

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

Сандалов Сергей Андреевич

## **Локальные модификации метода анализа сингулярного спектра**

Выпускная квалификационная работа  
Бакалавриат

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

Рецензент:  
к.ф.-м.н., программист, ИП Циркин А.О.  
Н. К. Звонарев

Saint Petersburg State University  
Applied Mathematics and Computer Science  
Computational Stochastics and Statistical Models

Sandalov Sergei Andreevich

LOCAL MODIFICATIONS OF SINGULAR SPECTRUM ANALYSIS

Graduation Project

Scientific Supervisor:  
Professor N. E. Golyandina

Reviewer:  
System programmer N. K. Zvonarev

Saint Petersburg  
2019

# Оглавление

<b>Введение</b> . . . . .	4
<b>Глава 1. Базовый алгоритм SSA для выделения сигнала и его недостатки</b>	5
1.1. Алгоритм . . . . .	5
1.2. Возникающие проблемы . . . . .	6
<b>Глава 2. Локальные модификации SSA: Overlap SSA, Sliding SSA, Averaging SSA</b> . . . . .	12
2.1. Overlap SSA . . . . .	13
2.2. Sliding SSA . . . . .	14
2.3. Averaging SSA . . . . .	15
2.4. Сравнение локальных алгоритмов . . . . .	16
<b>Глава 3. Численные эксперименты</b> . . . . .	17
3.1. Постановка задачи . . . . .	17
3.2. Полученные результаты . . . . .	20
<b>Заключение</b> . . . . .	26
<b>Список литературы</b> . . . . .	27

## Введение

Временной ряд — это последовательность измерений (вещественных чисел) в последовательные, обычно равноотстоящие моменты времени. Одна из моделей временного ряда — сумма детерминированной составляющей (сигнала) и случайной составляющей (шума).

Рассмотрим вещественный временной ряд  $\mathbb{F}_N = \mathbb{S}_N + \mathbb{R}_N$ , где  $\mathbb{S}_N = (s_1, s_2, \dots, s_N)$  с  $s_n = A(n) \cos(2\pi\omega(n) \cdot n + \phi)$  — сигнал, а  $\mathbb{R}_N = (r_1, r_2, \dots, r_N)$  — стационарный процесс.

В такой модели стандартные задачи — это оценка сигнала, его прогнозирование и оценка частоты сигнала  $\omega(n)$ . В своей работе я использую метод SSA (Singular Spectrum Analysis), подробно описанный в [1].

Известно [1], что если модуляция имеет вид:  $A(n) = Ae^{\gamma n}$ ,  $\omega(n) = \text{const}$  и  $\gamma, A \in \mathbf{R}$ , то метод хорошо справляется с выделением сигнала. Задача работы заключается в том, чтобы в рамках SSA предложить подход, который хорошо справляется с оценкой сигнала при более сложной модуляции как амплитуды, так и частоты.

Отчет является продолжением курсовой работы, и новый материал содержится в главах 2 и 3. В основе предлагаемых методов для оценки сигнала лежит идея применения стандартного SSA к скользящим отрезкам ряда.

Кратко опишем структуру работы. В главе 1 излагается стандартный алгоритм SSA для выделения сигнала. Также там обсуждается почему в случае более сложной модуляции, чем экспоненциальная, получается плохая оценка сигнала.

В главе 2 представлено описание локальных алгоритмов и их сравнение между собой. Описание дано в структурированном виде. Такое представление позволяет понять общее и отличия между методами.

В заключительной главе 3 обсуждается постановка численного эксперимента для сравнения методов, которые представлены в главе 2. Далее с помощью графиков и таблиц представлены результаты эксперимента.

## Глава 1

# Базовый алгоритм SSA для выделения сигнала и его недостатки

Стандартный алгоритм SSA для оценки сигнала состоит из двух этапов: разложения и восстановления. SSA принимает на вход вещественный временной ряд  $\mathbb{F}_N = (f_1, f_2, \dots, f_N)$  и два целых числа:  $L$ :  $1 < L < N$  — длина окна;  $r$ :  $r \leq \min(L, N - L + 1)$ .

### 1.1. Алгоритм

#### Первый этап: разложение

##### Шаг 1. Вложение

Процедура вложения переводит исходный временной ряд в последовательность многомерных векторов. Она образует  $K = N - L + 1$  векторов вложения

$$F_i = (f_i, \dots, f_{i+L-1})^T, \quad 1 \leq i \leq K,$$

имеющих размерность  $L$ .

Траекторная матрица ряда  $\mathbb{F}$

$$\mathbf{F} = [F_1 : \dots : F_K]$$

состоит из векторов вложения в качестве столбцов.

##### Шаг 2. Сингулярное разложение

Результатом этого шага является сингулярное разложение траекторной матрицы ряда:  $\mathbf{F} = \mathbf{F}_1 + \dots + \mathbf{F}_d$ ,  $\mathbf{F}_i = \sqrt{\lambda_i} U_i V_i^T$ . Здесь  $\lambda_i$  — положительные собственные числа матрицы  $\mathbf{F}\mathbf{F}^T$ , расположенные в порядке неубывания;  $U_i$  — ортонормированные собственные векторы матрицы  $\mathbf{F}\mathbf{F}^T$ , соответствующие собственным числам  $\lambda_i$ ; а  $V_i = \frac{1}{\sqrt{\lambda_i}} \mathbf{F}^T U_i$ . Числа  $\sqrt{\lambda_i}$  называются сингулярными числами матрицы  $\mathbf{F}$ .

#### Второй этап: восстановление

##### Шаг 3. Группировка

Из полученных  $d$  компонент разложения выбираются первые  $r \leq d$  и суммируются. Получается новая матрица  $\mathbf{F}_{(r)} = \mathbf{F}_1 + \dots + \mathbf{F}_r$

#### Шаг 4. Диагональное усреднение

Заключительным шагом полученная матрица  $\mathbf{F}_{(r)}$  переводится в новый временной ряд длины  $N$ .

Пусть  $\mathbf{Y}$  — некоторая  $L \times K$  матрица с элементами  $y_{ij}$ , где  $1 \leq i \leq L$ ,  $1 \leq j \leq K$ . Положим  $L^* = \min(L, K)$ ,  $K^* = \max(L, K)$  и  $N = L + K - 1$ . Пусть  $y_{ij}^* = y_{ij}$  если  $L < K$ , и  $y_{ij}^* = y_{ji}$  иначе. *Диагональное усреднение* переводит матрицу  $\mathbf{Y}$  в ряд  $g_1, \dots, g_N$  по формуле

$$g_k = \begin{cases} \frac{1}{k} \sum_{m=1}^k y_{m,k-m+1}^* & \text{для } 1 \leq k < L^*, \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m,k-m+1}^* & \text{для } L^* \leq k < K^* + 1, \\ \frac{1}{N-k+1} \sum_{m=k-K^*+1}^{N-K^*+1} y_{m,k-m+1}^* & \text{для } K^* + 1 \leq k \leq N. \end{cases} \quad (1.1)$$

## 1.2. Возникающие проблемы

В этом разделе рассмотрим несколько примеров, которые показывают, что в случаях амплитудной и частотной модуляций базовый алгоритм SSA нуждается в улучшении. Прежде чем перейти к рассмотрению примеров, введем несколько определений, которые понадобятся в дальнейшем. Здесь и далее первой жирной буквой ряда будет обозначена его траекторная матрица. Например  $\mathbf{S}$  — это траекторная матрица ряда  $\mathbb{S}_N$ .

