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

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

- $\mathsf{X}_{\text{Trend}}$ – тренд, медленно меняющаяся компонента;
- $\mathsf{X}_{\text{Periodics}}$ – сумма периодических компонент;
- $\mathsf{X}_{\text{Noise}}$ – 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.

По чему будем сравнивать?

TODO Оформить слайд

Нужно определиться с тем, по каким критериям будем сравнивать методы?

- Постановка задачи (для CiSSA она другая, решаем только с заранее заданными частотами)
- Выделение сигнала
- Разделимость

Разделимость

Определение 1

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 2

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.

Метод SSA. Алгоритм

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

1. **Построение траекторной матрицы:**
   $\mathbf{X} = \mathcal{H}(\mathsf{X}) = [\mathsf{X}_1 : \ldots : \mathsf{X}_K]$, $\mathsf{X}_i = (x_i, \ldots, x_{i+L-1})^T$,
   $1 \le i \le K, \quad K = N - L + 1.$
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:** $\tilde{\mathsf{X}}_{I_k} = \mathcal{H}^{-1}(\mathbf{X}_{I_k})$, $\mathsf{X} = \tilde{\mathsf{X}}_{I_1} + \cdots + \tilde{\mathsf{X}}_{I_m}$.

Вложенные варианты EOSSA, FOSSA

TODO Указать, что такое вложенный вариант, сказать, что таким образом улучшается разделимость. Так получается лучше разделять компоненты между собой

Замечание 1

*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.*

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 2

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

Определение 3

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

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

где $r_1, r_2 \in \mathbb{N}$ *— 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:

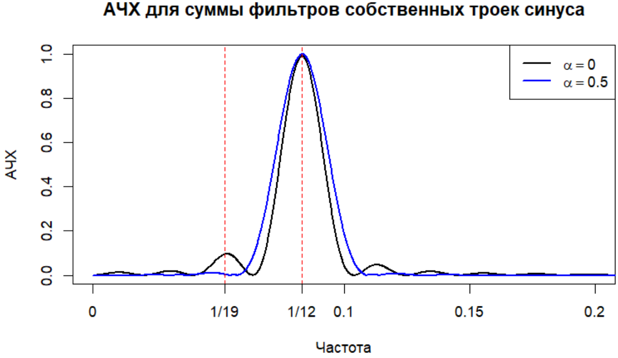
$$\widetilde{x}_s = \sum_{j=-(L-1)}^{L-1} \left(\sum_{k=1}^{L-|j|} u_k u_{k+|j|} / L \right) x_{s-j}, \quad L \le s \le K.$$

Аналогичное представление для GSSA:

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

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 $\mathsf{X} = \mathsf{X}_{\sin} + \mathsf{X}_{\cos}$

Метод/Ошибка	$\mathsf{X}_{\sin}$	$\mathsf{X}_{\cos}$	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 $\mathsf{X} = \mathsf{X}_{\sin} + \mathsf{X}_{\cos} + \varepsilon_n$, $\varepsilon_n \sim \mathrm{N}(0, 0.1^2)$

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

Получились наилучшие результаты.
TODO Дописать, что получился вложенный вариант алгоритма.

Метод CiSSA. Алгоритм

$X = (x_1, \ldots, x_N)$ — 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: $\mathbf{X} = \sum_{k=1}^{d} \mathbf{X}_{B_k}$, $d = \lfloor \frac{L+1}{2} \rfloor$ (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**.

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: $\mathbf{X} = \sum_{k=1}^{d} \mathbf{X}_{B_k}$, $d = \lfloor \frac{L+1}{2} \rfloor$ (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**.

TODO переписать третий шаг

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 4

Разложение

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

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

Замечание 3

*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 4

Разложение

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

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

Замечание 3

*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.*

TODO Переписать замечание (частично сделано)

Метод CiSSA. Свойства: нестационарный ряд

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

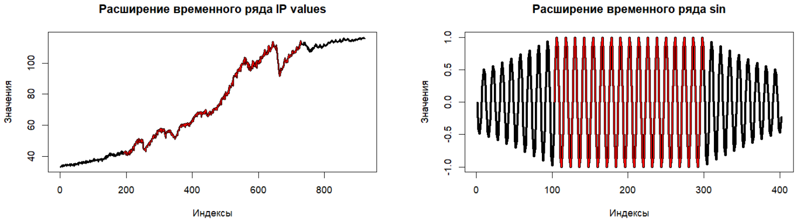
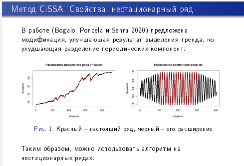


Рис. 1: Красный – настоящий ряд, черный – его расширение

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.

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:
$Nw_1, Nw_2 \in \mathbb{N}$, $w_1 \neq w_2$.
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**.

