

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

Отчет о научно-исследовательской работе

Сенов Михаил Андреевич

РОБАСТНЫЕ ВАРИАНТЫ МЕТОДА SSA ДЛЯ АНАЛИЗА
КОМПЛЕКСНЫХ ВРЕМЕННЫХ РЯДОВ

Научный руководитель:
к. ф.-м. н., доцент Н. Э. Голяндина

1. Введение

Временным рядом называется набор значений некоторой функции от времени, собранных в разные моменты времени.

Предположим, что временной ряд является суммой нескольких временных рядов, к примеру, тренда (медленно меняющейся составляющей), сезонной составляющей и шума. Для работы с таким рядом полезно выделить эти составляющие, поскольку работать с ними по отдельности может быть проще чем с исходным рядом, сделать это позволяет метод «Гусеница»-SSA (в дальнейшем просто SSA).

При подобном анализе возникает следующее затруднение. В данных часто возникают выделяющиеся ошибки, значительно большие, чем размер шума. Эти ошибки называются выбросами. Соответственно, возникает задача построения изначально устойчивых к выбросам модификаций SSA.

Решению данной задачи была посвящена работа [1]. Результаты были получены для вещественнозначных рядов. В реальности данные с многих приборов снимаются изначально в комплексном виде и, поэтому, задача анализа комплекснозначных временных рядов так же важна. Поэтому, целью данной работы является рассмотрение возможности переноса полученных ранее результатов на комплексный случай и их обобщение в случае неудачи.

2. Алгоритм SSA

В этом разделе рассмотрим базовый алгоритм SSA.

2.1. Описание алгоритма

Рассмотрим ненулевой ряд $X_N = (x_1, \dots, x_N)$, где $N > 2$. Базовый алгоритм SSA выполняет разложение исходного ряда на сумму из нескольких новых рядов и осуществляется в четыре этапа.

Вложение

Первым этапом алгоритма является построение траекторной матрицы.

Пусть L — некоторое целое число (*длина окна*), $1 < L < N$

Траекторная матрица — это матрица:

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & \dots & x_K \\ x_2 & x_3 & \dots & x_{K+1} \\ \vdots & \vdots & & \vdots \\ x_L & x_{L+1} & \dots & x_N \end{pmatrix}, K = N - L + 1.$$

Сингулярное разложение

Вторым этапом является сингулярное разложение (SVD) траекторной матрицы ряда, оно может быть записано как:

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_d$$

где $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$, λ_i — i -ое собственное число по убыванию матрицы $\mathbf{X}\mathbf{X}^T$, U_i — собственный вектор матрицы $\mathbf{X}\mathbf{X}^T$, соответствующий λ_i , V_i — собственный вектор матрицы $\mathbf{X}^T\mathbf{X}$, соответствующий λ_i , d — ранг матрицы \mathbf{X} .

Группировка

Третьим этапом является объединение в группы полученных матриц \mathbf{X}_i .

Матрица, соответствующая группе I :

$$\mathbf{X}_I = \mathbf{X}_{i_1} + \dots + \mathbf{X}_{i_r}.$$

И результат группировки:

$$\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_l}.$$

Диагональное усреднение

Последним этапом является перевод каждой матрицы, соответствующей группе, в новый ряд длины N .

Пусть \mathbf{Y} — некоторая матрица $L \times K$ с элементами y_{ij} . Положим $L^* = \min(L, K)$, $K^* = \max(L, K)$, $N = L + K - 1$. Пусть $y_{ij}^* = y_{ij}$, если $L < K$, и $y_{ij}^* = y_{ji}$ иначе.

Диагональное усреднение переводит матрицу \mathbf{Y} в ряд (y_0, \dots, y_{N-1}) по формуле:

$$y_k = \begin{cases} \frac{1}{k+1} \sum_{i=1}^{k+1} y_{i,k-i+2}^* & \text{для } 0 \leq i \leq L^* - 1 \\ \frac{1}{L^*} \sum_{i=1}^{L^*} y_{i,k-i+2}^* & \text{для } L^* - 1 \leq i \leq K^* . \\ \frac{1}{N-k} \sum_{i=k-K^*+2}^{N-K^*+1} y_{i,k-i+2}^* & \text{для } K^* \leq i \leq N-1 \end{cases}$$

Таким образом, мы разложили исходный ряд в сумму l новых рядов:

$$\mathbf{X}_N = \sum_{i=1}^l \mathbf{X}_{N_i}.$$

3. Робастные варианты SSA.

В данном разделе мы рассмотрим устойчивые к выбросам (робастные) комплексно-значные модификации SSA. Особенности методов (кроме замены транспонирования T на эрмитово сопряжение H) будем отмечать отдельно.

Будем рассматривать вариант метода SSA для выделения сигнала, когда группировка заключается в выборе первых r компонент. Для стандартного метода SSA это эквивалентно проекции по норме Фробениуса траекторной матрицы ряда на множество матриц ранга, не превосходящего r .