### 1.2.1. Основные определения

**Определение 1.** Если равенство  $\text{rk}(\mathbf{S}) = d < N/2$  имеет место для любого  $L: d \leq \min(L, K)$ , то будем говорить, что ряд  $\mathbb{S}_N$  имеет ранг  $d$  (обозн.  $\text{rk}(\mathbb{S}_N) = d$ ; говорят  $\mathbb{S}_N$  — ряд конечного ранга).

**Определение 2.** Пусть  $\mathbb{S}_N$  ряд ранга  $d$ . Числом обусловленности траекторной матрицы  $\mathbf{S}$  будем называть  $\text{cond}(\mathbf{S}) = \frac{\lambda_1^{\text{sig}}}{\lambda_d^{\text{sig}}}$ , где  $\lambda_i^{\text{sig}}$  — собственные числа матрицы  $\mathbf{S}$ , взятые в невозрастающем порядке.

**Определение 3.** Пусть  $\mathbb{F}_N = \mathbb{S}_N + \mathbb{R}_N$ , где  $\mathbb{S}_N = (s_1, s_2, \dots, s_N)$  — сигнал, а  $\mathbb{R}_N = (r_1, r_2, \dots, r_N)$  — стационарный процесс. Отношение сигнал/шум (сокр. SNR) — это безразмерная величина, равная  $\frac{\frac{1}{N} \sum_{i=1}^N s_i^2}{\text{Er}_1^2}$ .

Уровень шума будем мерить с помощью SNR. Точность восстановления сигнала будем измерять с помощью RMSE (root-mean-square error) — среднеквадратической

ошибки, которая вычисляется по формуле  $\text{RMSE}(\mathbb{S}_N, \tilde{\mathbb{S}}_N) = \sqrt{\sum_{i=1}^N (s_i - \tilde{s}_i)^2}$ , где  $\tilde{\mathbb{S}}_N$  — оценка сигнала. Отметим простой факт: чем больше SNR, тем уровень шума меньше.

Если в сингулярном разложении матрицы  $\mathbf{S}$  есть сингулярные числа, малые по величине относительно дисперсии шума, то компоненты сигнала могут смешаться с компонентами шума. Поэтому для полного восстановления сигнала нужно, чтобы в его разложении не было таких сингулярных чисел. Однако, если амплитуда шума имеет один порядок с амплитудой первый компонент сигнала, то указанное выше требование можно переформулировать в терминах числа обусловленности. Другими словами: разброс собственных чисел сигнала (точнее собственных чисел его траекторной матрицы) должен быть небольшой. Поэтому для полного восстановления сигнала важно, чтобы  $\text{cond}(\mathbf{S})$  было не слишком большое.

### 1.2.2. Примеры

Перейдем к рассмотрению примеров. Пусть имеется временной ряд  $\mathbb{F}_N = \mathbb{S}_N + \mathbb{R}_N$ , где  $\mathbb{S}_N = (s_1, s_2, \dots, s_N)$  — сигнал, а  $\mathbb{R}_N = (r_1, r_2, \dots, r_N)$  с  $r_n = \epsilon_n$  — белый гауссовский шум с фиксированным SNR. Отношение сигнал/шум задается с помощью выбора дисперсии шума.

### Экспоненциальная и квадратичная модуляции

Пусть сигнал  $\mathbb{S}_N = \mathbb{A}_N \mathbb{T}_N$ , где  $\mathbb{T}_N = (t_1, t_2, \dots, t_N)$  с  $t_n = \cos(2\pi \frac{1}{12}n)$ ,  $\text{rk}(\mathbb{T}_N) = 2$ . Сравним два вида амплитудной модуляции:

- Экспоненциальная:  $\mathbb{A}_N^{(1)} = e^{-\frac{n}{80}}$ ,  $\text{rk}(\mathbb{A}_N^{(1)}) = 1$
- Квадратичная:  $\mathbb{A}_N^{(2)} = 15 \cdot 10^{-5}n^2 - 0.02n + 1.01$ ,  $\text{rk}(\mathbb{A}_N^{(2)}) = 3$

Соответственно сигналы будут выглядеть следующим образом:

- Эксп.-модулир. сигнал:  $\mathbb{S}_N^{(1)} = e^{-\frac{n}{80}} \cos(2\pi \frac{1}{12}n)$ ,  $\text{rk}(\mathbb{S}_N^{(1)}) = 2$
- Квадр.-модулир. сигнал:  $\mathbb{S}_N^{(2)} = (15 \cdot 10^{-5}n^2 - 0.02n + 1.01) \cos(2\pi \frac{1}{12}n)$ ,  $\text{rk}(\mathbb{S}_N^{(2)}) = 6$

Поскольку ряды  $\mathbb{S}_N^{(1)}$  и  $\mathbb{S}_N^{(2)}$  являются произведениями временных рядов, то ранги  $\mathbb{S}_N^{(1)}$  и  $\mathbb{S}_N^{(2)}$  вычисляются как произведение рангов их сомножителей [1].

Полином  $\mathbb{A}_N^{(2)}$  был выбран таким образом, чтобы временной ряд  $\mathbb{S}_N^{(2)}$  имел большое число обусловленности  $\text{cond}(\mathbf{S}^{(2)}) = 126.187$ , в то время как  $\text{cond}(\mathbf{S}^{(1)}) = 1.008$ . Полином задает плавно меняющуюся в небольших пределах амплитуду сигнала. Выбор конкретных коэффициентов был сделан так, чтобы  $\mathbb{A}_N^{(2)}$  проходил через точки  $(n; \mathbb{A}_N^{(2)}(n))$ :  $(1, 0.99)$ ,  $(60, 0.35)$ ,  $(120, 0.77)$ . Длины обоих рядов возьмем  $N = 120$ . Таким образом, в указанном промежутке укладывается несколько периодов немодулированного сигнала  $\mathbb{T}_N$ . Согласно стандартным рекомендациям [1] в качестве длины окна возьмем  $L = 60$ , т.к. оно составляет половину длины ряда и кратно основному периоду  $\mathbb{T}_N$ .

Изобразим собственные числа обоих сигналов.  $L$  — длина окна,  $r$  — ранг сигнала,  $N$  — длина ряда. Звездочками отмечены ненулевые собственные числа.

