

Санкт-Петербургский государственный университет
Прикладная математика и информатика
Статистическое моделирование

Третьякова Александра Леонидовна

РОБАСТНЫЕ ВАРИАНТЫ МЕТОДА АНАЛИЗА СИНГУЛЯРНОГО СПЕКТРА

Отчет о научно-исследовательской работе

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

Санкт-Петербург
2020

Оглавление

Введение	4
Глава 1. Стандартный метод SSA и его свойства	7
1.1. Алгоритм метода SSA	7
1.1.1. Вложение	7
1.1.2. Сингулярное разложение	7
1.1.3. Группировка	8
1.1.4. Диагональное усреднение	8
1.2. Разделимость	8
1.3. Ранг ряда	9
Глава 2. Модификации метода SSA с проекторами по некоторой норме	11
2.1. Схема методов	11
2.2. Вид проекторов на пространство ганкелевых матриц по различным нормам	12
2.3. Построение проектора по норме в \mathbb{L}_1 на множество матриц ранга, не превосходящего r . Последовательный метод	14
2.4. Построение проектора по взвешенной норме в \mathbb{L}_2 на множество матриц ранга, не превосходящего r	16
2.4.1. Метод с итеративным обновлением весов	16
2.4.2. Выбор параметра σ_{ij}	20
2.4.3. Выбор параметра α	20
2.4.4. Модификация метода с итеративным обновлением весов	22
2.4.5. Модификация метода с итеративным обновлением весов с исполь- зованием метода Гаусса-Ньютона	24
2.5. Оценка трудоемкости методов	27
Глава 3. Вычислительные эксперименты	30
3.1. Пример 1	30
3.2. Пример 2	35
3.3. Пример 3	38
3.4. Выводы	39

3.5. Исследование числа итераций	42
Заключение	44
Список литературы	46

Введение

В реальной жизни часто возникают задачи исследования различных процессов с течением времени. Пусть имеется $x(t)$ — функция, описывающая некоторый процесс во времени. Если произвести измерения через одинаковые промежутки времени t_i , где $i = 1, \dots, N$, тогда $x_i = x(t_i)$ представляют собой временной ряд $\mathbf{X} = (x_1, \dots, x_N)$.

Для решения многих задач, к примеру, экономических, таких как планирование производства или инвестиций, оказывается полезным на основе данных за предшествующий период выделить основную динамику и тенденции, а также спрогнозировать развитие процесса. В данной работе для исследования временных рядов будет применен метод «Гусеница»-SSA (Singular Spectrum Analysis) [1, 2], который позволяет анализировать ряд без задания его параметрической модели. Метод нашел свое применение в задачах исследования климатических явлений [3], динамических систем [4, 5] и во многих других областях. Пусть имеется временной ряд $\mathbf{X} = (x_1, \dots, x_N)$ длины N , который представляет собой сумму сигнала и шума: $x_i = s_i + r_i$, $i = 1, \dots, N$. Данный метод позволяет получить разложение интересующего нас временного ряда \mathbf{X} на интерпретируемые аддитивные составляющие:

$$\mathbf{X} = \mathbf{S} + \mathbf{R},$$

где \mathbf{S} — сигнал, \mathbf{R} — шум, например, некоторый стационарный процесс.

На практике часто возникают выделяющиеся наблюдения или выбросы, которые можно интерпретировать как ошибки в данных или сбои измерительного прибора, значительно большие, чем размер шума. Отфильтровать их оказывается непростой задачей, необходимо сначала разобраться со структурой ряда, чтобы понять, что данное значение является выбросом. Поэтому разработка исходно устойчивых к выбросам методов представляет значительный интерес.

Ранее в работе [6] уже были предложены несколько устойчивых к выбросам вариантов метода. В данной работе стоит задача рассмотреть новые идеи построения устойчивых модификаций и предложить новые методы, которые бы подходили как для рядов с шумом постоянной дисперсии, так и для рядов с гетероскедастичным шумом, и при этом были не менее точные, чем уже предложенные методы.

В методе SSA при выделении сигнала используются два проектора, которые могут строиться по различным нормам. Один из проекторов — это проектор на пространство

ганкелевых матриц, второй — проектор на множество матриц ранга, не превосходящего r . В стандартном методе SSA используются проекторы в пространстве матриц по норме \mathbb{L}_2 (норма Фробениуса).

В качестве модификаций в работе [6] рассматривался стандартный прием использования аппроксимации (проекции) по норме в \mathbb{L}_1 вместо \mathbb{L}_2 . Если построение проектора на ганкелевы матрицы по норме \mathbb{L}_1 не представляет трудности, то вычисление проектора на матрицы ранга, не превосходящего r , по норме \mathbb{L}_1 не имеет решения в замкнутой форме. Имеются методы, численно решающие приближенные задачи, но не известно достаточно хороших методов для задачи, которую требуется решить при построении проектора на матрицы ранга, не превосходящего r .

