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., доц. Голяндина Н. Э.

Санкт-Петербург, 2025

Введение

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

$\mathsf{X} = \mathsf{X}_{\mathsf{Trend}} + \mathsf{X}_{\mathsf{Periodics}} + \mathsf{X}_{\mathsf{Noise}}$, где:

- $\mathsf{X}_{\mathsf{Trend}}$ – тренд, медленно меняющаяся компонента;
- $\mathsf{X}_{\mathsf{Periodics}}$ – сумма периодических компонент;
- $\mathsf{X}_{\mathsf{Noise}}$ – 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 **CiSSA** на основе циркулярной матрицы (Bogalo, Poncela и Senra 2020).

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

Метод SSA. Алгоритм

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

Алгоритм SSA:

1. **Построение траекторной матрицы:**
   $\mathbf{X} = \mathcal{H}(\mathsf{X}) = [\mathsf{X}_1 : \ldots : \mathsf{X}_K]$, $\mathsf{X}_i = (x_i, \ldots, x_{i+L-1})^T$,
   $1 \leq i \leq K, \quad K = N - L + 1.$
2. **Сингулярное разложение (SVD)** траекторной матрицы:
   $$\mathbf{X} = \sum_{i=1}^{d} \sqrt{\lambda_i} U_i V_i^T = \sum_{i=1}^{d} \mathbf{X}_i,\ d = \mathrm{rank}(\mathbf{X}).$$
   $\mathbf{X}_i$ — элементарные матрицы ранга 1.
   $(\sqrt{\lambda_i}, U_i, V_i^{\mathrm{T}})$ – i-ая собственная тройка.
3. **Группировка** индексов $1, \ldots, d$ на m непересекающихся подмножеств $I_1, \ldots, I_m$, $I_k = \{i_1^{(k)}, \ldots, i_{p_k}^{(k)}\}$.
   $\mathbf{X}_{I_k} = \mathbf{X}_{i_1^{(k)}} + \cdots + \mathbf{X}_{i_{p_k}^{(k)}}$. $\mathbf{X} = \mathbf{X}_{I_1} + \cdots + \mathbf{X}_{I_m}$.
4. **Восстановление:** $\tilde{\mathsf{X}}_{I_k} = \mathcal{H}^{-1}(\mathbf{X}_{I_k})$, $\mathsf{X} = \tilde{\mathsf{X}}_{I_1} + \cdots + \tilde{\mathsf{X}}_{I_m}$.

Метод GSSA. Алгоритм

$\mathsf{X} = (x_1, \ldots, x_N)$ — временной ряд, параметры L и $\alpha \geq 0$.

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

Шаг 1 алгорима GSSA:
$\mathbf{X}^{(\alpha)} = \mathcal{H}^{(\alpha)}(\mathsf{X}) = [\mathsf{X}_1^{\alpha} : \ldots : \mathsf{X}_K^{\alpha}]$,
$\mathsf{X}_i^{(\alpha)} = (w_1 x_{i-1}, \ldots, w_L x_{i+L-2})^{\mathrm{T}}$, $1 \leq i \leq K$.
Шаги 2-4: аналогичны **SSA**.

Замечание 1

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

Сравнение SSA и GSSA. Линейные фильтры 1

Определение 1

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

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

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

Пример. При применении фильтра Φ на $\mathsf{X}_{\cos} = \cos 2\pi\omega n$, получается ряд $y_j = A_\Phi(\omega)\cos\left(2\pi\omega j + \phi_\Phi(\omega)\right)$.
$\phi_\Phi(\omega)$ – фазово-частотная характеристика (ФЧХ).
$A_\Phi(\omega)$ – амплитудно-частотная характеристика (АЧХ).

Сравнение SSA и GSSA. Линейные фильтры 2

$\mathsf{X} = (x_1, \ldots, x_N)$, $(\sqrt{\lambda}, U, V)$ – собственная тройка **SSA**.
$U = (u_1, \ldots, u_L)$. $\widetilde{\mathsf{X}} = \mathcal{H}^{-1}(\sqrt{\lambda} U V^T)$.