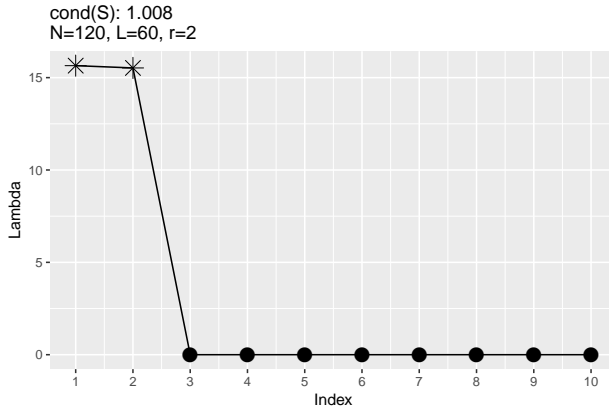


Рис. 1.1. Собственные числа  $\mathbb{S}_N^{(1)}$

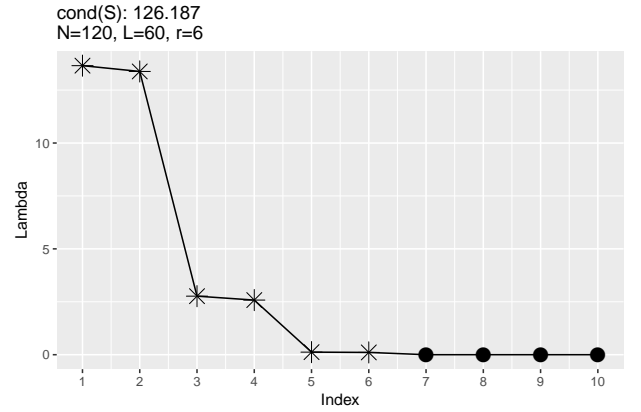
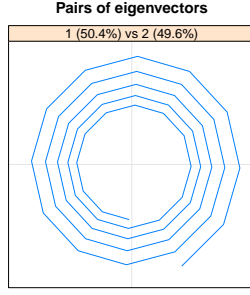
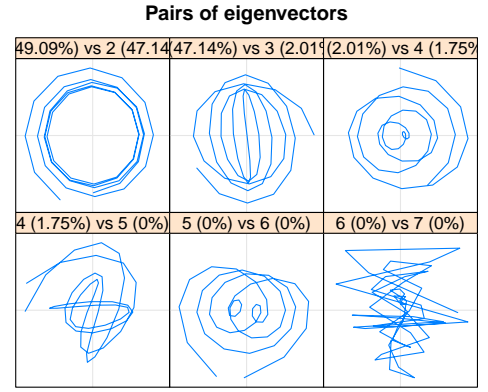
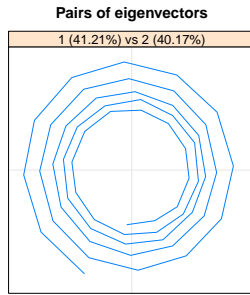
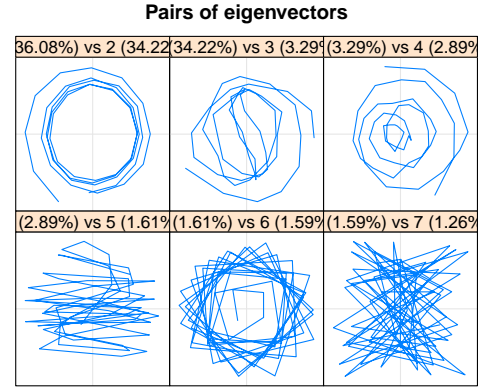


Рис. 1.2. Собственные числа  $\mathbb{S}_N^{(2)}$

Видим, что в случае экспоненциальной модуляции (см. рис. 1.1) получается небольшой разброс собственных чисел, что соответствует маленькому числу обусловленности. Напротив, у траекторной матрицы квадратично-модулированного сигнала — достаточно большое число обусловленности. Отсюда получается сильный разброс собственных чисел (см. рис. 1.2). Можно ожидать, что при добавлении шума собственные числа сигнала и шума перемешаются, и восстанавливая сигнал по  $r = 6$  компонентам мы будем брать шумовую составляющую.

Добавим шум к амплитудно-модулированным гармоникам и построим парные диаграммы рассеивания собственных векторов для сигналов  $\mathbb{S}_N^{(1)}$ ,  $\mathbb{S}_N^{(2)}$  и соответствующих рядов  $\mathbb{F}_N^{(1)}$ ,  $\mathbb{F}_N^{(2)}$ .



Рис. 1.3. Диаграмма с.в.  $\mathbb{S}_N^{(1)}$ Рис. 1.4. Диаграмма с.в.  $\mathbb{S}_N^{(2)}$ Рис. 1.5. Диаграмма с.в.  $\mathbb{F}_N^{(1)}$ Рис. 1.6. Диаграмма с.в.  $\mathbb{F}_N^{(2)}$ 

Отчетливо видно, что в случае экспоненциальной модуляции при добавлении шума структура собственных векторов сохраняется (см. рис. 1.3, 1.5), причем она регулярная. В случае с квадратичной модуляцией структура для 5-ой и 6-ой компонент явно не сохраняется при добавлении шума (см. рис.1.4, 1.6). Соответственно, выбор 5 и 6 компонент приведет к включению шумовой составляющей в оценку сигнала для этого вида модуляции.

Сравним теперь средние ошибки восстановления в случае оценки экспоненциально-модулированного сигнала по 2-ум компонентам, а квадратично-модулированного — по 4-м и 6-ти компонентам. Для этого сгенерируем  $k = 1000$  независимых реализаций белого гауссовского шума с фиксированным  $\text{SNR} = 5$  и добавим их к сигналу  $\mathbb{S}_N$ . Отношение сигнал/шум задается с помощью выбора дисперсии шума. Получим выборку из  $k$  временных рядов, на которой и оценим среднее значение ошибки.

Обозначим  $\sigma$ — стандартное отклонение шума. Стоит отметить, что при выбранном  $\text{SNR} = 5$ , значение  $\sigma$  равно 0.069. Для того, чтобы оценка сигнала получилась осмыс-

ленной необходимо, чтобы средняя по реализациям шума ошибка восстановления была значительно меньше, чем  $\sigma$ .

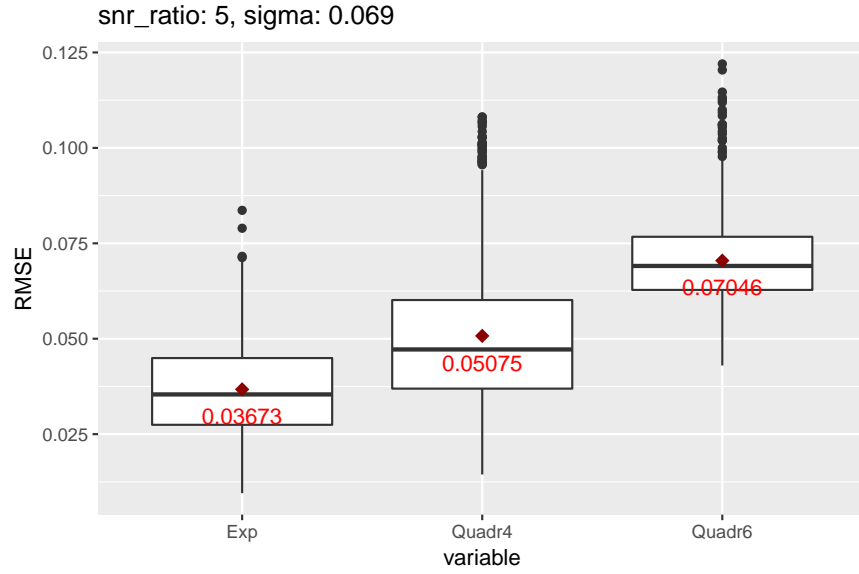


Рис. 1.7. Средняя ошибка оценки  $\mathbb{F}_N^{(1)}$  по 2 комп.,  $\mathbb{F}_N^{(2)}$  — по 4 и 6 комп.

