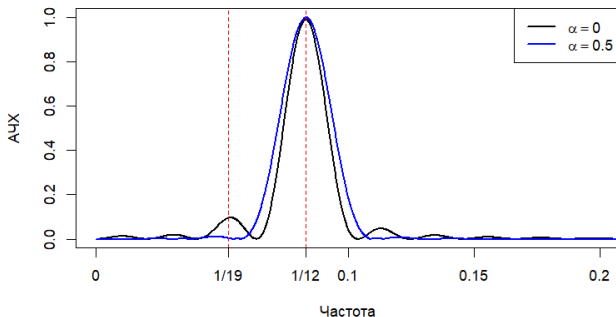
Замечание 1

Представление через линейные фильтры можно получить и для остальных точек ряда.

Сравнение SSA и GSSA. Пример

$X = X_{\sin} + X_{\cos} = \sin\left(\frac{2\pi}{12}n\right) + \frac{1}{2} \cos\left(\frac{2\pi}{19}n\right)$. $N = 96 \cdot 2 - 1$, $L = 48$.
Группировка: для X_{\sin} 1-2 SVD, для X_{\cos} 3-4 SVD.

АЧХ для суммы фильтров собственных троек синуса



$\alpha = 0.5$: шире полоса пропускания фильтра, чем при $\alpha = 0$, но нет волнообразного поведения на краях.

Сравнение SSA и GSSA. Пример, продолжение

Таблица 1: $X_{\sin} + X_{\cos}$, MSE оценок

Метод/Ошибка	X_{\sin}	X_{\cos}	X
SSA	5.15e-03	5.15e-03	6.01e-30
GSSA , $\alpha = 0.5$	3.68e-04	3.68e-04	9.53e-30

Без шума **GSSA** выдает результаты на порядок лучше **SSA**.

Таблица 2: $X_{\sin} + X_{\cos} + \varepsilon_n$, $\varepsilon_n \sim N(0, 0.1^2)$, MSE оценок

Метод/Ошибка	X_{\sin}	X_{\cos}	X
SSA	5.68e-03	5.44e-03	7.48e-04
GSSA , $\alpha = 0.5$	1.21e-03	1.25e-03	1.04e-03

С шумом выигрыш на порядок у **GSSA** пропал, но теперь **SSA** выделил сигнал на порядок лучше.

Вывод. Вложенный вариант SSA + GSSA

Можно объединить преимущества обоих алгоритмов, выделив сигнал с помощью **SSA**, а затем разделив компоненты друг от друга благодаря **GSSA**:

Метод/Ошибка	X_{\sin}	X_{\cos}	X
SSA	5.68e-03	5.44e-03	7.48e-04
GSSA , $\alpha = 0.5$	1.21e-03	1.25e-03	1.04e-03
SSA + GSSA , $\alpha = 0.5$	1.06e-03	1.12e-03	7.15e-04

Получается вложенный вариант **SSA**.

Метод CiSSA. Алгоритм

$\mathbf{X} = (x_1, \dots, x_N)$ — временной ряд. $1 < L < N$ — длина окна.

Алгоритм CiSSA:

❶ Построение траекторной матрицы: как в **SSA**.

❷ $l = 1 : L$,

$$U_l = L^{-1/2}(u_{l,1}, \dots, u_{l,L}), \quad u_{l,j} = \exp(-i2\pi(j-1)\frac{l-1}{L}).$$

Элементарное разложение: $\omega_k = \frac{k-1}{L}$, $k = 1 : \lfloor \frac{L+1}{2} \rfloor$

$$\mathbf{X}_{\omega_k} = U_k U_k^H \mathbf{X} + U_{L+2-k} U_{L+2-k}^H \mathbf{X};$$

$$\mathbf{X}_{\omega_{\frac{L}{2}+1}} = U_{\frac{L}{2}+1} U_{\frac{L}{2}+1}^H \mathbf{X}, \text{ если } L \bmod 2 = 0,$$

Разложение: $\mathbf{X} = \sum_{k=1}^d \mathbf{X}_{\omega_k}$, $d = \lfloor \frac{L+1}{2} \rfloor$ (или $\frac{L}{2} + 1$).

