

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

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 [ссылка]; **GSSA** – модификация **SSA** на основе добавления весов [ссылка]; **CiSSA** – модификация **CiSSA** на основе циркулярной матрицы [ссылка].

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

Модификации метода SSA

Введение

TODO Дописать ссылки

Введение

Пусть $X = (x_1, \dots, x_N)$ — произвольный ряд данных $N, x_i \in \mathbb{R}$ — наблюдение в момент времени i .
 $X = X_{\text{trend}} + X_{\text{seasonal}} + X_{\text{noise}}$ при:
• X_{trend} — тренд, медленно меняющаяся составляющая;
• X_{seasonal} — сезон, периодическая составляющая;
• X_{noise} — шум, случайная составляющая.
Методы SSA — метод, позволяющий раскладывать временной ряд в сумму экстремальных компонент [1, 2, 3, 4, 5].
GSSA — модификация SSA на основе добавления весов [2, 3, 4, 5].
CSSA — модификация CSSA на основе циркулярной матрицы [2, 3, 4, 5].
Задача: Описать модификации в контексте теории SSA, сравнить алгоритмы, реализовать их на языке R.

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

Алгоритм SSA:

- 1 Построение траекторной матрицы:

$$\mathbf{X} = \mathcal{H}(\mathbf{X}) = [\mathbf{X}_1 : \dots : \mathbf{X}_K], \mathbf{X}_i = (x_i, \dots, x_{i+L-1})^T, \\ 1 \leq i \leq K, \quad K = N - L + 1.$$

- 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 , $I_k = \{i_1^{(k)}, \dots, i_{p_k}^{(k)}\}$.

$$\mathbf{X}_{I_k} = \mathbf{X}_{i_1^{(k)}} + \dots + \mathbf{X}_{i_{p_k}^{(k)}}. \quad \mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_m}.$$

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

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

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

Шаг 1 алгоритма GSSA:

$$\mathbf{X}^{(\alpha)} = \mathcal{H}^{(\alpha)}(\mathbf{X}) = [\mathbf{X}_1^\alpha : \dots : \mathbf{X}_K^\alpha],$$

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

Шаги 2-4: аналогичны **SSA**.

Замечание 1

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

Модификации метода SSA

└ Метод GSSA. Алгоритм

TODO Сослаться на статью авторов.

Метод GSSA. Алгоритм

$X_0 = (x_1, \dots, x_N) \rightarrow$ инициал, параметры L и $\alpha \geq 0$.

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

Шаг 1 алгоритма GSSA:

$X^{(0)} = H^{(0)}(X) = [X_1^0; \dots; X_N^0],$

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

Шаг и 2-4: алгоритм SSA.

Замечание 1

При $\alpha = 0$, GSSA — это модифицированный алгоритм SSA.

Определение 1

Пусть бесконечный временной ряд $X = (\dots, x_{-1}, x_0, x_1, \dots)$.
Линейный конечный фильтр — это оператор Φ , который преобразует временной ряд X в новый по следующему правилу:

$$y_j = \sum_{i=-r_1}^{r_2} h_i x_{j-i}; \quad r_1, r_2 < \infty.$$

$A_\Phi(\omega) = |H_\Phi(e^{i2\pi\omega})|$ — амплитудно-частотная характеристика (АЧХ).

Пример. При применении фильтра Φ на $X_{\cos} = \cos 2\pi\omega n$, получается ряд $y_j = A_\Phi(\omega) \cos(2\pi\omega j + \phi_\Phi(\omega))$.

Модификации метода SSA

└ Сравнение SSA и GSSA. Линейные фильтры 1

Сравнение SSA и GSSA. Линейные фильтры 1

Определение 1

Пусть бесконечный вещественный ряд $X = \{x_1, x_2, x_3, \dots\}$.
Линейный конечный фильтр — это оператор Φ , который преобразует ряд X в новый по следующему правилу:

$$y_j = \sum_{i=r_1}^{r_2} b_i x_{j-i}, \quad r_1, r_2 < \infty.$$

$A_{\Phi}(\omega) = [H_{\Phi}(e^{i2\pi\omega})]$ — амплитудно-частотная характеристика или АЧХ.

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

TODO Переписать определение. Пояснить, что означает АЧХ на примере.

Сказать, что будем сравнивать эти методы с точки зрения линейных фильтров.