Из рис.1.7 видно, что в случае оценки второго сигнала по 6-ти компонентам средняя ошибка восстановления примерно равна  $\sigma$ . Это говорит о том, что в оценку сигнала попали шумовые компоненты.

Восстановление  $\mathbb{S}_N^{(2)}$  по 4 компонентам получилось лучше (см. рис 1.7), но мы взяли 4 компоненты для оценки сигнала, в то время как исходный сигнал имеет ранг 6. Этот пример показывает, что в случае более сложной модуляции применение стандартного SSA алгоритма дает или неполное, или зашумленное восстановление.

## Частотная модуляция

Пусть имеется временной ряд  $\mathbb{F}_N = \mathbb{S}_N + \mathbb{R}_N$ , где  $\mathbb{S}_N = (s_1, s_2, \dots, s_N)$  с  $s_n = \cos(2\pi(\frac{n}{60})^2)$  — сигнал (частотно-модулированная гармоника), а  $\mathbb{R}_N = (r_1, r_2, \dots, r_N)$  с  $r_n = \epsilon_n$  — белый гауссовский шум с фиксированным  $\text{SNR} = 5$ . Отношение сигнал/шум задается с помощью выбора дисперсии шума.  $N = 120$  — длина ряда.

Рассмотрим такое понятие как *оптимальные параметры алгоритма*. Обозначим за  $\tilde{\mathbb{S}}_N^{\mathcal{P}}$  оценку сигнала  $\mathbb{S}_N$ . Эта оценка получена применением алгоритма с набором параметров  $\mathcal{P}$ . Тогда под оптимальным набором параметров  $\tilde{\mathcal{P}}$  будем понимать такой набор

параметров алгоритма, который минимизирует  $E\|\mathbb{S}_N - \tilde{\mathbb{S}}_N^{\mathcal{P}}\|^2$ , т.е. MSE. Однако, это будет оптимальный набор параметров алгоритма только для конкретного примера.

Отметим, что в отличие от предыдущих двух случаев частотно-модулированный сигнал не является рядом конечного ранга. Поэтому, прежде чем говорить о том, насколько хорошо стандартный SSA умеет выделять такой вид сигнала, необходимо подобрать оптимальные  $L$  и  $r$ . Их будем искать методом полного перебора  $1 < L \leq \frac{N}{2}$ ,  $1 \leq r \leq L$ .

С помощью языка R были найдены следующие параметры:  $L = 34$ ,  $r = 5$ . При этих параметрах оценка сигнала выглядит следующим образом.

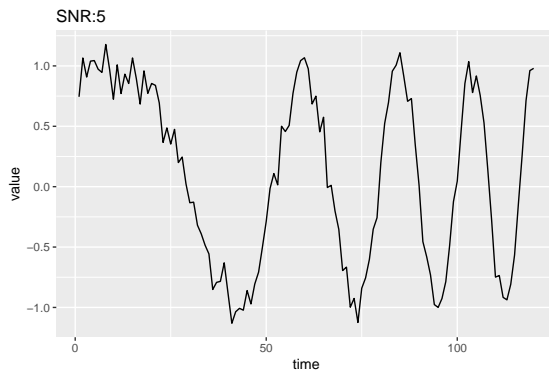


Рис. 1.8. Сигнал  $\mathbb{S}_N$

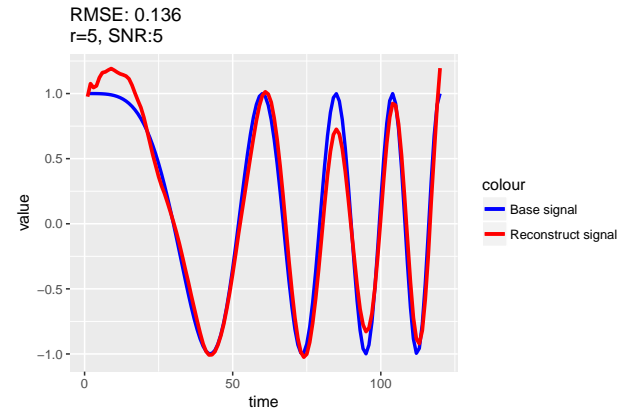


Рис. 1.9. Оценка сигнала.  $\mathbb{S}_N$

Рис. 1.9 показывает, что при найденных оптимальных параметрах базовый алгоритм SSA неприменим для такого вида модуляции.

Из разделов 1.2.2, 1.2.2 можно заключить, что базовый метод SSA плохо справляется с оценкой сигнала при рассмотренных амплитудной и частотной модуляциях.

## Глава 2

# Локальные модификации SSA: Overlap SSA, Sliding SSA, Averaging SSA

В главе 1 было показано, что стандартный алгоритм SSA нуждается в улучшении. В качестве нового подхода к оценке сигнала в случаях амплитудной или частотной модуляций будут рассматриваться локальные методы оценки сигнала. Локальность заключается в рассмотрении ряда на некотором небольшом промежутке, где структура ряда меняется несильно, и применении стандартного SSA алгоритма. То, каким образом выбирается локальный сегмент ряда, как происходит выбор параметров для SSA и как получается итоговая оценка сигнала, подробно разбирается ниже. В этой главе рассматриваются три алгоритма: Overlap SSA [2], Sliding SSA [3] и Averaging SSA. Первые два алгоритма аналогичны друг другу и отличаются только шагом, озаглавленным «Промежуточный шаг алгоритма». Последний алгоритм — Averaging SSA является модификацией алгоритма Overlap SSA.

У всех предлагаемых алгоритмов одинаковые входные параметры, отличающиеся только взаимными соотношениями. Поэтому вынесем их перед описанием алгоритмов. Далее  $N$  — длина ряда.

На вход алгоритмам подаются:

- $\mathbb{F}_N = (f_1, f_2, \dots, f_N)$  — временной ряд
- $Z$  — длина локального сегмента ряда
- $q$  — количество элементов внутри локального сегмента, используемое для восстановления
- $L$  — длина окна
- $r$  — количество компонент для оценки сигнала

Общие соотношения на параметры:

- $L, q < Z$
- $r \leq L$

Во всех алгоритмах рассматривается временной ряд  $\mathbb{F}_N = (f_1, f_2, \dots, f_N)$  и последовательность его скользящих отрезков (сегментов) длины  $Z$ , начало первого из которых совпадает с началом временного ряда. Эти отрезки пересекаются друг с другом. К каждому отрезку применяется SSA с длиной окна  $L$  и строится восстановленный временной ряд по первым  $r$  компонентам. Из результата берется  $q \leq Z$  точек по некоторому правилу. Оценка сигнала исходного временного ряда строится с помощью объединения результатов восстановления на отдельных сегментах по заданным алгоритмом правилам.