❸ Группировка по частотам:

$$\bigsqcup_{j=1}^m I_j = \bigsqcup_{j=1}^m \left[\omega_j^{(l)}, \omega_j^{(r)} \right] = [0, 0.5]. \quad \mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_m}.$$

❹ Диагональное усреднение: как в **SSA**.

- **SSA**: базис подпространства которого зависит от X, L, N (адаптивный).
CiSSA: базис зависит от L, N (фиксированный).
- В **CiSSA** группировка производится по диапазонам частот, поэтому алгоритм применим только в случае, когда заранее известны частоты интересующих компонент.

Определение 5

Разложение

$$x_n = c_0 + \sum_{k=1}^{\lfloor \frac{N+1}{2} \rfloor} (c_k \cos(2\pi nk/N) + s_k \sin(2\pi nk/N)),$$

где $1 \leq n \leq N$ и $s_{N/2} = 0$ для четного N , называется разложением Фурье ряда X .

Разложение Фурье ряда – разложение по базису из синусов и косинусов.

CiSSA – разложения Фурье для K векторов матрицы X с последующим диагональным усреднением слагаемых.

Сравнение SSA, Фурье, CiSSA. Точная разделимость

Фиксируем временной ряд $X = X_1 + X_2 =$
 $= A_1 \cos(2\pi\omega_1 n + \varphi_1) + A_2 \cos(2\pi\omega_2 n + \varphi_2).$

Условия точной разделимости X для разложения **Фурье**:
 $N\omega_1, N\omega_2 \in \mathbb{N}, \omega_1 \neq \omega_2.$

Условия точной разделимости X для **CiSSA**:
 $L\omega_1, L\omega_2 \in \mathbb{N}, \omega_1 \neq \omega_2.$

Условия точной разделимости X для **SSA**:
 $L\omega_1, L\omega_2, K\omega_1, K\omega_2 \in \mathbb{N}, \omega_1 \neq \omega_2, A_1 \neq A_2.$

Таким образом, условия на разделение косинусов, слабее у методов **CiSSA** и **Фурье**, чем у **SSA**.

Сравнение SSA, Фурье, CiSSA. Асимптотическая делимость

Асимптотически делимы в методе **SSA** полиномы, гармонические, экспоненциально-модулированные функции (Golyandina, Nekrutkin и Zhigljavsky 2001).

Для **разложения Фурье** асимптотически делимы только гармонические функции.

Для **CiSSA** асимптотически делимы гармонические, экспоненциально-модулированные функции.

Пример 1. Гармонические функции

Пример 1: $X = X_{\sin} + X_{\cos} = A_1 \sin(2\pi\omega_1 n) + A_2 \cos(2\pi\omega_2 n)$.

Группировка: $\delta = 1/L$,

для X_{\sin} 1-2 SVD или $(\omega_1 \pm 2\delta)$; для X_{\cos} 3-4 SVD или $(\omega_2 \pm 2\delta)$;

Метод	Параметры	MSE (X_{\sin})	MSE (X_{\cos})	MSE (X)
SSA	$L\omega_i \in \mathbb{N}, K\omega_i \in \mathbb{N}, A_1 \neq A_2$	6.8e-30	1.5e-29	1.8e-29
SSA EOSSA	$L\omega_i \in \mathbb{N}, K\omega_i \in \mathbb{N}, A_1 \neq A_2, r = 4$	8.2e-30	6.5e-30	5.5e-30
Fourier	$N\omega_i \in \mathbb{N}$	3.4e-28	9.8e-29	4.0e-28
CiSSA	$L\omega_i \in \mathbb{N}, A_1 \neq A_2$	1.1e-29	6.5e-30	7.8e-30
SSA	$L\omega_i \in \mathbb{N}, K\omega_i \in \mathbb{N}, A_1 = A_2$	3.8e-04	3.8e-04	6.0e-29
SSA	$L\omega_i \in \mathbb{N}, K\omega_i \notin \mathbb{N}, A_1 = A_2$	4.9e-03	3.4e-03	5.9e-29
SSA EOSSA	$L\omega_i \in \mathbb{N}, K\omega_i \notin \mathbb{N}, A_1 = A_2, r = 4$	1.4e-29	2.9e-29	1.1e-29
Fourier	$N\omega_i \notin \mathbb{N}$	7.6e-03	3.3e-03	5.6e-03

По таблице видно, что при нарушении условий точной разделимости, результаты значительно ухудшаются.

