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

Погребников Николай Вадимович, 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

Временные ряды представляют собой последовательность данных, собранных или измеренных в хронологическом порядке. Понимание эволюции явлений во времени является критическим для выявления тенденций, циклов и аномалий. В этих целях были созданы методы разложения временных рядов на сумму интерпретируемых компонент такие как **SSA** [3] и его модификации: **GSSA** [4], **CiSSA** [1].

Целью работы является описание модификаций в контексте теории **SSA** и на этой основе сравнение методов по теоретическим свойствам и численно.

Метод SSA. Алгоритм: разложение

Для временного ряда $\mathsf{X} = (x_1, \ldots, x_N)$ выбирается длина окна L, $1 < L < N$ и определяется $K = N - L + 1$. Строится L-траекторная матрица $\mathbf{X}$, состоящая из столбцов вида $\mathsf{X}_i = (x_{i-1}, \ldots, x_{i+L-2})^{\mathrm{T}}$, $1 \leq i \leq K$.
Пусть $\mathbf{S} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$, $\lambda_1, \ldots, \lambda_L$ — собственные числа матрицы $\mathbf{S}$, взятые в неубывающем порядке.

Определение 1

Сингулярным разложением называется представление матрицы в виде:

$$\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_d = \sum_{i=1}^{d} \sqrt{\lambda_i} U_i V_i^{\mathrm{T}}, \text{ где} \tag{1}$$

$U_1, \ldots, U_L$ *— ортонормированная система собственных векторов матрицы* $\mathbf{S}$, $d = \max\{i : \lambda_i > 0\}$ *и* $V_i = \mathbf{X}^{\mathrm{T}} U_i / \sqrt{\lambda_i}$.

Набор $(\sqrt{\lambda_i}, U_i, V_i^{\mathrm{T}})$ называется i-й собственной тройкой.

На основе разложения (1) производится процедура группировки, которая делит все множество индексов $\{1, \ldots, d\}$ на m непересекающихся подмножеств $I_1, \ldots, I_d$.
Пусть $I = \{i_1, \ldots, i_p\}$, тогда $\mathbf{X}_I = \mathbf{X}_{i_1} + \cdots + \mathbf{X}_{i_p}$. Такие матрицы вычисляются для каждого $I = I_1, \ldots, I_m$.

В результате получаются матрицы $\mathbf{X}_{I_1}, \ldots, \mathbf{X}_{I_m}$, для каждой из которых проводится операция диагонального усреднения, составляющая ряды длины N: $\mathsf{X}_1, \ldots, \mathsf{X}_m$.
При этом, $\mathsf{X}_1 + \cdots + \mathsf{X}_m = \mathsf{X}$.

Метод GSSA. Алгоритм

Алгоритм **GSSA** сильно схож с базовым **SSA**. Пусть $N > 2$, вещественнозначный временной ряд $\mathsf{X} = (x_1, \ldots, x_N)$ длины N. Фиксируется параметр $\alpha \geq 0$, отвечающий за веса:

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

Для временного ряда $\mathsf{X} = (x_1, \ldots, x_N)$ выбирается длина окна L, $1 < L < N$ и определяется $K = N - L + 1$. Строится L-траекторная матрица $\mathbf{X}^{(\alpha)}$, состоящая из столбцов вида $\mathsf{X}_i^{(\alpha)} = (w_1 x_{i-1}, \ldots, w_L x_{i+L-2})^{\mathrm{T}}$, $1 \leq i \leq K$.

Остальные действия те же самые, что и в **SSA**.

Замечание 1

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

Сравнение SSA и GSSA. Линейные фильтры 1

Определение 2

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

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

Связанные определения:

- h_i — импульсная характеристика фильтра;
- $H_\Phi(z) = \sum_{i=-r_1}^{r_2} h_i z^{-i}$ — передаточная функция;
- $A_\Phi(\omega) = \left|H_\Phi\left(e^{i2\pi\omega}\right)\right|$ — АЧХ;
- $\phi_\Phi(\omega) = \mathrm{Arg}\left(H_\Phi\left(e^{i2\pi\omega}\right)\right)$ — ФЧХ.

