

# Модификации метода анализа сингулярного спектра для анализа временных рядов: Circulant SSA

Погребников Николай Вадимович, гр. 21.Б04-мм

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

3 курс (бак.) «Производственная практика  
(научно-исследовательская работа)»  
(Семестр 6)

Санкт-Петербург, 2024

Модификации метода **SSA**

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

Погребников Николай Вадимович, гр. 21.Б04-мм  
Санкт-Петербургский государственный университет  
Прикладная математика и информатика  
Вычислительная стохастика и статистические модели

3 курс (бак.) «Производственная практика  
(научно-исследовательская работа)»  
(Семестр 6)  
Санкт-Петербург, 2024

Научный руководитель д. ф.-м. н., доц. Голяндина Нина Эдуардовна,  
кафедра статистического моделирования

Перед началом исследования были поставлены следующие цели:

- 1 Ознакомиться с алгоритмом **CiSSA**;
- 2 Реализовать алгоритм **CiSSA** на языке R;
- 3 Сравнить алгоритмы **SSA**, разложение Фурье и **CiSSA**.

## Модификации метода **SSA**

### └ Введение

#### Введение

Перед началом исследования были поставлены следующие цели:

- Ознакомиться с алгоритмом **CiSSA**;
- Реализовать алгоритм **CiSSA** на языке R;
- Сравнить алгоритмы **SSA**, разложение Фурье и **CiSSA**.

Сингулярный спектральный анализ (**SSA** [3]) — метод, целью которого является разложение оригинального ряда на сумму небольшого числа интерпретируемых компонент, таких как медленно изменяющаяся тенденция (тренд), колебательные компоненты (сезонность) и “структурный” шум. В данном исследовании рассматривается математическая составляющая вариации алгоритма **SSA** — circulant singular spectrum analysis (**CiSSA**), предложенная в статье [1], а также сравнение базового метода и циркулярного, применение их на языке R.

# Метод SSA. Алгоритм: разложение

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

## Определение 1

Сингулярным разложением называется представление матрицы в виде:

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_d = \sum_{i=1}^d \sqrt{\lambda_i} U_i V_i^T, \text{ где} \quad (1)$$

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

## Модификации метода SSA

└ Метод SSA. Алгоритм: разложение

### Метод SSA. Алгоритм: разложение

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

#### Определение 1

Сингулярным разложением называется представление матрицы в виде:

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_d = \sum_{i=1}^d \sqrt{\lambda_i} U_i V_i^T, \text{ где} \quad (1)$$

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

Полезным свойством является то, что матрица  $\mathbf{X}$  имеет одинаковые элементы на антидиагоналях. Таким образом,  $L$ -траекторная матрица является ганкелевой.

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

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

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

## Модификации метода SSA

### └ Метод SSA. Алгоритм: восстановление

#### Метод SSA. Алгоритм: восстановление

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

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

Диагональное усреднение для каждой антидиагонали усредняет значения элементов матрицы.

Применяя данную операцию к матрицам  $\mathbf{X}_{I_1}, \dots, \mathbf{X}_{I_m}$ , получаются  $m$  новых рядов:  $\mathbf{X}_1, \dots, \mathbf{X}_m$ . При этом,  $\mathbf{X}_1 + \dots + \mathbf{X}_m = \mathbf{X}$ .

# Метод SSA. Свойства: точная разделимость

Пусть временной ряд  $X = X^{(1)} + X^{(2)}$  и задачей является нахождение этих слагаемых.

Будем говорить, что ряд  $X$  точно разделим на  $X^{(1)}$  и  $X^{(2)}$ , если существует такое сингулярное разложение траекторной матрицы  $X$  ряда  $X$ , что его можно разбить на две части, являющиеся сингулярными разложениями траекторных матриц рядов  $X^{(1)}, X^{(2)}$  [3].

Таблица 1: Точная разделимость

	const	cos	exp	exp cos	ak+b
const	-	+	-	-	-
cos	+	+	-	-	-
exp	-	-	-	+	-
exp cos	-	-	+	+	-
ak+b	-	-	-	-	-

5/25Погребников Николай Вадимович, гр. 21.Б04-мм

Модификации метода SSA

## Модификации метода SSA

### └ Метод SSA. Свойства: точная разделимость

#### Метод SSA. Свойства: точная разделимость

Пусть временной ряд  $X = X^{(1)} + X^{(2)}$  и задачей является нахождение этих слагаемых.  
Будем говорить, что ряд  $X$  точно разделим на  $X^{(1)}$  и  $X^{(2)}$ , если существует такое сингулярное разложение траекторной матрицы  $X$  ряда  $X$ , что его можно разбить на две части, являющиеся сингулярными разложениями траекторных матриц рядов  $X^{(1)}, X^{(2)}$  [3].

Таблица 1: Точная разделимость

	const	cos	exp	exp cos	ak+b
const	-	+	-	-	-
cos	+	+	-	-	-
exp	-	-	-	+	-
exp cos	-	-	+	+	-
ak+b	-	-	-	-	-

Условия точной разделимости выводятся из понятий слабо L-разделимых рядов и сильно L-разделимых рядов [3]. Стоит отметить, что точная разделимость для cos достигается, если  $Lw \in \mathbb{N}$ ,  $Kw \in \mathbb{N}$ , где  $w$  — частота.

Однако, по таблице 1 видно, что условия точной разделимости достаточно жесткие и вряд ли выполнимы в реальных задачах. Тогда появляется такое понятие, как асимптотическая разделимость.