SSA EOSSA исправляет ситуацию для **SSA**.

Пример 1. Шум

Пример 1: $X = X_{\sin} + X_{\cos} + X_{\text{Noise}} =$

$$= A_1 \sin(2\pi\omega_1 n) + A_2 \cos(2\pi\omega_2 n) + \varepsilon_n, \quad \varepsilon_n \sim N(0, 0.1^2)$$

Группировка: $\delta = 1/L$,

для X_{\sin} 1-2 SVD или $(\omega_1 \pm 2\delta)$; для X_{\cos} 3-4 SVD или $(\omega_2 \pm 2\delta)$;

Метод	Параметры	MSE (X_{\sin})	MSE (X_{\cos})	MSE (X)
SSA	$L\omega_i \in \mathbb{N}, K\omega_i \in \mathbb{N}$	2.7e-04	3.3e-04	6.0e-04
SSA EOSSA	$L\omega_i \in \mathbb{N}, K\omega_i \in \mathbb{N}$	2.7e-04	3.3e-04	6.0e-04
Fourier	$N\omega_i \in \mathbb{N}$	1.5e-04	2.1e-04	3.6e-04
CiSSA	$L\omega_i \in \mathbb{N}$	1.6e-04	2.8e-04	4.3e-04
SSA	$L\omega_i \in \mathbb{N}, K\omega_i \in \mathbb{N}, A_1 = A_2$	2.5e-04	3.3e-04	6.0e-04
SSA	$L\omega_i \in \mathbb{N}, K\omega_i \notin \mathbb{N}, A_1 = A_2$	4.9e-03	3.4e-03	6.0e-04
SSA EOSSA	$L\omega_i \in \mathbb{N}, K\omega_i \notin \mathbb{N}, A_1 = A_2$	2.7e-04	3.4e-04	6.0e-04
Fourier	$N\omega_i \notin \mathbb{N}$	2.6e-02	7.3e-02	9.8e-02

Результаты ожидаемо ухудшились.

Пример 2. Экспоненциально-модулированные функции

Пример 2:

$$X = X_{e.\sin} + X_{e.\cos} = e^{A_1 n} \sin(2\pi\omega_1 n) + e^{A_2 n} \cos(2\pi\omega_2 n).$$

Группировка: $\delta = 1/L$,

для X_{\sin} 1-2 SVD или $(\omega_1 \pm 2\delta)$; для X_{\cos} 3-4 SVD или $(\omega_2 \pm 2\delta)$;

Метод	Параметры	MSE ($X_{e.\sin}$)	MSE ($X_{e.\cos}$)	MSE (X)
SSA	$L\omega_i \in \mathbb{N}, K\omega_i \in \mathbb{N}$	5.3e-05	5.3e-05	1.2e-27
SSA EOSSA	$L\omega_i \in \mathbb{N}, K\omega_i \in \mathbb{N}, r = 4$	3.0e-28	4.4e-28	7.4e-29
Fourier	$N\omega_i \in \mathbb{N}$	6.7e-02	1.4e-02	4.9e-02
CiSSA	$L\omega_i \in \mathbb{N}$	3.8e-03	2.6e-02	1.5e-02
SSA	$L\omega_i \in \mathbb{N}, K\omega_i \notin \mathbb{N}$	4.8e-04	4.8e-04	1.1e-27
SSA EOSSA	$L\omega_i \in \mathbb{N}, K\omega_i \notin \mathbb{N}, r = 4$	2.8e-28	4.2e-28	7.5e-29
Fourier	$N\omega_i \notin \mathbb{N}$	3.7e-02	1.1e-01	1.1e-01

При домножении на экспоненты периодик, все результаты ухудшились кроме **SSA EOSSA**. Фурье и **CiSSA** значительно ухудшились в точности разделения.