Под симметричностью алгоритма понимается принцип работы относительно исходного ряда и его перевернутой копии.

Программную реализацию на языке R рассматриваемых алгоритмов можно увидеть в [4].

## 2.1. Overlap SSA

Данный алгоритм был предложен M.C.R. Leles et al. и подробно разобран в [2].

Соотношения на параметры:

- $Z$  и  $q$  одинаковой четности

### 2.1.1. Алгоритм

**Сдвиг локального сегмента.** Для данного отрезка следующий получается путем смещения начала данного отрезка на  $q$  элементов ряда вправо. Исключение может составлять лишь последний отрезок: его начало выбирается таким образом, чтобы конец отрезка совпадал с концом ряда. Т.е. переход от предпоследнего отрезка к последнему осуществляется путем сдвига на  $\tilde{q}$ , которое может быть меньше, чем  $q$  (зависит от значений  $N, Z, q$ ).

**Промежуточный шаг алгоритма.** Рассмотрим сегмент длины  $Z$ . В общую оценку сигнала идет  $q$  его элементов из середины после восстановления.

**Начальная и конечная итерации.** На первом шаге алгоритма с первого локального отрезка берутся все точки, как и для промежуточного сегмента, но вдобавок еще и

все точки левее взятой середины. На последнем шаге берутся все точки, как и для промежуточного сегмента, только добавляются еще точки правее взятой середины.

**Способ объединения.** Полученные точки с каждого сегмента состыковываются друг с другом, т.к. они соответствуют разным моментам времени.

Стоит отметить, что алгоритм симметричен для промежуточных шагов: т.к. в качестве оценки с локального сегмента берется его середина длины  $q$ . Однако, при некоторых комбинациях значений  $N, Z, q$  с последнего сегмента может быть взято меньше элементов, чем с первого. Т.е. может наблюдаться несимметричность относительно граничных итераций.

## 2.2. Sliding SSA

Данный подход к оценке сигнала предложен Harmouche et.al. и подробно разобран в [3]. В статье рассматривается не только выделение сигнала, но еще и автоматическая группировка компонент. Однако для данной работы был взят только способ выбора  $q$  точек из скользящих отрезков длины  $Z$ .

Соотношения на параметры:

- $Z$  – нечетное
- $q \leq \frac{Z+1}{2}$

### 2.2.1. Алгоритм

**Сдвиг локального сегмента.** Для данного отрезка следующий получается путем смещения начала данного отрезка на  $q$  элементов ряда вправо. Исключение может составлять лишь последний отрезок: его начало выбирается таким образом, чтобы конец отрезка совпадал с концом ряда. Т.е. переход от предпоследнего отрезка к последнему осуществляется путем сдвига на  $\tilde{q}$ , которое может быть меньше, чем  $q$  (зависит от параметров  $N, Z, q$ ).

**Промежуточный шаг алгоритма.** Рассмотрим сегмент  $(f_p, f_{p+1}, \dots, f_{p+Z-1})$  длины  $Z$ . После восстановления, в общую оценку сигнала идет подвектор с индексами  $[n_c - q + 1, n_c - q + 2, \dots, n_c]$ , где  $n_c = p + \frac{Z-1}{2}$ . Т.е. берется  $q$  точек влево от средней точки.

**Начальная и конечная итерации.** На первом шаге алгоритма с первого локального отрезка берутся все точки, как и для промежуточного сегмента, но вдобавок еще и все точки левее уже взятых. На последнем шаге берутся все точки, как и для промежуточного сегмента, только добавляются еще точки правее уже взятых.

**Способ объединения.** Полученные точки с каждого сегмента состыковываются друг с другом, т.к. они соответствуют разным моментам времени.

Понятно, что алгоритм не симметричен, т.к. акцент в локальном сегменте делается на элементы, лежащие в его левой половине.

## 2.3. Averaging SSA

Данный подход является модификацией метода Overlap SSA, описанного в разделе 2.1.

Соотношения на параметры:

- $Z$  и  $q$  одинаковой четности

### 2.3.1. Алгоритм

**Сдвиг локального сегмента.** Для данного отрезка следующий получается путем смещения начала данного отрезка на 1 элемент ряда вправо. Всего локальных сегментов  $N - Z + 1$  штук.

**Промежуточный шаг алгоритма** Рассмотрим сегмент длины  $Z$ . В общую оценку сигнала идет  $q$  его элементов из середины после восстановления.

**Начальная и конечная итерации.** На первом шаге алгоритма с первого локального отрезка берутся все точки, как и для промежуточного сегмента, но вдобавок еще и все точки левее взятой середины. На последнем шаге берутся все точки, как и для промежуточного сегмента, только добавляются еще точки правее взятой середины.

**Способ объединения.** Для одной точки, вообще говоря, получается несколько оценок. Итоговый результат получается их усреднением.

Предложенный алгоритм полностью симметричен в виду простоты сдвига локальных отрезков: при любых допустимых значениях параметров алгоритма не возникает проблем с граничными итерациями.

## 2.4. Сравнение локальных алгоритмов

В этом разделе кратко сравним предложенные локальные алгоритмы между собой. По описанию алгоритмы Overlap SSA (см. раздел 2.1.1) и Sliding SSA (см. раздел 2.2.1) достаточно похожи. Разница заключается в способе выбора элементов из локального сегмента: в алгоритме Overlap SSA берется  $q$  средних элементов, а в алгоритме Sliding SSA берутся элементы, отстоящие не больше чем на  $q$  влево от середины. Тем самым, метод Sliding SSA имеет явную асимметрию.

Отметим два важных момента:

- Если положить  $q = 1$ , то все предложенные локальные алгоритмы совпадают между собой.
- Если положить  $Z = N$ , где  $N$  — длина ряда (нечетно), то все предложенные локальные алгоритмы совпадают со стандартным методом SSA (см. раздел 1.1).

Последний пункт выражает довольно естественную вещь: с ростом длины локального сегмента, локальный метод все больше походит на глобальный. В предельной ситуации весь ряд — один сегмент.

*Замечание:* при  $Z = N$  результаты локальных алгоритмов не зависят от  $q$ .



## Глава 3

**Численные эксперименты**

В главе 2 были рассказаны алгоритмы для оценки сигнала в нестационарном временном ряду. Задача этой главы описать численные эксперименты, которые показывают наличие или отсутствие преимуществ рассматриваемых подходов по сравнению со стандартным SSA.

**3.1. Постановка задачи**

Рассмотрим вещественный временной ряд  $\mathbb{F}_N = \mathbb{S}_N + \mathbb{R}_N$ , где  $\mathbb{S}_N = (s_1, s_2, \dots, s_N)$  с  $s_n = A(n) \cos(2\pi\omega(n) \cdot n + \phi)$  — сигнал, а  $\mathbb{R}_N = (r_1, r_2, \dots, r_N)$  — стационарный процесс. Стандартная задача любого метода заключается в оценке сигнала  $\mathbb{S}_N$  по имеющемуся ряду  $\mathbb{F}_N$ . Длину ряда  $N$  возьмем 121.