# Метод SSA. Свойства: асимптотическая разделимость

$$\rho_{i,j}^{(M)} = \frac{\left( X_{i,i+M-1}^{(1)}, X_{j,j+M-1}^{(2)} \right)}{\left\| X_{i,i+M-1}^{(1)} \right\| \left\| X_{j,j+M-1}^{(2)} \right\|}.$$

## Определение 2

Ряды  $X^{(1)}, X^{(2)}$  называются  $\varepsilon$ -разделимыми при длине окна  $L$ , если

$$\rho^{(L,K)} \stackrel{\text{def}}{=} \max \left( \max_{1 \leq i,j \leq K} |\rho_{i,j}^{(L)}|, \max_{1 \leq i,j \leq L} |\rho_{i,j}^{(K)}| \right) < \varepsilon \text{ [3]}.$$

## Определение 3

Если  $\rho^{(L(N),K(N))} \rightarrow 0$  при некоторой последовательности  $L = L(N), N \rightarrow \infty$ , то ряды  $X^{(1)}, X^{(2)}$  называются асимптотически  $L(N)$ -разделимыми [3].

## Модификации метода SSA

└ Метод SSA. Свойства: асимптотическая разделимость

Метод SSA. Свойства: асимптотическая разделимость

$$\rho_{i,j}^{(M)} = \frac{\left( X_{i,i+M-1}^{(1)}, X_{j,j+M-1}^{(2)} \right)}{\left\| X_{i,i+M-1}^{(1)} \right\| \left\| X_{j,j+M-1}^{(2)} \right\|}.$$

### Определение 2

Ряды  $X^{(1)}, X^{(2)}$  называются  $\varepsilon$ -разделимыми при длине окна  $L$ , если

$$\rho^{(L,K)} \stackrel{\text{def}}{=} \max \left( \max_{1 \leq i,j \leq K} |\rho_{i,j}^{(L)}|, \max_{1 \leq i,j \leq L} |\rho_{i,j}^{(K)}| \right) < \varepsilon \text{ [3]}.$$

### Определение 3

Если  $\rho^{(L(N),K(N))} \rightarrow 0$  при некоторой последовательности  $L = L(N), N \rightarrow \infty$ , то ряды  $X^{(1)}, X^{(2)}$  называются асимптотически  $L(N)$ -разделимыми [3].

Для любого ряда  $X$  длины  $N$  определим  $X_{i,j} = (x_{i-1}, \dots, x_{j-1}), 1 \leq i \leq j < N$ . Пусть  $X^{(1)} = (x_0^{(1)}, \dots, x_{N-1}^{(1)}), X^{(2)} = (x_0^{(2)}, \dots, x_{N-1}^{(2)})$ . Тогда определим коэффициент корреляции.

Таблица 2: Асимптотическая разделимость

	const	cos	exp	exp cos	ak+b
const	-	+	+	+	-
cos	+	+	+	+	+
exp	+	+	+	+	+
exp cos	+	+	+	+	+
ak+b	+	+	+	+	-

## Замечание 1

Для **SSA** существуют алгоритмы улучшения разделимости [2]. Они позволяют более точно отделять временные ряды друг от друга. В данной работе будут использоваться методы EOSSA и FOSSA.

## Модификации метода **SSA**

└ Метод SSA. Свойства: асимптотическая разделимость

Метод SSA. Свойства: асимптотическая разделимость

	const	cos	exp	exp cos	ak+b
const	-	+	+	+	-
cos	+	+	+	+	+
exp	+	+	+	+	+
exp cos	+	+	+	+	+
ak+b	+	+	+	+	-

### Замечание 1

Для **SSA** существуют алгоритмы улучшения разделимости [2]. Они позволяют более точно отделять временные ряды друг от друга. В данной работе будут использоваться методы EOSSA и FOSSA.

Как можно заметить по таблице 2, для гораздо большего класса функций асимптотическая разделимость имеет место [3].

## 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:

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

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

### Модификации метода **SSA**

└ Метод CiSSA. Алгоритм: разложение

#### Метод CiSSA. Алгоритм: разложение

Как и в **SSA** считается  $\mathbf{X}$ , 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:

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

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

Модификация **SSA** на основе циркулярной матрицы [1]. Авторы метода называют её автоматизированной. Причем автоматизированная в том смысле, что компоненты ряда группируются по частотам самим алгоритмом.



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  $w_0$  отвечает один собственный вектор —  $U_0$ . Если  $L$  — четное, то частоте  $w_{\frac{L}{2}+1}$  будет соответствовать один вектор  $U_{\frac{L}{2}+1}$ .

Следовательно, индексы группируются следующим образом:

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

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

Далее происходит группировка

$$\mathbf{X}_{B_k} = \mathbf{X}_k + \mathbf{X}_{L+2-k} = U_k U_k^H \mathbf{X} + U_{L+2-k} U_{L+2-k}^H \mathbf{X},$$

где  $U^H$  — это комплексное сопряжение и транспонирование вектора  $U$ . Затем диагональное усреднение.

## Модификации метода SSA

### └ Метод CiSSA. Алгоритм: восстановление

#### 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  $w_0$  отвечает один собственный вектор —  $U_0$ . Если  $L$  — четное, то частоте  $w_{\frac{L}{2}+1}$  будет соответствовать один вектор  $U_{\frac{L}{2}+1}$ . Следовательно, индексы группируются следующим образом:

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

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

Далее происходит группировка  $\mathbf{X}_{B_k} = \mathbf{X}_k + \mathbf{X}_{L+2-k} = U_k U_k^H \mathbf{X} + U_{L+2-k} U_{L+2-k}^H \mathbf{X}$ , где  $U^H$  — это комплексное сопряжение и транспонирование вектора  $U$ . Затем диагональное усреднение.