Сравнение SSA и GSSA. Линейные фильтры 2

Пусть $X = (x_1, \dots, x_N)$ — временной ряд длины N , $(\sqrt{\lambda}, U, V)$ — одна из собственных троек разложения методом **SSA**.

$U = (u_1, \dots, u_L)$.

Тогда компонента временного ряда \tilde{X} , восстановленная с использованием собственной тройки $(\sqrt{\lambda}, U, V)$, для средних точек (индексы от L до K) имеет вид:

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

Таким образом, имеется представление алгоритма **SSA** через линейные фильтры.

Аналогичное представления для **GSSA**:

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

Модификации метода SSA

Сравнение SSA и GSSA. Линейные фильтры 2

Сравнение SSA и GSSA. Линейные фильтры 2

Пусть $X = (x_1, \dots, x_N) \rightarrow$ в режиме \bar{V} ряд данных $N, (\sqrt{N}, U, V)$
 \rightarrow одна из собственных компонент разложения методом SSA.
 $U = (u_1, \dots, u_L)$
Тогда компоненты в режиме \bar{V} ряда \bar{X} представляются с использованием собственных \bar{V} -рядов (\sqrt{N}, U, V) , для каждого n (индекса от 1 до K) имеет вид:

$$\bar{X}_n = \sum_{j=n-(L-1)}^{L-1} \left(\sum_{k=1}^{L-|j|} u_k u_{n+|j|} / L \right) x_{n-j}, \quad L \leq n \leq K.$$

Таким образом, имеется представление алгоритма SSA в виде линейных фильтров.

Аналогично представляется и алгоритм GSSA:

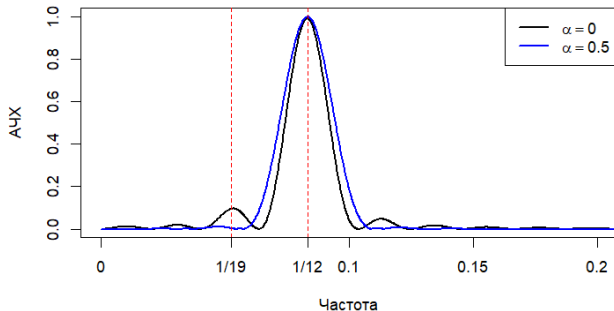
$$\bar{X}_n = \sum_{j=n-(L-1)}^{L-1} \left(\sum_{k=1}^{L-|j|} u_k^{(n)} u_{n+|j|}^{(n)} / \sum_{k=1}^L u_k \right) x_{n-j}, \quad L \leq n \leq K.$$

TODO Убрать весь текст, оставить только представление в виде фильтров

Сравнение SSA и GSSA. Пример

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

Фильтры для различных α

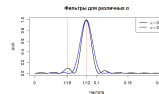


Модификации метода SSA

└ Сравнение SSA и GSSA. Пример

TODO Расписать, что **GSSA** хуже отделяет от шума, но лучше компоненты между собой, основываясь по рисунку. Допisać, по каким группам производилось объединение.

$$X = X_{\text{sig}} + X_{\text{noise}} = \sin\left(\frac{\pi}{10}n\right) + \frac{1}{2}\cos\left(\frac{\pi}{10}n\right), N = 96 \cdot 2 - 1, L = 48,$$



Сравнение SSA и GSSA. Пример, продолжение

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

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

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

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

Модификации метода SSA

└ Сравнение SSA и GSSA. Пример, продолжение

Таблица 1: МБ Естественный $X = X_{\text{obs}} + X_{\text{sim}}$

Метод/Параметр	X_{obs}	X_{sim}	X
SSA	5.33×10^{-1}	5.33×10^{-1}	1.07×10^{-1}
GSSA, $\alpha = 0.5$	1.41×10^{-14}	1.41×10^{-14}	1.33×10^{-1}

Таблица 2: МБ Естественный $X = X_{\text{obs}} + X_{\text{sim}} + \varepsilon_{\text{sim}}$, $\varepsilon_{\text{sim}} \sim N(0, 0.1^2)$

Метод	X_{obs}	X_{sim}	X
SSA	5.33×10^{-1}	5.46×10^{-1}	1.08×10^{-1}
GSSA, $\alpha = 0.5$	1.11×10^{-13}	1.23×10^{-13}	1.14×10^{-1}

TODO Подтвердить этими таблицами слова из предыдущего слайда про шум и разделимость компонент.

TODO Применение **SSA** для выделения сигнала, затем **GSSA** для разделения компонент для того же примера. Написать, что получилось объединить лучшие свойства двух методов. Оформить это как вывод.

