

Санкт-Петербургский государственный университет

ПОГРЕБНИКОВ Николай Вадимович

Выпускная квалификационная работа

**МОДИФИКАЦИИ МЕТОДА АНАЛИЗА СИНГУЛЯРНОГО СПЕКТРА ДЛЯ
АНАЛИЗА ВРЕМЕННЫХ РЯДОВ**

Уровень образования: бакалавриат

Направление 01.03.02 «Прикладная математика и информатика»

Основная образовательная программа 21.Б04-мм «Прикладная математика и
информатика»

Научный руководитель:

Профессор кафедры статистического
моделирования
д. ф.-м. н. Н. Э. Голяндина

Рецензент:

Лектор, университет Кардиффа
А. Н. Пепелышев

Санкт-Петербург

2025

Saint Petersburg State University
Applied Mathematics and Computer Science

POGREBNIKOV Nikolay Vadimovich

Graduation Project

**MODIFICATIONS OF SINGULAR SPECTRUM ANALYSIS FOR TIME
SERIES ANALYSIS**

Scientific Supervisor:
Professor, Department of Statistical
Modelling N. E. Golyandina

Reviewer:
Lecturer, University of Cardiff
A. N. Pepelyshev

Saint Petersburg
2025

Оглавление

Введение	5
Глава 1. Метод Singular spectrum analysis (SSA)	7
1.1. Алгоритм метода SSA	7
1.2. Свойства SSA	9
1.2.1. Ранг ряда	9
1.2.2. Точная разделимость	9
1.2.3. Асимптотическая разделимость	10
1.2.4. Частотная разделимость	11
1.2.5. Вложенные варианты SSA	12
1.2.6. SSA как линейный фильтр	13
1.3. Многомерные варианты SSA	15
1.3.1. MSSA	15
1.3.2. 2d-SSA	16
Глава 2. Метод Generalized singular spectrum analysis (GSSA)	18
2.1. Алгоритм метода GSSA	18
2.2. Свойства GSSA	20
2.2.1. Веса	20
2.2.2. Ранг ряда	20
2.2.3. GSSA как линейный фильтр	20
Глава 3. Метод Circulant singular spectrum analysis (CiSSA)	22
3.1. Алгоритм метода CiSSA	22
3.2. Свойства	27
3.2.1. Связь CiSSA с разложением Фурье	27
3.2.2. Проблемы фиксированного базиса в CiSSA	28
3.2.3. Точная разделимость	29
3.2.4. Асимптотическая разделимость	29
3.3. Обзор литературы	29
3.3.1. Cognitive Load Detection through EEG Lead-Wise Feature Optimization and Ensemble Classification	29
3.3.2. Application of visual stratigraphy from line-scan images to constrain chronology and melt features of a firn core from coastal Antarctica	30

3.3.3. Выводы	31
Глава 4. Метод Functional singular spectrum analysis (FSSA)	32
4.1. Алгоритм FSSA	32
4.2. Практическая реализация алгоритма FSSA	35
Глава 5. Сравнения алгоритмов	37
5.1. Сравнение SSA и GSSA	37
5.1.1. Линейные фильтры	37
5.1.2. Фильтры в различных точках	39
5.1.3. Влияние шума	40
5.1.4. Вложенный вариант	41
5.2. Различия SSA, CiSSA и Фурье	43
5.2.1. Автоматическая группировка	43
5.2.2. Собственные пространства	43
5.2.3. Выбор длины окна для CiSSA	43
5.2.4. Без шума	47
5.2.5. С шумом	52
5.2.6. Разделение трендовых составляющих между собой	56
5.2.7. Преимущества и недостатки методов SSA, Фурье и CiSSA	57
5.2.8. Демонстрация работы алгоритмов на реальных данных	57
5.3. Сравнение FSSA, MSSA, 2d-SSA	61
5.3.1. Численное сравнение	61
Заключение	63
Список литературы	64

Введение

Временные ряды представляют собой упорядоченную последовательность данных, собранных или измеренных в хронологическом порядке. Они играют ключевую роль в анализе и прогнозировании различных явлений в таких областях, как экономика, финансы, климатология и медицина. Понимание эволюции этих явлений во времени важно для выявления тенденций, циклов и аномалий.

Для уточнения терминологии, следует отметить, что *временной ряд* длины N представляет собой упорядоченную конечную последовательность значений, которая записывается как $X = (x_1, \dots, x_N)$, $x_i \in \mathbb{R}$. Одним из основных аспектов анализа временных рядов является разделение их на составляющие. Среди таких компонент важными являются *тренд*, который отражает медленно изменяющуюся долгосрочную динамику ряда, и *сезонность*, представляющая собой периодические колебания, вызванные повторяющимися факторами, такими как климатические или экономические циклы.

В данной работе будут рассмотрены следующие постановки задачи разделения временных рядов:

1. Разделение временного ряда на компоненты, соответствующие определенным частотным диапазонам;
2. Разделение временного ряда на компоненты без привязки к частотным характеристикам, то есть в их исходном виде.

Анализ сингулярного спектра (**SSA** [1]) — метод, целью которого является разложение исходного ряда на сумму небольшого числа интерпретируемых компонент, таких как медленно изменяющаяся тенденция (тренд), колебательные компоненты (сезонность) и шум. Позволяет решать как задачу в формулировке 1, так и её обобщение, представленное в 2. При этом базовый алгоритм метода **SSA** не требует стационарности ряда, знания модели тренда, а также сведений о наличии в ряде периодиках, а за счет своего адаптивного базиса позволяет подстраиваться под любой входной ряд.

В данном исследовании рассматриваются модификации **SSA**, предложенные другими авторами, а именно: **GSSA** [2], **CiSSA** [3] и **FSSA** [4].

GSSA отличается от базового **SSA** тем, что он добавляет веса на определенном этапе алгоритма **SSA**. В разделении компонент между собой это может оказаться полезным, однако также может повлиять на разделимость сигнала от шума в худшую сторону. Это исследование раскрывает смысловую ценность **GSSA** с точки зрения линейных фильтров и отмечает ситуации, где такой алгоритм предпочтительнее стандартного **SSA**.

В алгоритме **CiSSA** предложено решение задачи разделения временного ряда компонентам, отвечающим за заранее известные диапазоны частот (задача в постановке 1). За счет этого можно автоматически группировать компоненты по частотам, однако именно поэтому алгоритм лишается адаптивности, которая имеется в **SSA**.

FSSA рассматривается как многомерная модификация **SSA**, основанная на скрещивании подходов функционального анализа и теории **SSA**.

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

Существует два вида разделимости: *точная разделимость*, которая характеризует способность метода точно выделять отдельные компоненты ряда, и *асимптотическая разделимость*, которая показывает, что ошибка разделения уменьшается с получением новых данных.

Теперь формально. Пусть временной ряд состоит из двух компонент: $X_N = X_N^{(1)} + X_N^{(2)}$. M – метод разделения ряда на компоненты с параметрами Θ . $\hat{X}_N^{(1)}$ – оценка $X_N^{(1)}$.

Определение 1. Ряды $X_N^{(1)}$ и $X_N^{(2)}$ точно разделимы методом M , если существует такое Θ , что $\text{MSE}(\hat{X}_N^{(1)}, X_N^{(1)}) = 0$.

Определение 2. Ряды $X_N^{(1)}$ и $X_N^{(2)}$ асимптотически разделимы методом M , если существует последовательность $\Theta(N)$, $N \rightarrow \infty$, что $\text{MSE}(\hat{X}_N^{(1)}, X_N^{(1)}) \rightarrow 0$.

Замечание 1. $\hat{X}^{(2)} = X - \hat{X}^{(1)}$ является оценкой для $X^{(2)}$, в условиях асимптотической разделимости выполнено $\text{MSE}(\hat{X}^{(2)}, X^{(2)}) \rightarrow 0$.

Далее кратко изложим структуру работы. В главе 1 рассматривается базовый метод **SSA**, его ключевые свойства, а также обобщения на многомерный случай — **MSSA** и **2d-SSA**. Главы 2, 3 и 4 посвящены алгоритмам **GSSA**, **CiSSA** и **FSSA** соответственно, где описываются их особенности и отличия от классического подхода. В каждом разделе главы 5 проводится теоретическое и численное сравнение модификаций с базовым методом. Наконец, в секции 5.3.1 подводятся итоги исследования.

Глава 1

Метод Singular spectrum analysis (SSA)

Рассмотрим базовый метод сингулярного спектрального анализа [1], его свойства, а также многомерные модификации MSSA и 2d-SSA.

1.1. Алгоритм метода SSA

Фиксируем временной ряд длины N : $\mathbf{X} = (x_1, \dots, x_N)$, $x_1, \dots, x_N \in \mathbb{R}$.

Базовый алгоритм SSA состоит из четырех шагов.

Вложение

Параметром этого шага является L — некоторое целое число (длина окна), $1 < L < N$. Строится L -траекторная матрица \mathbf{X} , состоящая из $K = N - L + 1$ векторов вложения:

$$\mathcal{T}(\mathbf{X}) = \mathcal{T}_{SSA}(\mathbf{X}) = \mathbf{X} = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_K \\ x_2 & x_3 & x_4 & \dots & x_{K+1} \\ x_3 & x_4 & x_5 & \dots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \dots & x_N \end{pmatrix}. \quad (1.1)$$

Полезным свойством является то, что матрица \mathbf{X} имеет одинаковые элементы на антидиагоналях. Таким образом, L -траекторная матрица является ганкелевой. \mathcal{T}_{SSA} называется оператором вложения. Обозначим множество всех возможных траекторных матриц как $\mathcal{M}_{L,K}^{(H)}$. Буква H используется, чтобы подчеркнуть, что эти матрицы имеют ганкелеву структуру.

Сингулярное разложение (SVD)

Результатом этого шага является сингулярное разложение (SVD) траекторной матрицы ряда.

Пусть $\mathbf{S} = \mathbf{X}\mathbf{X}^T$, $\lambda_1, \dots, \lambda_L$ — собственные числа матрицы \mathbf{S} , взятые в неубывающем порядке, и U_1, \dots, U_L — ортонормированная система собственных векторов, соответствующих собственным числам матрицы \mathbf{S} .

Определим $d = \max \{i : \lambda_i > 0\}$ и $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$. Тогда сингулярным разложением

называется представление матрицы в виде:

$$\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_d = \sum_{i=1}^d \sqrt{\lambda_i} U_i V_i^T. \quad (1.2)$$

Набор $(\sqrt{\lambda_i}, U_i, V_i^T)$ называется i -й собственной тройкой разложения (1.2).

Группировка

На основе разложения (1.2) производится процедура группировки, которая делит все множество индексов $\{1, \dots, d\}$ на m непересекающихся подмножеств I_1, \dots, I_m . Это разбиение является параметром шага группировки.

Пусть $I = \{i_1, \dots, i_p\}$, тогда $\mathbf{X}_I = \mathbf{X}_{i_1} + \cdots + \mathbf{X}_{i_p}$. Такие матрицы вычисляются для каждого $I = I_1, \dots, I_m$. В результате получаются матрицы $\mathbf{X}_{I_1}, \dots, \mathbf{X}_{I_m}$. Тем самым разложение (1.2) может быть записано в сгруппированном виде:

$$\mathbf{X} = \mathbf{X}_{I_1} + \cdots + \mathbf{X}_{I_m}.$$

Диагональное усреднение

Пусть \mathbf{Y} — матрица размерности $L \times K$. $L^* = \min(L, K)$, $K^* = \max(L, K)$. Диагональное усреднение переводит матрицу \mathbf{Y} в временной ряд g_1, \dots, g_N :

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y_{m,k-m+2}^* & \text{для } 1 \leq k < L^*, \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m,k-m+2}^* & \text{для } L^* \leq k < K^* + 1, \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} y_{m,k-m+2}^* & \text{для } K^* + 1 \leq k \leq N. \end{cases}$$

После применения этой операции к матрицам $\mathbf{X}_{I_1}, \dots, \mathbf{X}_{I_m}$, получаются m новых рядов: $\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_m$. Результатом данного шага и всего алгоритма является разложение временного ряда $\mathbf{X} = \tilde{\mathbf{X}}_1 + \cdots + \tilde{\mathbf{X}}_m$.

То же самое можно переписать в операторном виде:

$$\tilde{\mathbf{X}} = \mathcal{T}^{-1} \circ \Pi_{\mathbf{H}}(\tilde{\mathbf{X}}^{(k)}),$$

$\Pi_{\mathbf{H}}$ — оператор проектирования на множество $\mathcal{M}_{L,K}^{(\mathbf{H})}$.

1.2. Свойства SSA

1.2.1. Ранг ряда

Зафиксируем ряд $\mathbf{X} = (x_1, \dots, x_N)$ длины $N > 3$ и длину окна L .

Рассмотрим базовый SSA. В процессе процедуры вложения получаем последовательность векторов вложения:

$$\mathbf{X}_i^{(L)} = \mathbf{X}_i = (x_{i-1}, \dots, x_{i+L-2}), \quad i = 1, \dots, K,$$

$\mathcal{L}^{(L)} = \mathcal{L}^{(L)}(\mathbf{X}) \stackrel{\text{def}}{=} \text{span}(\mathbf{X}_1, \dots, \mathbf{X}_K)$ — траекторное пространство ряда \mathbf{X} . При этом, если $\dim \mathcal{L}^{(L)} = \text{rank } \mathbf{X} = d$, то будем говорить, что ряд \mathbf{X} имеет L -ранг d и записывать это как $\text{rank}_L = d$.

1.2.2. Точная разделимость

Ранее было дано более общее определение 1 точной разделимости. Рассмотрим его с другой стороны.

Пусть временной ряд $\mathbf{X} = \mathbf{X}^{(1)} + \mathbf{X}^{(2)}$ и задачей является нахождение этих слагаемых. В результате базового алгоритма SSA при $m = 2$ также получаем 2 ряда. Возникает вопрос: в каких случаях мы можем так выбрать параметр алгоритма L и так сгруппировать собственные тройки, чтобы получить исходные ряды без смешиваний? При выборе длины окна L каждый из рядов $\mathbf{X}^{(1)}$, $\mathbf{X}^{(2)}$, \mathbf{X} порождает траекторную матрицу $\mathbf{X}^{(1)}$, $\mathbf{X}^{(2)}$, \mathbf{X} .

Определение 3 ([1]). Будем говорить, что ряды $\mathbf{X}^{(1)}$ и $\mathbf{X}^{(2)}$ слабо L -разделимы, если пространства, порождаемые строками $\mathbf{X}^{(1)}$ и $\mathbf{X}^{(2)}$ ортогональны.

Если выполняется условие слабой L -разделимости, тогда существует такое сингулярное разложение траекторной матрицы \mathbf{X} ряда \mathbf{X} , что его можно разбить на две части, являющиеся сингулярными разложениями траекторных матриц рядов $\mathbf{X}^{(1)}$ и $\mathbf{X}^{(2)}$ [1].

Определение 4 ([1]). Будем говорить, что ряды $\mathbf{X}^{(1)}$ и $\mathbf{X}^{(2)}$ сильно L -разделимы, если они слабо L -разделимы и множества сингулярных чисел траекторных матриц рядов не имеют совпадений.

Если выполняется условие сильной L -разделимости, тогда любое сингулярное разложение траекторной матрицы \mathbf{X} ряда \mathbf{X} можно разбить на две части, являющиеся сингулярными разложениями траекторных матриц рядов $\mathbf{X}^{(1)}$ и $\mathbf{X}^{(2)}$ [1]. Это будет означать,

что для разложения ряда базовым методом SSA с $m = 2$ и таким L будет выполняться $\text{MSE}(\mathbf{X}^{(1)}, \hat{\mathbf{X}}^{(1)}) = 0$ (а значит и $\text{MSE}(\mathbf{X}^{(2)}, \hat{\mathbf{X}}^{(2)}) = 0$).

Рассмотрим таблицу, в которой знаком $+$ отмечены пары рядов, для которых существуют группировка и параметр L , при которых они разделимы (точно разделимы). Данная таблица 1.1 и условия разделимости с доказательствами взяты из книги [1].

Таблица 1.1. Точная разделимость.

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

Отметим, что $+$ в таблице 1.1 для $x_n^{(\cos_1)} = A_1 \cos(2\pi\omega_1 n + \varphi_1)$, $x_n^{(\cos_2)} = A_2 \cos(2\pi\omega_2 n + \varphi_2)$ достигается, если $L\omega_1 \in \mathbb{N}$, $K\omega_1 \in \mathbb{N}$ или $L\omega_2 \in \mathbb{N}$, $K\omega_2 \in \mathbb{N}$, $\omega_1 \neq \omega_2$ [1].

Однако, по таблице 1.1 видно, что условия точной разделимости достаточно жесткие и вряд ли выполнимы в реальных задачах. Тогда появляется такое понятие, как асимптотическая разделимость.

1.2.3. Асимптотическая разделимость

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

$$\rho_{i,j}^{(M)} = \frac{(\mathbf{X}_{i,i+M-1}^{(1)}, \mathbf{X}_{j,j+M-1}^{(2)})}{\|\mathbf{X}_{i,i+M-1}^{(1)}\| \|\mathbf{X}_{j,j+M-1}^{(2)}\|}.$$

Определение 5 ([1]). Ряды $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}$ называются ε -разделимыми при длине окна L , если

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

Определение 6 ([1]). Если $\rho^{(L(N), K(N))} \rightarrow 0$ при некоторой последовательности $L = L(N)$, $N \rightarrow \infty$, то ряды $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}$ называются асимптотически $L(N)$ -разделимыми.

Как можно заметить по таблице 1.2, для гораздо большего класса функций асимптотическая разделимость имеет место [1].

Таблица 1.2. Асимптотическая разделимость.

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

1.2.4. Частотная разделимость

Следуя [1, Глава 1, Раздел 1.4], опишем понятие разделимости на основе частот.

Для начала введем определение разложения Фурье. Пусть $\mathbf{X} = (x_1, \dots, x_N)$ — временной ряд

Определение 7. Разложение

$$x_n = c_0 + \sum_{k=1}^{\lfloor \frac{N+1}{2} \rfloor} (c_k \cos(2\pi nk/N) + s_k \sin(2\pi nk/N)), \quad (1.3)$$

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

Таким образом, можно выделить компоненту ряда, отвечающую за частоту $w_k = \frac{k-1}{L}$, $k = 1 : \lfloor \frac{N+1}{2} \rfloor$.

Теперь на основе (1.3) введем понятие периодограммы как объекта, характеризующего вклад частот в разложение Фурье ряда.

Определение 8. Периодограммой ряда $\mathbf{X} = (x_1, \dots, x_N)$ будем называть функцию Π_f^N , которая задана на множестве частот $\{k/N, k = 0, \dots, [N/2]\}$ формулой

$$\Pi_f^N(k/N) = \frac{N}{2} \begin{cases} 2c_0^2 & \text{для } k = 0, \\ c_k^2 + s_k^2 & \text{для } 0 < k < N/2, \\ 2c_{N/2}^2 & \text{для } k = N/2. \end{cases}$$

Последний случай имеет место только при четном N .

Фактически, $\Pi_f^N(k/N)$ показывает мощность частоты $\omega_k = \frac{k}{N}$.

Определение 9. $\Pi_f^N(k/N)$ называется мощностью частоты $\omega_k = \frac{k}{N}$. Набор частот ω_k с положительными мощностями называется носителем периодограммы. Если носитель периодограммы принадлежит некоторому интервалу $[a, b]$, то этот интервал будем называть диапазоном частот периодограммы.