Группировка будет производиться на непересекающиеся подгруппы по частотам от 0 до 0.5, поскольку частоты выше 0.5 представляют собой зеркальное отражение частот ниже 0.5. Именно поэтому объединяются матрицы  $\mathbf{X}_{B_k} = \mathbf{X}_k + \mathbf{X}_{L+2-k}$ .

## Определение 4

Разложение

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

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  $w_k = \frac{k-1}{L}$ . Это пространство соответствует компонентам синусоидальной структуры временного ряда, связанных с конкретной частотой, выделяемой методом.

## Модификации метода SSA

└ Метод CiSSA. Свойства: связь с разложением Фурье

Метод CiSSA. Свойства: связь с разложением Фурье

Определение 4  
Разложение

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

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  $w_k = \frac{k-1}{L}$ . Это пространство соответствует компонентам синусоидальной структуры временного ряда, связанных с конкретной частотой, выделяемой методом.

По замечанию 2 видно, что при вычислении  $\mathbf{X}_{B_k} = \mathbf{X}_k + \mathbf{X}_{L+2-k} = U_k U_k^H \mathbf{X} + U_{L+2-k} U_{L+2-k}^H \mathbf{X}$ , воспроизводится разложение Фурье для  $K$  векторов матрицы  $X$ . Затем вычисляется диагональное усреднение  $*X_{B_k}$ .

**Точная разделимость.** Поскольку данный метод является аналогом разложения Фурье, то в смысле сильной разделимости можно точно разделить ряд, в котором одной из компонент является  $\cos(2\pi w + \varphi)$  с частотой  $w$  такой, что  $Lw = k \in \mathbb{N}$ , или константа.

**Асимптотическая разделимость.** Асимптотическая разделимость в данном случае будет означать, что при увеличении  $L$  разбиение сетки будет увеличиваться, а значит, и частоты в сетке начнут сближаться к истинным частотам периодических компонентов (либо становиться равными им), что будет снижать ошибку вычислений.

## Модификации метода SSA

### └ Метод CiSSA. Свойства: разделимость

#### Метод CiSSA. Свойства: разделимость

**Точная разделимость.** Поскольку данный метод является аналогом разложения Фурье, то в смысле сильной разделимости можно точно разделить ряд, в котором одной из компонент является  $\cos(2\pi w + \varphi)$  с частотой  $w$  такой, что  $Lw = k \in \mathbb{N}$ , или константа.

**Асимптотическая разделимость.** Асимптотическая разделимость в данном случае будет означать, что при увеличении  $L$  разбиение сетки будет увеличиваться, а значит, и частоты в сетке начнут сближаться к истинным частотам периодических компонентов (либо становиться равными им), что будет снижать ошибку вычислений.

Как можно заметить, классов точной разделимости меньше, чем у SSA

## Определение 5

Будем говорить, что методы  $M_1$  и  $M_2$  асимптотически эквивалентны, если их матрицы вложения  $S_1, S_2$  асимптотически эквивалентны в смысле  $\lim_{L \rightarrow \infty} \frac{\|S_1 - S_2\|_F}{\sqrt{L}} = 0$ , где  $\|\cdot\|_F$  — норма Фробениуса. Тогда  $M_1 \sim M_2, S_1 \sim S_2$ .

## Теорема 1

Дана  $L \times K$  траекторная матрица  $\mathbf{X}$ . Пусть  $S_B = \mathbf{X}\mathbf{X}^T/K$ ,  $S_C$  — матрица, определенная в (8). Тогда  $S_B \sim S_C$ .

## Доказательство.

Доказательство в источнике [1]. ☐

## Модификации метода SSA

└ Метод CiSSA. Свойства: эквивалентность методов

Метод CiSSA. Свойства: эквивалентность методов

### Определение 5

Будем говорить, что методы  $M_1$  и  $M_2$  асимптотически эквивалентны, если их матрицы вложения  $S_1, S_2$  асимптотически эквивалентны в смысле  $\lim_{L \rightarrow \infty} \frac{\|S_1 - S_2\|_F}{\sqrt{L}} = 0$ , где  $\|\cdot\|_F$  — норма Фробениуса. Тогда  $M_1 \sim M_2, S_1 \sim S_2$ .

### Теорема 1

Дана  $L \times K$  траекторная матрица  $\mathbf{X}$ . Пусть  $S_B = \mathbf{X}\mathbf{X}^T/K$ ,  $S_C$  — матрица, определенная в (8). Тогда  $S_B \sim S_C$ .

### Доказательство.

Доказательство в источнике [1]. ☐

В статье [1] говорится, что асимптотически методы **SSA** и **CiSSA** эквивалентны и в доказательство приводится теорема.

# Метод CiSSA. Свойства: применимость к нестационарным рядам

Алгоритм **CiSSA** применим для стационарных рядов. Однако, как утверждается авторами статьи [1], для использования на нестационарных временных рядах, нужно выполнить процедуру расширения ряда.

## Модификации метода **SSA**

### └ Метод CiSSA. Свойства: применимость к нестационарным рядам

Метод CiSSA. Свойства: применимость к нестационарным рядам

Алгоритм CiSSA применим для стационарных рядов. Однако, как утверждается авторами статьи [1], для использования на нестационарных временных рядах, нужно выполнить процедуру расширения ряда.

Формальное определение стационарности ряда можно увидеть в отчёте данной работы [4]. Стационарный ряд — это такой временной ряд, в котором изменения происходят вокруг некоторого среднего значения, и это среднее остаётся более-менее постоянным на протяжении всего ряда.