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

Сравнение SSA и GSSA. Линейные фильтры 2

Пусть $\mathsf{X} = (x_1, \ldots, x_N)$ — временной ряд длины N, $(\sqrt{\lambda}, U, V)$ — одна из собственных троек разложения методом **SSA**.
$U = (u_1, \ldots, u_L)$.
Тогда компонента временного ряда $\widetilde{\mathsf{X}}$, восстановленная с использованием собственной тройки $(\sqrt{\lambda}, U, V)$, для средних точек (индексы от L до K) имеет вид:

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

Таким образом, имеется представление алгоритма **SSA** через линейные фильтры.
Аналогичное представления для **GSSA**:

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

Сравнение SSA и GSSA. Пример

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

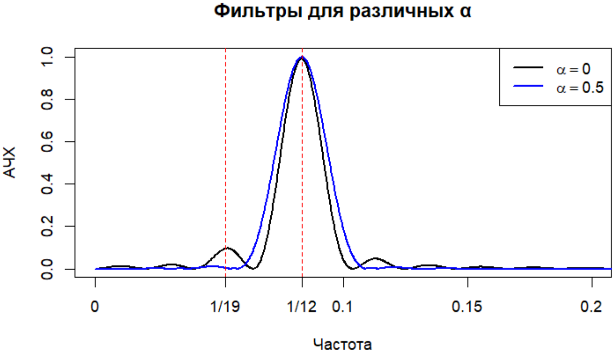

Рис. 1: Ряд $\mathsf{X} = \mathsf{X}_{\sin} + \mathsf{X}_{\cos}$. АЧХ фильтров, отвечающих за $\mathsf{X}_{\sin} = \sin\left(\frac{2\pi}{12}n\right)$, при разных α

Сравнение SSA и GSSA. Пример, продолжение

Метод/Ошибка	$X_{\sin}$	$X_{\cos}$	X
SSA	5.15e-03	5.15e-03	6.01e-30
GSSA, $\alpha = \frac{1}{2}$	3.68e-04	3.68e-04	9.53e-30

Таблица 1: MSE разложений ряда $X = X_{\sin} + X_{\cos}$ для **SSA** и **GSSA** с $\alpha = \frac{1}{2}$

Добавим к X шумовую компоненту:
$X = X_{\sin} + X_{\cos} + X_{\text{noise}} = \sin\left(\frac{2\pi}{12}x\right) + \frac{1}{2}\cos\left(\frac{2\pi}{19}x\right) + \varepsilon_n$, где $\varepsilon_n \sim \mathrm{N}(0, 0.1^2)$.

Метод	$X_{\sin}$	$X_{\cos}$	X
SSA	5.68e-03	5.44e-03	7.48e-04
GSSA, $\alpha = \frac{1}{2}$	1.21e-03	1.25e-03	1.04e-03

Таблица 2: MSE разложений ряда $X = X_{\sin} + X_{\cos} + X_{\text{noise}}$ для **SSA** и **GSSA** с $\alpha = \frac{1}{2}$

Сравнение SSA и GSSA. Выводы

По теоретическим результатам и примерам можно сделать понять, что **GSSA** позволяет улучшить разделимость периодических компонент ряда. Однако, вместе с тем, разложение будет захватывать больше шума в сравнении с базовым **SSA**.

Метод CiSSA. Алгоритм: разложение

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

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

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

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

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

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

Метод CiSSA. Алгоритм: восстановление

Для каждой частоты $w_k = \frac{k-1}{L}$, $k = 2 : \lfloor \frac{L+1}{2} \rfloor$, есть два собственных вектора: U_k и U_{L+2-k}. За частоту w_1 отвечает один собственный вектор — U_1. Если L — четное, то частоте $w_{\frac{L}{2}+1}$ будет соответствовать один вектор $U_{\frac{L}{2}+1}$.

Следовательно, индексы разбиваются на элементарную группировку следующим образом:

$$B_1 = \{1\};\ B_k = \{k, L+2-k\},\ \text{для } k = 2 : \lfloor \frac{L+1}{2} \rfloor;$$

$$B_{\frac{L}{2}+1} = \left\{ \frac{L}{2} + 1 \right\},\ \text{если } L \mid 2.$$

$\mathbf{X}_{B_k} = \mathbf{X}_k + \mathbf{X}_{L+2-k} = U_k U_k^H \mathbf{X} + U_{L+2-k} U_{L+2-k}^H \mathbf{X}$,
где U^H — это комплексное сопряжение и транспонирование вектора U. Далее идет группировка по диапазонам интересующих частот, после чего следует диагональное усреднение.

Метод CiSSA. Свойства: связь с разложением Фурье

Определение 3

Разложение

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

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.*

Метод CiSSA. Свойства: нестационарный ряд

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

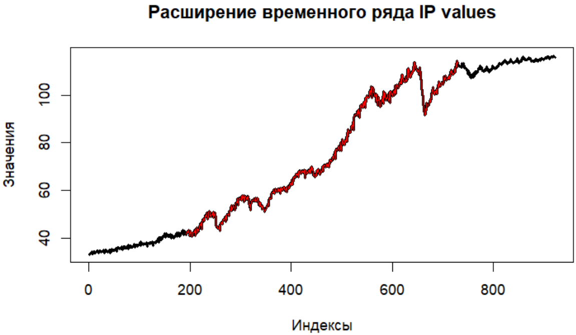

Рис. 2: Красный — настоящий ряд, черный — расширеннный

Так, алгоритм лучше выделяет нелинейную составляющую.