Метод CiSSA. Алгоритм: разложение

Как и в **SSA** считается \mathbf{X} , по которой строится $\hat{\mathbf{C}}_L$:

$$\hat{c}_m = \frac{L-m}{L}\hat{\gamma}_m + \frac{m}{L}\hat{\gamma}_{L-m}, \quad \hat{\gamma}_m = \frac{1}{N-m} \sum_{t=1}^{N-m} x_t x_{t+m}, \quad m = 0 : L-1.$$

$$\hat{\mathbf{C}}_L = \begin{pmatrix} \hat{c}_1 & \hat{c}_2 & \dots & \hat{c}_L \\ \hat{c}_2 & \hat{c}_1 & \dots & \hat{c}_{L-1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{c}_L & \hat{c}_{L-1} & \dots & \hat{c}_1 \end{pmatrix}.$$

Собственные числа и вектора матрицы $\hat{\mathbf{C}}_L$, задаются по формулам:

$$U_k = L^{-1/2}(u_{k,1}, \dots, u_{k,L}), \quad \text{где } u_{k,j} = \exp\left(-i2\pi d(j-1)\frac{k-1}{L}\right),$$

$$\lambda_{L,k} = \sum_{m=0}^{L-1} \hat{c}_m \exp\left(i2\pi m\frac{k-1}{L}\right), \quad k = 1 : L.$$

Модификации метода SSA

└ Метод CiSSA. Алгоритм: разложение

Метод CiSSA. Алгоритм: разложение

Как и в SSA считается X , по которой строится \tilde{C}_L :

$$\tilde{c}_{mn} = \frac{k-m}{L} \tilde{c}_{mn} + \frac{m}{L} \tilde{c}_{L-m, n}, \quad \tilde{c}_{mn} = \frac{1}{N-m} \sum_{k=0}^{N-m} x_k x_{k+m}, \quad m=0:L-1,$$

$$C_L = \begin{pmatrix} \tilde{c}_1 & \tilde{c}_2 & \dots & \tilde{c}_L \\ \tilde{c}_2 & \tilde{c}_3 & \dots & \tilde{c}_{L-1} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{c}_L & \tilde{c}_{L-1} & \dots & \tilde{c}_1 \end{pmatrix}.$$

Собственные числа и векторы матрицы C_L заданы по формулам:

$$U_k = L^{-1/2} (u_{k,1}, \dots, u_{k,L}), \quad \text{где } u_{k,j} = \exp \left(-i 2\pi d(j-1) \frac{k-1}{L} \right),$$

$$\lambda_{k,k} = \sum_{m=0}^{L-1} \tilde{c}_{mm} \exp \left(i 2\pi m \frac{k-1}{L} \right), \quad k=1:L.$$

TODO Полностью переписать алгоритм по аналогии того, как описан **SSA**. Написать, что группировка производится по частотам, указать, что из-за этого можно работать с **CiSSA** только тогда, когда заранее знаем частоты. Указать про фиксированный базис **CiSSA**

Метод CiSSA. Свойства: нестационарный ряд

Для использования на нестационарных временных рядах, нужно выполнить расширения ряда (экстраполировать) [1].

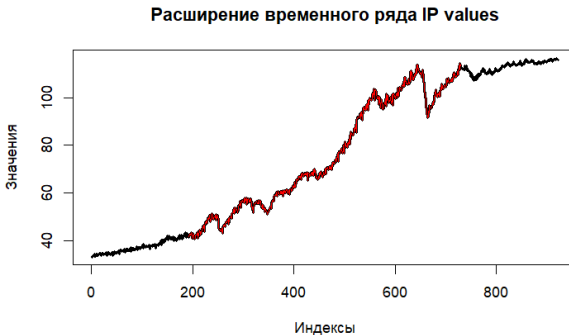


Рис. 1: Красный — настоящий ряд, черный — расширенный

Так, алгоритм лучше выделяет нелинейную составляющую.

Модификации метода SSA

└ Метод CiSSA. Свойства: нестационарный ряд

TODO Написать алгоритм расширения ряда, сказать, что изначально алгоритм пригоден только для стационарных рядов.

Метод CiSSA. Свойства: нестационарный ряд

Для использования метода CiSSA на нестационарных временных рядах, нужно выполнять расширение ряда [источник: 14].