Сама процедура расширения ряда  $X$  производится с использованием авторегрессионной (AR) модели. Эта процедура позволяет предсказать значения временного ряда за его пределами (экстраполяция) как в правом, так и в левом направлениях на заданное число шагов  $H$ . Таким образом, трендовая (нелинейная) компонента ряда будет выделяться заметно лучше.

Будем сравнивать пять различных методов: базовый **SSA**, **SSA** с использованием EOSSA для улучшения разделимости, разложения Фурье, базового **CiSSA** и **CiSSA** с расширением ряда. Для наглядного отображения преимуществ каждого из этих методов составлена таблица 3

Метод/Условие	$\cos, Lw = k \in \mathbb{N}, Kw = k \in \mathbb{N}$	$\cos, Lw = k \in \mathbb{N}, Kw = k \notin \mathbb{N}$	$\cos, Lw = k \notin \mathbb{N}, Kw = k \notin \mathbb{N}$	$X_{np1}$	$X_{np}$	group
SSA	+	→	→	→	→	—
SSA EOSSA	+	→	→	→	→	+
Fourier	+	+	→	—	—	+
CiSSA	+	+	→	—	—	+
CiSSA extended	+	+	→	→	—	+

Таблица 3: Преимущества и недостатки пяти методов

## Модификации метода SSA

### — Сравнение алгоритмов. SSA, разложение Фурье, CiSSA

#### Сравнение алгоритмов. SSA, разложение Фурье, CiSSA

Будем сравнивать пять различных методов: базовый **SSA**, **SSA** с использованием EOSSA для улучшения разделимости, разложения Фурье, базового **CiSSA** и **CiSSA** с расширением ряда. Для наглядного отображения преимуществ каждого из этих методов составлена таблица 3

Метод/Условие	$\cos, Lw = k \in \mathbb{N}, Kw = k \in \mathbb{N}$	$\cos, Lw = k \in \mathbb{N}, Kw = k \notin \mathbb{N}$	$\cos, Lw = k \notin \mathbb{N}, Kw = k \notin \mathbb{N}$	$X_{np1}$	$X_{np}$	group
SSA	+	→	→	→	→	—
SSA EOSSA	+	→	→	→	→	+
Fourier	+	+	→	—	—	+
CiSSA	+	+	→	—	—	+
CiSSA extended	+	+	→	→	—	+

Таблица 3: Преимущества и недостатки пяти методов

На пересечении строк и столбцов указан знак, показывающий, достигается ли разделение компоненты: плюс (+) обозначает точное выполнение, знак стремления указывает на асимптотическое выполнение, а минус (—) — на отсутствие разделимости. Для разложения Фурье подразумевается, что  $L = N$ .

Обозначения:

- $\cos$  — в ряде присутствуют только периодические компоненты вида  $\cos(2\pi\omega x + \varphi)$ ;
- $X_{np1}$  — одна непериодическая компонента в ряде, остальные имеют период;
- $X_{np}$  — несколько непериодических компонент в ряде, остальные имеют период, интересует разделение между непериодическими компонентами;
- group — автоматическая группировка по заданным частотам.

# Сравнение алгоритмов. Пример 1

$X = X_{\sin} + X_{\cos} = \sin \frac{2\pi}{12}x + \frac{1}{2} \cos \frac{2\pi}{3}x$ ,  $L = 96$ ,  $N = 96 \cdot 2$  для разложения Фурье и  $N = 96 \cdot 2 - 1$  для остальных, чтобы выполнялись условия выполнения разделимости частот.

Сравним результаты по среднеквадратичной ошибке:

Метод/Компонента	$X_{\sin}$	$X_{\cos}$
SSA	6.8e-30	1.5e-29
SSA EOSSA	1.5e-29	7.5e-30
Fourier	1.7e-28	3.5e-28
CiSSA	1.9e-29	5.3e-30
CiSSA extended	2.0e-04	8.6e-04

Таблица 4: MSE разложений ряда  $X = X_{\sin} + X_{\cos}$  пяти методов

## Модификации метода SSA

### — Сравнение алгоритмов. Пример 1

#### Сравнение алгоритмов. Пример 1

$X = X_{\sin} + X_{\cos} = \sin \frac{2\pi}{12}x + \frac{1}{2} \cos \frac{2\pi}{3}x$ ,  $L = 96$ ,  $N = 96 \cdot 2$  для разложения Фурье и  $N = 96 \cdot 2 - 1$  для остальных, чтобы выполнялись условия выполнения разделимости частот. Сравним результаты по среднеквадратичной ошибке:

Метод/Компонента	$X_{\sin}$	$X_{\cos}$
SSA	6.8e-30	1.5e-29
SSA EOSSA	1.5e-29	7.5e-30
Fourier	1.7e-28	3.5e-28
CiSSA	1.9e-29	5.3e-30
CiSSA extended	2.0e-04	8.6e-04

Таблица 4: MSE разложений ряда  $X = X_{\sin} + X_{\cos}$  пяти методов

Таблица 4 показывает, что первые четыре разложения сделали правильное (с точностью до вычислений с помощью компьютера) разделение компонент ряда. Однако расширение в методе **CiSSA** ухудшило разделимость периодических частей.

## Сравнение алгоритмов. Пример 2

$X = X_{\sin} + X_{\cos} + X_{\text{noise}} = \sin \frac{2\pi}{12}x + \frac{1}{2} \cos \frac{2\pi}{3}x + \varepsilon_n$ , где  $\varepsilon_n \sim N(0, 0.1)$ ,  $L = 96$ ,  $N = 96 \cdot 2$  для разложения Фурье и  $N = 96 \cdot 2 - 1$  для остальных.