Рассматриваются следующие методы оценки сигнала:

- SSA
- Overlap SSA
- Sliding SSA
- Averaging SSA

В качестве типичных примеров сигналов, структура которых меняется с течением времени выступают:

- Экспоненциально-модулированный косинус (см. рис. 3.1)

$$s_n = e^{-\frac{n}{80}} \cos(2\pi \frac{1}{12}n)$$

- Квадратично-модулированный косинус (см. рис. 3.2)

$$s_n = (15 \cdot 10^{-5}n^2 - 0.02n + 1.01) \cos(2\pi \frac{1}{12}n)$$

- Частотно-модулированный косинус (см. рис. 3.3)

$$s_n = \cos(2\pi \cdot (\frac{n}{60})^2)$$

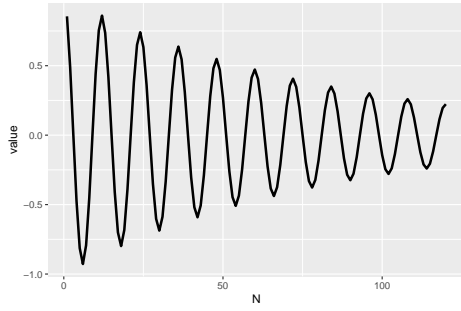


Рис. 3.1. Эксп.-модулир. сигнал

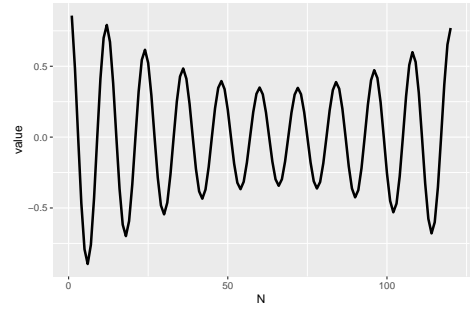


Рис. 3.2. Квадр.-модулир. сигнал

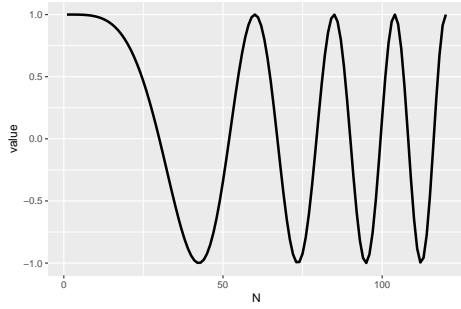


Рис. 3.3. Частотно-модулир. сигнал

Чтобы корректно сравнивать методы между собой, нужно для каждого из них найти оптимальные параметры при выбранном виде сигнала. Подбор оптимальных параметров осуществляется с помощью метода полного перебора всех допустимых параметров алгоритма.

Поэтапно опишем процедуру сравнения методов:

1. Выбрать вид выделяемого сигнала.
2. Зафиксировать последовательность  $\text{SNR}_n = \{3, 4, 5\}$  и для каждого элемента:
  - а. Промоделировать  $k = 5000$  реализаций белого гауссовского шума при выбранном SNR и добавить их к сигналу.  $\mathbb{F}_N = \mathbb{S}_N + \mathbb{R}_N$ , где  $\mathbb{S}_N = (s_1, s_2, \dots, s_N)$  с  $s_n = A(n) \cos(2\pi\omega(n) \cdot n + \phi)$  — выбранный сигнал, а  $\mathbb{R}_N = (r_1, r_2, \dots, r_N)$  с  $r_n = \epsilon_n$  — белый гауссовский шум. Получится выборка из временных рядов.
  - б. Для каждого из указанных методов найти оптимальные параметры на полученной выборке.
  - в. Сравнить методы при найденных оптимальных параметрах на выборке из временных рядов, полученной в 2а. Сравнить результаты двух методов можно с помощью t-test для зависимых выборок.

Для более наглядного представления уровня зашумленности ряда изобразим рассматриваемые сигналы (см. рис 3.1, 3.2, 3.3) с шумом. Отношение сигнал/шум равно 4.

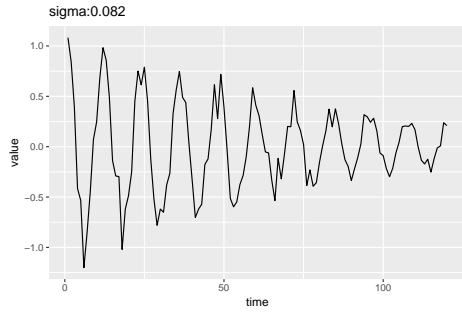


Рис. 3.4. Эксп.-модулир. сигнал с шумом

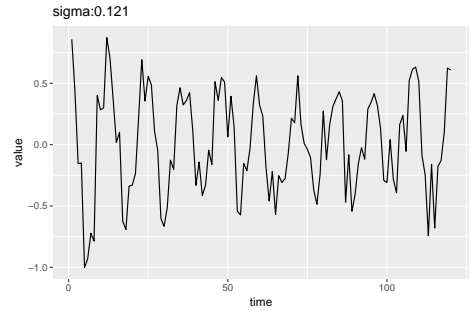


Рис. 3.5. Квадр.-модулир. сигнал с шумом

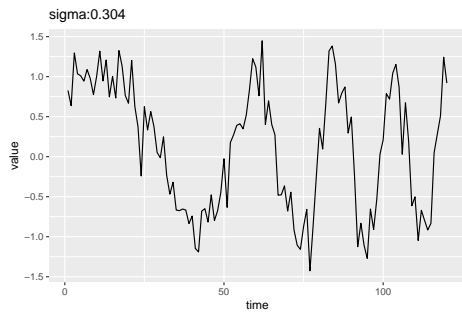


Рис. 3.6. Частотно-модулир. сигнал с шумом

Для того, чтобы найти оптимальные параметры для каждого из алгоритмов при фиксированном уровне шума применялся метод сеток. Сначала на все множество допустимых значений параметров алгоритма накладывалась сетка с большим шагом для каждого параметра: для  $Z$  от 1 до 121 с шагом 10, для  $q$  от 1 до  $Z$  с шагом 5, для  $L$  от 2 до  $\frac{Z}{2}$  с шагом 5 и для  $r$  от 1 до  $L$  с шагом 1. Затем вокруг найденных параметров накладывалась мелкая сетка с шагом 1 по каждому параметру и длиной, равной шагу крупной сетки. Окончательные оптимальные параметры вычислялись на мелкой сетке.