Теперь можно перейти к разделимости двух рядов с точки зрения периодограмм. Пусть $X_N = X_N^{(1)} + X_N^{(2)}$ – временной ряд длины N , L – длина окна, $K = N - L + 1$.

Предложение 1 ([1, Глава 1, Раздел 1.4]). Если носители периодограмм отрезков ряда $X_N^{(1)}$ длины L не пересекаются с носителями периодограмм отрезков ряда $X_N^{(2)}$ длины L и аналогичное утверждение справедливо для отрезков рядов длины K , то ряды $X_N^{(1)}$ и $X_N^{(2)}$ слабо разделимы.

Теперь можно перейти к аналогичному предложению для асимптотической разделимости. Но для начала определим меру спектральной ортогональности рядов $X_N^{(1)}$ и $X_N^{(2)}$.

Определение 10. Величину

$$\rho_{12}^{(\Pi)} = \frac{\sum_{k=0}^{[N/2]} \sqrt{\Pi_{f_1}^N(k/N) \Pi_{f_2}^N(k/N)}}{\sqrt{\sum_{k=0}^{[N/2]} \Pi_{f_1}^N(k/N)} \sqrt{\sum_{k=0}^{[N/2]} \Pi_{f_2}^N(k/N)}} \quad (13)$$

будем называть *спектральным коэффициентом корреляции* рядов $X_N^{(1)}$ и $X_N^{(2)}$.

Предложение 2 ([1, Глава 1, Раздел 1.4]). Если спектральные коэффициенты корреляции отрезков ряда $X^{(1)}$ длины $L = L(N)$ и отрезков ряда $X^{(2)}$ длины L равномерно стремятся к нулю при $N \rightarrow \infty$ и аналогичное утверждение справедливо для отрезков рядов длины $K = K(N)$, то ряды $X^{(1)}$ и $X^{(2)}$ асимптотически разделимы.

1.2.5. Вложенные варианты SSA

В алгоритме базового SSA используется SVD-разложение траекторной матрицы. Поскольку оно обладает наилучшими аппроксимационными свойствами среди методов разложения матриц на матрицы меньшего ранга, применение базового SSA обеспечивает наилучшее разделение сигнала и шума. Однако, как показывают условия точной и асимптотической разделимости, компоненты сигнала удаётся разделить лишь при достаточно строгих требованиях. Поэтому можно рассматривать различные альтернативные алгоритмы, способным разлагать сигнал на компоненты при менее жёстких условиях. Отсюда возникает идея вложенных вариантов SSA.

Определение 11 ([5]). Пусть временной ряд состоит из двух компонент сигнала и шума: $X = S + X_{\text{Noise}} = S^{(1)} + S^{(2)} + X_{\text{Noise}}$.

Вложенный вариант SSA — двухэтапный метод:

1. Задаётся r . \tilde{S} – сумма первых r слагаемых SVD разложения траекторной матрицы сигнала S с помощью базового SSA.

2. Применение другого метода к $\tilde{\mathbf{S}}$ для улучшения разделимости: $\tilde{\mathbf{S}} = \tilde{\mathbf{S}}_1 + \tilde{\mathbf{S}}_2$.

Таким образом, получается наилучшим образом разделить как сигнал от шума, так и компоненты сигнала между собой.

В данной работе будут использоваться вложенные варианты SSA EOSSA и SSA FOSSA. Подробнее про них можно почитать в [6]. Кроме того, благодаря ним можно автоматически группировать компоненты в соответствии с заранее заданными частотами.

1.2.6. SSA как линейный фильтр

Разложение временного ряда методом SSA можно интерпретировать как применение линейных фильтров. Для дальнейшего исследования введем следующие определения.

Определение 12. Рассмотрим бесконечный временной ряд $\mathbf{X} = (\dots, x_{-1}, x_0, x_1, \dots)$. Линейный конечный фильтр — это оператор Φ , который преобразует временной ряд \mathbf{X} в новый по следующему правилу:

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

Набор коэффициентов h_i — импульсная характеристика фильтра.

Там, где не оговорено обратного, будем называть линейный конечный фильтр просто линейным фильтром.

Определение 13. Передаточная функция линейного фильтра Φ :

$$H_\Phi(z) = \sum_{i=-r_1}^{r_2} h_i z^{-i}.$$

Определение 14. Амплитудно-частотная характеристика (АЧХ) линейного фильтра Φ :

$$A_\Phi(\omega) = |H_\Phi(e^{i2\pi\omega})|.$$

АЧХ фильтра — это график или функция, которая показывает, как фильтр изменяет амплитуды (силу) разных частот входного сигнала.

Определение 15. Фазово-частотная характеристика (ФЧХ) линейного фильтра Φ :

$$\phi_\Phi(\omega) = \text{Arg}(H_\Phi(e^{i2\pi\omega})).$$

Посмотрим, как это выглядит для косинуса. Пусть исходный ряд $x_n^{\cos} = \cos 2\pi\omega n$. Тогда:

$$y_j = A_\Phi(\omega) \cos(2\pi\omega j + \phi_\Phi(\omega))$$

Теперь рассмотрим алгоритм SSA с точки зрения линейных фильтров [7, Глава 3, Раздел 3.9]. Пусть $\mathbf{X} = (x_1, \dots, x_N)$ — временной ряд длины N , $K = N - L + 1$, $L^* = \min(L, K)$. Пусть L будет длиной окна, а $(\sqrt{\lambda}, U, V)$ — одной из собственных троек. Определим диагональную матрицу $N \times N$:

$$\mathbf{D} = \text{diag}(1, 2, 3, \dots, L^* - 1, L^*, L^*, \dots, L^*, L^* - 1, \dots, 2, 1)$$

и матрицу $K \times N$

$$\mathbf{W} = \begin{pmatrix} u_1 & u_2 & u_3 & \cdots & u_L & 0 & \cdots & 0 & 0 & 0 \\ 0 & u_1 & u_2 & u_3 & \cdots & u_L & 0 & \cdots & 0 & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \cdots & \ddots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & u_1 & u_2 & u_3 & \cdots & u_L & 0 & \vdots \\ 0 & 0 & \cdots & 0 & u_1 & u_2 & u_3 & \cdots & u_L & 0 \\ 0 & 0 & 0 & \cdots & 0 & u_1 & u_2 & u_3 & \cdots & u_L \end{pmatrix}.$$

Здесь $U = (u_1, \dots, u_L)$ — собственный вектор матрицы \mathbf{S} .

Теорема 1 ([7, Глава 3, Раздел 3.9]). Компонента временного ряда $\tilde{\mathbf{X}}$, восстановленная с использованием собственной тройки $(\sqrt{\lambda}, U, V)$, имеет вид:

$$\tilde{\mathbf{X}}^T = \mathbf{D}^{-1} \mathbf{W}^T \mathbf{W} \mathbf{X}^T.$$

Таким образом, для восстановления методом SSA средних точек (индексы от L до K) имеем следующий фильтр:

$$\tilde{x}_s = \sum_{j=-(L-1)}^{L-1} \left(\sum_{k=1}^{L-|j|} u_k u_{k+|j|/L} \right) x_{s-j}, \quad L \leq s \leq K. \quad (1.4)$$

Похожим образом можно переписать SSA через линейные фильтры для точек в начале и конце.

1.3. Многомерные варианты SSA

Временные ряды могут быть не только одномерными, но и многомерными, то есть представлять собой наборы связанных наблюдений. В данной работе рассматриваются две модификации базового SSA: MSSA и 2d-SSA.

Обе эти модификации имеют схожую структуру с одномерным SSA, только нужно переопределить оператор вложения \mathcal{T} и понять, какую структуру будет иметь множество траекторных матриц $\mathcal{M}_{L,K}^{(H)}$.

Таким образом, имеем общую схему алгоритмов SSA:

1. Вложение:

Для временного ряда \mathbf{X} строится траекторная матрица:

$$\mathbf{X} = \mathcal{T}(\mathbf{X})$$

2. Разложение:

Траекторная матрица \mathbf{X} разлагается в сумму матриц ранга один:

$$\mathbf{X} = \sum_{j=1}^d \mathbf{X}_j$$

3. Группировка:

Элементарные матрицы \mathbf{X}_j разбиваются на группы:

$$\mathbf{X} = \mathbf{X}_{I_1} + \cdots + \mathbf{X}_{I_m}, \quad \text{где } \mathbf{X}_{I_k} = \sum_{j \in I_k} \mathbf{X}_j$$

4. Реконструкция:

К сгруппированным матрицам применяется диагональное усреднение:

$$\tilde{\mathbf{X}} = \tilde{\mathbf{X}}_1 + \cdots + \tilde{\mathbf{X}}_m, \quad \tilde{\mathbf{X}}_k = \mathcal{T}^{-1} \circ \Pi_H(\mathbf{X}_{I_k})$$

1.3.1. MSSA

Рассмотрим многомерный временной ряд, то есть набор $\{\mathbf{X}^{(p)} = (x_j^{(p)})_{j=1}^{N_p}, p = 1, \dots, s\}$ из s временных рядов длины N_p , где $p = 1, \dots, s$.

Обозначим $\mathbf{X} = (\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(s)})$ как исходные данные для алгоритма **MSSA** [8, Глава 4, Раздел 4.2]. Пусть L – длина окна, $1 < L < \min(N_p, p = 1, \dots, s)$. Для каждого временного ряда $\mathbf{X}^{(p)}$ формируем $K_p = N_p - L + 1$ векторов $\mathbf{X}_j^{(p)} = (x_j^{(p)}, \dots, x_{j+L-1}^{(p)})^T$, где $1 \leq j \leq K_p$. Обозначим $K = \sum_{p=1}^s K_p$. Траекторная матрица многомерного ряда \mathbf{X} – это матрица размера $L \times K$ следующего вида:

$$\mathcal{T}_{\text{MSSA}}(\mathbf{X}) = \mathbf{X} = [\mathbf{X}_1^{(1)} : \dots : \mathbf{X}_{K_1}^{(1)} : \dots : \mathbf{X}_1^{(s)} : \dots : \mathbf{X}_{K_s}^{(s)}] = [\mathbf{X}^{(1)} : \dots : \mathbf{X}^{(s)}],$$

$\mathbf{X}^{(p)} = \mathcal{T}_{\text{SSA}}(\mathbf{X}^{(p)})$ – траекторная матрица одномерного ряда $\mathbf{X}^{(p)}$. Таким образом, траекторная матрица системы временных рядов имеет блочно-ганкелеву структуру.

Комментарии

Собственные векторы U_i из SVD траекторной матрицы $\mathbf{X} = \sum_i \sqrt{\lambda_i} U_i V_i^T$ формируют общее базисное пространство для всех временных рядов. Векторы V_i содержат подкомпоненты $V_i^{(p)}$, соответствующие каждому ряду:

$$V_i = \begin{pmatrix} V_i^{(1)} \\ \vdots \\ V_i^{(s)} \end{pmatrix},$$

$V_i^{(p)} \in \mathbb{R}^{K_p}$ принадлежит строковому пространству p -го ряда. U_i отражают общую структуру, $V_i^{(p)}$ – её проявление в отдельных рядах.

Кроме того, преимущество **MSSA** проявляется, когда ряды имеют общую с точки зрения **SSA** структуру.

1.3.2. 2d-SSA

2d-SSA – это расширение базового **SSA** для двумерных массивов [8, Глава 5, Раздел 5.1] (например, изображений). Данные представляются в виде матрицы:

$$\mathbf{X} = (x_{ij})_{i,j=1}^{N_x, N_y}, \quad \text{где } N_x \times N_y \text{ – размер массива.}$$

Матрица содержит наблюдения, упорядоченные по двум пространственным или иным измерениям (например, изображение).

Параметрами метода являются размеры окна (L_x, L_y) , где $1 \leq L_x \leq N_x$, $1 \leq L_y \leq N_y$.

Определим \mathcal{T} :

1. Из массива \mathbf{X} выделяются все возможные подматрицы размера $L_x \times L_y$ с помощью скользящего окна.
2. Каждая подматрица $\mathbf{X}_{k,l}^{(L_x, L_y)}$ преобразуется в столбец: $\mathbf{X}_{k+(l-1)K_x} = \text{vec}(\mathbf{X}_{k,l}^{(L_x, L_y)})$.
3. Траекторная матрица строится как объединение этих столбцов:

$$\mathcal{T}_{2D\text{-SSA}}(\mathbf{X}) = \mathbf{X} = [\mathbf{X}_1 : \dots : \mathbf{X}_{K_x K_y}].$$

Матрица \mathbf{X} имеет блочно-ганкелеву структуру:

$$\mathbf{X} = \begin{pmatrix} \mathbf{H}_1 & \mathbf{H}_2 & \dots & \mathbf{H}_{K_y} \\ \mathbf{H}_2 & \mathbf{H}_3 & \dots & \mathbf{H}_{K_y+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{L_y} & \mathbf{H}_{L_y+1} & \dots & \mathbf{H}_{N_y} \end{pmatrix},$$

где каждая \mathbf{H}_j — ганкелевская матрица, построенная из столбцов \mathbf{X} .

Комментарии

Связь с **MSSA**: если $L_x = 1$ или $L_y = 1$, **2d-SSA** эквивалентен **MSSA** для временных рядов одинаковой длины. Поэтому его можно назвать обобщением **MSSA**.

Также при применении **2d-SSA** важен порядок следования строк и столбцов в \mathbf{X} , иными словами, предполагается регулярное поведение, как по строкам, так и по столбцам. В отличие от **MSSA**, где регулярное поведение нужно только по времени.

Глава 2

Метод Generalized singular spectrum analysis (GSSA)

В этом разделе описана модификация **SSA** на основе добавления специальных весов к строкам L -траекторной матрицы \mathbf{X} [2]. Это делается для уменьшения растекания частоты (spectral leakage). Авторы метода называют его обобщенным, поскольку базовый **SSA** является частным случаем **GSSA** с параметром $\alpha = 0$.

2.1. Алгоритм метода GSSA

Алгоритм **GSSA** сильно схож с базовым **SSA**. Фиксируем временной ряд длины N : $\mathbf{X} = (x_1, \dots, x_N)$, $x_1, \dots, x_N \in \mathbb{R}$. Выбирается параметр $\alpha \geq 0$, отвечающий за веса:

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

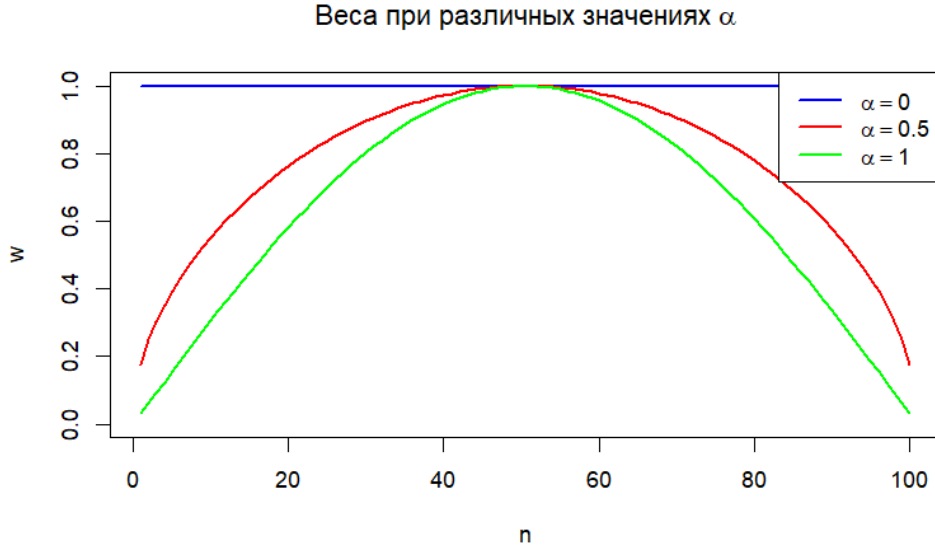


Рис. 2.1. График весов для различных значений α .

Вложение

L — некоторое целое число (длина окна), $1 < L < N$. Строится L -траекторная матрица $\mathbf{X}^{(\alpha)}$:

$$\mathbf{X}^{(\alpha)} = \begin{pmatrix} w_1 x_1 & w_1 x_2 & w_1 x_3 & \dots & w_1 x_K \\ w_2 x_2 & w_2 x_3 & w_2 x_4 & \dots & w_2 x_{K+1} \\ w_3 x_3 & w_3 x_4 & w_3 x_5 & \dots & w_3 x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_L x_L & w_L x_{L+1} & w_L x_{L+2} & \dots & w_L x_N \end{pmatrix}. \quad (2.1)$$

Сингулярное разложение (SVD)

Этот шаг такой же, как и в SSA, только матрица \mathbf{X} заменяется на $\mathbf{X}^{(\alpha)}$. Будем обозначать собственные тройки в этом случае так: $(\sqrt{\lambda^{(\alpha)}}, U^{(\alpha)}, V^{(\alpha)})$.

Группировка

В точности как в SSA. Тем самым, разложение может быть записано в сгруппированном виде:

$$\mathbf{X}^{(\alpha)} = \mathbf{X}_{I_1}^{(\alpha)} + \dots + \mathbf{X}_{I_m}^{(\alpha)}.$$

Взвешенное диагональное усреднение

Поскольку траекторная матрица была изменена весами, то диагональное усреднение тоже будет зависеть от весов.

Пусть \mathbf{Y} — матрица размерности $L \times K$. Взвешенное диагональное усреднение переводит матрицу \mathbf{Y} в временной ряд g_1, \dots, g_N :

$$g_k = \begin{cases} \frac{1}{\sum_{n=1}^k w_n} \sum_{m=1}^{k+1} y_{m,k-m+2}^* & \text{для } 1 \leq k < L, \\ \frac{1}{\sum_{n=1}^L w_n} \sum_{m=1}^L y_{m,k-m+2}^* & \text{для } L \leq k < K+1, \\ \frac{1}{\sum_{n=k-K+1}^L w_n} \sum_{m=k-K+2}^{N-K+1} y_{m,k-m+2}^* & \text{для } K+1 \leq k \leq N. \end{cases}$$

Применяя данную операцию к матрицам $\mathbf{X}_{I_1}^{(\alpha)}, \dots, \mathbf{X}_{I_m}^{(\alpha)}$, получаются m новых рядов: $\mathbf{X}_1^{(\alpha)}, \dots, \mathbf{X}_m^{(\alpha)}$. Результатом данного шага и всего алгоритма является разложение временного ряда $\mathbf{X}_1^{(\alpha)} + \dots + \mathbf{X}_m^{(\alpha)} = \mathbf{X}^{(\alpha)}$.

2.2. Свойства GSSA

2.2.1. Веса

Для минимизации эффекта спектрального размывания (spectral leakage), связанного с конечностью временного интервала наблюдений, к исходному ряду применялось оконное преобразование (tapering [9]). В качестве оконной функции используются степенные синус-косинусные веса (power-of-sine/cosine window).

Данное преобразование выполняет две ключевые функции:

1. Снижение краевых эффектов: умножение исходного ряда на убывающую к краям функцию $w(t)$;
2. Сглаживание периодограммы: веса используются для усреднения значений.

Такой подход позволяет более точно отделять компоненты ряда друг от друга.

2.2.2. Ранг ряда

Зафиксируем ряд $\mathbf{X} = (x_1, \dots, x_N)$ длины $N > 3$ и длину окна L .