Метод/Компонента	$X_{\sin}$	$X_{\cos}$
SSA	2.9e-04	3.1e-04
SSA EOSSA	2.9e-04	3.1e-04
Fourier	1.0e-04	1.1e-04
CiSSA	1.6e-04	1.8e-04
CiSSA extended	6.6e-04	1.9e-03

Таблица 5: MSE разложений ряда  $X = X_{\sin} + X_{\cos} + X_{\text{noise}}$  пяти методов

### Модификации метода SSA

#### — Сравнение алгоритмов. Пример 2

#### Сравнение алгоритмов. Пример 2

$X = X_{\sin} + X_{\cos} + X_{\text{noise}} = \sin \frac{2\pi}{12}x + \frac{1}{2} \cos \frac{2\pi}{3}x + \varepsilon_n$ , где  $\varepsilon_n \sim N(0, 0.1)$ ,  $L = 96$ ,  $N = 96 \cdot 2$  для разложения Фурье и  $N = 96 \cdot 2 - 1$  для остальных.

Метод/Компонента	$X_{\sin}$	$X_{\cos}$
SSA	2.9e-04	3.1e-04
SSA EOSSA	2.9e-04	3.1e-04
Fourier	1.0e-04	1.1e-04
CiSSA	1.6e-04	1.8e-04
CiSSA extended	6.6e-04	1.9e-03

Таблица 5: MSE разложений ряда  $X = X_{\sin} + X_{\cos} + X_{\text{noise}}$  пяти методов

Проводилось 100 тестов, в таблице 5 указаны средние значения ошибки для одних и тех же реализаций шума.

По таблице 5 видно, что зашумление ряда дало негативный эффект на ошибку. Также был проведен парный t-критерий для зависимых выборок с целью проверки гипотезы о равенстве средних значений ошибки для каждой компоненты, попарно для всех методов. В качестве нулевой гипотезы ( $H_0$ ) предполагалось, что средние значения двух сравниваемых выборок равны. Критический уровень значимости был установлен на уровне  $\alpha = 0.05$ . Результаты анализа показали, что во всех случаях  $p$ -значение оказалось меньше 0.05, что позволяет отвергнуть нулевую гипотезу.



## Сравнение алгоритмов. Пример 3

$$X = X_{\sin} + X_{\cos} + X_c + X_e = \sin \frac{2\pi}{12}x + \frac{1}{2} \cos \frac{2\pi}{3}x + 1 + e^{\frac{x}{100}},$$

$L = 96$ ,  $N = 96 \cdot 2$  для разложения Фурье и  $N = 96 \cdot 2 - 1$

Метод/Компонента	$X_c + X_e$	$X_{\sin}$	$X_{\cos}$
SSA	5.0e-03	8.9e-07	5.2e-05
SSA EOSSA	1.7e-28	1.6e-29	8.7e-30
Fourier	1.1e-01	6.1e-04	6.8e-03
CiSSA	5.3e-02	1.6e-05	4.9e-04
CiSSA extended	5.0e-04	2.1e-04	1.1e-03

Таблица 6: MSE разложений ряда  $X = X_{\sin} + X_{\cos} + X_c + X_e$  четырех методов

### Модификации метода SSA

#### — Сравнение алгоритмов. Пример 3

#### Сравнение алгоритмов. Пример 3

$$X = X_{\sin} + X_{\cos} + X_c + X_e = \sin \frac{2\pi}{12}x + \frac{1}{2} \cos \frac{2\pi}{3}x + 1 + e^{\frac{x}{100}},$$

$L = 96$ ,  $N = 96 \cdot 2$  для разложения Фурье и  $N = 96 \cdot 2 - 1$

Метод/Компонента	$X_c + X_e$	$X_{\sin}$	$X_{\cos}$
SSA	5.0e-03	8.9e-07	5.2e-05
SSA EOSSA	1.7e-28	1.6e-29	8.7e-30
Fourier	1.1e-01	6.1e-04	6.8e-03
CiSSA	5.3e-02	1.6e-05	4.9e-04
CiSSA extended	5.0e-04	2.1e-04	1.1e-03

Таблица 6: MSE разложений ряда  $X = X_{\sin} + X_{\cos} + X_c + X_e$  четырех методов

Непериодические компоненты будут отвечать низким частотам. Проблема лишь в том, что с помощью методов разложения Фурье **CiSSA** невозможно различить между собой две непериодические компоненты, поскольку группировка работает по частотам, элементы разложения неизбежно смешаются между собой. Будем искать экспоненту и константу по низким частотам, назовем это трендовой составляющей ряда. По таблице 3 лучше всех должен справиться **SSA** с улучшением разделимости EOSSA. Хуже всех — разложение Фурье, поскольку он никаким образом не сможет вычленить из ряда экспоненту.

Результаты таблицы 6 повторяют вышеизложенные рассуждения. Также заметно, что периодические компоненты лучше выделились с помощью **CiSSA** без процедуры расширения ряда в сравнении с **CiSSA** с расширением.

## Сравнение алгоритмов. Пример 4

$X = X_{\sin} + X_{\cos} + X_c + X_e + X_{\text{noise}} = \sin \frac{2\pi}{12}x + \frac{1}{2} \cos \frac{2\pi}{3}x + 1 + e^{\frac{x}{100}} + \varepsilon_n$ , где  $\varepsilon_n \sim N(0, 0.1)$ ,  $L = 96$ ,  $N = 96 \cdot 2$  для разложения Фурье и  $N = 96 \cdot 2 - 1$ .