Пусть имеется временной ряд $\mathbf{X}_N = (x_1, \dots, x_N)$.

$\mathcal{M}_{\mathcal{H}}$ — пространство ганкелевых матриц $L \times K$,

\mathcal{M}_r — пространство матриц ранга, не превосходящего r , размера $L \times K$.

Оператор вложения $\mathcal{T} : \mathbb{R}^N(\mathbb{C}^N) \rightarrow \mathcal{M}_{\mathcal{H}} : \mathcal{T}(\mathbf{X}_N) = \mathbf{X}$,

$\Pi_r : \mathcal{M} \rightarrow \mathcal{M}_r$ — проектор на множество матриц ранга, не превосходящего r , по некоторой норме в пространстве матриц,

$\Pi_{\mathcal{H}} : \mathcal{M} \rightarrow \mathcal{M}_{\mathcal{H}}$ — проектор на пространство ганкелевых матриц по некоторой норме в пространстве матриц.

В результате применения данных операторов получаем оценку сигнала:

$$\tilde{S} = \mathcal{T}^{-1} \Pi_{\mathcal{H}} \Pi_r \mathcal{T}(\mathbf{X}_N).$$

В случае, когда проекторы Π_r и $\Pi_{\mathcal{H}}$ берутся по норме в пространстве \mathbb{L}_2 , оценка сигнала соответствует алгоритму SSA, для случая, когда восстановление производится по одной группе, состоящей из первых r компонент.

Существует два известных подхода к построению устойчивых к выбросам модификаций SSA:

- Проекторы Π_r и $\Pi_{\mathcal{H}}$ строятся по норме в пространстве \mathbb{L}_1 ,
- Проекторы Π_r и $\Pi_{\mathcal{H}}$ строятся по взвешенной норме в пространстве \mathbb{L}_2 .

В работе [1] были предложены реализации обоих подходов, приведём адаптированные на комплексный случай алгоритмы ниже.

3.1. Проекция по норме \mathbb{L}_1

Пусть $\mathbf{Y} \in \mathbb{C}^{L \times K}$ — траекторная матрица ряда. Необходимо решить задачу

$$\|\mathbf{Y} - \mathbf{U}\mathbf{V}^H\|_1 \longrightarrow \min_{\mathbf{U}, \mathbf{V}}, \mathbf{U} \in \mathbb{C}^{L \times r}, \mathbf{V} \in \mathbb{C}^{K \times r}.$$

Алгоритм 1: Последовательный метод построения \mathbb{L}_1 -проектора на множество матриц ранга, не превосходящего r

Входные данные: $\mathbf{Y} \in \mathbb{C}^{L \times K}$ — траекторная матрица ряда, r — ранг сигнала;

параметры критерия остановки: $\varepsilon = 10^{-4}$, максимальное

число итераций $N_{iter} = 10$

Выходные данные: $\hat{\mathbf{Y}} = \mathbf{U}\mathbf{V}^H$ — проекция траекторной матрицы на множество матриц ранга, не превосходящего r

Инициализация $\mathbf{V}(0) \in \mathbb{C}^{L \times r}$, нормировка столбцов $\mathbf{V}(0)$;

$t := 0$;

до тех пор, пока $\max_{\substack{i=1,\dots,L \\ j=1,\dots,r}} |u_{ij}(t) - u_{ij}(t-1)| > \varepsilon$ **и** $t < N_{iter}$ **выполнять**

$t := t + 1$;

$\mathbf{U}(t) = \operatorname{argmin}_{\mathbf{U} \in \mathbb{C}^{L \times r}} \|\mathbf{Y} - \mathbf{U}\mathbf{V}^H(t-1)\|_1$;

$\mathbf{V}(t) = \operatorname{argmin}_{\mathbf{V} \in \mathbb{C}^{K \times r}} \|\mathbf{Y} - \mathbf{U}(t)\mathbf{V}^H\|_1$;

Нормировка столбцов $\mathbf{V}(t)$;

конец

$\mathbf{U} := \mathbf{U}(t)$; $\mathbf{V} := \mathbf{V}(t)$;

В приведённой реализации $\mathbf{V}(0)$ инициализируется при помощи сингулярного разложения, но, согласно [3], инициализация может быть произведена при помощи любой матрицы требуемого размера с сохранением сходимости.