Сравнение SSA, Фурье, CiSSA. Разделимость

Определение 4

Есть метод разделения ряда на компоненты с параметрами Θ, *ряд* $\mathsf{X} = \mathsf{X}^{(1)} + \mathsf{X}^{(2)}$. $\exists$ *набор параметров* $\hat{\Theta}$, 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 5

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 $\mathsf{X} = \mathsf{X}_1 + \mathsf{X}_2 =$
$= A_1 \exp(\alpha_1 n) \cos(2\pi w_1 n + \varphi_1) + A_2 \exp(\alpha_2 n) \cos(2\pi w_2 n + \varphi_2)$.

Условия точной разделимости X для разложения Фурье:
$N w_1, N w_2 \in \mathbb{N}$, $w_1 \neq w_2$, $\alpha_1 = \alpha_2 = 0$.
Условия точной разделимости X для **CiSSA**:
$L w_1, L w_2 \in \mathbb{N}$, $w_1 \neq w_2$, $\alpha_1 = \alpha_2 = 0$.
Условия точной разделимости X для **SSA**:
$L w_1, L w_2, K w_1, K w_2 \in \mathbb{N}$, $w_1 \neq w_2$, $A_1 \neq A_2$, $\alpha_1 \neq \alpha_2$.

Таким образом, условия на разделение косинусов, слабее у методов **CiSSA** и Фурье, чем у **SSA**. Однако **SSA** может точно отличать друг от друга больше классов функций.

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.

Сравнение SSA, Фурье, CiSSA. Выделение тренда

Любая непериодическая компонента будет отвечать частотам, близким к нулю. Из-за этого алгоритмы **CiSSA** и разложение Фурье не смогут отличить друг от друга две непериодики.

Пример. Рассмотрим ряд
$\mathsf{X} = \mathsf{X}_c + \mathsf{X}_e + \mathsf{X}_{\sin} + \mathsf{X}_{\cos} = 1 + e^{\frac{x}{100}} + \sin\frac{2\pi}{12}x + \frac{1}{2}\cos\frac{2\pi}{3}x.$

Метод	Параметры	$\mathrm{MSE}(\mathsf{X}_c + \mathsf{X}_e)$	$\mathrm{MSE}(\mathsf{X}_{\sin})$	$\mathrm{MSE}(\mathsf{X}_{\cos})$	$\mathrm{MSE}(\mathsf{X})$
SSA	$L = 96, K = 96$	5.0e-03	8.9e-07	5.2e-05	4.4e-03
SSA EOSSA, $r = 7$	$L = 96, K = 96$	1.7e-28	1.6e-29	8.7e-30	1.6e-28
Fourier	$N = 96 \cdot 2$	1.1e-01	6.1e-04	6.8e-03	1.1e-01
Fourier extended	$N = 96 \cdot 2$	1.4e-03	1.3e-03	8.4e-03	9.6e-03
CiSSA	$L = 96$	5.3e-02	1.6e-05	4.9e-04	4.4e-02
CiSSA extended	$L = 96$	5.0e-04	2.1e-04	1.1e-03	6.0e-04

Таблица 3: MSE разложений ряда $\mathsf{X} = \mathsf{X}_c + \mathsf{X}_e + \mathsf{X}_{\sin} + \mathsf{X}_{\cos}$

По таблице 3 видно, что расширение ряда негативно повлияло на выделение периодики и положительно на трендовую составляющую (непериодику).

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

Таблица 4: 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 5: Преимущества и недостатки методов Fourier

Сравнение SSA, Фурье, CiSSA. Выводы 2

По полученным результатам, можно следующие выводы:

1. Алгоритм **CiSSA** работает лучше разложения Фурье;
2. 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;
3. Если данные зашумлены или имеется непериодичность, алгоритм **SSA** с улучшением разделимости справляется в среднеквадратичном лучше **CiSSA** с расширением ряда или без.

Список литературы

[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.

[4] Jialiang Gu, Kevin Hung, Bingo Wing-Kuen Ling, Daniel Hung-Kay Chow, Yang Zhou, Yaru Fu, and Sio Hang Pun. Generalized singular spectrum analysis for the decomposition and analysis of non-stationary signals. *Journal of the Franklin Institute*, Accepted/In Press, 2024.