В секции 1.2.1 было введено понятие ранга ряда для базового SSA. Теперь рассмотрим GSSA и поймем, что для того же ряда $\text{rank } \mathbf{X}^{(\alpha)} = \text{rank } \mathbf{X}$, а значит, что для GSSA также применимы понятия L -ранга ряда. Из вида (2.1) $\mathbf{X}^{(\alpha)}$ можно получить, что $\mathbf{X}^{(\alpha)} = \text{diag}(w_1, w_2, \dots, w_L) \mathbf{X} = \text{diag}(\mathbf{w}^{(a)}) \mathbf{X}$. Поскольку матрица $\text{diag}(\mathbf{w}^{(a)})$ имеет ранг равный L , она диагональна, то и $\text{rank } \mathbf{X}^{(\alpha)} = \text{rank } \text{diag}(\mathbf{w}^{(a)}) \mathbf{X} = \text{rank } \mathbf{X}$.

2.2.3. GSSA как линейный фильтр

Аналогично SSA, метод GSSA можно переписать с помощью линейных фильтров. Пусть $\mathbf{X} = (x_1, \dots, x_N)$ — временной ряд длины N , $K = N - L + 1$, $L^* = \min(L, K)$. $(\sqrt{\lambda^{(\alpha)}}, U^{(\alpha)}, V^{(\alpha)})$ — одна из собственных троек. Определим диагональную матрицу $N \times N$:

$$\mathbf{D}^{(\alpha)} = \text{diag}(w_1, w_1 + w_2, \dots, \sum_{i=1}^{L^*-1} w_i, \sum_{i=1}^{L^*} w_i, \sum_{i=1}^{L^*} w_i, \dots, \sum_{i=1}^{L^*} w_i, \sum_{i=2}^{L^*} w_i, \dots, w_{L^*-1} + w_{L^*}, w_{L^*})$$

и две матрицы $K \times N$:

$$\mathbf{W}^{(\alpha)} = \begin{pmatrix} u_1^{(\alpha)} & u_2^{(\alpha)} & u_3^{(\alpha)} & \cdots & u_L^{(\alpha)} & 0 & \cdots & 0 & 0 & 0 \\ 0 & u_1^{(\alpha)} & u_2^{(\alpha)} & u_3^{(\alpha)} & \cdots & u_L^{(\alpha)} & 0 & \cdots & 0 & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \cdots & \ddots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & u_1^{(\alpha)} & u_2^{(\alpha)} & u_3^{(\alpha)} & \cdots & u_L^{(\alpha)} & 0 & \vdots \\ 0 & 0 & \cdots & 0 & u_1^{(\alpha)} & u_2^{(\alpha)} & u_3^{(\alpha)} & \cdots & u_L^{(\alpha)} & 0 \\ 0 & 0 & 0 & \cdots & 0 & u_1^{(\alpha)} & u_2^{(\alpha)} & u_3^{(\alpha)} & \cdots & u_L^{(\alpha)} \end{pmatrix},$$

$$\mathbf{W}_w^{(\alpha)} = \begin{pmatrix} w_1 u_1^{(\alpha)} & w_2 u_2^{(\alpha)} & w_3 u_3^{(\alpha)} & \cdots & w_L u_L^{(\alpha)} & 0 & \cdots & 0 & 0 & 0 \\ 0 & w_1 u_1^{(\alpha)} & w_2 u_2^{(\alpha)} & w_3 u_3^{(\alpha)} & \cdots & w_L u_L^{(\alpha)} & 0 & \cdots & 0 & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \cdots & \ddots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & w_1 u_1^{(\alpha)} & w_2 u_2^{(\alpha)} & w_3 u_3^{(\alpha)} & \cdots & w_L u_L^{(\alpha)} & 0 & \vdots \\ 0 & 0 & \cdots & 0 & w_1 u_1^{(\alpha)} & w_2 u_2^{(\alpha)} & w_3 u_3^{(\alpha)} & \cdots & w_L u_L^{(\alpha)} & 0 \\ 0 & 0 & 0 & \cdots & 0 & w_1 u_1^{(\alpha)} & w_2 u_2^{(\alpha)} & w_3 u_3^{(\alpha)} & \cdots & w_L u_L^{(\alpha)} \end{pmatrix}.$$

Здесь $U = (u_1, \dots, u_L)$ — собственный вектор матрицы \mathbf{S} .

Теорема 2. Компонента временного ряда $\tilde{\mathbf{X}}$, восстановленная с использованием собственной тройки $(\sqrt{\lambda^{(\alpha)}}, U^{(\alpha)}, V^{(\alpha)})$, имеет вид:

$$\tilde{\mathbf{X}}^T = \mathbf{D}^{(\alpha)-1} \mathbf{W}^{(\alpha)T} \mathbf{W}_w^{(\alpha)} \mathbf{X}^T.$$

Доказательство. Доказательство проводится аналогично доказательству теоремы 1. \square

Таким образом, для восстановления методом **GSSA** средних точек (индексы от L до K) имеем следующий фильтр:

$$\tilde{x}_s = \sum_{j=-(L-1)}^{L-1} \left(\sum_{k=1}^{L-|j|} u_k^{(\alpha)} u_{k+|j|}^{(\alpha)} w_k / \sum_{i=1}^L w_i \right) x_{s-j}, \quad L \leq s \leq K. \quad (2.2)$$

Похожим образом можно переписать **GSSA** через линейные фильтры для точек в начале и конце (то представление можно взять из матричной записи в теореме 2).

Глава 3

Метод Circulant singular spectrum analysis (CiSSA)

В этом разделе описана модификация SSA на основе циркулярной матрицы [3]. В отличие от базового SSA, в CiSSA для каждого конкретного L базис разложения остается одинаковым для любого входного временного ряда. Поскольку из-за этого повышается интерпретируемость каждой компоненты в разложении, авторы метода назвали CiSSA автоматизированной версией SSA в том смысле, что компоненты ряда группируются по частотам самим алгоритмом. Сначала будет рассмотрен метод только для стационарного случая, затем показана применимость модифицированной версии CiSSA при использовании нестационарного ряда.

Стационарность подразумевает неизменность определенных свойств ряда во времени. Определим это понятие формально [1].

Определение 16. Пусть $X = (x_1, \dots, x_n, \dots)$ — временной ряд. Ряд X называется стационарным, если существует функция $R_X(k)$ ($-\infty < k < +\infty$) такая, что для любых $k, l \geq 1$

$$R_X^{(N)}(k, l) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{m=1}^N x_{k+m} x_{l+m} \xrightarrow{N \rightarrow \infty} R_X(k - l). \quad (3.1)$$

Если (3.1) выполняется, тогда R_X называется ковариационной функцией стационарного ряда X .

Теорема 3 ([1]). Пусть R_X — ковариационная функция стационарного ряда X . Тогда существует конечная мера m_X , определенная на борелевских подмножествах $(-1/2, 1/2]$, такая, что

$$R_X(k) = \int_{(-\frac{1}{2}, \frac{1}{2}]} e^{i2\pi k\omega} m_X(d\omega).$$

Мера m_X называется спектральной мерой ряда X .

3.1. Алгоритм метода CiSSA

Данный алгоритм, как и SSA, состоит из четырех основных шагов.

Зафиксируем стационарный временной ряд X состоящий из N элементов и выберем длину окна L .

Вложение

Такой же, как и в SSA. Считаем матрицу \mathbf{X} , заданную в (1.1).

Разложение

Для каждого $k = 1 : L$ вычисляются собственные векторы U_k :

$$U_k = L^{-1/2}(u_{k,1} \cdot \dots, u_{k,L}), \text{ где } u_{k,j} = \exp\left(-i2\pi(j-1)\frac{k-1}{L}\right), \quad (3.2)$$

причем $U_k = U_{L+2-k}^*$, U^* — комплексное сопряжение вектора U .

Элементарное разложение

Для каждой частоты $\omega_k = \frac{k-1}{L}$, $k = 1 : \lfloor \frac{L+1}{2} \rfloor$, есть два собственных вектора: U_k и U_{L+2-k} . За частоту ω_0 отвечает один собственный вектор — U_0 . Если же L — четное, то частоте $\omega_{\frac{L}{2}+1}$ будет соответствовать один вектор $U_{\frac{L}{2}+1}$.

Следовательно, индексы группируются следующим образом:

$$B_k = \begin{cases} \{1\}, & \text{если } k = 1, \\ \{k, L+2-k\}, & \text{если } 2 \leq k \leq \lfloor \frac{L+1}{2} \rfloor \text{ и } k \neq \frac{L}{2} + 1, \\ \{\frac{L}{2} + 1\}, & \text{если } k = \frac{L}{2} + 1 \text{ и } L \bmod 2 = 0. \end{cases}$$

Таким образом, получается элементарная группировка по частотам ω_k :

$$\mathbf{X}_{B_k} = \begin{cases} U_1 U_1^H \mathbf{X}, & \text{если } k = 1, \\ U_k U_k^H \mathbf{X} + U_{L+2-k} U_{L+2-k}^H \mathbf{X}, & \text{если } 2 \leq k \leq \lfloor \frac{L+1}{2} \rfloor \text{ и } k \neq \frac{L}{2} + 1, \\ U_{\frac{L}{2}+1} U_{\frac{L}{2}+1}^H \mathbf{X}, & \text{если } k = \frac{L}{2} + 1 \text{ и } L \bmod 2 = 0. \end{cases}$$

где U^H — это комплексное сопряжение и транспонирование вектора U .

Результатом данного шага будет разложение исходной матрицы \mathbf{X} в сумму матриц \mathbf{X}_{B_k} , отвечающих периодикам с определенными частотами ω_k :

$$\mathbf{X} = \sum_{k=1}^d \mathbf{X}_{B_k}.$$

Группировка

Такой же шаг, как и в базовом SSA. Однако группировка будет производиться по m непересекающимся диапазонам частот Ω_j , которые находятся в диапазоне от 0 до 0.5.

То есть, $\bigsqcup_{j=1}^m \Omega_j = \bigsqcup_{j=1}^m [\omega_j^{(l)}, \omega_j^{(r)}] = [0, 0.5]$. $\mathbf{X}_{\Omega_j} = \sum_{\omega_k \in \Omega_j} \mathbf{X}_{\omega_k}$.

Диагональное усреднение

Такой же шаг, как и в базовом SSA.

Замечание 2. U_k можно получить по аналогии с SSA.

Определим автоковариации:

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

На основе $\hat{\gamma}_m$ определим матрицу:

$$\hat{\gamma}_L = \begin{pmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \dots & \hat{\gamma}_{L-1} \\ \hat{\gamma}_1 & \hat{\gamma}_2 & \dots & \hat{\gamma}_{L-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}_{L-1} & \hat{\gamma}_{L-2} & \dots & \hat{\gamma}_0 \end{pmatrix}. \quad (3.3)$$

Данная матрица $L \times L$ называется Теплицевой и используется в методе Toeplitz SSA (подробнее про данный метод можно прочитать в книге [1]). На ее основе составим циркулярную матрицу для алгоритма Circulant SSA [3]:

$$\hat{\mathbf{C}}_L = \begin{pmatrix} \hat{c}_0 & \hat{c}_1 & \dots & \hat{c}_{L-1} \\ \hat{c}_1 & \hat{c}_0 & \dots & \hat{c}_{L-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{c}_{L-1} & \hat{c}_{L-2} & \dots & \hat{c}_0 \end{pmatrix}, \quad (3.4)$$

где $\hat{c}_m = \frac{L-m}{L} \hat{\gamma}_m + \frac{m}{L} \hat{\gamma}_{L-m}$, $m = 0 : L-1$. Собственные числа матрицы $\hat{\mathbf{C}}_L$, определенной в (3.4), задаются по формуле:

$$\lambda_{L,k} = \sum_{m=0}^{L-1} \hat{c}_m \exp\left(i2\pi m \frac{k-1}{L}\right), \quad k = 1 : L, \text{ причем } \lambda_{L,k} = \lambda_{L,L+2-k},$$

а собственные вектора, связанные с $\lambda_{L,k}$ — это векторы U_k из (3.2).

Замечание 3. $U_k U_k^H + U_{L+2-k} U_{L+2-k}^H$ является оператором проектирования на подпространство, которое порождено синусами и косинусами с частотой $\omega_k = \frac{k-1}{L}$.

Замечание 4. В разделе 3.2.1 рассмотрена связь между матрицей \mathbf{X}_{B_k} и разложениями Фурье для векторов вложения.

Нестационарный случай

Для применения данного алгоритма на нестационарных временных рядах, авторами предлагается применить процедуру расширения ряда. Как утверждается авторами статьи [3], после расширения, **CiSSA** можно применить к нестационарному ряду. Сама процедура расширения ряда \mathbf{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. В ходе работы алгоритм выполняет следующие шаги:

1. **Построение дифференцированного ряда:** Временной ряд \mathbf{X} сначала преобразуется в дифференцированный ряд $d\mathbf{X}$, чтобы удалить трендовые компоненты;
2. **Определение порядка AR-модели:** Метод определяет порядок p AR-модели как целую часть от деления длины ряда N на 3. Это значение порядка модели p будет использовано для построения авторегрессионной модели на дифференцированном временном ряде;
3. **Построение AR-модели:** После этого для дифференцированного ряда вычисляются коэффициенты авторегрессионной модели A с использованием метода Юла-Уокера, основываясь на определенном ранее порядке p ;
4. **Правое расширение ряда:** С помощью AR-модели ряд $d\mathbf{X}$ прогнозируется на H шагов вправо. Затем возвращается к своему изначальному состоянию путем интегрирования $d\mathbf{X}$. Получается расширение исходного ряда \mathbf{X} на H шагов вправо;
5. **Левое расширение ряда:** Аналогично предыдущему пункту, ряд прогнозируется на H шагов влево;
6. **Возвращение расширенного ряда:** В конце метод возвращает расширенный временной ряд $\mathbf{X}_{\text{extended}}$, который содержит как левое, так и правое расширение на H шагов от исходного ряда \mathbf{X} .

Таким образом, алгоритм расширения ряда позволяет выполнять предсказания временного ряда по обе стороны от его границ, основываясь на авторегрессионной модели, построенной на дифференцированном ряде, что полезно для выделения тренда.



Рис. 3.1. Расширение временного ряда IP values. Красным показан настоящий ряд, черным — его расширение

Однако поскольку мы рассматриваем расширенный ряд, то и периодические компоненты будут строиться по нему. Поэтому в угоду лучшего выделения трендовой составляющей, будет несколько жертвоваться точность разделения периодических компонентов.

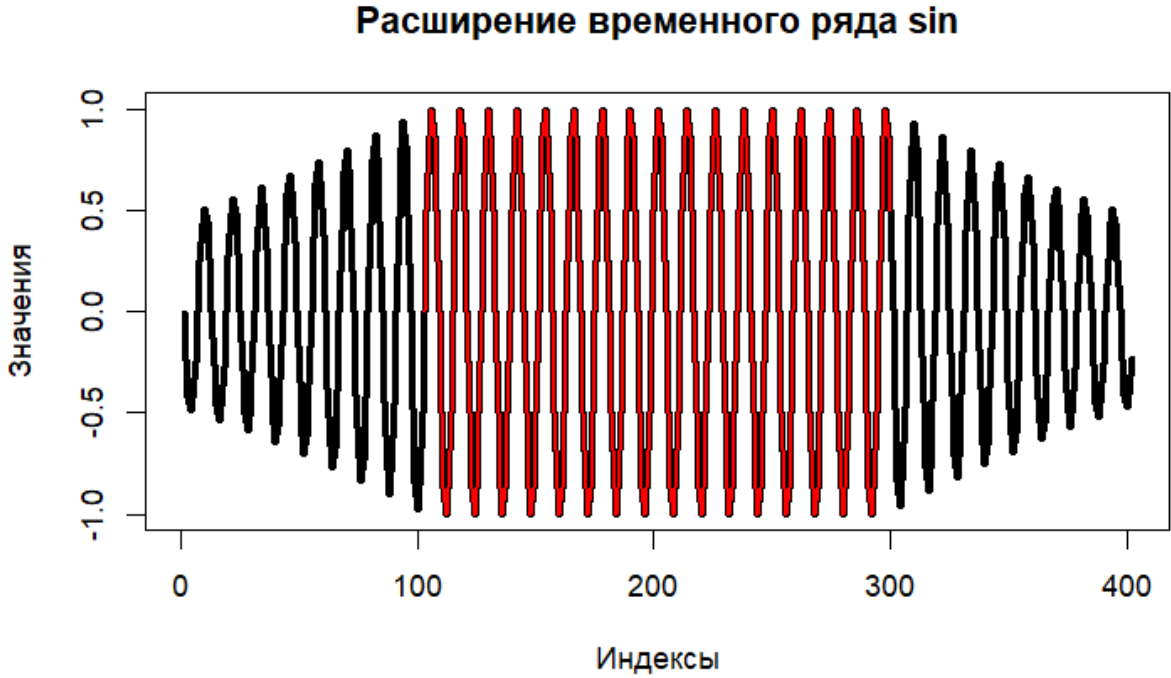


Рис. 3.2. Расширение временного ряда синуса. Красным показан настоящий ряд, черным — его расширение

На рисунке 3.2 видно, что синус расширился неправильно, от концов настоящего ряда до концов расширенного значения постепенно уменьшались. Как будет показано в секции 5.2, это повлияет на значения ошибки выделения сигнала и его компонент.

3.2. Свойства

3.2.1. Связь CiSSA с разложением Фурье

Для описания конечных, но достаточно длинных рядов можно использовать разложение Фурье. Определение разложения Фурье было дано в (1.3).

Алгоритм CiSSA тесно связан с разложением Фурье. По замечанию 3 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 \mathbf{X} . Затем вычисляется диагональное усреднение \mathbf{X}_{B_k} . А именно, CiSSA можно представить так:

1. Вычисляем разложение Фурье для каждого вектора вложения L -траекторной матрицы \mathbf{X} , состоящей из $K = N - L + 1$ векторов. Получается K разложений Фурье по частотам $\omega_k = \frac{k-1}{L}$, $k = 1 : \lfloor \frac{L+1}{2} \rfloor$;

2. По получившимся разложениям Фурье усредняем значения для соответствующих x_i и частот ω_k .

3.2.2. Проблемы фиксированного базиса в CiSSA

Одной из особенностей метода CiSSA является использование фиксированного базиса, определенного на пространстве гармоник с частотами, расположенными на регулярной сетке. В отличие от базового SSA, где базис формируется адаптивно в результате сингулярного разложения траекторной матрицы и зависит от структуры входного ряда, CiSSA априори фиксирует частотное представление сигнала. Это принципиальное различие приводит к ряду проблем.

Одна из таких проблем – растекание спектра. Оно возникает в ситуациях, когда действительная частота сигнальной компоненты не совпадает с узлом частотной решётки CiSSA. В таком случае амплитуда истинной частоты оказывается распределённой между несколькими соседними частотами. Это снижает точность восстановления сигнала. Аналогичный эффект наблюдается и при наличии модуляции в частотах, например, когда амплитуда компоненты изменяется во времени. Такие примеры будут рассмотрены в разделе 5.2.

Для того, чтобы исправить этот недостаток, в CiSSA используется группировка частотных компонент по интервалам $[\omega \pm \delta]$. Однако выбор параметра δ оказывается сложной задачей и представляет собой компромисс между смещением и дисперсией оценки. Если δ слишком мало, то компоненту не удаётся полностью захватить, что ведёт к смещению. Если δ велико, в группу попадает значительное количество шума, что увеличивает дисперсию и приводит к ухудшению оценки.

Таким образом, точность восстановления сигнала в CiSSA зависит от:

- ширины интервалов (δ);
- наличия модуляции;
- расположения частоты сигнальной компоненты относительно частотной решётки базиса.

Эти факторы — прямое следствие использования неадаптивного базиса.