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

Метод	sin_err	cos_err	all_err
SSA, Lw, Kw in N	6.8e-30	1.5e-29	1.8e-29
SSA EOSSA, Lw, Kw in N	8.2e-30	6.5e-30	5.5e-30
Fourier, Nw in N	3.4e-28	9.8e-29	4.0e-28
Fourier extended, Nw in N	8.4e-07	1.4e-06	5.9e-07
CiSSA, Lw in N	1.1e-29	6.5e-30	7.8e-30
CiSSA extended, Lw in N	1.6e-06	5.5e-06	3.7e-06
SSA, Lw, Kw in N a1= a2	3.8e-04	3.8e-04	6.0e-29
SSA EOSSA, Lw, Kw in N, a1 = a2	1.4e-29	2.9e-29	1.1e-29

Таблица 3: a1 * sin + a2 * cos

Сравнение SSA, Фурье, CiSSA. Асимптотическая разделимость

Асимптотически разделимы в методе **SSA** полиномы, гармонические функции (Golyandina, Nekrutkin и Zhigljavsky 2001).
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.

Сравнение SSA, Фурье, CiSSA. Асимптотическая разделимость

Асимптотически разделимы в методе **SSA** полиномы, гармонические функции (Golyandina, Nekrutkin и Zhigljavsky 2001).
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.

TODO Переформулировать с меньшим количеством слов

TODO Пример, когда условия нарушаются (добавление тренда).

Метод	exp_err	sin_err	cos_err	all_err
SSA, Lw, Kw in N	6.1e-05	5.2e-05	8.9e-07	2.9e-28
SSA EOSSA, Lw, Kw in N	3.3e-28	7.3e-30	2.0e-29	3.2e-28
Fourier, Nw in N	1.1e-01	2.1e-02	1.7e-02	7.8e-02
Fourier extended, Nw in N	1.4e-03	1.8e-03	2.2e-04	1.0e-03
CiSSA, Lw in N	5.3e-02	4.9e-04	3.3e-03	2.8e-02
CiSSA extended, Lw in N	5.0e-04	1.1e-03	6.9e-05	4.3e-04
SSA, Lw, Kw in N, a1 = a2	6.1e-05	1.0e-02	9.9e-03	4.4e-28
SSA EOSSA, Lw, Kw in N, a1 = a2	5.1e-28	1.6e-29	6.6e-29	4.9e-28

Таблица 4: 1 + exp(x/100) + a1 * sin + a2 * cos

TODO Пример, когда шум.

Метод	sin_err	cos_err	all_err
SSA, Lw, Kw in N	2.4e-04	3.0e-04	5.4e-04
SSA EOSSA, Lw, Kw in N	2.4e-04	3.0e-04	5.4e-04
Fourier, Nw in N	1.4e-04	2.1e-04	3.5e-04
Fourier extended, Nw in N	3.1e-03	1.3e-03	4.4e-03
CiSSA, Lw in N	1.5e-04	2.7e-04	4.1e-04
CiSSA extended, Lw in N	1.9e-03	3.6e-04	2.3e-03

Таблица 5: a1 * sin + a2 * cos + noise

Метод	exp_err	sin_err	cos_err	all_err
SSA, Lw, Kw in N	2.9e-04	2.4e-04	4.3e-04	7.1e-04
SSA EOSSA, Lw, Kw in N	9.7e-04	1.7e-04	4.2e-04	1.6e-03
Fourier, Nw in N	1.1e-01	5.2e-03	1.2e-04	1.0e-01
Fourier extended, Nw in N	7.1e-03	9.5e-03	2.0e-03	7.5e-03
CiSSA, Lw in N	5.5e-02	5.6e-04	3.0e-04	4.8e-02
CiSSA extended, Lw in N	2.8e-03	1.5e-03	2.8e-04	2.5e-03

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

Таблица 7: Преимущества и недостатки методов **SSA**, **CiSSA**

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

Таблица 8: Преимущества и недостатки методов Fourier

Сравнение SSA, Фурье, CiSSA. Выводы 2

По полученным результатам, можно следующие выводы:

1. **CiSSA** и разложение Фурье работает лучше базового **SSA** только в том случае, когда периодики имеют одинаковые амплитуды. При этом, **SSA** с EOSSA исправляет этот недостаток. Во всех остальных случаях **CiSSA** и разложение Фурье показывают результаты не лучше **SSA**.
2. Если и использовать алгоритм **CiSSA**, то как вложенный вариант, предварительно выделив трендовую составляющую, шум, и внутри разделять компоненты.

Сравнение SSA, Фурье, CiSSA. Выводы 2

По полученным результатам, можно следующие выводы:

1. **CiSSA** и разложение Фурье работает лучше базового **SSA** только в том случае, когда периодики имеют одинаковые амплитуды. При этом, **SSA** с EOSSA исправляет этот недостаток. Во всех остальных случаях **CiSSA** и разложение Фурье показывают результаты не лучше **SSA**.
2. Если и использовать алгоритм **CiSSA**, то как вложенный вариант, предварительно выделив трендовую составляющую, шум, и внутри разделять компоненты.

TODO Переписать выводы в соответствии с примерами.

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.