Запись SSA через линейный фильтр:

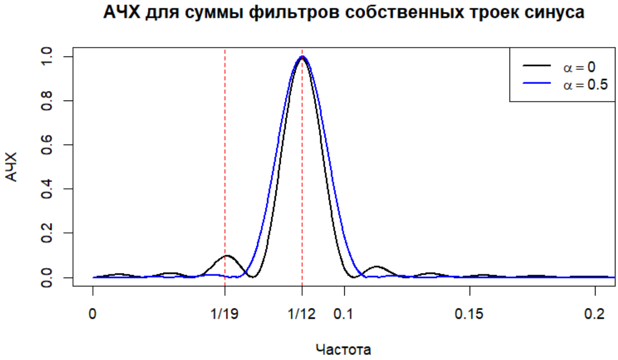
$$\widetilde{x}_s = \sum_{j=-(L-1)}^{L-1} \left(\sum_{k=1}^{L-|j|} u_k u_{k+|j|} / L \right) x_{s-j}, \quad L \le s \le K.$$

Аналогичное представление для GSSA:

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

Сравнение SSA и GSSA. Пример

$$\mathsf{X} = \mathsf{X}_{\sin} + \mathsf{X}_{\cos} = \sin\left(\frac{2\pi}{12}n\right) + \frac{1}{2}\cos\left(\frac{2\pi}{19}n\right).\ N = 96\cdot 2 - 1,\ L = 48.$$



При $\alpha = 0.5$ АЧХ без волн, но с широкой областью около частоты синуса, что ухудшает отделение от шума, но улучшает разделение компонентов.

Сравнение SSA и GSSA. Пример, продолжение

Таблица 1: MSE разложений $X = X_{\text{sin}} + X_{\text{cos}}$

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: MSE разложений $X = X_{\text{sin}} + X_{\text{cos}} + \varepsilon_n$, $\varepsilon_n \sim \mathrm{N}(0, 0.1^2)$

Метод	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.

Сравнение SSA и GSSA. Выводы

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

Метод	$\mathsf{X}_{\mathrm{sin}}$	$\mathsf{X}_{\mathrm{cos}}$	X
SSA + **GSSA**, $\alpha = 0.5$	**1.06e-03**	**1.12e-03**	**7.15e-04**

Получились наилучшие результаты.

Метод CiSSA. Алгоритм

$\mathsf{X} = (x_1, \ldots, x_N)$ — временной ряд. $1 < L < N$ – длина окна.
Алгоритм SSA:

1. **Построение траекторной матрицы:** как в **SSA**.
2. $l = 1 : L$,
   $U_l = L^{-1/2}(u_{l,1}, \ldots, u_{l,L})$, $u_{l,j} = \exp\left(-\mathrm{i}2\pi(j-1)\frac{l-1}{L}\right)$.
   Элементарное разложение: $w_k = \frac{k-1}{L}$, $k = 1 : \lfloor\frac{L+1}{2}\rfloor$

   $$\mathbf{X}_{w_k} = U_k U_k^H \mathbf{X} + U_{L+2-k} U_{L+2-k}^H \mathbf{X};$$
   $$\mathbf{X}_{w_{\frac{L}{2}+1}} = U_{\frac{L}{2}+1} U_{\frac{L}{2}+1}^H \mathbf{X}, \text{ если } L \mod 2 = 0,$$

   Разложение: $\mathbf{X} = \sum\limits_{k=1}^{d} \mathbf{X}_{B_k}$, $d = \lfloor\frac{L+1}{2}\rfloor$ (или $\frac{L}{2}+1$).
3. **Группировка** по частотам $w_k \in [0, 0.5]$ на непересекающиеся диапазоны $I_i = [w_{i0}, w_{i1}]$.
4. **Диагональное усреднение:** как в **SSA**.

Замечания:

- В отличие от **SSA**, базис подпространства которого зависит от X, L, N (адаптивный), базис в **CiSSA** зависит только от L, N (фиксированный).
- Поскольку группировка производится по частотам, а частоты зависят от L, то алгоритм применим только в случае, когда заранее известны интересующие частоты.