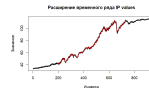


Рис. 2. Красный — исходный ряд, черный — расширенный

Так, алгоритм лучше выделяет иле в 2-х компонентах

Определение 2

Разложение

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

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

Замечание 2

$U_k U_k^H + U_{L+2-k} U_{L+2-k}^H$ является оператором проектирования на подпространство, которое порождено синусами и косинусами с частотой $w_k = \frac{k-1}{L}$. То есть, воспроизводится разложение Фурье для K векторов матрицы X . Затем вычисляется диагональное усреднение.

Модификации метода SSA

└ Метод CiSSA. Свойства: связь с разложением Фурье

TODO Переписать замечание

Метод CiSSA. Свойства: связь с разложением Фурье

Отметим, что

Возьмем

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

где $1 \leq n \leq N$ и $a_k, c_k = 0$ для четного N , и мы имеем разложение Фурье ряда X .

Замечание 1

$U_k U_{-k}^{H/2} + U_{L-k+2-k} U_{L-k+2-k}^{H/2}$ является оператором проектирования на подпространство, которое порождено синусом и косинусом с частотой $\omega_k = \frac{2\pi k}{N}$. Таким образом, разложение Фурье для X эквивалентно разложению X . За тем является дилеммой линейного ускорения.

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

Определение 3

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

Определение 4

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

Фиксируем временной ряд $X = X_1 + X_2 =$
 $= A_1 \cos(2\pi w_1 n + \varphi_1) + A_2 \cos(2\pi w_2 n + \varphi_2).$

Условия точной разделимости X для разложения Фурье:
 $Nw_1, Nw_2 \in \mathbb{N}, w_1 \neq w_2.$

Условия точной разделимости X для **CiSSA**:
 $Lw_1, Lw_2 \in \mathbb{N}, w_1 \neq w_2.$

Условия точной разделимости X для **SSA**:
 $Lw_1, Lw_2, Kw_1, Kw_2 \in \mathbb{N}, w_1 \neq w_2, A_1 \neq A_2.$

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 [3].

Замечание 3

*Для **SSA** существуют алгоритмы улучшения разделимости, например, EOSSA и FOSSA [2]. 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, учитывая соседние частоты.

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, а не полиномы [1].

Заметки [1]

Для SSA существуют алгоритмы улучшения разделимости на примере, EOSSA и FOSSA [2]. По заданному набору коэффициентов они позволяют более точно выделять компоненты.

В алгоритме разложения CiSSA [Фурье] увеличивая длину окна L [N] за мысленную частоту. Это означает, что даже если не удастся подобрать та же L [N], все равно какое-то будет точно отделять, что постепенно увеличивая шаг можно приближать частоты к частоте компонентов. В итоге, можно избежать ошибки выделения нужной компоненты, учитывая все другие частоты.

TODO Переформулировать с меньшим количеством слов, ссылки переработать

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	+	→	→	→	→	—
SSA EOSSA	+	→	→	→	→	+
CiSSA	+	+	→	—	—	+
CiSSA extended	+	+	→	→	—	+

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	+	→	—	—	+
Fourier extended	+	→	→	—	+

Таблица 4: Преимущества и недостатки методов Fourier

По полученным результатам, можно следующие выводы:

- ❶ 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;
- ❷ Если данные зашумлены или имеется непериодичность, алгоритм **SSA**, чем **CiSSA** с расширением ряда или без.

Модификации метода SSA

└ Сравнение SSA, Фурье, CiSSA. Выводы 2

TODO Переписать выводы в соответствии с примерами.

По полученным результатам, можно следующие выводы:

- Если известно, что ряд содержит только неперемежаемые элементы, следует использовать CiSSA без процедуры расширения, поскольку она делает ошибку в разделении неперемежаемых элементов. И наоборот, если есть неперемежаемость, лучше расширить ряд;
- Если данные зашумлены или имеются неперемежаемые элементы SSA, чем CiSSA с расширенным рядом или без.

- [1] Juan Bogalo, Pilar Poncela, and Eva Senra. Circulant singular spectrum analysis: A new automated procedure for signal extraction. *Signal Processing*, 177, 2020.
- [2] Nina Golyandina, Pavel Dudnik, and Alex Shlemov. Intelligent identification of trend components in singular spectrum analysis. *Algorithms*, 16(7):353, 2023.
- [3] Nina Golyandina, Vladimir Nekrutkin, and Anatoly Zhigljavsky. *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman and Hall/CRC, 2001.