Рассмотрим подробнее решение задачи

$$\mathbf{U}(t) = \operatorname{argmin}_{\mathbf{U} \in \mathbb{C}^{K \times r}} \|\mathbf{Y} - \mathbf{U}\mathbf{V}^H(t-1)\|_1. \quad (1)$$

Целевую функцию можно представить в виде

$$\|\mathbf{Y} - \mathbf{U}\mathbf{V}^H(t-1)\|_1 = \sum_{i=1}^L \|\mathbf{y}_i^H - \mathbf{V}(t-1)\mathbf{u}_i^H\|_1,$$

где $\mathbf{y}_i \in \mathbb{C}^K$ — строки \mathbf{Y} , $\mathbf{v}_i \in \mathbb{C}^r$ — строки \mathbf{U} . Согласно [3], задача (1) может быть разбита на L независимых подзадач

$$\mathbf{u}_i(t) = \operatorname{argmin}_u \|\mathbf{y}_i^H - \mathbf{V}(t-1)\mathbf{u}^H\|_1. \quad (2)$$

Подзадача (3) в свою очередь может быть разбита на r подзадач

$$\mathbf{u}_{ic}(t) = \underset{u_c}{\operatorname{argmin}} \|\mathbf{y}_i^H - \mathbf{v}_c(t-1)\mathbf{u}_c^H\|_1. \quad (3)$$

Решение каждой из которых является взвешенной медианой вектора $\frac{\mathbf{y}_i}{\mathbf{v}_c(t-1)}$ с вектором весов $|\mathbf{v}_c(t-1)|$.

3.2. Проекция по взвешенной норме \mathbb{L}_2 с итеративным обновлением весов

Пусть $\mathbf{Y} \in \mathbb{C}^{L \times K}$ — траекторная матрица ряда. Необходимо решить задачу

$$\|\mathbf{W}^{1/2} \odot (\mathbf{Y} - \mathbf{U}\mathbf{V}^H)\|_F^2 \longrightarrow \min_{\mathbf{U}, \mathbf{V}}, \mathbf{U} \in \mathbb{C}^{L \times r}, \mathbf{V} \in \mathbb{C}^{K \times r}.$$

Для начала рассмотрим алгоритм с фиксированной матрицей весов.

Алгоритм 2: Алгоритм решения задачи взвешенной аппроксимации для фиксированной матрицы весов \mathbf{W}

Входные данные: $\mathbf{Y} \in \mathbb{C}^{L \times K}$ — траекторная матрица ряда, r — ранг сигнала,

$\mathbf{W} \in \mathbb{R}^{L \times K}$ — матрица весов;

параметры критерия останковки: $\varepsilon = 10^{-4}$,

максимальное число итераций $N_\alpha = 5$

Выходные данные: $\hat{\mathbf{Y}} = \mathbf{U}\mathbf{V}^H$ — решение задачи взвешенной аппроксимации при фиксированной матрице весов \mathbf{W}

1. $t := 0$;

2. до тех пор, пока $\|\mathbf{W}^{1/2} \odot (\mathbf{Y} - \mathbf{U}\mathbf{V}^H)\|_F^2 > \varepsilon$ и $t < N_\alpha$ выполнять

а. Вычисление матрицы $\mathbf{U} \in \mathbb{C}^{L \times r}$ с помощью решения задачи

$$(y_i^H - \mathbf{V}u_i^H)^H \mathbf{W}_i (y_i^H - \mathbf{V}u_i^H) \rightarrow \min_{u_i}, \quad i = 1, \dots, L, \quad (4)$$

где $\mathbf{W}_i = \operatorname{diag}(w_i) \in \mathbb{R}^{K \times K}$ — матрица, составленная из i -ой строки \mathbf{W} ;

б. Вычисление матрицы $\mathbf{V} \in \mathbb{C}^{K \times r}$ с помощью решения задачи

$$(y_j - \mathbf{U}v_j^H)^H \mathbf{W}^j (y_j - \mathbf{U}v_j^H) \rightarrow \min_{v_j}, \quad j = 1, \dots, K, \quad (5)$$

где $\mathbf{W}^j = \operatorname{diag}(W_j) \in \mathbb{R}^{L \times L}$ — матрица, составленная из j -го столбца \mathbf{W} ;

в. $t := t + 1$.

конец

Задачи (4), (5) решаются при помощи QR-разложения матриц $\mathbf{V}^H \mathbf{W}_i \mathbf{V}$ и $\mathbf{U}^H \mathbf{W}^j \mathbf{U}$ соответственно, алгоритм решения представлен в [4].

У авторов этого алгоритма в [5] допущена ошибка в его описании. Дело в том, что в задаче 4 решение линейного уравнения ищется по эрмитово-сопряжённой системе, а не изначальной, а в задаче 5 находится сразу \mathbf{V}^H , а не \mathbf{V} . Эта ошибка была несущественной в случае вещественной реализации в [1], так как вещественный аналог эрмитового сопряжения — транспонирование, не меняет элементы, но оказалась существенной в комплексном случае.

Теперь рассмотрим алгоритм с итеративным обновлением весов.