Кроме того, на спектральную утечку влияет и выбор окна L . При меньших значениях L решётка становится более грубой, что может уменьшить спектральное растекание, но при этом ухудшается способность отделения сигнала от шума. Таким образом, возникает дополнительный компромисс между точностью частотного выделения и устойчивостью к шуму.

3.2.3. Точная разделимость

В разделе 1.2.4 описана частотная разделимость, являющаяся достаточным условием для точной разделимости базового SSA. Для CiSSA введем точную разделимость также через периодограмму.

Предложение 3. Если носители периодограмм отрезков ряда $X_N^{(1)}$ длины L не пересекаются с носителями периодограмм отрезков ряда $X_N^{(2)}$ длины L , то ряды $X_N^{(1)}$ и $X_N^{(2)}$ точно разделимы методом CiSSA в смысле определения 1.

Предложение 3 объяснимо тем, что поскольку данный метод является аналогом разложения Фурье, то в смысле сильной разделимости можно точно разделить ряд, в котором одной из компонент является $\cos(2\pi\omega + \varphi)$ с частотой ω такой, что $L\omega = k \in \mathbb{N}$, или константа. Для сравнения, при применении базового SSA, условие накладывалось не только на $L\omega \in \mathbb{N}$, но и на $K\omega \in \mathbb{N}$.

Поэтому до применения алгоритма необходимо выделить интересующие частоты, то есть знать их заранее, и, исходя из них, выбирать значение L .

3.2.4. Асимптотическая разделимость

Аналогично точной разделимости, рассмотрим асимптотическую для CiSSA через периодограмму.

Предложение 4. Если спектральные коэффициенты корреляции отрезков ряда $X^{(1)}$ длины $L = L(N)$ и отрезков ряда $X^{(2)}$ длины L равномерно стремятся к нулю при $N \rightarrow \infty$, то ряды $X^{(1)}$ и $X^{(2)}$ асимптотически разделимы методом CiSSA в смысле определения 2.

3.3. Обзор литературы

В данном разделе рассмотрены применения CiSSA на практике.

3.3.1. Cognitive Load Detection through EEG Lead-Wise Feature Optimization and Ensemble Classification

В статье [10] рассматриваются несколько наборов данных ЭЭГ под различной нагрузкой, два из которых наиболее значимы:

- **MAT (Mental Arithmetic Task):** участие студентов в задаче на математический счёт.

- **STEW (Simultaneous Task EEG Workload)**: параллельное выполнение нескольких заданий.

Задача исследования заключалась в следующем:

- С помощью метода CiSSA (и других различных методов) разлагают исходные сигналы ЭЭГ на несколько компонент, каждая из которых несёт информацию о разных частотных диапазонах и временных структурах.
- После разложения из полученных компонент (или их комбинаций) вычисляют числовые признаки: энергетические, энтропийные и другие.
- Эти признаки затем подают на вход классификатору (например, k NN, SVM или другому алгоритму машинного обучения), чтобы автоматически определить уровень когнитивной нагрузки. Классификация может быть как бинарной (наличие/отсутствие нагрузки), так и многоуровневой (лёгкая/средняя/высокая нагрузка).

Таким образом, метод CiSSA был выбран в качестве одного из подходов разложения сигнала по частотам.

Выводом исследования является таблица 5 статьи [10], в которой авторы сравнивают по метрикам различные решения задачи. Подходы, связанные с CiSSA, являются не самыми наилучшими.

3.3.2. Application of visual stratigraphy from line-scan images to constrain chronology and melt features of a firn core from coastal Antarctica

В работе [11] исследовалось таяние ледников, а также анализировались данные визуальной стратиграфии (VS) для построения хронологии фирнового керна из прибрежной Антарктиды. Основной задачей было отделение долгосрочного тренда, связанного с изменением плотности фирна, от сезонных сигналов, обусловленных включениями пыли и морской соли.

Для разложения сигнала по частотам был выбран алгоритм CiSSA. Длина окна $L = 10$, поскольку «дальнейшее её увеличение не оказывало существенного влияния на результаты». Хотя профиль визуальной стратиграфии является функцией глубины, а не времени, последовательность данных вдоль керна интерпретируется как время. Это обусловлено тем, что слои снега и льда откладываются ежегодно, и глубина в этом контексте отражает хронологическую последовательность. Измерения производились с

частотой 5 сантиментов по глубине от 0 до 50 метров. Таким образом, методы обработки сигналов, разработанные для временных рядов, включая CiSSA, применимы и к профилям по глубине.

Ключевые этапы анализа:

- **Первая компонента (RC1):** отвечала за долгосрочный тренд, связанный с постепенным увеличением плотности фирна с глубиной;
- **Компоненты со второй по пятую (RC2-RC5):** отражали сезонность, обусловленную изменениями в содержании пыли и морских солей;
- **Остальные компоненты:** содержали шумовой сигнал и не учитывались в дальнейшем анализе.

Таким образом, метод CiSSA был использован как инструмент частотного разложения.

3.3.3. Выводы

В рассмотренных работах CiSSA используется как прикладной инструмент для разделения сигнала по частотам. В работе [10] это конкретные частоты, а для [11] это инструмент для выделения тренда. По целям использования алгоритмов, можно сделать вывод, что с таким же успехом можно было применить базовый SSA с автоматической группировкой по компонентам, как сказано в секции 1.2.5.

Глава 4

Метод Functional singular spectrum analysis (FSSA)

Functional SSA – метод для анализа функциональных временных рядов [4]. Предлагается объединить теорию многомерного функционального анализа главных компонент (MFPCA) и SSA.

Авторы метода FSSA сравнивают его с MSSA и динамическим функциональным анализом главных компонент (dFPCA), показывая преимущество использования FSSA в разделении компонент функциональных данных, особенно для нестационарных рядов.

В данном разделе преимущественно используются обозначения из статьи [4] авторов метода. Таким образом, например, \mathbf{X} заменяется на \mathbf{x} .

Рассмотрим, основные определения. Функциональным временным рядом будем называть $\mathbf{x} \in \mathbb{H}^k$: $\mathbf{x}(s) = (x_1(s), \dots, x_N(s))^\top$, $x_i \in \mathbb{H} = \mathcal{L}^2([0, 1])$, $s \in [0, 1]$. Скалярное произведение: $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{H}^k} = \sum_{i=1}^k \langle x_i, y_i \rangle$, $\langle x, y \rangle = \int_0^1 x(s)y(s)ds$. Также определим тензорное произведение: для $x \in \mathbb{H}_1$, $y \in \mathbb{H}_2$, оператор $x \otimes y : \mathbb{H}_1 \rightarrow \mathbb{H}_2$, $(x \otimes y)h = \langle x, h \rangle y$, $h \in \mathbb{H}_1$.

4.1. Алгоритм FSSA

Алгоритм схож по структуре с базовым SSA.

Пусть $\mathbf{x}(s) = (x_1(s), \dots, x_N(s))^\top$ – функциональный временной ряд длины N . Для целого числа $1 \leq L \leq N/2$ положим $K = N - L + 1$ и определим множество многомерных функциональных векторов в \mathbb{H}^L как

$$\mathbf{x}_j(s) := (y_j(s), y_{j+1}(s), \dots, y_{j+L-1}(s))^\top, \quad j = 1, \dots, K, \quad (4.1)$$

где \mathbf{x}_j обозначают функциональные векторы. Следующий алгоритм предоставляет результаты FSSA в четыре этапа.

Вложение

Траекторный оператор $\mathcal{X} : \mathbb{R}^K \rightarrow \mathbb{H}^L$.

$$\mathcal{X}\mathbf{a} := \sum_{j=1}^K a_j \mathbf{x}_j = \begin{pmatrix} \sum_{j=1}^K a_j y_j \\ \sum_{j=1}^K a_j y_{j+1} \\ \vdots \\ \sum_{j=1}^K a_j y_{j+L-1} \end{pmatrix}, \quad \mathbf{a} = (a_1, \dots, a_K)^\top \in \mathbb{R}^K. \quad (4.2)$$

Кроме того, $\mathcal{X} = \mathcal{T}\mathbf{x}$, где \mathcal{T} — оператор вложения. Вычисление $\mathcal{X}\mathbf{a}$ в заданной точке $s \in [0, 1]$ эквивалентно матричному произведению $\mathbf{X}(s)\mathbf{a}$, где $\mathbf{X}(s)$ — это $L \times K$ ганкелева матрица, заданная как

$$\mathbf{X}(s) = [\mathbf{x}_1(s), \dots, \mathbf{x}_K(s)]. \quad (4.3)$$

Оператор \mathcal{X} является ограниченным линейным оператором. Если определить $\mathcal{X}^* : \mathbb{H}^L \rightarrow \mathbb{R}^K$ как

$$\mathcal{X}^*\mathbf{z} = \begin{pmatrix} \sum_{i=1}^L \langle y_i, z_i \rangle \\ \sum_{i=1}^L \langle y_{i+1}, z_i \rangle \\ \vdots \\ \sum_{i=1}^L \langle y_{i+K-1}, z_i \rangle \end{pmatrix}, \quad \mathbf{z} = (z_1, \dots, z_L)^\top \in \mathbb{H}^L, \quad (4.4)$$

то \mathcal{X}^* является сопряжённым оператором для \mathcal{X} .

Разложение

Определим оператор $\mathcal{S} : \mathbb{H}^L \rightarrow \mathbb{H}^L$ как $\mathcal{S} := \mathcal{X}\mathcal{X}^*$. Следовательно, для заданного $\mathbf{z} \in \mathbb{H}^L$ это означает, что

$$\mathcal{S}\mathbf{z} = \sum_{j=1}^K \sum_{i=1}^L \langle y_{i+j-1}, z_i \rangle \mathbf{x}_j = \sum_{j=1}^K \langle \mathbf{x}_j, \mathbf{z} \rangle_{\mathbb{H}^L} \mathbf{x}_j = \sum_{j=1}^K (\mathbf{x}_j \otimes \mathbf{x}_j) \mathbf{z}. \quad (4.5)$$

Предложение 5 ([4, Раздел 3.1]). Существует ортонормированная система базисных векторов $\{\boldsymbol{\psi}_i, i \in \mathbb{N}\}$ в пространстве \mathbb{H}^L , такая что

$$\mathcal{S}\boldsymbol{\psi}_i = \lambda_i \boldsymbol{\psi}_i, \quad \text{причём } \lambda_i \rightarrow 0 \text{ при } i \rightarrow \infty.$$

Кроме того,

$$\mathcal{S} = \sum_{i=1}^{\infty} \lambda_i \boldsymbol{\psi}_i \otimes \boldsymbol{\psi}_i.$$

Для любого положительного i определим оператор $\mathcal{X}_i : \mathbb{R}^K \rightarrow \mathbb{H}^L$, заданный как

$$\mathcal{X}_i \mathbf{a} := \sum_{j=1}^K a_j (\boldsymbol{\psi}_i \otimes \boldsymbol{\psi}_i) \mathbf{x}_j = (\boldsymbol{\psi}_i \otimes \boldsymbol{\psi}_i) \sum_{j=1}^K a_j \mathbf{x}_j. \quad (4.6)$$

Вычисление $\mathcal{X}_i \mathbf{a}$ в заданной точке $s \in [0, 1]$ эквивалентно матричному произведе-

нию $\mathbf{X}_i(s)\mathbf{a}$, где $\mathbf{X}_i(s)$ — это $L \times K$ матрица, заданная как

$$\begin{aligned}\mathbf{X}_i(s) &:= \left[\langle \boldsymbol{\psi}_i, \mathbf{x}_1 \rangle_{\mathbb{H}^L} \boldsymbol{\psi}_i(s), \dots, \langle \boldsymbol{\psi}_i, \mathbf{x}_K \rangle_{\mathbb{H}^L} \boldsymbol{\psi}_i(s) \right] \\ &= \left[(\boldsymbol{\psi}_i \otimes \boldsymbol{\psi}_i) \mathbf{x}_1(s), \dots, (\boldsymbol{\psi}_i \otimes \boldsymbol{\psi}_i) \mathbf{x}_K(s) \right].\end{aligned}\quad (4.7)$$

Элементарные операторы \mathcal{X}_i раскладывают траекторный оператор \mathcal{X} как

$$\mathcal{X} = \sum_{i=1}^{\infty} \mathcal{X}_i. \quad (4.8)$$

Следующая теорема предоставляет сингулярное разложение (SVD) траекторного оператора \mathcal{X} для получения соответствующих собственных троек $(\sqrt{\lambda_i}, \mathbf{v}_i, \boldsymbol{\psi}_i)$ на этапе декомпозиции.

Теорема 4 ([4]). Пусть $\{\boldsymbol{\psi}_i\}_{i=1}^{\infty}$ и $\{\lambda_i\}_{i=1}^{\infty}$ — собственные элементы \mathcal{S} , и

$$\mathcal{S}^\dagger := \mathcal{X}^* \mathcal{X} = \begin{bmatrix} \sum_{i=1}^L \langle y_i, y_i \rangle & \cdots & \sum_{i=1}^L \langle y_i, y_{i+K-1} \rangle \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^L \langle y_{i+K-1}, y_i \rangle & \cdots & \sum_{i=1}^L \langle y_{i+K-1}, y_{i+K-1} \rangle \end{bmatrix}. \quad (4.9)$$

Тогда сингулярное разложение траекторного оператора \mathcal{X} можно записать как

$$\mathcal{X} = \sum_{i=1}^{\infty} \sqrt{\lambda_i} \mathbf{v}_i \otimes \boldsymbol{\psi}_i, \quad (4.10)$$

где $\mathbf{v}_i = \left(\frac{\langle \boldsymbol{\psi}_i, \mathbf{x}_1 \rangle_{\mathbb{H}^L}}{\sqrt{\lambda_i}}, \dots, \frac{\langle \boldsymbol{\psi}_i, \mathbf{x}_K \rangle_{\mathbb{H}^L}}{\sqrt{\lambda_i}} \right)^\top$. Кроме того, для любого $\mathbf{a} \in \mathbb{R}^K$, используя (4.10), мы имеем

$$\mathcal{X}\mathbf{a} = \sum_{i=1}^{\infty} \sqrt{\lambda_i} \langle \mathbf{v}_i, \mathbf{a} \rangle_{\mathbb{R}^K} \boldsymbol{\psi}_i, \quad (4.11)$$

где

- i) $\{\lambda_i\}_{i=1}^{\infty}$ — множество неубывающих собственных значений \mathcal{S}^\dagger , и
- ii) \mathbf{v}_i — соответствующие ортонормированные собственные векторы \mathcal{S}^\dagger , удовлетворяющие $\mathcal{X}\mathbf{v}_i = \sqrt{\lambda_i} \boldsymbol{\psi}_i$.

Группировка

Аналогично базовому SSA. Получается

$$\mathcal{X} = \mathcal{X}_{I_1} + \mathcal{X}_{I_2} + \dots + \mathcal{X}_{I_m}. \quad (4.12)$$

Диагональное усреднение

Для заданного q ($1 \leq q \leq m$) используется \mathcal{T}^{-1} для преобразования оператора \mathcal{X}_{I_q} из (4.12) в $\tilde{\mathbf{x}}^q$.

Элементы \mathcal{X}_{I_q} и $\tilde{\mathcal{X}}_{I_q}$ обозначаются $[x_{i,j}^q]$ и $[\tilde{x}_{i,j}^q]$. Метод диагонального усреднения обобщается на $\mathbb{H}^{L \times K}$, где

$$\tilde{x}_{i,j}^q = \frac{1}{n_s} \sum_{(l,k): l+k=s} x_{l,k}^q, \quad (4.13)$$

$s = i + j$, n_s — число пар (l, k) , таких что $l + k = s$.

Как итог: $\tilde{\mathbf{x}}^q = \mathcal{T}^{-1} \tilde{\mathcal{X}}_{I_q}$.

4.2. Практическая реализация алгоритма FSSA

Для использования на практике алгоритма FSSA, нужно понять, как использовать дискретные данные с теорией, основанной на непрерывности по s .

Преобразование

На вход алгоритму приходят дискретные данные $\{x_j(s_k)\}_{j=1, \dots, N}^{k=1, \dots, n}$ в виде матрицы. После чего они преобразуются в функции $x_j(s) \in \mathbb{H}$ на $[0, 1]$ с использованием базисного разложения:

$$x_j(s) = \sum_{i=1}^d a_{i,j} \nu_i(s),$$

где $\{\nu_i\}_{i=1}^d$ — базис (например, В-сплайны), d является параметром алгоритма. Получается функциональный временной ряд $\mathbf{x} = \{x_1(s), \dots, x_N(s)\}$.

Вложение

Задаётся L . Формируются векторы:

$$\mathbf{x}_j(s) = (x_j(s), \dots, x_{j+L-1}(s))^\top, \quad j = 1, \dots, K, \quad K = N - L + 1.$$

Траекторный оператор $\mathcal{X} : \mathbb{R}^K \rightarrow \mathbb{H}^L$ представлен матрицей $\mathbf{X}(s) = [\mathbf{x}_1(s), \dots, \mathbf{x}_K(s)]$, вычисляемой через базисные коэффициенты $a_{i,j}$.

Разложение

Вычисляется оператор $\mathcal{S} = \mathcal{X}\mathcal{X}^*$, представленный матрицей \mathbf{S} в базисе \mathbb{H}_d^L . Формируется базис $\{\phi_k\}_{k=1}^{Ld}$, где ϕ_k — вектор с ν_{q_k} на позиции r_k , $k = (q_k - 1)L + r_k$. Вычисляются матрицы:

$$\mathbf{G} = [\delta_{r_i, r_j} \langle \nu_{q_i}, \nu_{q_j} \rangle]_{i,j=1}^{Ld}, \quad \mathbf{S}_0 = \left[\sum_{m=1}^K \langle x_{r_i+m-1}, \nu_{q_i} \rangle \langle x_{r_j+m-1}, \nu_{q_j} \rangle \right]_{i,j=1}^{Ld}.$$

$$\mathbf{S} = \mathbf{G}^{-1} \mathbf{S}_0.$$

Находятся собственные значения λ_i и векторы \mathbf{c}_{ψ_i} матрицы \mathbf{S} . Собственные функции: $\psi_i = \sum_{k=1}^{Ld} (\mathbf{c}_{\psi_i})_k \phi_k$. Вычисляются элементарные операторы $\mathcal{X}_i = (\psi_i \otimes \psi_i) \mathcal{X}$.

Группировка

Аналогично теоретической группировке (4.12).

$$\mathcal{X} = \mathcal{X}_{I_1} + \mathcal{X}_{I_2} + \dots + \mathcal{X}_{I_m}.$$

Реконструкция

Для \mathcal{X}_{I_q} выполняется диагональное усреднение:

$$\tilde{x}_{i,j}^q = \frac{1}{n_s} \sum_{(l,k): l+k=i+j} x_{l,k}^q,$$

где $x_{l,k}^q$ — элементы \mathcal{X}_{I_q} , n_s — количество пар (l, k) , равных $i + j$. Получаем $\tilde{\mathcal{X}}_{I_q} = \Pi_{\mathbb{H}} \mathcal{X}_{I_q}$. Затем применяется \mathcal{T}^{-1} : $\tilde{\mathbf{x}}_N^q = \mathcal{T}^{-1} \tilde{\mathcal{X}}_{I_q}$, выраженное через базисные коэффициенты.

Глава 5

Сравнения алгоритмов

5.1. Сравнение SSA и GSSA

В данном разделе сравниваются алгоритмы базового SSA и GSSA с параметром $\alpha \neq 0$. Все вычисления, а также код метода GSSA можно найти в репозитории [12].

5.1.1. Линейные фильтры

Чтобы понять отличие, рассмотрим методы с точки зрения линейных фильтров: по представлениям (1.4) и (2.2) можно построить амплитудно-частотные характеристики.