## 3.2. Полученные результаты

### 3.2.1. Экспоненциально-модулированный косинус

Таблица 3.1. Оптимальные параметры

	Z	q	L	r	RMSE
SSA			60	2	0.4522
Overlap SSA	121	$\forall$	60	2	0.4522
Sliding SSA	121	$\forall$	60	2	0.4522
Averaging SSA	121	$\forall$	60	2	0.4522

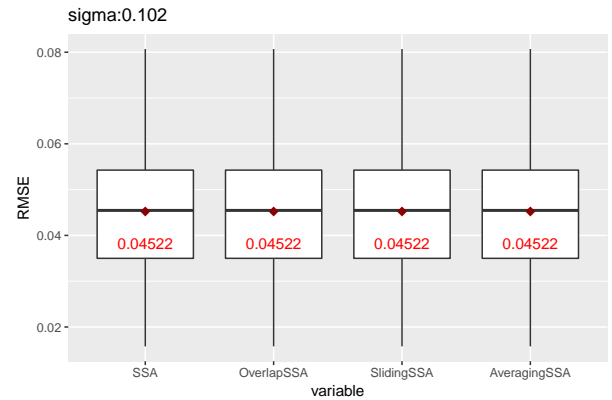


Рис. 3.7.  $\text{SNR}_n = 3$

Таблица 3.2. Оптимальные параметры

	Z	q	L	r	RMSE
SSA			60	2	0.3906
Overlap SSA	121	$\forall$	60	2	0.3906
Sliding SSA	121	$\forall$	60	2	0.3906
Averaging SSA	121	$\forall$	60	2	0.3906

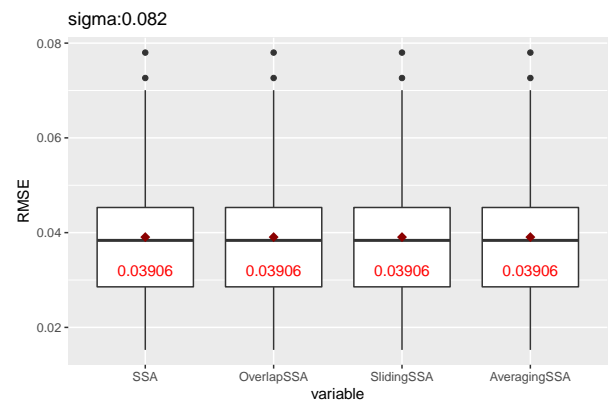
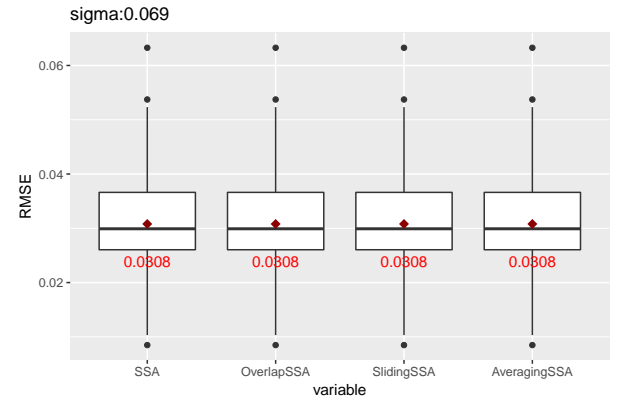


Рис. 3.8.  $\text{SNR}_n = 4$

Таблица 3.3. Оптимальные параметры

	Z	q	L	r	RMSE
SSA			60	2	0.0308
Overlap SSA	121	$\forall$	60	2	0.0308
Sliding SSA	121	$\forall$	60	2	0.0308
Averaging SSA	121	$\forall$	60	2	0.0308

Рис. 3.9.  $\text{SNR}_n = 5$ 

Ожидалось, что глобальный метод должен был показать лучший результат, поэтому параметры локальных методов должны быть такими, чтобы локальные совпадали с глобальным. В результате численного моделирования так и получилось. Согласно замечанию 2.4 в столбце  $q$  стоит  $\forall$ .

### 3.2.2. Квадратично-модулированный косинус

Таблица 3.4. Оптимальные параметры

	Z	q	L	r	RMSE
SSA			29	2	0.07084
Overlap SSA	43	1	22	2	0.06681
Sliding SSA	43	1	22	2	0.06681
Averaging SSA	43	1	22	2	0.06681

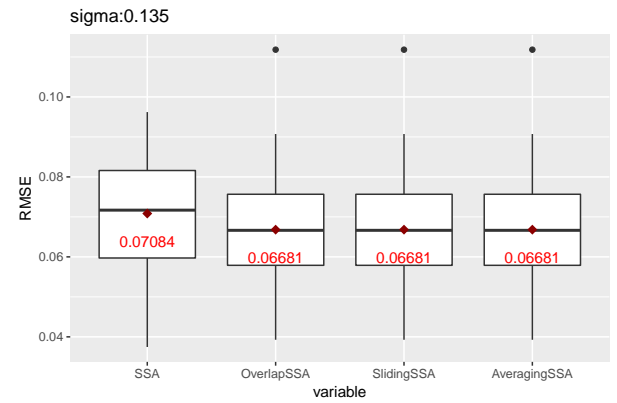


Рис. 3.10.  $\text{SNR}_n = 3$

Таблица 3.5. Оптимальные параметры

	Z	q	L	r	RMSE
SSA			27	2	0.06784
Overlap SSA	41	1	20	2	0.05587
Sliding SSA	41	1	20	2	0.05887
Averaging SSA	41	1	20	2	0.05887

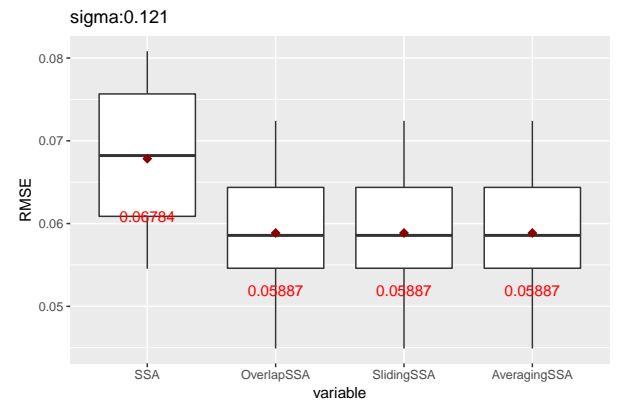
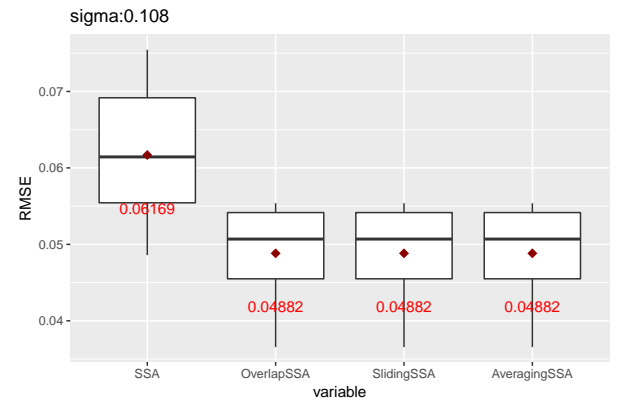


Рис. 3.11.  $\text{SNR}_n = 4$

Таблица 3.6. Оптимальные параметры

	Z	q	L	r	RMSE
SSA			27	2	0.06169
Overlap SSA	41	1	20	2	0.04882
Sliding SSA	41	1	20	2	0.04882
Averaging SSA	41	1	20	2	0.04882

Рис. 3.12.  $\text{SNR}_n = 5$ 

Как и предполагалось для локальных методов, оптимальное значение  $q = 1$  и (см. 2.4) локальные алгоритмы совпадают. Более того, гипотеза о равенстве средних при уровне значимости  $\alpha = 0.05$  отвергается. Следовательно локальный метод дает значимое преимущество для оценки квадратично-модулированного сигнала.