Алгоритм 3: Метод с итеративным обновлением весов для нахождения проекции на множество матриц ранга, не превосходящего r

Входные данные: $\mathbf{Y} \in \mathbb{C}^{L \times K}$ — траекторная матрица ряда, r — ранг сигнала;

параметр весовой функции $\alpha = 4.685$; параметры критерия остановки: $\varepsilon = 10^{-4}$, максимальное число итераций

$$N_{iter} = 10$$

Выходные данные: $\hat{\mathbf{Y}} = \mathbf{U}\mathbf{V}^H$ — проекция траекторной матрицы на множество матриц ранга, не превосходящего r

Инициализация $\mathbf{U} \in \mathbb{C}^{L \times r}$ и $\mathbf{V}(0) \in \mathbb{C}^{K \times r}$ (например, с помощью сингулярного разложения матрицы \mathbf{Y});

$$t := 0;$$

до тех пор, пока $\|\mathbf{W}^{\frac{1}{2}} \odot (\mathbf{Y} - \mathbf{U}\mathbf{V}^H)\|_F^2 > \varepsilon$ и $t < N_{iter}$ выполнять

Вычисление матрицы остатков $\mathbf{R} = \{r_{ij}\}_{i,j=1}^{n,p} = \mathbf{Y} - \mathbf{U}\mathbf{V}^H$;

Обновление матрицы $\Sigma = \{\sigma_{ij}\}_{i,j=1}^{L,K}$;

Вычисление матрицы весов $\mathbf{W} = \{w_{ij}\}_{i,j=1}^{L,K} = \{w(\frac{r_{ij}}{\sigma_{ij}})\}_{i,j=1}^{L,K}$, используя

$$w(x) = \begin{cases} (1 - (\frac{|x|}{\alpha})^2)^2 & |x| \leq \alpha \\ 0 & |x| > \alpha \end{cases};$$

Решение задачи взвешенной аппроксимации (обновление матриц \mathbf{U} , \mathbf{V})

$$\|\mathbf{W}^{\frac{1}{2}} \odot (\mathbf{Y} - \mathbf{U}\mathbf{V}^H)\|_F^2 \longrightarrow \min_{\mathbf{U}, \mathbf{V}},$$

при помощи алгоритма 2;

конец

Данный алгоритм был предложен в [5], авторы предложили взять $\alpha = 4.685$, $N_\alpha = 5$ и $N_{iter} = 10$, ссылаясь на численные эксперименты.

Параметр сигма предлагается взять равным $\sigma_{ij} = \sigma = 1.4826 \text{ med } |\mathbf{R} - \text{med } |\mathbf{R}||$, где \mathbf{R} — это вектор, составленный из всех элементов матрицы остатков $\mathbf{R} = \{r_{ij}\}_{i,j=1}^{L,K}$, то есть

$$\mathbf{R} = (r_{11}, \dots, r_{1K}; r_{21}, \dots, r_{2K}; \dots; r_{L1}, \dots, r_{LK}).$$

Данная оценка предлагается авторами ввиду её робастности.

3.3. Модификация метода с итеративным обновлением весов

У представленного выше алгоритма есть одна важная проблема, а именно, выбор параметра σ_{ij} не зависящим от i и j . В случае не стационарных рядов выявление выбросов может происходить неверно. Например, если шум растёт к концу ряда, то выбросы в начале ряда могут получить больший вес, чем не выбросы в конце ряда. В [1] была приведена модификация алгоритма, призванная решить эту проблему. Здесь же мы рассмотрим её комплексную адаптацию.

Ключевая задача параметра σ_{ij} — приписывание определённого веса определённому элементу ряда, чем элемент больше похож на выброс, тем больше сигма и наоборот. Ввиду такой интерпретации логично рассматривать сигмы как ряд, идущий дополнением к данному $\sigma = (\sigma_1, \dots, \sigma_N)$, а после, ганкелизацией привести этот ряд к матричному виду $\Sigma = \{\sigma_{ij}\}_{i,j=1}^{L,K}$. Сам же ряд σ автор [1] предлагает взять равным тренду (математическому ожиданию) ряда из модулей остатков. Это предложение справедливо и для комплексного случая, так как в комплексном случае выброс характеризуется величиной модуля.

Теперь рассмотрим сам алгоритм.

Алгоритм 4: Модификация метода с итеративным обновлением весов для нахождения проекции на множество матриц ранга, не превосходящего r

Входные данные: $\mathbf{Y} \in \mathbb{C}^{L \times K}$ — траекторная матрица ряда, r — ранг сигнала;

параметр весовой функции $\alpha = 4.685$; параметры критерия

остановки: $\varepsilon = 10^{-4}$, максимальное число итераций $N_{iter} = 10$

Выходные данные: $\hat{\mathbf{Y}} = \mathbf{U}\mathbf{V}^T$ — проекция траекторной матрицы на множество матриц ранга, не превосходящего r

1. Инициализация $\mathbf{U} \in \mathbb{C}^{L \times r}$ и $\mathbf{V} \in \mathbb{C}^{K \times r}$ (например, с помощью сингулярного разложения матрицы \mathbf{Y});

2. $t := 0$;