Рассмотрим временной ряд

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

Построим АЧХ для первых двух компонент при α , равных 0 (базовый SSA), $\frac{1}{2}$, 1, 2.

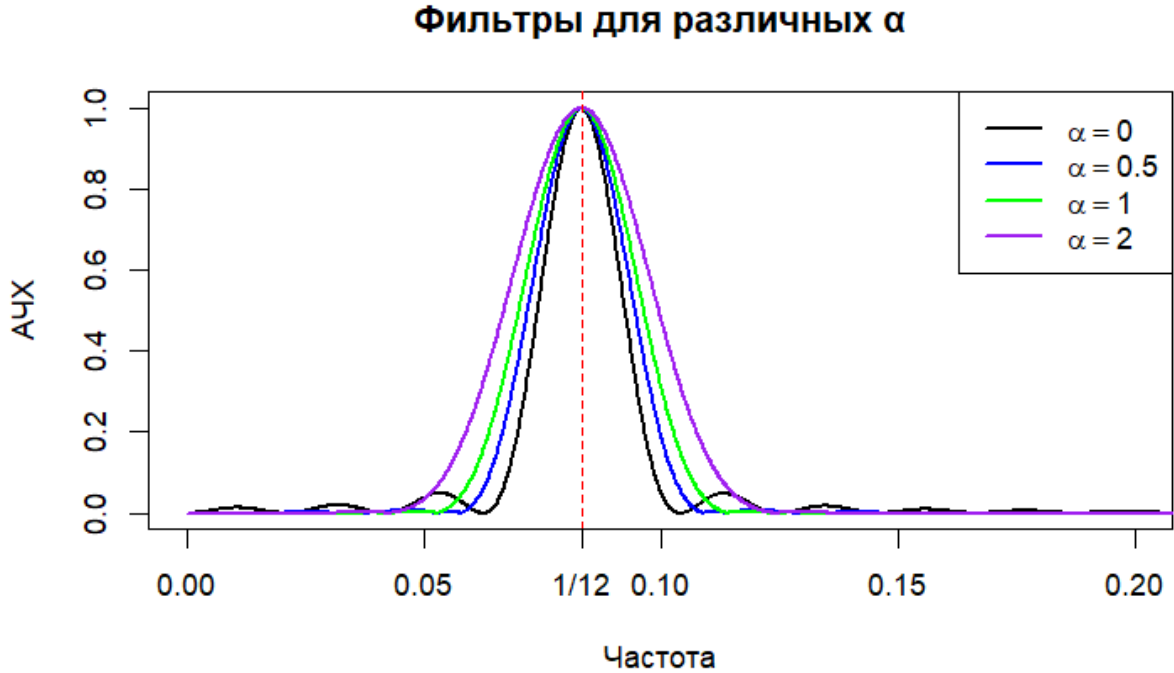


Рис. 5.1. АЧХ фильтров для первых двух компонент SVD ряда $X = \left\{ \sin \left(\frac{2\pi}{12} n \right), n = 1, \dots, N \right\}$ при разных α .

На рисунке 5.1 показано, как фильтры ведут себя для различных значений параметра α . Для всех рассмотренных значений α фильтры подавляют частоты, значительно отличающиеся от частоты синуса $\omega = \frac{1}{12}$. При малых значениях α , таких как $\alpha = 0$, наблюдается волнообразное поведение фильтра, что указывает на частичное захватывание соседних частот, хотя и не близких к частоте синуса. С увеличением α это волнообразное поведение уменьшается, но фильтр начинает захватывать больше частот, близких к $\frac{1}{12}$.

Таким образом, метод **GSSA** должен работать лучше **SSA** в случае, когда временной ряд содержит пара периодических функций, частота одной из которых попадает в вершину волны АЧХ фильтра для другой функции. Например, добавим к $X_{\sin} = \{\sin(\frac{2\pi}{12}n), n = 1, \dots, N\}$ косинус с частотой $\frac{1}{19}$. Тогда

$$X = X_{\sin} + X_{\cos} = \left\{ \sin\left(\frac{2\pi}{12}n\right) + \frac{1}{2} \cos\left(\frac{2\pi}{19}n\right), \quad n = 1, \dots, N \right\},$$

и можем рассмотреть АЧХ для первых двух компонент разложения (синуса) при базовом **SSA** ($\alpha = 0$) и **GSSA** при $\alpha = \frac{1}{2}$. $N = 96 \cdot 2 - 1$, $L = 48$.

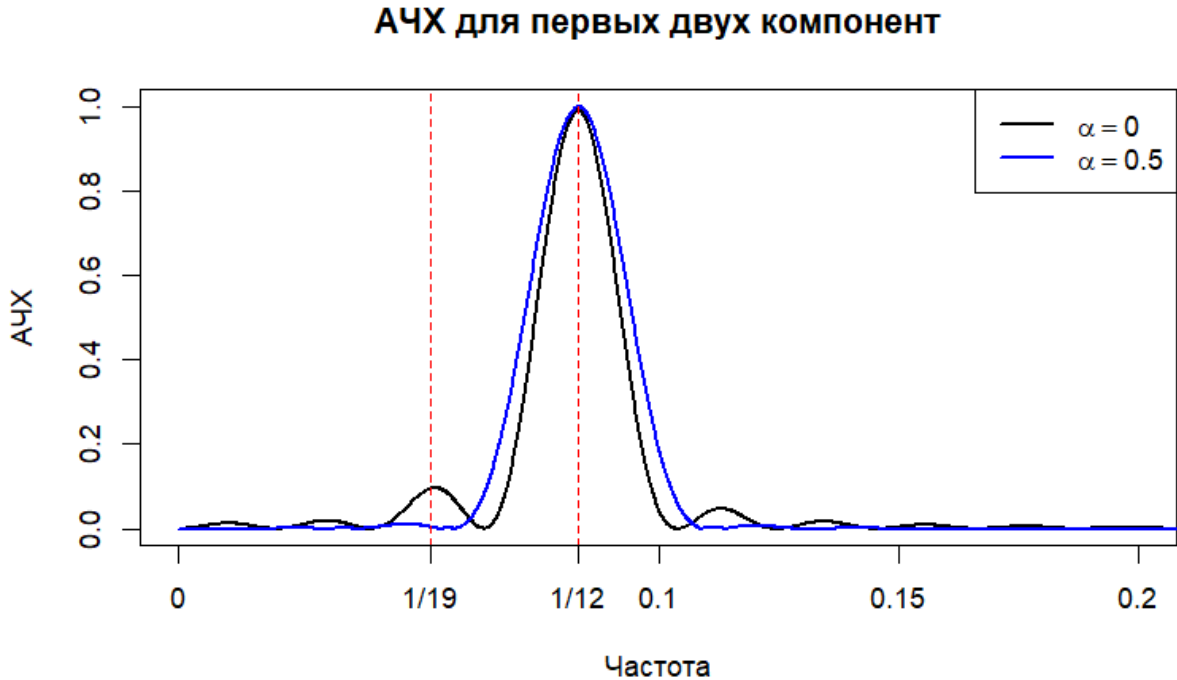


Рис. 5.2. Ряд $X = X_{\sin} + X_{\cos}$. АЧХ фильтров для первых двух компонент (синуса), при разных α .

По рисунку 5.2 заметно, что фильтр для синуса в базовом SSA также частично захватит периодику с частотой $\frac{1}{19}$, в то время, как GSSA не будет испытывать таких проблем.

5.1.2. Фильтры в различных точках

В зависимости от точек ряда, линейные фильтры будут отличаться друг от друга. Рассмотрим тот же пример $X = X_{\sin} + X_{\cos} + X_{\text{noise}} = \sin\left(\frac{2\pi}{12}x\right) + \frac{1}{2}\cos\left(\frac{2\pi}{19}x\right) + \varepsilon_n$, где $\varepsilon_n \sim N(0, 0.1^2)$.

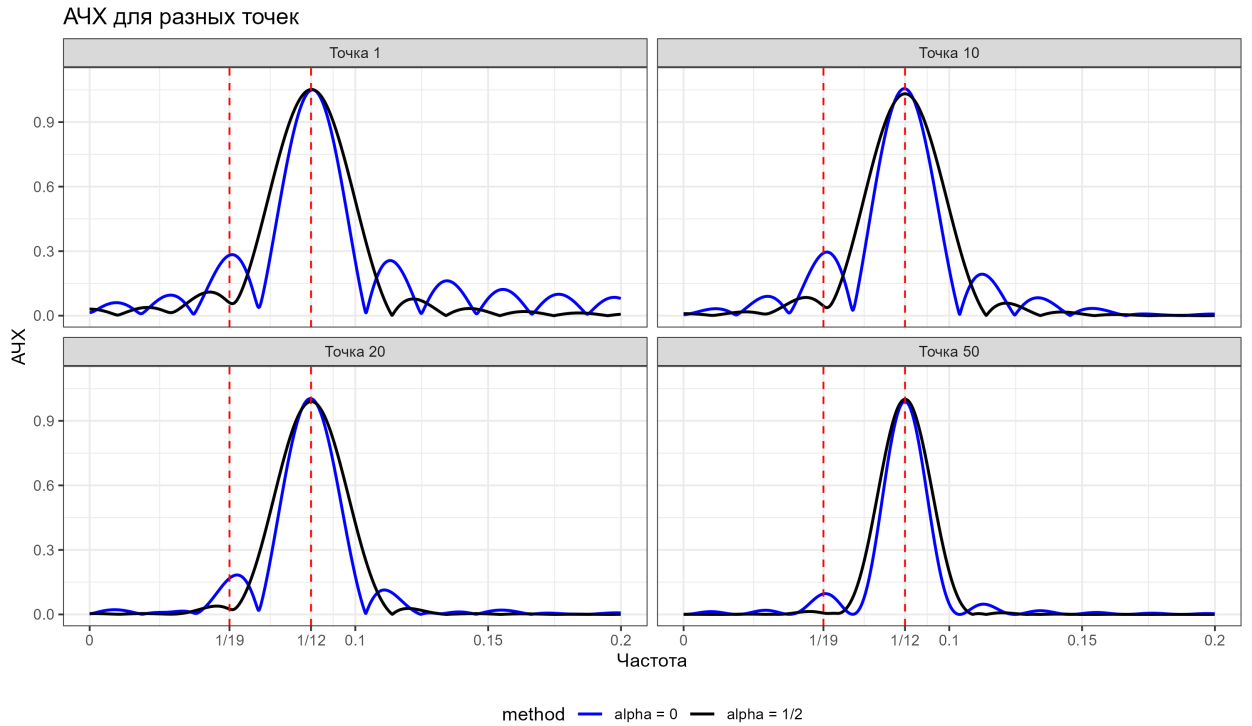


Рис. 5.3. Ряд $X = X_{\sin} + X_{\cos} + X_{\text{noise}}$. АЧХ фильтров в разных точках, для первых двух компонент при разных α .

По рисунку 5.3 видно, что когда точка s приближается по времени к средним точкам временного ряда ($L \leq s \leq K$), полоса пропускания фильтра становится уже, а также фильтр начинает все меньше и меньше захватывать соседние частоты.

Для этого примера также можно посмотреть на график средней MSE ошибки в зависимости от точки ряда. Эксперимент проводился 1000 раз.

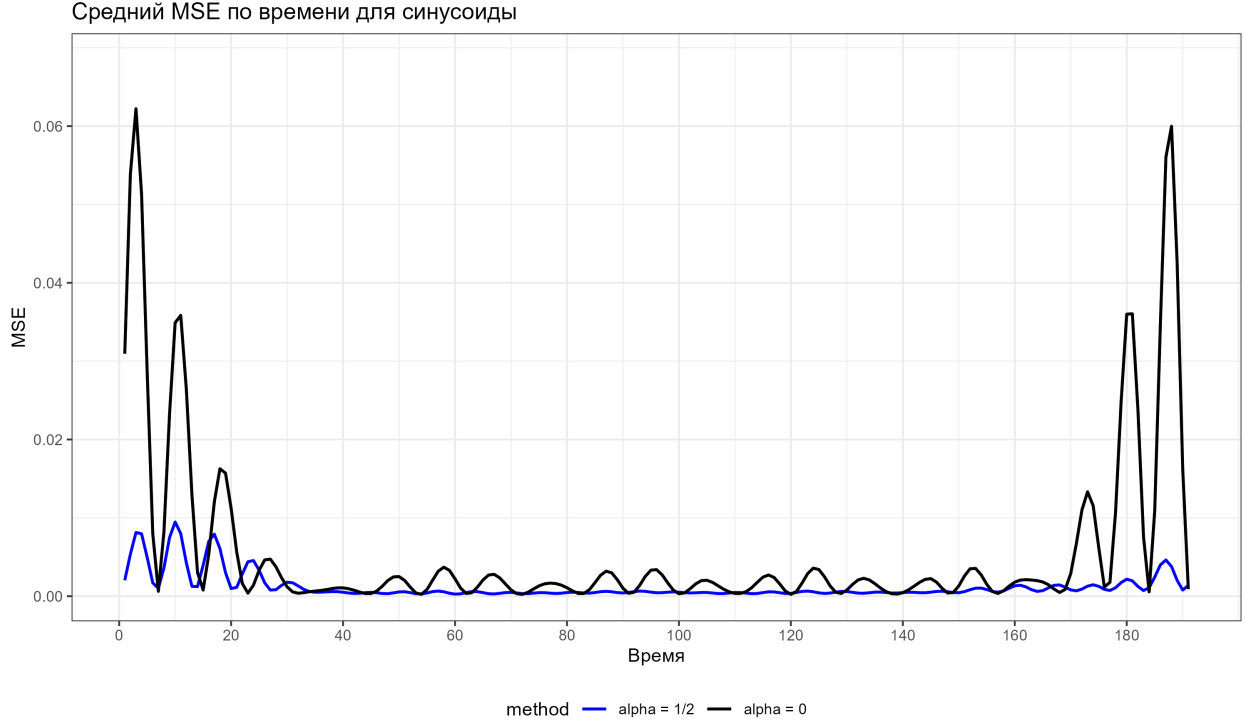


Рис. 5.4. Ряд $X = X_{\sin} + X_{\cos} + X_{\text{noise}}$. MSE ошибки для первых двух компонент в зависимости от точки ряда.

Таким образом, можно сделать вывод, что разделимость также зависит от точки ряда. В средних точках достигаются наилучшие значения ошибки. Однако это не означает, что нужно брать маленькое L , поскольку чем больше длина окна, тем лучше происходит разделение компонент между собой в целом [1].

5.1.3. Влияние шума

Теперь посмотрим на общую ошибку для примера $X = X_{\sin} + X_{\cos}$.

Таблица 5.1. MSE разложений ряда $X = X_{\sin} + X_{\cos}$ для SSA и GSSA с $\alpha = \frac{1}{2}$.

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

В таблице 5.1 жирным выделены наилучшие результаты ошибки. По ней видно, что GSSA справился с разделением на порядок лучше SSA.

Однако, у GSSA есть другая проблема. Если добавить к ряду шум, то оба алгоритма будут относить к синусу часть шума, пропускаемую фильтром, а для GSSA диапазон

пропускаемых частот шире, поэтому для **GSSA** больше шума попадет в компоненты, отвечающие за периодики.

Добавим к \mathbf{X} шумовую компоненту:

$$\mathbf{X} = \mathbf{X}_{\sin} + \mathbf{X}_{\cos} + \mathbf{X}_{\text{noise}} = \left\{ \sin\left(\frac{2\pi}{12}n\right) + \frac{1}{2} \cos\left(\frac{2\pi}{19}n\right) + \varepsilon_n, n = 1, \dots, N \right\},$$

где $\varepsilon_n \sim N(0, 0.1^2)$, $N = 96 \cdot 2 - 1$, $L = 48$. Проводилось 100 тестов, в таблице 5.2 указаны средние значения ошибки для одних и тех же реализаций шума.

Таблица 5.2. MSE разложений ряда $\mathbf{X} = \mathbf{X}_{\sin} + \mathbf{X}_{\cos} + \mathbf{X}_{\text{noise}}$ для **SSA** и **GSSA** с $\alpha = \frac{1}{2}$.

Метод	\mathbf{X}_{\sin}	\mathbf{X}_{\cos}	\mathbf{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

В таблице 5.2 жирным отмечены наилучшие результаты ошибки. По ней видно, что **GSSA** все же справился лучше **SSA**, однако порядок ошибки теперь одинаковый для рассмотрения косинуса или синуса. Но при этом, отделение сигнала от шума получилось лучше у **SSA**. Также был проведен парный t-критерий для зависимых выборок с целью проверки гипотезы о равенстве средних значений ошибки для каждой компоненты. В качестве нулевой гипотезы (H_0) предполагалось, что средние значения сравниваемых выборок из величин ошибок равны. Уровень значимости был установлен на уровне $\alpha_{\text{hypothesis}} = 0.05$. Результаты анализа показали, что во всех случаях p -значение оказалось меньше 0.05, что позволяет отвергнуть нулевую гипотезу.

5.1.4. Вложенный вариант

Сингулярное разложение матрицы обладает наилучшими аппроксимационными свойствами в смысле минимизации нормы Фробениуса (или спектральной нормы) для заданного ранга. Из этого следует, что разложение матрицы $\mathbf{S} = \mathbf{X}\mathbf{X}^T$ методом SVD будет наилучшим образом отделять сигнал от шума.

По результатам сравнений методов из таблиц 5.1 и 5.2, а также предыдущих рассуждений, получается, что есть смысл использовать базовый **SSA** для выделения сигнала, а уже сам сигнал разделять на различные компоненты с помощью **GSSA**.

Рассмотрим

$$\mathbf{X} = \mathbf{X}_{\sin} + \mathbf{X}_{\cos} + \mathbf{X}_{\text{noise}} = \left\{ \sin\left(\frac{2\pi}{12}n\right) + \frac{1}{2} \cos\left(\frac{2\pi}{19}n\right) + \varepsilon_n, n = 1, \dots, N \right\},$$

где $\varepsilon_n \sim N(0, 0.1^2)$, только теперь сначала применим **SSA**, а затем **GSSA**. Получится вложенный вариант. $N = 96 \cdot 2 - 1$, $L = 48$ для обоих методов.

Таблица 5.3. $X_{\sin} + X_{\cos} + \varepsilon_n$, $\varepsilon_n \sim N(0, 0.1^2)$, MSE оценок.

Метод/Ошибка	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
SSA + GSSA , $\alpha = 0.5$	1.06e-03	1.12e-03	7.15e-04

Анализ данных, представленных в таблице 5.3, позволяет сделать вывод, что комбинирование алгоритмов привело к улучшению как в выделении полезного сигнала, так и в разделении компонент между собой. Полученные результаты демонстрируют более высокую точность по сравнению с данными, приведёнными в таблице 5.2.

Таким образом, по приведенным примерам можно сделать вывод, что **GSSA** позволяет улучшить разделимость периодических компонент ряда. Однако, вместе с тем, разложение будет захватывать больше шума в сравнении с базовым **SSA**.

5.2. Различия SSA, CiSSA и Фурье

В данной секции проводится сравнение методов разложения временного ряда: базовый SSA, SSA с использованием EOSSA для улучшения разделимости, разложения Фурье и Фурье с расширением ряда, базового CiSSA и CiSSA с расширением ряда. Все вычисления, а также код метода CiSSA можно найти в репозитории [12]

5.2.1. Автоматическая группировка

Авторы статьи [3] выделяют главным преимуществом то, что CiSSA автоматически разделяет компоненты ряда по частотам. Однако есть метод, позволяющий сделать автоматическое объединение частот по периодограмме в методе SSA [13]. При этом, прежде чем применять его, стоит выполнить процедуру улучшения разделимости. В данной работе будут использоваться методы EOSSA и FOSSA [6]. Отсюда следует, что все рассматриваемые в данной секции алгоритмы могут по заранее заданным диапазонам частотам выдать временные ряды, отвечающие за эти диапазоны.