Метод CiSSA. Свойства: связь с разложением Фурье

Определение 2

Разложение

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

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

Замечание 2

В разложении Фурье производится проекция всего ряда на пространства, порожденные синусами и косинусами. В ***CiSSA*** *производится разложение Фурье для* K *векторов матрицы* X. *Затем соответствующие элементы усредняются.*

Метод CiSSA. Свойства: связь с разложением Фурье

Определение 2

Разложение

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

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

Замечание 2

*В разложении Фурье производится проекция всего ряда на пространства, порожденные синусами и косинусами. В **CiSSA** производится разложение Фурье для K векторов матрицы $\mathbf{X}$. Затем соответствующие элементы усредняются.*

TODO Переписать замечание (частично сделано)

TODO Написать, что будем сравнивать алгоритмы **SSA** и **GSSA** по разделимости компонент между собой, когда заранее знаем, на какие частоты будем разделять. Для этого нужны определения разделимости.

Сравнение SSA, Фурье, CiSSA. Разделимость

Определение 3

Есть метод разделения ряда на компоненты с параметрами Θ, *ряд* $\mathsf{X} = \mathsf{X}^{(1)} + \mathsf{X}^{(2)}$. $\exists$ *набор параметров* $\hat{\Theta}$, L, N, *что при разделении ряда на компоненты этим методом,* $\hat{\mathsf{X}}^{(1)}$ *является оценкой* $\mathsf{X}^{(1)}$, *при этом,* $\mathrm{MSE}\left(\mathsf{X}^{(1)}, \hat{\mathsf{X}}^{(1)}\right) = 0$. *Тогда ряды* $\mathsf{X}^{(1)}$ *и* $\mathsf{X}^{(2)}$ *точно разделимы данным методом.*

Определение 4

Есть метод разделения ряда на компоненты с параметрами Θ, *ряд* $\mathsf{X} = \mathsf{X}^{(1)} + \mathsf{X}^{(2)}$. $\exists$ *набор параметров* $\hat{\Theta}$ *и* $L = L(N)$, $N \to \infty$, *что при разделении ряда на компоненты этим методом,* $\hat{\mathsf{X}}^{(1)}$ *является оценкой* $\mathsf{X}^{(1)}$, *при этом,* $\mathrm{MSE}\left(\mathsf{X}^{(1)}, \hat{\mathsf{X}}^{(1)}\right) \to 0$. *Тогда ряды* $\mathsf{X}^{(1)}$ *и* $\mathsf{X}^{(2)}$ *называются асимптотически* $L(N)$*-разделимыми данным методом.*

Сравнение SSA, Фурье, CiSSA. Точная разделимость

Фиксируем временной ряд $\mathsf{X} = \mathsf{X}_1 + \mathsf{X}_2 =$
$= A_1 \cos(2\pi w_1 n + \varphi_1) + A_2 \cos(2\pi w_2 n + \varphi_2)$.

Условия точной разделимости X для разложения Фурье:
$N w_1, N w_2 \in \mathbb{N}$, $w_1 \neq w_2$.
Условия точной разделимости X для **CiSSA**:
$L w_1, L w_2 \in \mathbb{N}$, $w_1 \neq w_2$.
Условия точной разделимости X для **SSA**:
$L w_1, L w_2, K w_1, K w_2 \in \mathbb{N}$, $w_1 \neq w_2$, $A_1 \neq A_2$.

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

Сравнение SSA, Фурье, CiSSA. Асимптотическая разделимость

Асимптотически разделимы в методе **SSA** полиномы, гармонические функции, не удовлетворяющие условиям точной разделимости, экспоненты (Golyandina, Nekrutkin и Zhigljavsky 2001).

Замечание 3

*Для **SSA** существуют алгоритмы улучшения разделимости, например, EOSSA и FOSSA (Golyandina, Dudnik и Shlemov 2023). По заданному набору компонент, они позволяют более точно отделять компоненты.*

В алгоритме разложения **CiSSA** (Фурье) увеличение длины окна L (N) изменяет сетку частот. Это означает, что даже если не удастся подобрать такое L (N), при котором косинус будет точно отделим, его постепенное увеличение позволит приблизить частоты сетки к частоте компоненты. В итоге,

Модификации метода **SSA**

Сравнение SSA, Фурье, CiSSA. Асимптотическая разделимость

Сравнение SSA, Фурье, CiSSA. Асимптотическая разделимость

Асимптотически разделимы в методе **SSA** полиномы, гармонические функции, не удовлетворяющие условиям точной разделимости, экспоненты (Golyandina, Nekrutkin и Zhigljavsky 2001).