3. **повторять**

a. Вычисление матрицы остатков $\mathbf{R} = \{r_{ij}\}_{i,j=1}^{n,p} = \mathbf{Y} - \mathbf{U}\mathbf{V}^H$;

b. Ганкелизация матрицы \mathbf{R} и получение ряда длины N из остатков:

$$\mathbf{R} = \mathcal{T}^{-1}\Pi_{\mathcal{H}}(\mathbf{R}) = (r_1, \dots, r_N);$$

c. Пусть $\mathbf{R}_+ = (|r_1|, \dots, |r_N|)$ — ряд из модулей остатков. Вычисление

$\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_N)$ как оценки мат. ожидания $\mathbb{E}(\mathbf{R}_+)$ некоторым методом;

d. Получение матрицы $\boldsymbol{\Sigma} = \{\sigma_{ij}\}_{i,j=1}^{L,K} = \mathcal{T}(\boldsymbol{\sigma})$;

e. Вычисление матрицы весов $\mathbf{W} = \{w_{ij}\}_{i,j=1}^{L,K} = \{w(\frac{r_{ij}}{\sigma_{ij}})\}_{i,j=1}^{L,K}$, используя

$$w(x) = \begin{cases} (1 - (\frac{|x|}{\alpha})^2)^2, & |x| \leq \alpha \\ 0, & |x| > \alpha \end{cases}$$

f. Решение задачи взвешенной аппроксимации (обновление матриц \mathbf{U} и \mathbf{V})

$$\left\| \mathbf{W}^{1/2} \odot (\mathbf{Y} - \mathbf{U}\mathbf{V}^H) \right\|_F^2 \rightarrow \min_{\mathbf{U}, \mathbf{V}}$$

g. $t := t + 1$.

до тех пор, пока $\left\| \mathbf{W}^{1/2} \odot (\mathbf{Y} - \mathbf{U}\mathbf{V}^H) \right\|_F^2 > \varepsilon$ и $t < N_{IRLS}$;

Преимуществом данного алгоритма, является то, что пользователь сам может выбрать, каким методом он хочет вычислять матожидание ряда. В приведённой реализации представлены три метода: локальная регрессия loess, скользящая медиана и взвешенная локальная регрессия lowess.

4. Примеры работы алгоритмов

В данном разделе мы приведём несколько примеров комплексных временных рядов и сравним результаты работы методов.

Сравнение будет проводиться по величине среднеквадратичной ошибки, согласованной с \mathbb{L}_2 , которая вычисляется по формуле

$$\text{MSE} = \mathbb{E} \left(\frac{1}{N} \sum_{i=1}^N (s_i - \hat{s}_i)^2 \right), \quad (6)$$

где $\mathbf{S} = (s_1, \dots, s_N)^T$ — сигнал, $\hat{\mathbf{S}} = (\hat{s}_1, \dots, \hat{s}_N)^T$ — его оценка. Будем вычислять

$$\text{RMSE} = \sqrt{\text{MSE}}.$$

Так же будем проверять значимость сравнения, для этого будем использовать гипотезу, что MSE для некоторых методов равны между собой.

$H_0 : \mathbb{E}(\xi_1 - \xi_2) = 0$. Имеем две выборки $X = (x_1, \dots, x_M)$ и $Y = (y_1, \dots, y_M)$ объема M . Обозначим \bar{X} и \bar{Y} — их выборочные средние, s_x^2 и s_y^2 — выборочные дисперсии, $\hat{\rho}$ — коэффициент корреляции. Статистика критерия

$$t = \frac{\sqrt{M}(\bar{X} - \bar{Y})}{\sqrt{s_x^2 + s_y^2 - 2s_x s_y \hat{\rho}}}$$

имеет асимптотически нормальное распределение.

Шум в примерах будет иметь стандартное комплексное нормальное распределение. Определим, что это значит.

Определение. Комплексная случайная величина Z имеет стандартное комплексное нормальное распределение, если

1. $\Re(Z)$ и $\Im(Z)$ независимы,
2. $\Re(Z), \Im(Z) \sim N(0, 1/2)$.

И обозначается $Z \sim CN(0, 1)$.

4.1. Синтетический пример №1

Рассмотрим ряд с постоянной амплитудой и шумом постоянной дисперсии. Длину ряда возьмём $N = 240$

$$x_n = e^{2n\pi/30i} + \varepsilon_n, \quad \varepsilon_n \sim CN(0, 1).$$

Рассмотрим результаты работы методов для такого ряда (1). В таблице 2 представлены p-value для сравнения методов с лучшим. Ранг ряда равен 1.

Таблица 1. Оценки RMSE различных методов для $M = 30$ реализаций ряда без выбросов.

Method	CSSA	L1	weighted L2	loess L2	median L2	lowess L2
RMSE	0.1016	0.125	0.1017	0.103	0.105	0.104

Таблица 2. p-value для сравнения различных методов с наилучшим без выбросов.