5.2.2. Собственные пространства

Каждый алгоритм после группировки порождает собственные подпространства. В случае базового SSA базис подпространств является адаптивным, то есть зависящим от X, L, N .

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

На базис разложения Фурье влияет только N .

5.2.3. Выбор длины окна для CiSSA

Фиксируем временной ряд $X = (x_1, \dots, x_N)$ длины N , а также набор интересующих частот $\Omega = \{\omega_k : \omega_k \in [0, 0.5]\}$ и соответствующие диапазоны Ω_j : $\bigsqcup_{j=1}^m \Omega_j = \bigsqcup_{j=1}^m [\omega_j^{(l)}, \omega_j^{(r)}] = [0, 0.5], \omega_k \in \Omega_k$.

Возникает вопрос: какое L лучше выбрать?

Без шума

Если ряд состоит только из синусов и косинусов, то по условиям точной разделимости разницы нет, ошибка всегда будет нулевая. Однако, если это не так, то из всех

возможных L нужно выбирать наименьшее подходящее, поскольку чем оно меньше, тем более локальные свойства CiSSA будет извлекать из ряда.

В подтверждение этому рассмотрим пример

$$X = X_{e\cdot\sin} + X_{e\cdot\cos} = \{e^{A_1 n} \sin(2\pi\omega_1 n) + e^{A_2 n} \cos(2\pi\omega_2 n), n = 1, \dots, N\}$$

с $A_1 = \frac{1}{200}$, $A_2 = \frac{1}{100}$, $\omega_1 = \frac{1}{12}$, $\omega_2 = \frac{1}{3}$, $N = 96 \cdot 2$. Будем следить за ошибкой в зависимости от длины окна. В качестве L рассмотрим значения, кратные частотам компонентам сигнала. Конкретнее, $L = 12, 24, 36, 48, 60, 72, 84, 96, 192$.

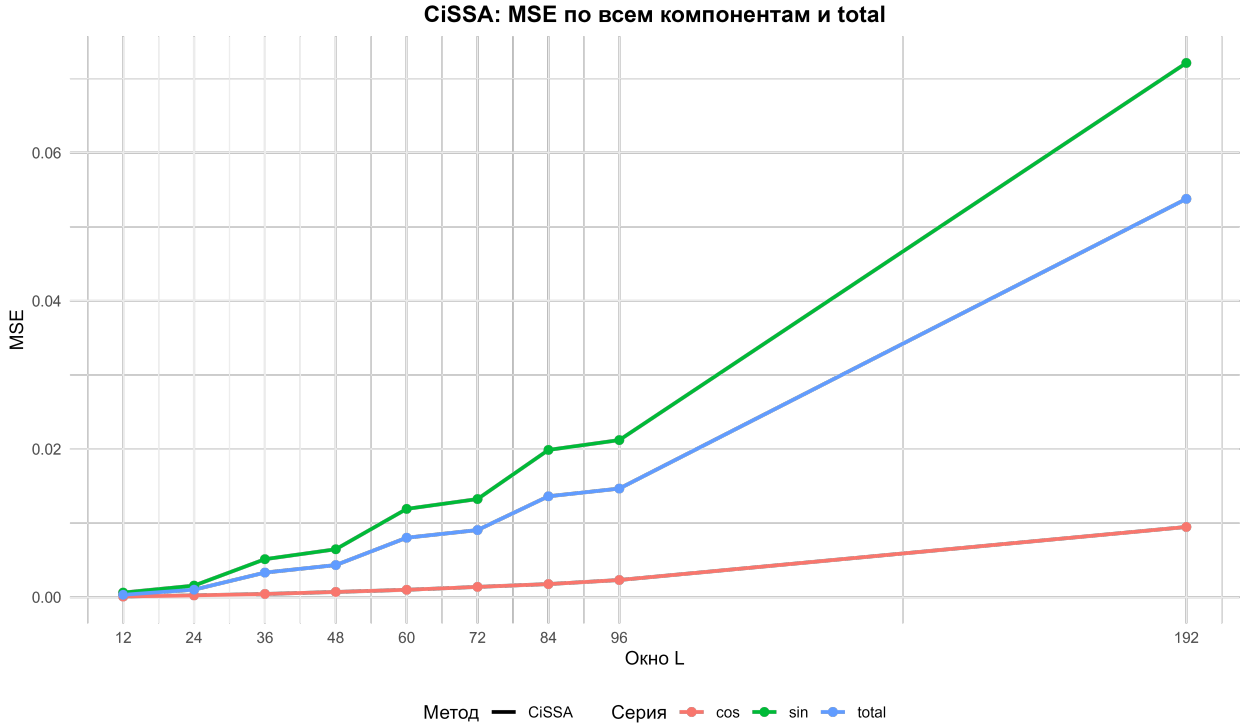


Рис. 5.5. MSE разделения CiSSA ряда $X = X_{e\cdot\sin} + X_{e\cdot\cos}$ в зависимости от длины окна L .

Рисунок 5.5 наглядно показывает, что при фиксированном N и с ростом L ошибка увеличивается. При этом максимума ошибка достигает при $L = 192$, что соответствует обычному применению преобразования Фурье. Кроме того, здесь не важно, по каким диапазонам частот будет произведена группировка, поскольку нет шума. Рассмотрим теперь его влияние.

С шумом

Теперь ошибка для суммы двух гармоник не такие однозначные. Рассмотрим пример с гауссовским шумом.

$$\mathbf{X} = \mathbf{X}_{\sin} + \mathbf{X}_{\cos} + \mathbf{X}_{\text{noise}} = \{A_1 \sin(2\pi\omega_1 n) + A_2 \cos(2\pi\omega_2 n) + \varepsilon_n, n = 1, \dots, N\},$$

где $\varepsilon_n \sim N(0, 1)$, $A_1 = 1$, $A_2 = 0.5$, $\omega_1 = \frac{1}{12}$, $\omega_2 = \frac{1}{3}$, $N = 96 \cdot 2$. Восстановление проводилось по диапазонам $[\omega_i \pm 2 \cdot \frac{1}{193}]$. Было проведено 100 экспериментов и результаты усреднялись для соответствующих компонент и L . В качестве L рассмотрим значения, кратные частотам компонентам сигнала. Конкретнее, $L = 12, 24, 36, 48, 60, 72, 84, 96, 192$.

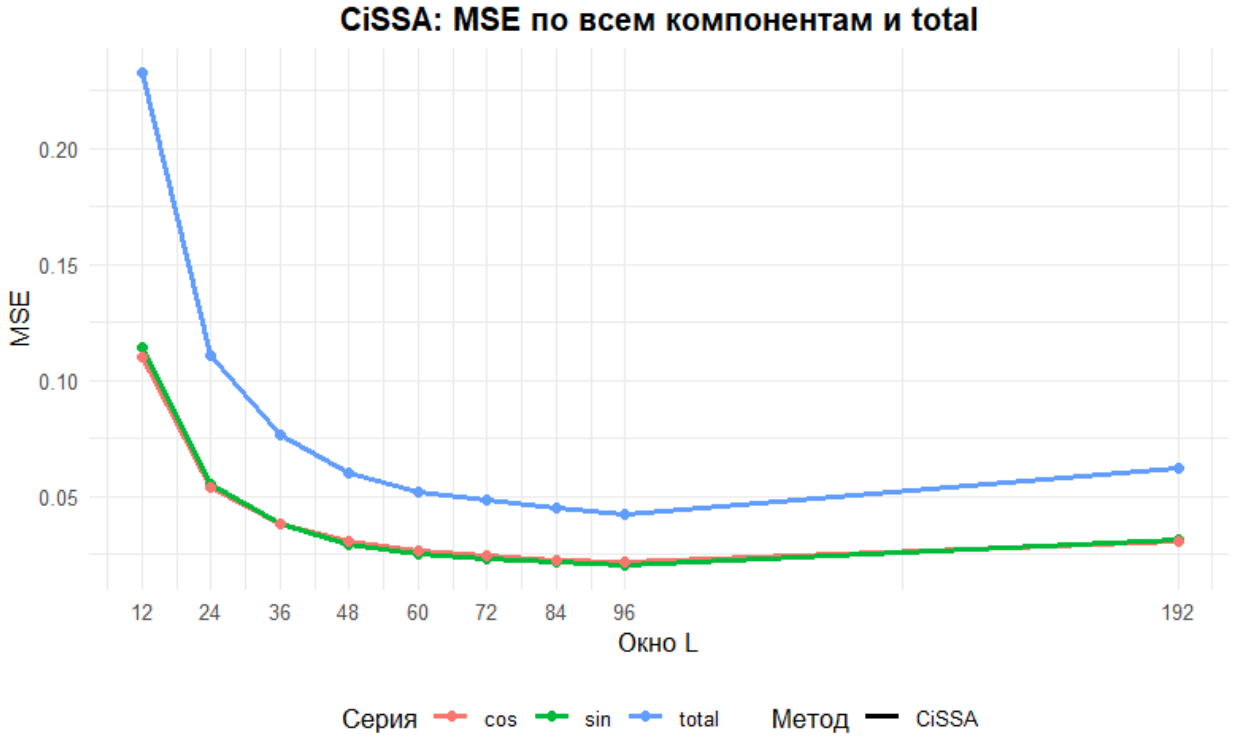


Рис. 5.6. MSE разделения CiSSA ряда $\mathbf{X} = \mathbf{X}_{\sin} + \mathbf{X}_{\cos} + \mathbf{X}_{\text{noise}}$ в зависимости от длины окна L .

Как видно по рисунку 5.6, ситуация поменялась. Ошибка начинает снижаться с ростом длины окна. Это основано на смещении и дисперсии оценки (3.2.2). Дисперсия оценки снижается, а смещение в данном случае никак не меняется с ростом окна.

Теперь рассмотрим пример с экспоненциальной модуляцией:

$$\mathbf{X} = \mathbf{X}_{e\text{-sin}} + \mathbf{X}_{e\text{-cos}} = \{e^{A_1 n} \sin(2\pi\omega_1 n) + e^{A_2 n} \cos(2\pi\omega_2 n) + \varepsilon_n, n = 1, \dots, N\}.$$

$\varepsilon_n \sim N(0, 1)$ с $A_1 = \frac{1}{200}$, $A_2 = \frac{1}{100}$, $\omega_1 = \frac{1}{12}$, $\omega_2 = \frac{1}{3}$, $N = 96 \cdot 2$. Восстановление проводилось по диапазонам $[\omega_i \pm 4 \cdot \frac{1}{193}]$. Было проведено 100 экспериментов и результаты усреднялись для соответствующих компонент и L . В качестве L рассмотрим значения, кратные

частотам компонентам сигнала. Конкретнее, $L = 12, 24, 36, 48, 60, 72, 84, 96, 192$.

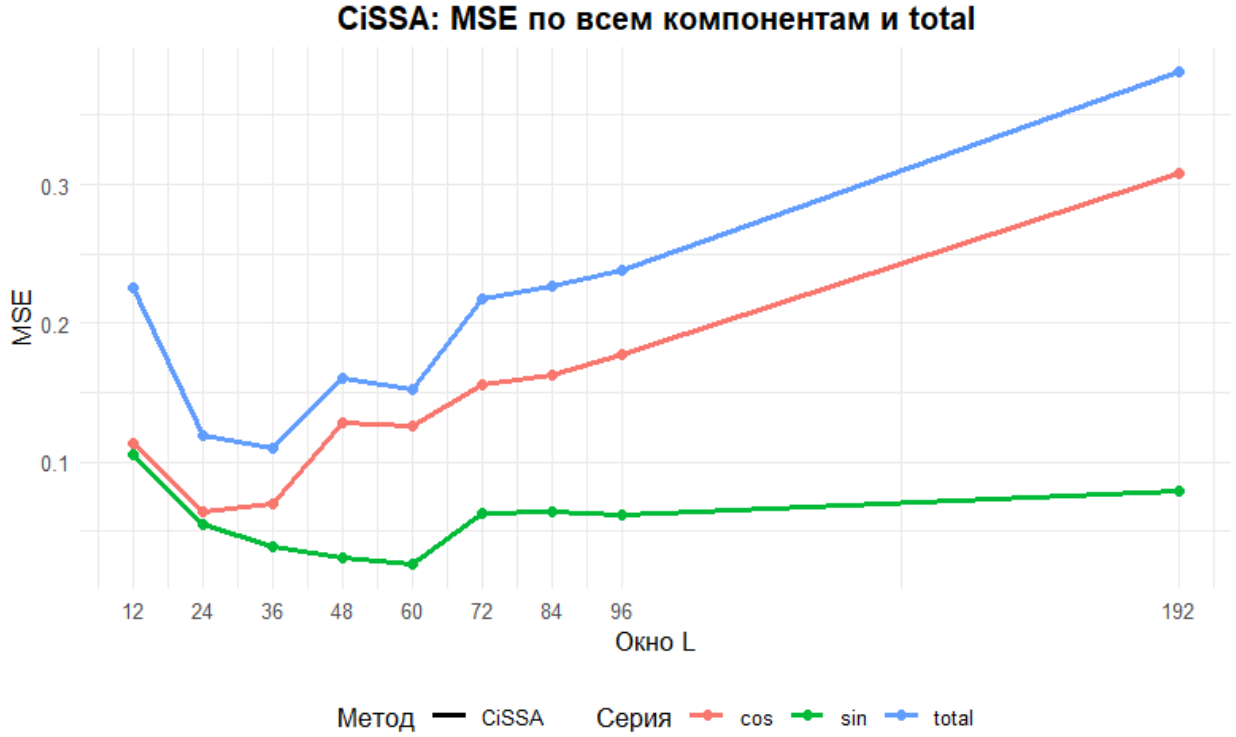


Рис. 5.7. MSE разделения CiSSA ряда $X = X_{e\text{-sin}} + X_{e\text{-cos}} + \varepsilon_n$ в зависимости от длины окна L с $[\omega_i \pm 4 \cdot \frac{1}{193}]$.

По рисунку 5.7 можно увидеть, как борются два эффекта — увеличение смещения и уменьшение дисперсии оценки при росте L .

Кроме того, здесь проявляется проблема фиксированного базиса из раздела 3.2.2. Если взять тот же пример

$$X = X_{e\text{-sin}} + X_{e\text{-cos}} = \{e^{A_1 n} \sin(2\pi\omega_1 n) + e^{A_2 n} \cos(2\pi\omega_2 n) + \varepsilon_n, n = 1, \dots, N\},$$

где $\varepsilon_n \sim N(0, 1)$ с $A_1 = \frac{1}{200}$, $A_2 = \frac{1}{100}$, $\omega_1 = \frac{1}{12}$, $\omega_2 = \frac{1}{3}$, $N = 96 \cdot 2$, но восстановление проводить по диапазонам $[\omega_i \pm 2 \cdot \frac{1}{193}]$, то картина ошибки будет выглядеть иначе.

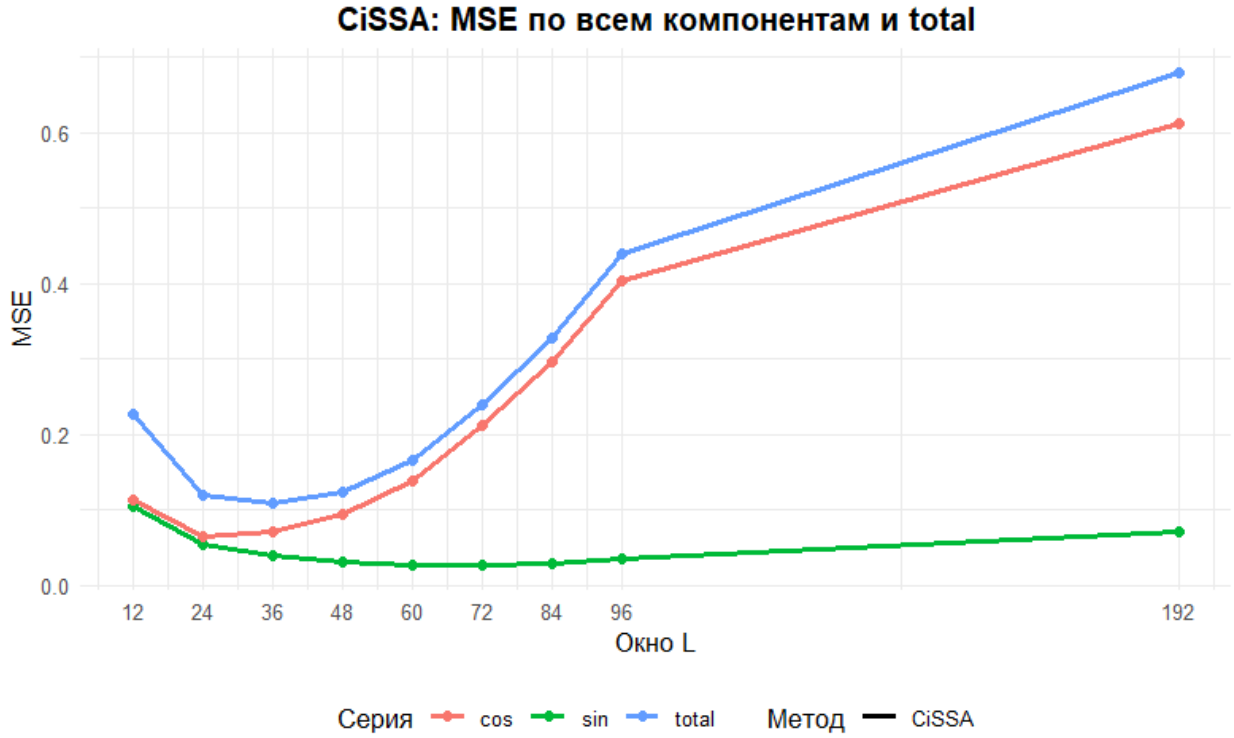


Рис. 5.8. MSE разделения CiSSA ряда $X = X_{e,\sin} + X_{e,\cos} + \varepsilon_n$ в зависимости от длины окна L с $[\omega_i \pm 2 \cdot \frac{1}{193}]$.

Из рисунков 5.7 и 5.8 можно сделать вывод, что в силу неадаптивности базиса исследователю нужно очень аккуратно выбирать диапазоны частот, по которым следует объединять компоненты.

5.2.4. Без шума

В данном разделе будем рассматривать результаты сравнений для временного ряда, состоящего только из сигнала.

Гармонические функции

В свойствах методов были приведены классы функций и условия, при которых методы могут безошибочно разделить два ряда друг от друга. Сравним эти условия.

Фиксируем временной ряд, состоящий из двух периодических компонент с постоянной модуляцией $X = X_1 + X_2 = A_1 \cos(2\pi\omega_1 n + \varphi_1) + A_2 \cos(2\pi\omega_2 n + \varphi_2)$.

Таблица 5.4. Условия точной разделимости для SSA, SSA EOSSA, Fourier, CiSSA.

Метод	Условия точной разделимости
SSA	$L\omega_1, L\omega_2, K\omega_1, K\omega_2 \in \mathbb{N}, \omega_1 \neq \omega_2, A_1 \neq A_2$
SSA EOSSA	$\omega_1 \neq \omega_2$
Fourier	$N\omega_1, N\omega_2 \in \mathbb{N}, \omega_1 \neq \omega_2$
CiSSA	$L\omega_1, L\omega_2 \in \mathbb{N}, \omega_1 \neq \omega_2$

Из таблицы 5.4 видно, что для самые жесткие условия для SSA, а наилучшие у SSA EOSSA. Для Fourier и CiSSA условия схожи, однако тут стоит отметить, что длину ряда N далеко не всегда можно подобрать, в отличие от L . В этом смысле лучше себя показывает CiSSA в сравнении с Fourier.