Метод/Компонента	$X_{\sin}$	$X_{\cos}$	$X_c + X_e$
SSA	2.9e-04	3.6e-04	5.2e-03
SSA EOSSA	2.9e-04	3.1e-04	9.4e-04
Fourier	6.9e-04	7.2e-03	1.2e-01
CiSSA	1.7e-04	7.0e-04	5.5e-02
CiSSA extended	6.8e-04	2.1e-03	2.7e-03

Таблица 7: MSE разложений ряда  $X = X_{\sin} + X_{\cos} + X_c + X_e + X_{\text{noise}}$  четырех методов

### Модификации метода SSA

#### — Сравнение алгоритмов. Пример 4

#### Сравнение алгоритмов. Пример 4

$X = X_{\sin} + X_{\cos} + X_c + X_e + X_{\text{noise}} = \sin \frac{2\pi}{12}x + \frac{1}{2} \cos \frac{2\pi}{3}x + 1 + e^{\frac{x}{100}} + \varepsilon_n$ , где  $\varepsilon_n \sim N(0, 0.1)$ ,  $L = 96$ ,  $N = 96 \cdot 2$  для разложения Фурье и  $N = 96 \cdot 2 - 1$ .

Метод/Компонента	$X_{\sin}$	$X_{\cos}$	$X_c + X_e$
SSA	2.9e-04	3.6e-04	5.2e-03
SSA EOSSA	2.9e-04	3.1e-04	9.4e-04
Fourier	6.9e-04	7.2e-03	1.2e-01
CiSSA	1.7e-04	7.0e-04	5.5e-02
CiSSA extended	6.8e-04	2.1e-03	2.7e-03

Таблица 7: MSE разложений ряда  $X = X_{\sin} + X_{\cos} + X_c + X_e + X_{\text{noise}}$  четырех методов

Как видно из таблицы 7, разделения ухудшились, однако **SSA** с улучшением разделимости EOSSA отработал лучше всех. Также был проведен двухвыборочный t-критерий для зависимых выборок с целью проверки гипотезы о равенстве средних значений ошибки для каждой компоненты, попарно для всех методов. В качестве нулевой гипотезы ( $H_0$ ) предполагалось, что средние значения двух сравниваемых выборок равны. Критический уровень значимости был установлен на уровне  $\alpha = 0.05$ . Результаты анализа показали, что во всех случаях  $p$ -значение оказалось меньше 0.05, что позволяет отвергнуть нулевую гипотезу.

Каждый алгоритм после группировки порождает построенными матрицами собственные подпространства. В случае базового **SSA** алгоритма базис подпространств является адаптивным, то есть зависящим от  $X, L, N$ . Таким образом, **SSA** может отличить, например, произведение полиномов, экспонент и косинусов друг от друга.

В случае **CiSSA** базис зависит только от  $L, N$ . Если зафиксировать данные параметры, и менять  $X$ , базис никак не поменяется.

## Модификации метода **SSA**

### └ Сравнение алгоритмов. Собственные пространства

#### Сравнение алгоритмов. Собственные пространства

Каждый алгоритм после группировки порождает построенными матрицами собственные подпространства. В случае базового **SSA** алгоритма базис подпространств является адаптивным, то есть зависящим от  $X, L, N$ . Таким образом, **SSA** может отличить, например, произведение полиномов, экспонент и косинусов друг от друга.  
В случае **CiSSA** базис зависит только от  $L, N$ . Если зафиксировать данные параметры, и менять  $X$ , базис никак не поменяется.

От собственных пространств зависит то, какие компоненты временно-го ряда будут разделимы между собой.

Теперь рассмотрим реальные данные — месячные ряды промышленного производства (Industrial Production, IP), index 2010 = 100, в США. Размер выборки составляет  $N = 537$ .

Применим как **CiSSA**, так и **SSA** с автоматическим определением частот и улучшением разделимости по следующим группам:

- 1 Трендовой составляющей должны отвечать низкие частоты, поэтому диапазон:  $[0, \frac{1}{192}]$ ;
- 2 Циклы бизнеса по диапазонам:  $[\frac{2}{192}, \frac{10}{192}]$ ;
- 3 Сезонность по частотам  $\omega_k = 1/12, 1/6, 1/4, 1/3, 5/12, 1/2$ ;

На основе предыдущих требований взято  $L = 192$ .

## Модификации метода SSA

### — Сравнение алгоритмов. Реальные данные

#### Сравнение алгоритмов. Реальные данные

Теперь рассмотрим реальные данные — месячные ряды промышленного производства (Industrial Production, IP), index 2010 = 100, в США. Размер выборки составляет  $N = 537$ .  
Применим как **CiSSA**, так и **SSA** с автоматическим определением частот и улучшением разделимости по следующим группам:

- 1 Трендовой составляющей должны отвечать низкие частоты, поэтому диапазон:  $[0, \frac{1}{192}]$ ;
- 2 Циклы бизнеса по диапазонам:  $[\frac{2}{192}, \frac{10}{192}]$ ;
- 3 Сезонность по частотам  $\omega_k = 1/12, 1/6, 1/4, 1/3, 5/12, 1/2$ ;

На основе предыдущих требований взято  $L = 192$ .

Данные промышленного производства полезны, поскольку оно указывается в определении рецессии Национальным бюро экономических исследований (NBER), как один из четырех ежемесячных рядов индикаторов, которые необходимо проверять при анализе делового цикла. Эти показатели демонстрируют различные тенденции, сезонность и цикличность (периодические компоненты, которые соответствуют циклам бизнеса).