Method	L1	weighted L2	loess L2	median L2	lowess L2
CSSA	1.5e-06	0.91	0.69	0.24	0.43

В случае отсутствия выбросов лучший результат показывают метод Complex SSA, но сравнение значимо только с методом проекции на \mathbb{L}_1 .

Теперь добавим к ряду 5% выбросов с величиной выброса $5x_i$. Графики ряда представлены на Рис. 1, 2.

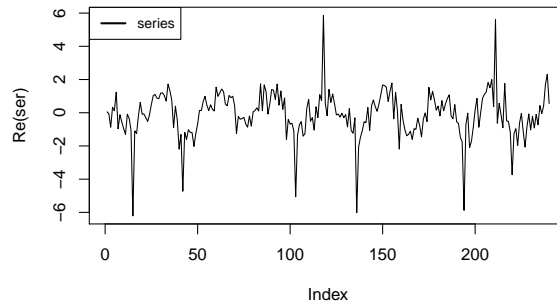


Рис. 1. График вещественной части ряда.

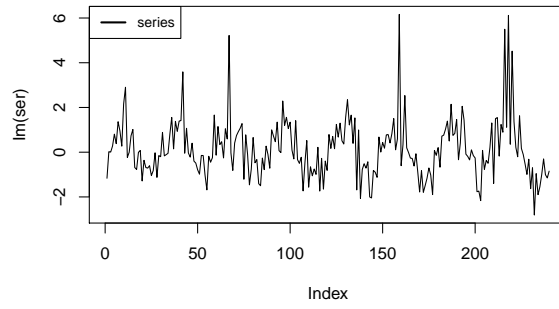


Рис. 2. График мнимой части ряда.

Графики результатов анализа представлены в Рис. 3, 4. В таблице 3 представлены сравнения ошибок для различных методов. В таблице 4 представлены p-value для сравнения методов с лучшим. Длина окна взята $L = 120$.

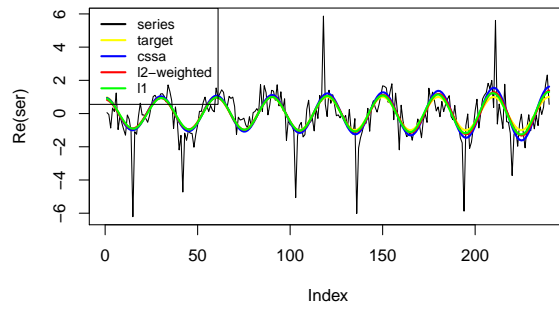


Рис. 3. Вещественная часть выделения тренда несколькими способами.

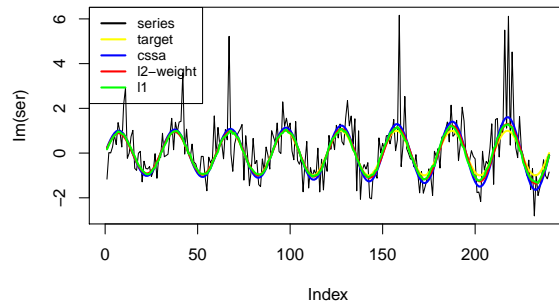


Рис. 4. Мнимая часть выделения тренда несколькими способами.

Таблица 3. Оценки RMSE различных методов для $M = 30$ реализаций ряда с выбросами.

Method	CSSA	L1	weighted L2	loess L2	median L2	lowess L2
RMSE	0.285	0.147	0.158	0.112	0.114	0.114

Таблица 4. p-value для сравнения различных методов с наилучшим с выбросами.

Method	CSSA	L1	weighted L2	median L2	lowess L2
loess L2	0	7.7e-13	4.1e-10	0.134	0.262

В случае наличия выбросов метод loess проекции на \mathbb{L}_2 показал наилучший результат, за исключением того, что сравнения с другими вариациями модифицированной взвешенной проекции незначимы.

4.2. Синтетический пример №2

Рассмотрим ряд с растущей амплитудой и шумом непостоянной дисперсии. Длину ряда возьмём $N = 240$

$$x_n = e^{4n/N} e^{2n\pi/30i} + \frac{1}{2} e^{4n/N} \varepsilon_n, \quad \varepsilon_n \sim CN(0, 1).$$

Рассмотрим результаты работы методов для такого ряда (5). В таблице 6 представлены p-value для сравнения методов с лучшим. Ранг ряда равен 1.

Таблица 5. Оценки RMSE различных методов для $M = 30$ реализаций ряда без выбросов.

Method	CSSA	L1	weighted L2	loess L2	median L2	lowess L2
RMSE	1.28	1.52	1.90	1.36	1.43	1.37

Таблица 6. p-value для сравнения различных методов с наилучшим без выбросов.