Теперь посмотрим численно, на следующие ситуации: параметры рассматриваемых методов подобраны верно, параметры подобраны плохо.

Пример. Будем разделять временной ряд

$$\mathbf{X} = \mathbf{X}_{\sin} + \mathbf{X}_{\cos} = \left\{ \sin \frac{2\pi}{12}n + \frac{1}{2} \cos \frac{2\pi}{3}n, n = 1, \dots, N \right\}.$$

Таблица 5.5. MSE разложений $\mathbf{X} = \mathbf{X}_{\sin} + \mathbf{X}_{\cos}$.

Метод	Условия	MSE(\mathbf{X}_{\sin})	MSE(\mathbf{X}_{\cos})	MSE(\mathbf{X})
SSA	(1)	6.8e-30	1.5e-29	1.8e-29
SSA	(2)	2.2e-06	2.2e-06	2.0e-29
SSA EOSSA	(3)	7.5e-30	1.5e-29	2.0e-29
SSA EOSSA	(4)	8.8e-30	1.5e-29	1.9e-29
Fourier	(5)	3.4e-28	1.2e-28	4.3e-28
Fourier	(6)	1.2e-03	3.1e-03	4.1e-03
CiSSA	(7)	5.5e-30	1.4e-29	8.6e-30
CiSSA	(8)	5.5e-30	1.6e-29	1.4e-29
CiSSA	(9)	2.2e-03	5.7e-03	6.2e-03

Таблица 5.6. Параметры для таблицы 5.5.

Условия	Параметры
(1)	$L = 96, N = 191$. Группы: $X_{\sin} - [1, 2]; X_{\cos} - [3, 4]$.
(2)	$L = 96, N = 192$. Группы: $X_{\sin} - [1, 2]; X_{\cos} - [3, 4]$.
(3)	$L = 96, N = 191, r = 4$. Частоты: $X_{\sin} - [\omega_1 \pm \varepsilon]; X_{\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 6/192$.
(4)	$L = 96, N = 192, r = 4$. Частоты: $X_{\sin} - [\omega_1 \pm \varepsilon]; X_{\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 6/192$.
(5)	$N = 192$. Частоты: $X_{\sin} - [\omega_1 \pm \varepsilon]; X_{\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 6/192$.
(6)	$N = 192$. Частоты: $X_{\sin} - [\omega_1 \pm \varepsilon]; X_{\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 6/192$.
(7)	$L = 12, N = 192$. Частоты: $X_{\sin} - [\omega_1 \pm \varepsilon]; X_{\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 6/192$.
(8)	$L = 96, N = 192$. Частоты: $X_{\sin} - [\omega_1 \pm \varepsilon]; X_{\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 6/192$.
(9)	$L = 97, N = 192$. Частоты: $X_{\sin} - [\omega_1 \pm \varepsilon]; X_{\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 6/192$.

По результатам таблицы 5.5 видно, что все алгоритмы при нарушении условий точной разделимости перестают отделять компоненты друг от друга с точностью машинного нуля. Однако в следствии частотной разделимости, ошибка все равно мала. Меньше всего при нарушении условий точной разделимости ошибка для **SSA**, на 3 порядка хуже у **CiSSA** и **Fourier**.

Экспоненциально-модулированные функции

Как следствие асимптотической разделимости можно рассмотреть следующий пример.

Пример. Будем разделять временной ряд

$$X = X_{e\text{-}\sin} + X_{e\text{-}\cos} = \{e^{n/200} \sin(2\pi n/12) + e^{n/100} \cos(2\pi n/3), n = 1, \dots, N\}.$$

Таблица 5.7. MSE разложений $X = X_{e.\sin} + X_{e.\cos}$.

Метод	Условия	MSE(X_{\sin})	MSE(X_{\cos})	MSE(X)
SSA	(1)	5.3e-05	5.3e-05	1.2e-27
SSA	(2)	1.8e-04	1.8e-04	1.1e-27
SSA EOSSA	(3)	8.3e-29	1.2e-27	1.1e-27
SSA EOSSA	(4)	8.4e-29	1.2e-27	1.1e-27
Fourier	(5)	1.5e-02	8.6e-02	7.9e-02
Fourier	(6)	9.4e-03	1.2e-01	1.3e-01
CiSSA	(7)	5.7e-05	1.2e-03	1.0e-03
CiSSA	(8)	3.1e-03	4.2e-02	3.5e-02
CiSSA	(9)	1.2e-02	1.5e-01	1.1e-01

Таблица 5.8. Параметры для таблицы 5.7.

Условия	Параметры
(1)	$L = 96, N = 191$. Группы: $X_{e.\sin} - [3, 4]; X_{e.\cos} - [1, 2]$.
(2)	$L = 96, N = 192$. Группы: $X_{e.\sin} - [3, 4]; X_{e.\cos} - [1, 2]$.
(3)	$L = 96, N = 191, r = 4$. Частоты: $X_{e.\sin} - [\omega_1 \pm \varepsilon]; X_{e.\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 10/192$.
(4)	$L = 96, N = 192, r = 4$. Частоты: $X_{e.\sin} - [\omega_1 \pm \varepsilon]; X_{e.\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 10/192$.
(5)	$N = 192$. Частоты: $X_{e.\sin} - [\omega_1 \pm \varepsilon]; X_{e.\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 10/192$.
(6)	$N = 191$. Частоты: $X_{e.\sin} - [\omega_1 \pm \varepsilon]; X_{e.\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 10/192$.
(7)	$L = 12, N = 192$. Частоты: $X_{e.\sin} - [\omega_1 \pm \varepsilon]; X_{e.\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 10/192$.
(8)	$L = 97, N = 192$. Частоты: $X_{e.\sin} - [\omega_1 \pm \varepsilon]; X_{e.\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 10/192$.
(9)	$L = 96, N = 192$. Частоты: $X_{e.\sin} - [\omega_1 \pm \varepsilon]; X_{e.\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 10/192$.

По таблице 5.7 видно, что CiSSA все же разделяет ряды, но не так хорошо, как это делает SSA или тем более SSA с EOSSA. Совсем плохо показал себя Fourier. Также видим, что с маленьким L CiSSA справляется с разделением лучше, чем с большим.

Функции с трендом

Рассмотрим, влияние непериодических компонент на разложение ряда.

Базовый алгоритм **SSA** может выделять трендовую составляющую за счет своего адаптивного базиса. Для алгоритмов **CiSSA** и разложения Фурье нужно применять процедуры расширения временного ряда, чтобы использовать их для выделения тренда.

Пример. Рассмотрим ряд из предыдущего примера и добавим к нему тренд.

$$X = X_c + X_e + X_{\sin} + X_{\cos} = \left\{ 1 + e^{\frac{n}{100}} + \sin \frac{2\pi}{12}n + \frac{1}{2} \cos \frac{2\pi}{3}n, n = 1, \dots, N \right\},$$

$$X_{\text{trend}} = X_c + X_e.$$

Таблица 5.9. MSE разложений X на три компоненты.

Метод	Условия	MSE(X_{trend})	MSE(X_{\sin})	MSE(X_{\cos})	MSE(X)
SSA	(1)	6.1e-05	5.2e-05	8.9e-07	2.9e-28
SSA	(2)	7.3e-05	6.2e-05	4.2e-06	8.8e-28
SSA EOSSA	(3)	8.2e-28	1.1e-29	1.8e-29	8.1e-28
SSA EOSSA	(4)	3.3e-28	7.3e-30	1.8e-29	3.1e-28
Fourier	(5)	2.0e-01	5.0e-02	4.3e-03	1.4e-01
Fourier	(6)	2.2e-01	6.3e-02	8.8e-03	1.5e-01
Fourier _{ext}	(7)	4.8e-03	2.6e-03	1.9e-04	7.7e-03
Fourier _{ext}	(8)	5.1e-03	3.0e-03	1.8e-04	7.6e-03
CiSSA	(9)	2.7e-03	8.5e-04	3.1e-05	5.6e-04
CiSSA	(10)	9.6e-02	4.4e-03	1.5e-04	5.8e-02
CiSSA	(11)	1.1e-01	1.4e-02	9.9e-03	6.5e-02
CiSSA _{ext}	(12)	1.0e-04	7.6e-05	2.9e-06	5.2e-06
CiSSA _{ext}	(13)	4.7e-04	2.0e-04	2.6e-05	3.1e-04
CiSSA _{ext}	(14)	4.9e-04	2.3e-04	6.5e-04	8.8e-04

Таблица 5.10. Параметры для таблицы 5.9.

Условия	Параметры
(1)	$L = 96, N = 191$. Группы: $X_{\text{trend}} = [1, 6], X_{\text{sin}} = [2, 3], X_{\text{cos}} = [4, 5]$.
(2)	$L = 96, N = 192$. Группы: $X_{\text{trend}} = [1, 6], X_{\text{sin}} = [2, 3], X_{\text{cos}} = [4, 5]$.
(3)	$L = 96, N = 192, r = 6$; Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(4)	$L = 96, N = 191, r = 6$; Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(5)	$N = 192$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(6)	$N = 191$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(7)	$N = 192$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(8)	$N = 191$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(9)	$L = 12, N = 192$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(10)	$L = 96, N = 192$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(11)	$L = 97, N = 192$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(12)	$L = 12, N = 192$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(13)	$L = 96, N = 192$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(14)	$L = 97, N = 192$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.

По таблице 5.9 видно, что алгоритмы **CiSSA** и Фурье без модификаций достаточно плохо определяют тренд. Ситуация с выделением тренда улучшается при использовании процедуры расширения ряда или использовании маленькой длины окна.

При этом, лучше всего себя показывают **SSA** и **SSA EOSSA**.

5.2.5. С шумом

Теперь пройдемся по тем же примерам, но добавим к ним шумовую составляющую

Гармонические функции

Пример. Будем разделять временной ряд

$$\mathbf{X} = \mathbf{X}_{\sin} + \mathbf{X}_{\cos} + \mathbf{X}_{\text{noise}} = \left\{ \sin \frac{2\pi}{12}n + \frac{1}{2} \cos \frac{2\pi}{3}n + \varepsilon_n, n = 1, \dots, N \right\},$$

где $\varepsilon_n \sim N(0, 0.1^2)$. Проводилось 100 экспериментов с усреднением результата.

Таблица 5.11. MSE разложений $\mathbf{X} = \mathbf{X}_{\sin} + \mathbf{X}_{\cos} + \mathbf{X}_{\text{noise}}$.

Метод	Условия	MSE(\mathbf{X}_{\sin})	MSE(\mathbf{X}_{\cos})	MSE(\mathbf{X})
SSA	(1)	2.7e-04	3.4e-04	6.2e-04
SSA	(2)	2.7e-04	3.4e-04	6.1e-04
SSA EOSSA	(3)	3.1e-04	4.3e-04	7.5e-04
SSA EOSSA	(4)	2.8e-04	4.0e-04	6.8e-04
Fourier	(5)	4.1e-04	6.3e-04	1.0e-03
Fourier	(6)	2.8e-03	7.0e-03	9.7e-03
CiSSA	(9)	1.2e-03	1.3e-03	2.5e-03
CiSSA	(10)	4.3e-04	6.1e-04	1.0e-03
CiSSA	(11)	3.8e-03	9.2e-03	1.2e-02

Таблица 5.12. Параметры для таблицы 5.11.

Условия	Параметры
(1)	$L = 96, N = 191$. Группы: $\mathbf{X}_{\sin} - [1, 2]; \mathbf{X}_{\cos} - [3, 4]$.
(2)	$L = 96, N = 192$. Группы: $\mathbf{X}_{\sin} - [1, 2]; \mathbf{X}_{\cos} - [3, 4]$.
(3)	$L = 96, N = 191, r = 4$. Частоты: $\mathbf{X}_{\sin} - [\omega_1 \pm \varepsilon]; \mathbf{X}_{\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 3/192$.
(4)	$L = 96, N = 192, r = 4$. Частоты: $\mathbf{X}_{\sin} - [\omega_1 \pm \varepsilon]; \mathbf{X}_{\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 3/192$.
(5)	$N = 192$. Частоты: $\mathbf{X}_{\sin} - [\omega_1 \pm \varepsilon]; \mathbf{X}_{\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 3/192$.
(6)	$N = 192$. Частоты: $\mathbf{X}_{\sin} - [\omega_1 \pm \varepsilon]; \mathbf{X}_{\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 3/192$.
(7)	$L = 12, N = 192$. Частоты: $\mathbf{X}_{\sin} - [\omega_1 \pm \varepsilon]; \mathbf{X}_{\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 3/192$.
(8)	$L = 96, N = 192$. Частоты: $\mathbf{X}_{\sin} - [\omega_1 \pm \varepsilon]; \mathbf{X}_{\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 3/192$.
(9)	$L = 97, N = 192$. Частоты: $\mathbf{X}_{\sin} - [\omega_1 \pm \varepsilon]; \mathbf{X}_{\cos} - [\omega_2 \pm \varepsilon], \varepsilon = 3/192$.

По таблице 5.11 можно сделать те же выводы, что и без шума.

Экспоненциально-модулированные функции

Пример. Будем разделять временной ряд

$$\mathbf{X} = \mathbf{X}_{e\cdot\sin} + \mathbf{X}_{e\cdot\cos} + \mathbf{X}_{\text{noise}} = \{e^{n/200} \sin(2\pi n/12) + e^{n/100} \cos(2\pi n/3) + \varepsilon_n, n = 1, \dots, N\},$$

где $\varepsilon_n \sim N(0, 0.1^2)$. Проводилось 100 экспериментов с усреднением результата.

Таблица 5.13. MSE разложений $\mathbf{X} = \mathbf{X}_{e\cdot\sin} + \mathbf{X}_{e\cdot\cos} + \mathbf{X}_{\text{noise}}$.

Метод	Условия	MSE(\mathbf{X}_{\sin})	MSE(\mathbf{X}_{\cos})	MSE(\mathbf{X})
SSA	(1)	2.3e-04	2.3e-04	3.8e-04
SSA	(2)	3.4e-04	3.3e-04	3.7e-04
SSA EOSSA	(3)	1.9e-04	2.9e-04	4.8e-04
SSA EOSSA	(4)	1.7e-04	2.8e-04	4.5e-04
Fourier	(5)	1.6e-02	2.2e-01	2.3e-01
Fourier	(6)	1.8e-02	1.5e-01	1.6e-01
CiSSA	(7)	1.3e-03	2.5e-03	3.7e-03
CiSSA	(8)	6.5e-03	8.8e-02	8.8e-02
CiSSA	(9)	1.5e-02	2.8e-01	2.6e-01

Таблица 5.14. Параметры для таблицы 5.13.

Условия	Параметры
(1)	$L = 96, N = 191$. Группы: $\mathbf{X}_{e\cdot\sin} - [3, 4]$; $\mathbf{X}_{e\cdot\cos} - [1, 2]$.
(2)	$L = 96, N = 192$. Группы: $\mathbf{X}_{e\cdot\sin} - [3, 4]$; $\mathbf{X}_{e\cdot\cos} - [1, 2]$.
(3)	$L = 96, N = 191, r = 4$. Частоты: $\mathbf{X}_{e\cdot\sin} - [\omega_1 \pm \varepsilon]$; $\mathbf{X}_{e\cdot\cos} - [\omega_2 \pm \varepsilon]$, $\varepsilon = 6/192$.
(4)	$L = 96, N = 192, r = 4$. Частоты: $\mathbf{X}_{e\cdot\sin} - [\omega_1 \pm \varepsilon]$; $\mathbf{X}_{e\cdot\cos} - [\omega_2 \pm \varepsilon]$, $\varepsilon = 6/192$.
(5)	$N = 192$. Частоты: $\mathbf{X}_{e\cdot\sin} - [\omega_1 \pm \varepsilon]$; $\mathbf{X}_{e\cdot\cos} - [\omega_2 \pm \varepsilon]$, $\varepsilon = 6/192$.
(6)	$N = 191$. Частоты: $\mathbf{X}_{e\cdot\sin} - [\omega_1 \pm \varepsilon]$; $\mathbf{X}_{e\cdot\cos} - [\omega_2 \pm \varepsilon]$, $\varepsilon = 6/192$.
(7)	$L = 12, N = 192$. Частоты: $\mathbf{X}_{e\cdot\sin} - [\omega_1 \pm \varepsilon]$; $\mathbf{X}_{e\cdot\cos} - [\omega_2 \pm \varepsilon]$, $\varepsilon = 6/192$.
(8)	$L = 97, N = 192$. Частоты: $\mathbf{X}_{e\cdot\sin} - [\omega_1 \pm \varepsilon]$; $\mathbf{X}_{e\cdot\cos} - [\omega_2 \pm \varepsilon]$, $\varepsilon = 6/192$.
(9)	$L = 96, N = 192$. Частоты: $\mathbf{X}_{e\cdot\sin} - [\omega_1 \pm \varepsilon]$; $\mathbf{X}_{e\cdot\cos} - [\omega_2 \pm \varepsilon]$, $\varepsilon = 6/192$.

С добавлением шума результаты, ожидаемо, ухудшились. Однако можно заметить,

что SSA наилучшим образом разделяет как сигнал от шума, так и компоненты между собой. Остальные методы имеют точность меньше.

Кроме того, в сравнении с этим же примером, но без шума, видно, что маленькое L для CiSSA уже не дает значительного улучшения ошибки разделения.

Функции с трендом

Пример. Рассмотрим ряд из предыдущего примера и добавим к нему тренд.

$$\mathbf{X} = \mathbf{X}_c + \mathbf{X}_e + \mathbf{X}_{\sin} + \mathbf{X}_{\cos} + \mathbf{X}_{\text{noise}} = \{1 + e^{\frac{n}{100}} + \sin \frac{2\pi}{12}n + \frac{1}{2} \cos \frac{2\pi}{3}n + \varepsilon_n, n = 1, \dots, N\},$$

где $\varepsilon_n \sim N(0, 0.1^2)$. Проводилось 100 экспериментов с усреднением результата.

Таблица 5.15. MSE разложений \mathbf{X} на три компоненты.

Метод	Условия	MSE($\mathbf{X}_{\text{trend}}$)	MSE(\mathbf{X}_{\sin})	MSE(\mathbf{X}_{\cos})	MSE(\mathbf{X})
SSA	(1)	2.0e-04	2.8e-04	1.8e-04	5.8e-04
SSA	(2)	1.9e-04	3.0e-04	1.8e-04	5.5e-04
SSA EOSSA	(3)	2.4e-04	2.1e-04	2.6e-04	7.1e-04
SSA EOSSA	(4)	2.0e-04	2.1e-04	2.5e-04	6.6e-04
Fourier	(5)	2.0e-01	5.1e-02	5.1e-03	1.5e-01
Fourier	(6)	2.2e-01	6.4e-02	9.3e-03	1.5e-01
Fourier _{ext}	(7)	5.0e-03	3.1e-03	9.1e-04	8.7e-03
Fourier _{ext}	(8)	5.9e-03	4.2e-03	8.5e-04	9.5e-03
CiSSA	(9)	3.0e-03	2.2e-03	1.1e-03	3.5e-03
CiSSA	(10)	9.5e-02	5.1e-03	5.9e-04	5.9e-02
CiSSA	(11)	1.1e-01	1.5e-02	1.1e-02	6.6e-02
CiSSA _{ext}	(12)	6.9e-04	1.7e-03	1.1e-03	3.3e-03
CiSSA _{ext}	(13)	1.3e-03	1.1e-03	5.6e-04	2.1e-03
CiSSA _{ext}	(14)	1.3e-03	1.4e-03	1.5e-03	3.1e-03

Таблица 5.16. Параметры для таблицы 5.15.