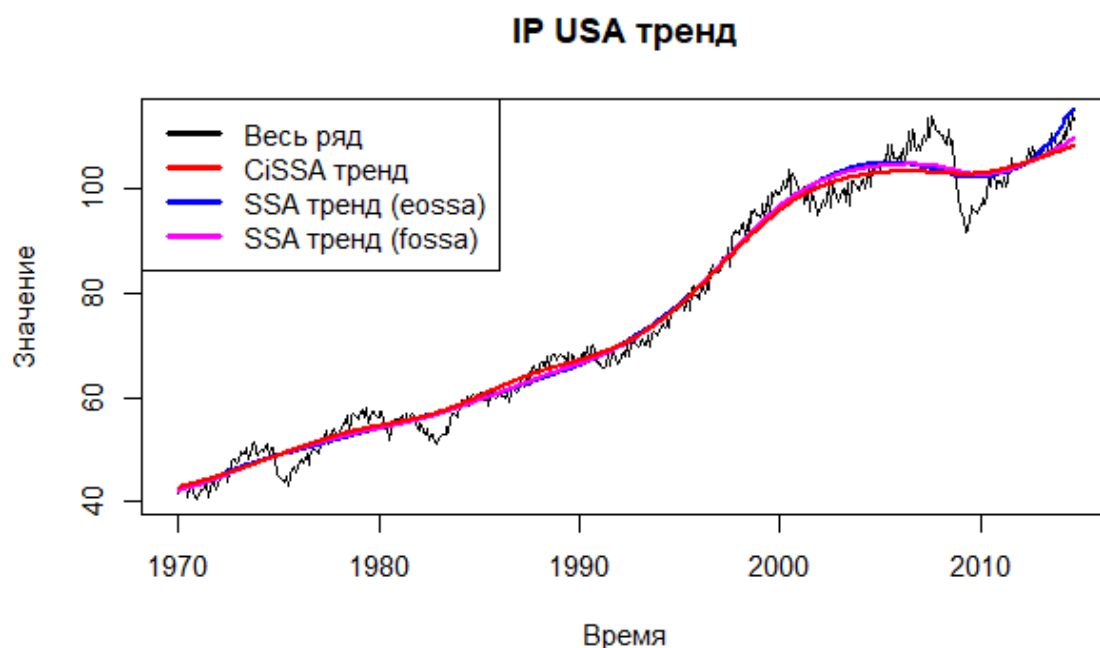


Рис. 1: Трендовая составляющая данных IP USA

## Модификации метода **SSA**

### └ Сравнение алгоритмов. Реальные данные

Сравнение алгоритмов. Реальные данные

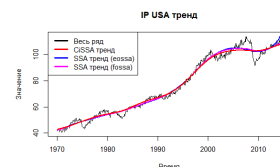


Рис. 1: Трендовая составляющая данных IP USA

При применении FOSSA улучшения разделимости алгоритм **SSA** выделяет тренд довольно похоже с **CiSSA**. Весь график **SSA** тренд EOSSA выглядит более изогнутым при визуальном сравнении с остальными.

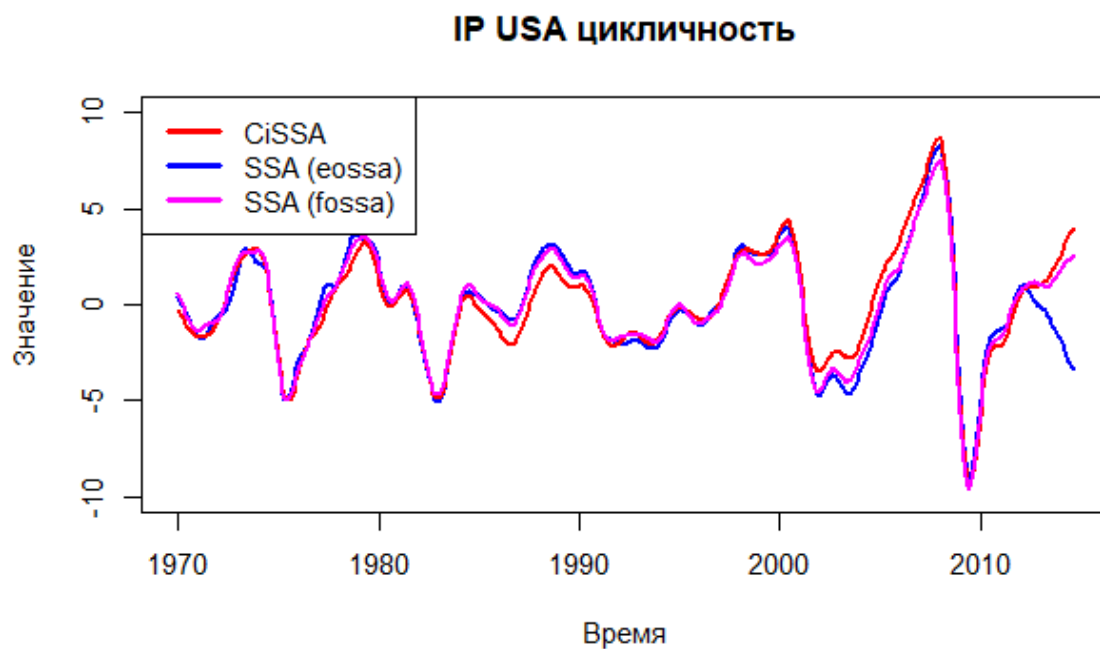


Рис. 2: Циклическая составляющая данных IP USA

## Модификации метода SSA

— Сравнение алгоритмов. Реальные данные

Сравнение алгоритмов. Реальные данные

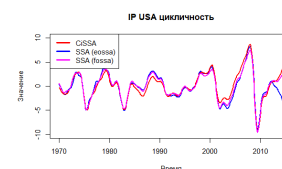


Рис. 2: Циклическая составляющая данных IP USA

Аналогичная тренду ситуация происходит с цикличностью. В случае EOSSA правый хвост (значения ряда после 2010-ого года) смешался между цикличностью и трендом.