Method	L1	weighted L2	loess L2	median L2	lowess L2
CSSA	0.005	0.0001	0.001	7.6e-5	0.0007

В случае отсутствия выбросов лучший результат показывает Complex SSA. Здесь же видно, что модифицированный метод взвешенной проекции справляется с не стац-

онарным рядом куда лучше чем классический, к примеру, сравнивая классический и loess, сравнение является значимым с $p\text{-value} = 0.0005$.

Теперь добавим к ряду 5% выбросов с величиной выброса $5x_i$. Графики ряда представлены на Рис. 5, 6.

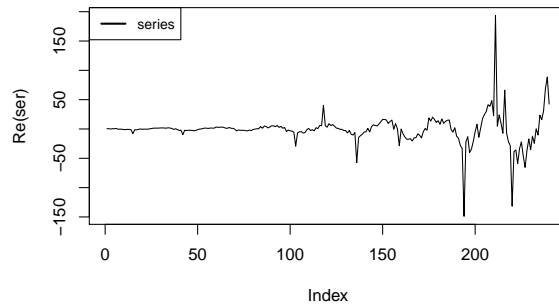


Рис. 5. График вещественной части ряда.

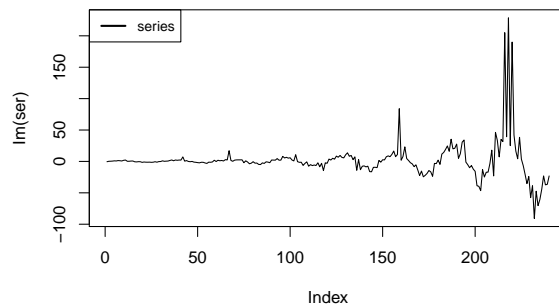


Рис. 6. График мнимой части ряда.

Графики результатов анализа представлены на Рис. 7, 8. В таблице 7 представлены сравнения ошибок для различных методов. В таблице 8 представлены $p\text{-value}$ для сравнения методов с лучшим. Длина окна взята $L = 120$.

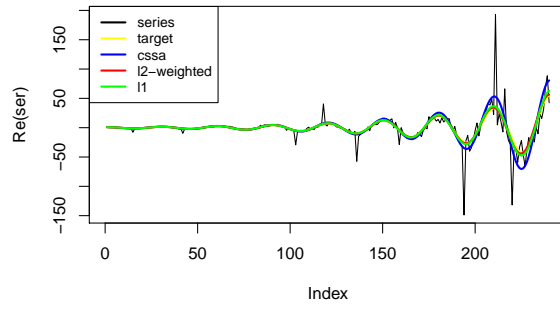


Рис. 7. Вещественная часть выделения тренда несколькими способами.

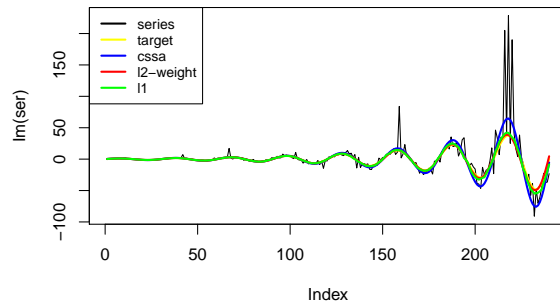


Рис. 8. Мнимая часть выделения тренда несколькими способами.

Таблица 7. Оценки RMSE различных методов для $M = 30$ реализаций ряда с выбросами.

Method	CSSA	L1	weighted L2	loess L2	median L2	lowess L2
RMSE	6.14	1.78	1.66	1.48	1.50	1.49

Таблица 8. p-value для сравнения различных методов с наилучшим с выбросами.

Method	CSSA	L1	weighted L2	median L2	lowess L2
loess L2	3.5e-16	0.005	0.13	0.387	0.28

В случае присутствия выбросов метод loess проекции на \mathbb{L}_2 показывает себя наилучшим образом на данном примере, однако сравнения с взвешенным \mathbb{L}_2 , median \mathbb{L}_2 и lowess \mathbb{L}_2 не являются значимыми.

4.3. Синтетический пример №3

При построении комплексных робастных методов возникает вопрос: Что считать выбросом? В данной работе выброс считается как элемент с аномально большим модулем. В связи с этим возникает другой вопрос: Не могут ли возникнуть проблемы в случае несимметричности распределения модуля выброса по вещественной и мнимой части? То есть не может ли получиться такой ситуации, что алгоритм посчитает выбросом точку, имеющую очень большое отклонение по вещественной оси, но на мнимой оси эта точка выбросом не является, что приведёт к ухудшению выделения тренда по мнимой оси.

Для рассмотрения примера на тему возьмём прошлый ряд, но выбросы сосредоточим только на вещественной оси, их вещественную часть оставив прежней. Так же в данном примере будет осмыслено посчитать помимо совместных ошибок ещё и отдельно ошибки вещественной и мнимой частей.