Замечание 3

*Для **SSA** существуют алгоритмы улучшения разделимости, например, EOSSA и FOSSA (Golyandina, Dudnik и Shlemov 2023). По заданному набору компонент, они позволяют более точно отделять компоненты.*

В алгоритме разложения **CiSSA** (Фурье) увеличение длины окна L (N) изменяет сетку частот. Это означает, что даже если не удастся подобрать такое L (N), при котором косинус будет точно отделим, его постепенное увеличение позволит приблизить частоты сетки к частоте компоненты. В итоге, можно снизить ошибку выделения нужной компоненты.

TODO Переформулировать с меньшим количеством слов, ссылки переделать

В работе (Bogalo, Poncela и Senra 2020) предложена модификация, улучшающая результат выделения тренда, но ухудшающая разделение периодических компонент:

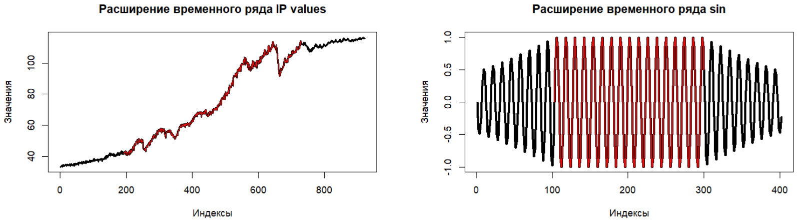
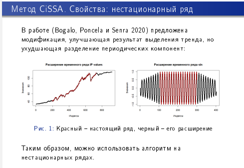


Рис. 1: Красный – настоящий ряд, черный – его расширение

Таким образом, можно использовать алгоритм на нестационарных рядах.



TODO Написать алгоритм расширения ряда в конце как приложение в конце. сказать, что изначально алгоритм пригоден только для стационарных рядов.

TODO Пример, когда известны частоты, попадаем / не попадаем в решетку.

TODO Пример, когда условия нарушаются (добавление тренда).

TODO Пример, когда шум.

Сравнение SSA, Фурье, CiSSA. Выводы 1

Метод/Условие	cos, $Lw \in \mathbb{N}$, $Kw \in \mathbb{N}$	cos, $Lw \in \mathbb{N}$, $Kw \notin \mathbb{N}$	cos, $Lw \notin \mathbb{N}$, $Kw \notin \mathbb{N}$	X_{np1}	X_{np}	group
SSA	$+$	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\rightarrow$	$-$
SSA EOSSA	$+$	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\rightarrow$	$+$
CiSSA	$+$	$+$	$\rightarrow$	$-$	$-$	$+$
CiSSA extended	$+$	$+$	$\rightarrow$	$\rightarrow$	$-$	$+$

Таблица 3: Преимущества и недостатки методов **SSA**, **CiSSA**

Метод/Условие	cos, $Nw \in \mathbb{N}$	cos, $Nw \notin \mathbb{N}$	X_{np1}	X_{np}	group
Fourier	$+$	$\rightarrow$	$-$	$-$	$+$
Fourier extended	$+$	$\rightarrow$	$\rightarrow$	$-$	$+$

Таблица 4: Преимущества и недостатки методов Fourier

Сравнение SSA, Фурье, CiSSA. Выводы 2

По полученным результатам, можно следующие выводы:

1. Если понятно, что ряд состоит только из периодических компонент, стоит использовать **CiSSA** без процедуры расширения, поскольку она делает ошибки разделений периодики больше. И напротив, если есть непериодичность, лучше расширять ряд;
2. Если данные зашумлены или имеется непериодичность, алгоритм **SSA**, чем **CiSSA** с расширением ряда или без.

TODO Переписать выводы в соответствии с примерами.

Сравнение SSA, Фурье, CiSSA. Выводы 2

По полученным результатам, можно следующие выводы:

1. Если понятно, что ряд состоит только из периодических компонент, стоит использовать **CiSSA** без процедуры расширения, поскольку она делает ошибки разделений периодики больше. И напротив, если есть непериодичность, лучше расширять ряд;
2. Если данные зашумлены или имеется непериодичность, алгоритм **SSA**, чем **CiSSA** с расширением ряда или без.

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.

Список литературы II

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.