Условия	Параметры
(1)	$L = 96, N = 191$. Группы: $X_{\text{trend}} = [1, 6], X_{\text{sin}} = [2, 3], X_{\text{cos}} = [4, 5]$.
(2)	$L = 96, N = 192$. Группы: $X_{\text{trend}} = [1, 6], X_{\text{sin}} = [2, 3], X_{\text{cos}} = [4, 5]$.
(3)	$L = 96, N = 192$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(4)	$L = 96, N = 191$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(5)	$N = 192$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(6)	$N = 191$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(7)	$N = 192$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(8)	$N = 191$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(9)	$L = 12, N = 192$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(10)	$L = 96, N = 192$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(11)	$L = 97, N = 192$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(12)	$L = 12, N = 192$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(13)	$L = 96, N = 192$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.
(14)	$L = 97, N = 192$. Частоты: $X_{\text{trend}} = [0, 1/24], X_{\text{sin}} = [1/12 \pm 4\varepsilon], X_{\text{cos}} = [1/3 \pm 4\varepsilon]$.

Заметно, что с трендом хуже всего справляется **Fourier**. А лучше всех себя показывает **SSA** и **SSA EOSSA** как в смысле разделимости компонент между собой, так и в разделении сигнала от шума.

Кроме того, аналогично экспоненциально-модулированному примеру с шумом видно, что уменьшение L не дает значительного улучшения в разделимости.

5.2.6. Разделение трендовых составляющих между собой

Как удалось выяснить, все рассматриваемые алгоритмы могут выделять трендовую составляющую из ряда. Однако лишь **SSA** способен различить между собой две неперiodических компоненты. Про скорость асимптотической разделимости для **SSA**

можно подробнее узнать в книге [1]. Методы **CiSSA** и Фурье никаким образом не смогут отличить две непериодики между друг другом, поскольку они объединяют компоненты только по частотам. А двум непериодикам соответствуют одинаковые (низкие) наборы частот.

5.2.7. Преимущества и недостатки методов **SSA**, Фурье и **CiSSA**

Из ранее рассмотренных примеров и теоретических результатов можно сделать следующие выводы:

- Для **CiSSA** условия точной разделимости слабее, чем для **SSA** (3.2.3);
- Процедура расширения ряда улучшает выделение тренда для **CiSSA** и **Fourier**, однако **SSA** справляется с этой задачей лучше (5.2.5);
- **CiSSA** показывает результаты лучше, чем обычный **Fourier** (5.2.3);
- За счет фиксированного базиса в **CiSSA** возникают такие проблемы, как утечка спектра и трудность при выборе диапазонов частот для объединения (3.2.2);
- Базовый **SSA** численно не хуже, чем **CiSSA**, а шум выделяет лучше (5.2.4, 5.2.5);
- **SSA EOSSA** исправляет недостатки базового **SSA**, значительно улучшая результаты (5.2.4, 5.2.5).

5.2.8. Демонстрация работы алгоритмов на реальных данных

Теперь рассмотрим реальные данные — месячные ряды промышленного производства (Industrial Production, IP), index 2010 = 100, в США. 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. Выборка охватывает период с января 1970 года по сентябрь 2014 года, поэтому размер выборки составляет $N = 537$. Источником данных является база данных IMF. Эти показатели демонстрируют различные тенденции, сезонность и цикличность (периодические компоненты, которые соответствуют циклам бизнеса). Данные IP также рассматривались в статье [3]. Применим как **CiSSA** с расширением ряда, так и **SSA** с автоматическим определением частот и улучшениями разделимости **EOSSA** и **FOSSA** с параметром $r = 30$ 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 $L = 192$.

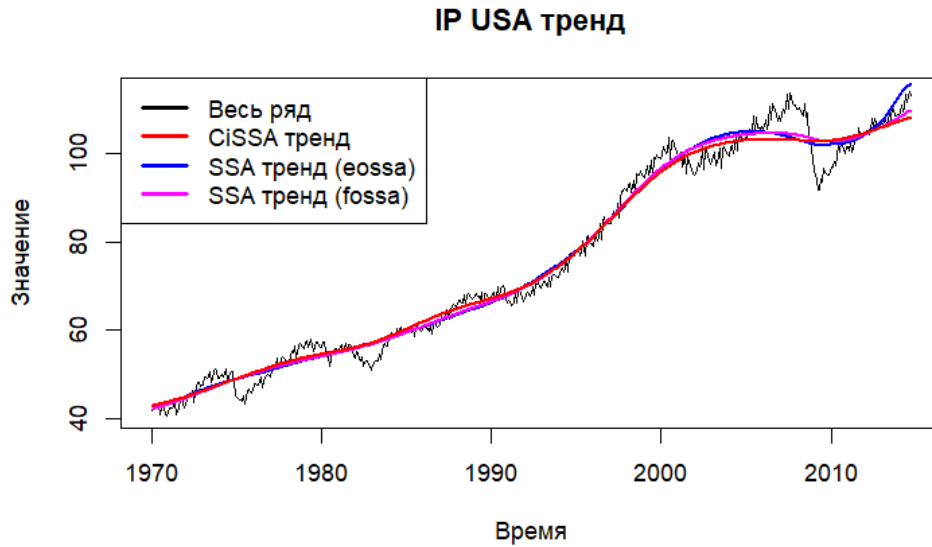


Рис. 5.9. Трендовая составляющая данных IP USA.

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.

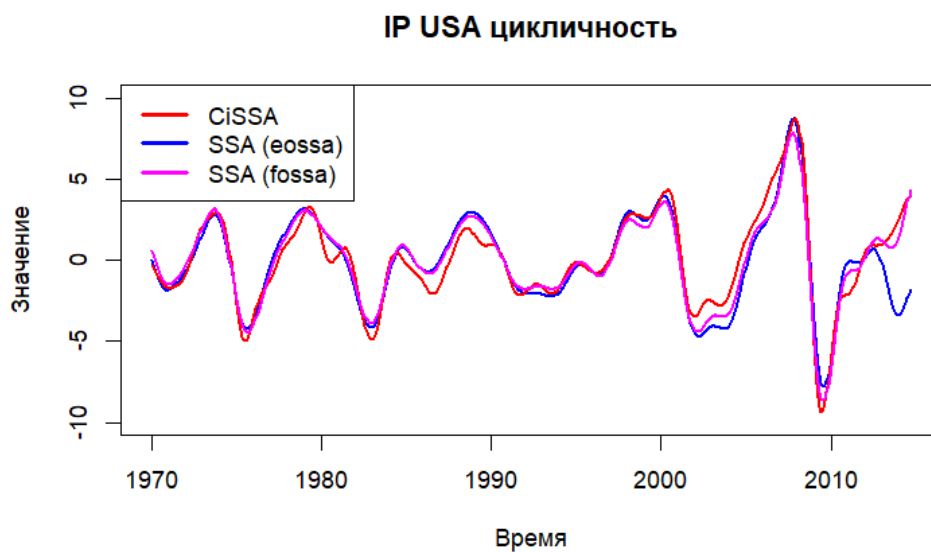


Рис. 5.10. Циклическая составляющая данных IP USA.

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

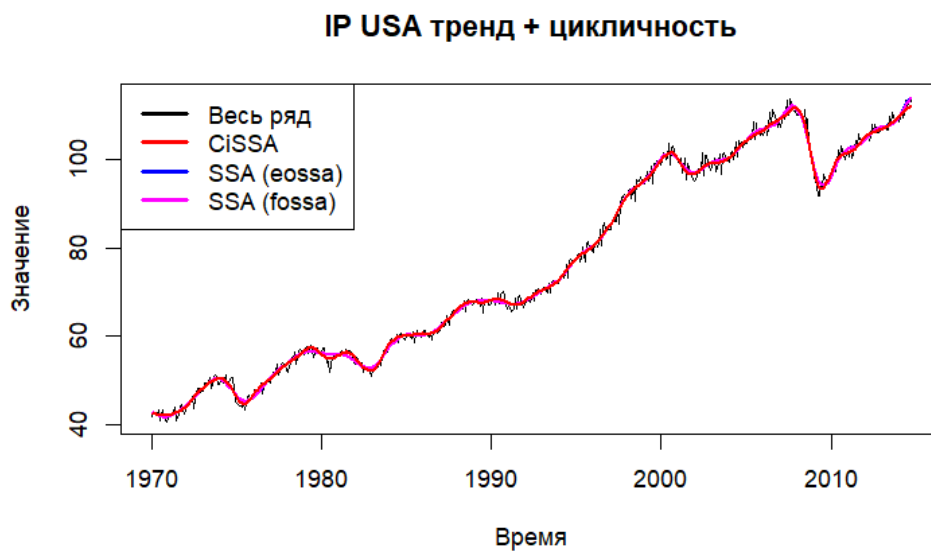


Рис. 5.11. Объединение тренда и цикличности IP USA.

Как видно из графика 5.11, объединив тренд и цикличность получаем одинаковые результаты для всех рассматриваемых алгоритмов.

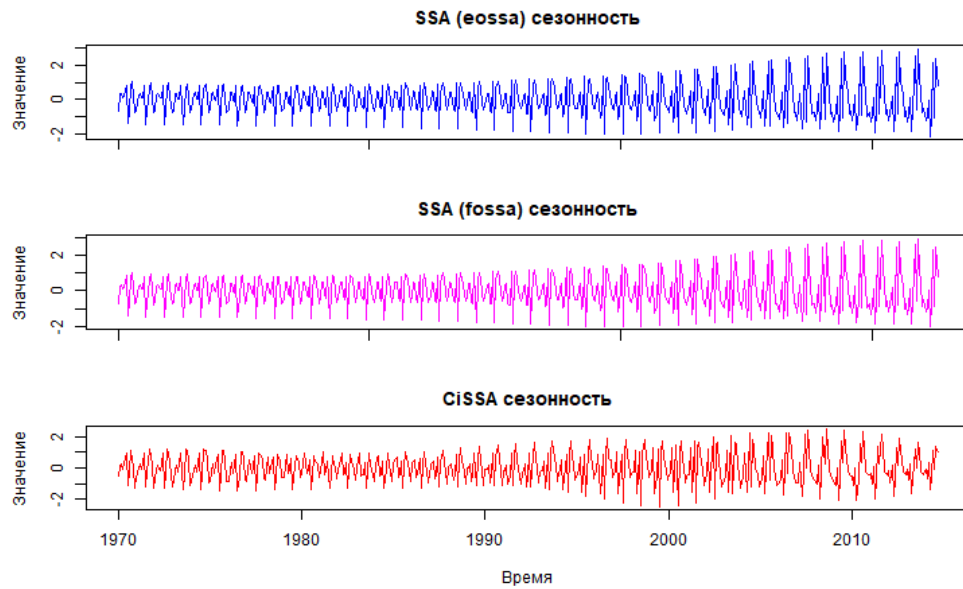


Рис. 5.12. Сезонная составляющая данных IP USA.

Сезонность выглядит схоже для EOSSA и FOSSA. Несколько иначе для CiSSA.

Шум же является нормальным во всех случаях.

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.

5.3. Сравнение FSSA, MSSA, 2d-SSA

В оригинальной работе FSSA [4] рассматривалось сравнение алгоритма с dFPCA, MSSA. Однако, 2d-SSA в данном сравнении не было, хотя он может показать результаты лучше, чем MSSA, за счет того, что в 2d-SSA можно рассматривать зависимость по второй переменной.

5.3.1. Численное сравнение

Рассматриваются функциональные временные ряды длины $N = 50, 100, 150$ и 200 , наблюдаемые в $M = 100$ фиксированных равноудаленных дискретных точках

$$s_i = \frac{i-1}{M-1}, \quad i = 1, 2, \dots, M$$

на $[0, 1]$ по следующей модели:

$$X_t(s_i) = m_t(s_i) + \varepsilon_t(s_i), \quad s_i \in [0, 1], i = 1, \dots, M \text{ и } t = 1, \dots, N. \quad (5.1)$$

Для преобразования $\{X_t(s_i)\}$ в гладкие (непрерывные) функциональные кривые применяется кубическая В-сплайн базисная функция с 15 степенями свободы. В данной модели $m_t(s)$ является периодической компонентой, определенной как

$$m_t(s) = e^{s^2} \cos(2\pi\omega t) + \cos(4\pi s) \sin(2\pi\omega t), \quad (5.2)$$

где ω — частота модели с двумя значениями ($\omega = 0.1$ и 0.25).

В этом случае $\{\varepsilon_t(s_i), t = 1, \dots, N \text{ и } i = 1, \dots, n\}$ генерируется из независимого гауссовского белого шума (GWN) с нулевым средним и стандартным отклонением равным 0.1.

Для сравнения производительности методов FSSA и MSSA рассматриваются три значения длины окна вдоль времени: $L = 20, 30$ и 40 . Для 2d-SSA вдоль s длина окна равна 50. Также в алгоритмах MSSA и FSSA используются первые две собственные компоненты, а для 2d-SSA первые 8 собственных компонент. Разделения сравниваются по RMSE:

$$RMSE = \sqrt{\frac{1}{N \times n} \sum_{t=1}^N \sum_{i=1}^n \left(X_t(s_i) - \hat{X}_t(s_i) \right)^2},$$

где $\hat{X}_t(s_i)$ — функциональный временной ряд, реконструированный каждым методом. Эксперимент повторяется 100 раз, и считается среднее значение RMSE.

Таблица 5.17. Результаты сравнения методов 2d-SSA, MSSA и FSSA при различных параметрах ω и N .

ω	N	$L = 20$			$L = 30$			$L = 40$		
		2D-SSA	MSSA	FSSA	2D-SSA	MSSA	FSSA	2D-SSA	MSSA	FSSA
0.10	50	0.009	0.028	0.009	0.010	0.026	0.010	0.011	0.021	0.013
	100	0.009	0.027	0.008	0.007	0.023	0.006	0.006	0.020	0.006
	150	0.008	0.027	0.005	0.007	0.023	0.006	0.006	0.019	0.005
	200	0.007	0.026	0.005	0.006	0.022	0.005	0.006	0.019	0.005
0.25	50	0.009	0.029	0.010	0.009	0.023	0.010	0.010	0.021	0.014
	100	0.008	0.027	0.006	0.008	0.024	0.007	0.006	0.020	0.007
	150	0.007	0.026	0.005	0.007	0.021	0.006	0.006	0.020	0.006
	200	0.008	0.026	0.005	0.007	0.022	0.005	0.006	0.019	0.005

По результатам таблицы 5.17 можно сказать, что **FSSA** местами справляется лучше в данной модели, чем **2d-SSA**, а местами наоборот. **2d-SSA** показывает тот же порядок ошибки, что и **FSSA**, разница незначительна. **2d-SSA** в свою очередь показывает более хорошие результаты, в отличие от **MSSA**.

Заключение

Работа посвящена изучению трех существующих подходов, которые предлагаются для построений модификаций метода анализ сингулярного спектра (SSA) для анализа временных рядов.

Целью работы является анализ рассмотренных подходов с точки зрения теории SSA, выявление их сильных и слабых сторон, а также использование их сильных сторон в новых модификациях. Были рассмотрены следующие модификации. Первая — Generalized SSA (GSSA), которая введением весов улучшает разделимость в отсутствии шума. Это свойство было объяснено с помощью рассмотрения применения SSA как системы линейных адаптивных фильтров. Была предложена новая модификация, объединяющая аппроксимационные свойства SSA и преимущества GSSA.

Вторая модификация — это Circular SSA (CiSSA), где адаптивный базис, который строится в SSA, заменен на фиксированный базис Фурье в траекторном пространстве ряда. Естественные проблемы в CiSSA связаны с выделением тренда и амплитудно-модулированных гармоник, однако авторы метода делали акцент на простоте автоматического выделения компонент с заданными частотными диапазонами. В работе показано, что современные методы автоматической идентификации компонент в базовом варианте SSA справляются с выделением компонент, как минимум, не хуже, причем точность лучше.

И третья модификация — Functional SSA (FSSA), которая оказалась, по сути, методом для анализ двумерных данных. При сравнении с 2d-SSA в работе показано, методы сравнимы по точности, а существенное преимущество FSSA, демонстрируемое его авторами, связано с тем, что авторы его сравнивали с SSA для многомерных рядов, игнорируя регулярное поведение по аргументу, рассматриваемому авторами как непрерывный параметр.

Работа дополнена большим количеством численных экспериментов для демонстрации результатов.

Список литературы

1. Golyandina N., Nekrutkin V., Zhigljavsky A. Analysis of Time Series Structure: SSA and Related Techniques. — Chapman and Hall/CRC, 2001.
2. Generalized singular spectrum analysis for the decomposition and analysis of non-stationary signals / Gu J., Hung K., Ling B.W.-K., Chow D.H.-K., Zhou Y., Fu Y., and Pun S.H. // Journal of the Franklin Institute. — 2024. — Vol. 361, no. 6. — P. 106696. — Access mode: <https://www.sciencedirect.com/science/article/pii/S0016003224001170>.
3. Bogalo J., Poncela P., Senra E. Circulant singular spectrum analysis: A new automated procedure for signal extraction // Signal Processing. — 2021. — Vol. 179. — P. 107824. — Access mode: <https://www.sciencedirect.com/science/article/pii/S0165168420303686>.
4. Functional singular spectrum analysis / Haghbin H., Najibi S.M., Mahmoudvand R., Trinka J., and Maadooliat M. // Stat. — 2021. — Vol. 10, no. 1.
5. Golyandina N., Shlemov A. Variations of singular spectrum analysis for separability improvement: non-orthogonal decompositions of time series // Statistics and Its Interface. — 2015. — Vol. 8, no. 3. — P. 277–294.
6. Golyandina N., Dudnik P., Shlemov A. Intelligent identification of trend components in singular spectrum analysis // Algorithms. — 2023. — Vol. 16, no. 7.
7. Golyandina N., Zhigljavsky A. Singular Spectrum Analysis for Time Series. Springer-Briefs in Statistics. — 2 ed. — Springer Berlin Heidelberg, 2020.
8. Golyandina N., Korobeynikov A., Zhigljavsky A. Singular Spectrum Analysis with R. — Berlin, Heidelberg : Springer-Verlag, 2018.
9. Weisstein E.W. CRC Concise Encyclopedia of Mathematics. — CRC Press, 2002.
10. Cognitive load detection through EEG lead wise feature optimization and ensemble classification / Yedukondalu J., Sunkara K., Radhika V., Kondaveeti S., Anumothu M., and Krishna Y. M. // Scientific Reports. — 2025. — Vol. 15, no. 1. — P. 842. — Access mode: <https://doi.org/10.1038/s41598-024-84429-6>.
11. Application of visual stratigraphy from line-scan images to constrain chronology and melt features of a firn core from coastal Antarctica / Dey R., Thamban M., Laluraj C.M., Mahalinganathan K., Redkar B.L., Kumar S., and Matsuoka K. // Journal of Glaciology. — 2023. — Vol. 69, no. 273. — P. 179–190.
12. Pogrebnikov N. SPbSU SSA Coursework: Time Series Analysis. — https://github.com/xSICHx/spbu_ssa_methods_coursework/tree/main. — 2024.
13. Golyandina N., Zhornikova P. On automated identification in singular spectrum analysis

for different types of objects. — 2023. — <https://arxiv.org/abs/2302.08993>.