Графики выделения тренда представлены на Рис. 9, 10. Результаты RMSE для примера представлены в таблице 9. Так же посчитаем RMSE отдельно для вещественной и мнимой части для прошлого примера, для которого выбросы по модулю сделаем равными текущим, результат в таблице ??.

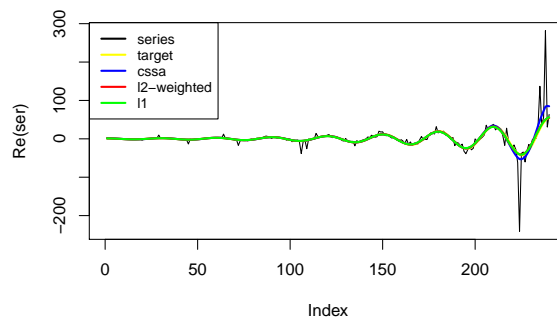


Рис. 9. Вещественная часть выделения тренда несколькими способами.

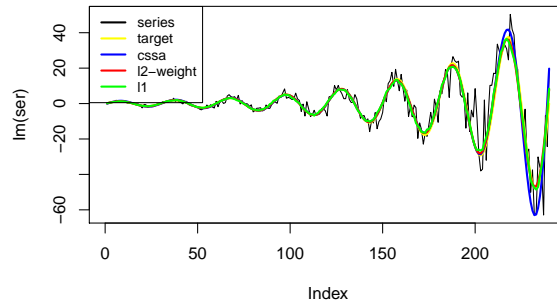


Рис. 10. Мнимая часть выделения тренда несколькими способами.

Таблица 9. Оценки RMSE различных методов для $M = 30$ реализаций ряда с вещественными выбросами.

Method	CSSA	L1	weighted L2	loess L2	median L2	lowess L2
RMSE (совм.)	3.53	1.55	1.72	1.45	1.48	1.46
RMSE (Re)	2.63	1.06	1.23	1.05	1.07	1.06
RMSE (Im)	2.33	1.14	1.2	0.98	1.02	1

Таблица 10. Оценки RMSE различных методов для $M = 30$ реализаций прошлого примера.

Method	CSSA	L1	weighted L2	loess L2	median L2	lowess L2
RMSE (совм.)	3.53	1.29	1.34	0.97	1	0.98
RMSE (Re)	2.5	0.85	0.95	0.7	0.71	0.7
RMSE (Im)	2.5	0.97	0.95	0.69	0.71	0.7

Получаем, что общая ошибка в случае, когда выбросы равномерно распределены по вещественной и по мнимой оси ниже, но распределение ошибок по вещественной и мнимой осям сохраняется. Что свидетельствует в пользу нашего предположения и кажется разумным для данных методов, как для методов, приближающих весь комплексный ряд, а не вещественную и мнимую части по отдельности.

5. Заключение

В работе были приведены и исследованы два обобщения робастных вариантов метода SSA на комплексно-значный случай.

Был проведён обзор двух известных подходов к построению робастных версий SSA: замена проекции по норме \mathbb{L}_2 на проекцию по норме \mathbb{L}_1 и на взвешенную проекцию по норме \mathbb{L}_2 и их имплементация на комплексный случай.

Работа методов была показана на нескольких примерах, подтверждающих эффективность робастных модификаций, в сравнении с Complex SSA для рядов с выбросами. Все рассматриваемые модификации были реализованы на R.

Список литературы

1. Третьякова А. Л. Робастные варианты метода анализа сингулярного спектра : квалификационная работа магистра ; науч. рук. к. ф.-м. н., доцент Н. Э. Голяндина. — 2020.
2. Голяндина Н. Э. Метод «Гусеница»-SSA: анализ временных рядов: Учеб. пособие. — Санкт-Петербург : ВММ, 2004.
3. Ke Qifa, Kanade Takeo. Robust L1 Norm Factorization in the Presence of Outliers and Missing Data by Alternative Convex Programming // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2005). — 2005. — June.
4. Solving Weighted Least Squares (WLS) problems on ARM-based architectures / Belloch Jose A., Bank Balázs, Igual Francisco D., Quintana-Ortí Enrique S., and Vidal Antonio M. // The Journal of Supercomputing. — 2017. — 01. — Vol. 71. — P. 530–542.
5. Chen Ke, Sacchi Mauricio. Robust reduced-rank filtering for erratic seismic noise attenuation // GEOPHYSICS. — 2015. — 01. — Vol. 80. — P. V1–V11.
6. Nina Golyandina, Anton Korobeynikov, Anatoly Zhigljavsky. Singular Spectrum Analysis with R. — Springer, 2018.
7. Brooks J. Paul, Jot Sapan. pcaL1: An Implementation in R of Three Methods for L1-Norm Principal Component Analysis. — 2012. — unpublished.
8. Tsagkarakis N., Markopoulos P. P. and Sklivanitis G., Pados D. A. L1-Norm Principal-Component Analysis of Complex Data // IEEE Transactions on Signal Processing. — 2018. — 06.