Пример 2. Шум

Пример 2: $X = X_{e \cdot \sin} + X_{e \cdot \cos} + X_{\text{Noise}} =$
 $= e^{A_1 n} \sin(2\pi w_1 n) + e^{A_2 n} \cos(2\pi w_2 n) + \varepsilon_n, \varepsilon_n \sim N(0, 0.1^2)$

Метод	Параметры	MSE ($X_{e \cdot \sin}$)	MSE ($X_{e \cdot \cos}$)	MSE (X)
SSA	$Lw \in \mathbb{N}, Kw \in \mathbb{N}$	3.1e-04	3.6e-04	5.6e-04
SSA EOSSA	$Lw \in \mathbb{N}, Kw \in \mathbb{N}$	2.2e-04	3.4e-04	5.6e-04
Fourier	$Nw \in \mathbb{N}$	1.5e-02	7.2e-02	7.2e-02
CiSSA	$Lw \in \mathbb{N}$	5.2e-03	3.4e-02	3.3e-02
SSA	$L\omega_i \in \mathbb{N}, K\omega_i \notin \mathbb{N}$	7.7e-04	8.7e-04	5.6e-04
SSA EOSSA	$L\omega_i \in \mathbb{N}, K\omega_i \notin \mathbb{N}, r = 4$	5.8e-04	5.6e-04	7.1e-04
Fourier	$N\omega_i \notin \mathbb{N}$	4.2e-02	3.3e-01	3.5e-01

Результаты ожидаемо ухудшились.

По полученным результатам, можно следующие выводы:

- 1 Изменение амплитуды периодических компонент временного ряда со временем значительно затрудняет выделение сигнала и разделение компонент при использовании методов **CiSSA** и разложения **Фурье**;
- 2 **CiSSA** показывает себя лучше **Фурье**;
- 3 На разделение периодических компонент для базового **SSA** накладываются более строгие ограничения относительно **CiSSA**. В остальных случаях **SSA** работает лучше;
- 4 **SSA EOSSA** исправляет недостатки базового **SSA**.
- 5 Имеет смысл вложенный вариант с **CiSSA**.

Результаты данного исследования:

- Выявлены сильные и слабые стороны методов;
- Предложены собственные вложенные модификации;
- Методы реализованы на языке R.

Последующие действия:

- Рассмотрение других модификаций;
- Добавление к списку примеров ряды с трендом;
- Реализация вложенного варианта с **CiSSA**.

Список литературы I

-  Bogalo, Juan, Pilar Poncela и Eva Senra (2020). «Circulant singular spectrum analysis: A new automated procedure for signal extraction». В: *Signal Processing* 177. ISSN: 0165-1684. DOI: 10.1016/j.sigpro.2020.107750. URL: <http://www.sciencedirect.com/science/article/pii/S0165168420303264>.
-  Golyandina, Nina, Pavel Dudnik и Alex Shlemov (2023). «Intelligent Identification of Trend Components in Singular Spectrum Analysis». В: *Algorithms* 16.7, с. 353. DOI: 10.3390/a16070353. URL: <https://doi.org/10.3390/a16070353>.
-  Golyandina, Nina, Anton Korobeynikov и Anatoly Zhigljavsky (янв. 2018). *Singular Spectrum Analysis with R*. ISBN: 978-3-662-57378-5. DOI: 10.1007/978-3-662-57380-8.



Golyandina, Nina, Vladimir Nekrutkin и Anatoly Zhigljavsky (2001). *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman и Hall/CRC. URL: https://www.academia.edu/34626051/Analysis_of_Time_Series_Structure_-_SSA_and_Related_Techniques.



Gu, Jialiang и др. (2024). «Generalized singular spectrum analysis for the decomposition and analysis of non-stationary signals». В: *Journal of the Franklin Institute Accepted/In Press*. ISSN: 0016-0032. DOI: 10.1016/j.jfranklin.2024.106696. URL: <https://doi.org/10.1016/j.jfranklin.2024.106696>.