### 3.2.3. Частотно-модулированный косинус

Таблица 3.7. Оптимальные параметры

	Z	q	L	r	RMSE
SSA			36	5	0.18797
Overlap SSA	31	1	11	2	0.16664
Sliding SSA	31	1	11	2	0.16664
Averaging SSA	31	1	11	2	0.16664

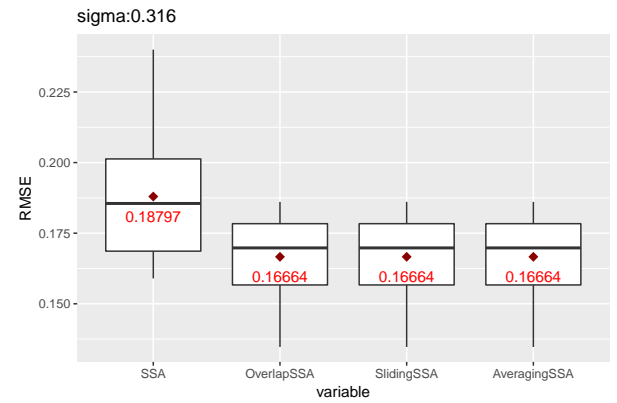


Рис. 3.13.  $\text{SNR}_n = 3$

Таблица 3.8. Оптимальные параметры

	Z	q	L	r	RMSE
SSA			34	5	0.15664
Overlap SSA	29	1	13	2	0.13849
Sliding SSA	29	1	13	2	0.13849
Averaging SSA	29	1	13	2	0.13849

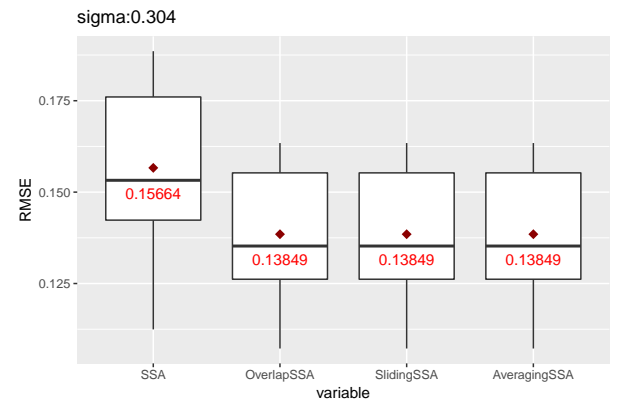
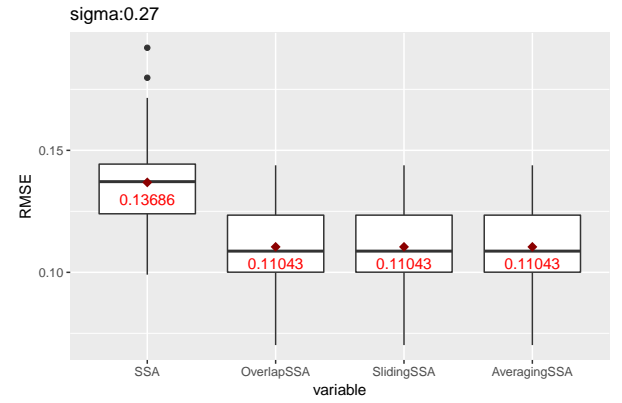


Рис. 3.14.  $\text{SNR}_n = 4$



Таблица 3.9. Оптимальные параметры

	Z	q	L	r	RMSE
SSA			34	5	0.13686
Overlap SSA	29	1	13	2	0.11043
Sliding SSA	29	1	13	2	0.11043
Averaging SSA	29	1	13	2	0.11043

Рис. 3.15.  $\text{SNR}_n = 5$ 

В случае частотной модуляции при поиске параметров значение  $q = 1$  оказалось оптимальным для всех локальных алгоритмов. Поэтому, согласно пункту 2.4, все локальные алгоритмы совпадают между собой. Проверка гипотезы о равенстве средних между локальным и глобальным методами при уровне значимости  $\alpha = 0.05$  дает значимое различие. Отсюда можно заключить, что рассматриваемые локальные алгоритмы дают улучшение в случае оценки частотной модуляции по сравнению со стандартным SSA.

## Заключение

Таким образом, в работе мною проделано следующее:

- Изучена теория метода SSA
- Разобраны и реализованы локальные алгоритмы Overlap SSA, Sliding SSA и Averaging SSA на языке R

Также в ходе численных экспериментов были получены выводы о том, какой метод дает лучший результат для рассматриваемого вида сигнала.

В случае экспоненциальной модуляции с оценкой сигнала лучше всех справился стандартный SSA. В случае более сложной амплитудной модуляции — квадратичной, локальные методы показали значимый результат, который оказался немного лучше, чем у стандартного SSA. В случае с частотной модуляцией результат оценки локальными методами оказался намного лучше, чем стандартным SSA.

Оптимальными параметрами для всех локальных методов оказались такие параметры, при которых методы совпали, а именно, использование только центральной точки восстановления и сдвиг на одну точку ( $q = 1$ ). Это логично, так как ошибка восстановления в центральной точке минимальна, а степень локальности регулируется выбором длины сегмента  $Z$ . Поэтому можно считать, что есть всего один локальный метод и не рассматривать его варианты (если речь не идет о трудоемкости, которая, чем меньше  $q$ , тем больше). Что касается выбора  $Z$ , то моделирование показало, что чем сложнее сигнал для SSA, тем больше должна быть степень локальности метода (меньше  $Z$ ). В примере с частотной модуляцией  $Z$  меньше. Про оптимальное  $L$  можно прочитать в [5], где написано, что  $L$  должно составлять от 0.3 до 0.5 от длины ряда.

В работе не был сделан более широкий выбор примеров сигналов для численного исследования, т.к. для этого нужен более эффективный способ выбора оптимальных параметров.

## Список литературы

1. Golyandina N., Nekrutkin V., Zhigljavsky A. Analysis of Time Series Structure: SSA and Related Techniques. — London : Chapman&Hall/CRC, 2001.
2. Improving reconstruction of time-series based in Singular Spectrum Analysis: A segmentation approach / M. Leles, J. Sansão, A. Mozelli, H. Guimarães // Digital Signal Processing. — 2018. — Vol. 77. — P. 63–76.
3. Harmouche J., Fourier D. The Sliding Singular Spectrum Analysis: A Data-Driven Non-stationary Signal Decomposition Tool // Transaction on signal processing. — 2018. — Vol. 66. — P. 251–262.
4. Sandalov S. Local algorithms. — 2019. — may. — URL: <http://doi.org/10.5281/zenodo.3235301>.
5. Golyandina N. On the choice of parameters in Singular Spectrum Analysis and related subspace-based methods // Stat. Interface. — 2010. — Vol. 3, no. 3. — P. 259–279.