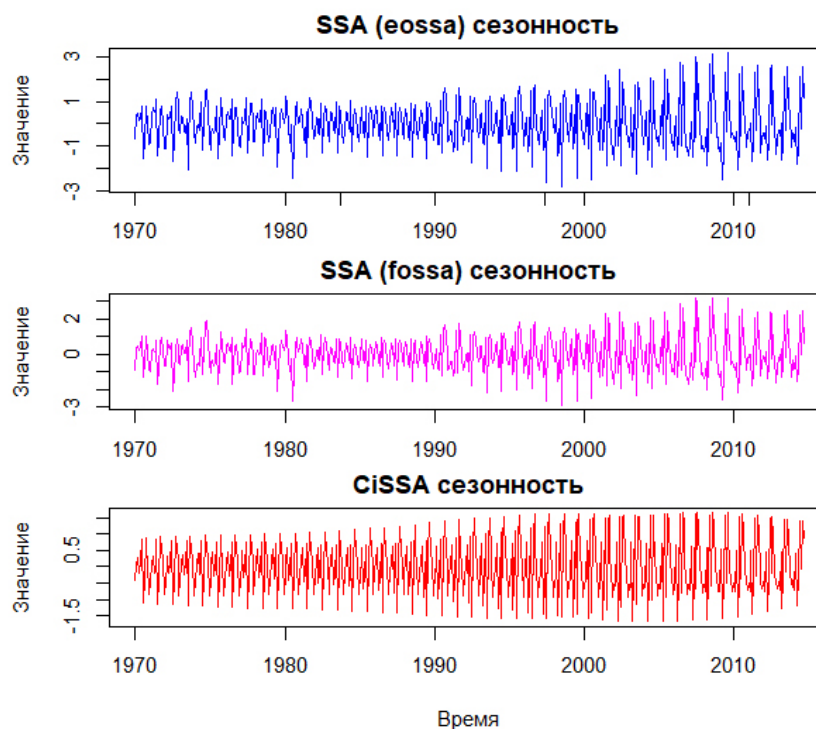
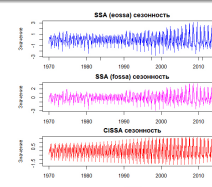


Рис. 3: Сезонная составляющая данных IP USA

## Модификации метода SSA

### — Сравнение алгоритмов. Реальные данные

Поскольку в базовом **SSA** адаптивный базис, сезонность является менее систематичной, разброс значений выше по сравнению с **CiSSA**. Таким образом, получились довольно похожие результаты в выделении тренда и цикличности при использовании **SSA** с FOSSA и **CiSSA**. Несколько иные результаты при **SSA** с EOSSA. Сезонная составляющая в силу неадаптивного базиса более строго выглядит для метода **CiSSA**.



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 делимости справляется в среднеквадратичном лучше **CiSSA** с расширением ряда или без.

## Модификации метода SSA

### — Заключение

#### Заключение

По полученным результатам, можно следующие выводы:

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

В данной работе исследован алгоритм **CiSSA**, сравнены методы **CiSSA** и **SSA**, и полученные знания были проверены на реальных и смоделированных примерах с помощью языка R. Оба алгоритма справляются с поставленными задачами, существенным различием является то, что алгоритм **SSA** является более гибким: в нем адаптивный базис, есть дополнительные алгоритмы, которые довольно похоже приближают этот алгоритм к **CiSSA**, а также методы для автоматического выбора компонентов по частотам. Метод **CiSSA** является простым в использовании.

Дальнейшими действиями является рассмотрение других модификаций метода **SSA**.

Все вычисления, а также код **CiSSA** можно найти в github репозитории [4].



# Список литературы



Juan Bogalo, Pilar Poncela, and Eva Senra.

Circulant singular spectrum analysis: A new automated procedure for signal extraction.

*Signal Processing*, 177, 2020.



Nina Golyandina, Pavel Dudnik, and Alex Shlemov.

Intelligent identification of trend components in singular spectrum analysis.

*Algorithms*, 16(7):353, 2023.



Nina Golyandina, Vladimir Nekrutkin, and Anatoly Zhigljavsky.

*Analysis of Time Series Structure: SSA and Related Techniques*.

Chapman and Hall/CRC, 2001.



Nikolay Pogrebnikov.

SPbSU CISSA coursework: Time series analysis.

[https://github.com/xSICHx/spbu\\_cissa\\_coursework](https://github.com/xSICHx/spbu_cissa_coursework), 2024.

25/25Погребников Николай Вадимович, гр. 21.Б04-мм Модификации метода SSA

## Модификации метода SSA

### Список литературы

#### 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. Nikolay Pogrebnikov. SPbSU CISSA coursework: Time series analysis. [https://github.com/xSICHx/spbu\\_cissa\\_coursework](https://github.com/xSICHx/spbu_cissa_coursework), 2024.

На данном слайде представлен список основных источников, используемых в моей работе. Спасибо за внимание.