Еще одной идеей для достижения устойчивости метода к выделяющимся наблюдениям является присвоение значениям в точках, содержащих выбросы, меньший вес. Рассмотрим алгоритм, описанный в статье [7], включающий в себя итеративное обновление весов. Метод имеет параметры: $\{\sigma_{ij}, i = 1, \dots, L, j = 1, \dots, K\}$ (нормировка для остатков) и α (значение, начиная с которого точку считать выбросом). В оригинальном методе из статьи [7] параметры σ_{ij} полагаются одинаковыми для всех элементов траекторной матрицы, то есть $\sigma_{ij} = \sigma \forall i, j$. Данный метод оказывается неподходящим для рядов, где шум имеет непостоянную дисперсию. Поэтому стоит задача предложить его модификацию, которая бы оставалась устойчивой как в случае рядов с шумом постоянной дисперсии, так и в случае рядов с гетероскедастичным шумом. В модификации метода предполагается замена параметра σ_{ij} на элементы траекторной матрицы тренда (оценки математического ожидания) из ряда из модулей остатков.

Структура работы следующая. В главе 1 опишем базовый алгоритм метода SSA, введем необходимые понятия и обозначения, обсудим выбор параметров метода на основе теории метода SSA.

В начале главы 2 рассмотрим общую схему методов без указания конкретной нормы. Приведем вид проекторов на пространство ганкелевых матриц по трем нормам: по норме в пространствах \mathbb{L}_2 , \mathbb{L}_1 и взвешенной норме в \mathbb{L}_2 . В данной главе представим нахождение \mathbb{L}_1 -проектора на множество матриц ранга, не превосходящего r . В работе идет речь о последовательном методе проектирования. Этот метод рассматривается в R-пакете в рамках построения устойчивого к выбросам анализа главных компонент [8]. Далее опишем метод с итеративным обновлением весов и его модификацию, подходя-

щую для рядов с гетероскедастичным шумом. Приведем алгоритмы для каждого из методов.

Глава 3 будет содержать численные сравнения с исследованием влияние выброса на результат восстановления сигнала. В данной главе опишем сравнение последовательного метода с классическим методом L2-SSA, а также с методом из статьи [7] и с его модификацией. Сравнение будем проводить при отсутствии выделяющихся наблюдений, при 1% выбросов в случайных точках ряда и при 5% выбросов.

Проведем сравнение на нескольких примерах. Один из них уже был представлен в работе [6], но мы добавим большее количество выбросов. Второй пример — ряд с растущей амплитудой с шумом, имеющим непостоянную дисперсию. Третий пример — ряд с сильно растущей амплитудой и большим разбросом значений, но с шумом постоянной дисперсии.

В заключении опишем основные результаты работы, подведем итоги. На основе проведенных исследований дадим рекомендации, какой из рассмотренных методов следует использовать в том или ином случае.

Глава 1

Стандартный метод SSA и его свойства

1.1. Алгоритм метода SSA

Кратко опишем базовый алгоритм метода «Гусеница»-SSA, следуя [2].

1.1.1. Вложение

На первом шаге алгоритма выбирается некоторое целое число L : $1 < L < N$, называемое *длиной окна*. Исходный временной ряд переводится в последовательность многомерных векторов длины L . В результате образуются $K = N - L + 1$ векторов вложения

$$X_i = (x_i, \dots, x_{i+L-1})^T, 1 \leq i \leq K.$$

Траекторной матрицей ряда \mathbf{X} называется матрица

$$\mathbf{X} = [X_1 : \dots : X_K] = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_K \\ x_2 & x_3 & x_4 & \dots & x_{K+1} \\ x_3 & x_4 & x_5 & \dots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \dots & x_N \end{pmatrix}.$$

Заметим, что построенная таким образом траекторная матрица \mathbf{X} является *ганке-левой*, т.е. элементы, находящиеся на диагоналях $i + j = \text{const}$, равны между собой.

1.1.2. Сингулярное разложение

Пусть $\mathbf{S} = \mathbf{X}\mathbf{X}^T$, обозначим $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d > 0$ — ненулевые собственные числа матрицы \mathbf{S} , U_1, \dots, U_d — ортонормированная система собственных векторов матрицы \mathbf{S} , соответствующих собственным числам. *Сингулярным разложением* матрицы \mathbf{X} называется разложение

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

где $\sqrt{\lambda_i}$ — *сингулярные числа*, U_i — *левые сингулярные вектора*, $V_i = \frac{1}{\sqrt{\lambda_i}} \mathbf{X}^T U_i$ — *правые сингулярные вектора*.

Набор $(\sqrt{\lambda_i}, U_i, V_i)$ назовем i -ой *собственной тройкой* сингулярного разложения.

1.1.3. Группировка

Разделим множество индексов $\{1, \dots, d\}$ на m дизъюнктивных подмножеств I_1, \dots, I_m . Пусть $I = \{i_1, \dots, i_p\}$. Тогда *резльтирующая матрица* \mathbf{X}_I имеет вид

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

Таким образом, получаем разложение матрицы \mathbf{X} в сгруппированном виде

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

1.1.4. Диагональное усреднение

На последнем шаге каждая матрица \mathbf{X}_{I_i} переводится в новый ряд с помощью усреднения элементов матрицы вдоль антидиагоналей $i + j = k + 1$. Применяя диагональное усреднение к результирующим матрицам, получаем ряды $\tilde{\mathbf{X}}^{(k)} = (\tilde{x}_1^k, \dots, \tilde{x}_N^k)$.

В результате получаем разложение исходного ряда (x_1, \dots, x_N) в сумму m рядов:

$$x_n = \sum_{k=1}^m \tilde{x}_n^{(k)}.$$

1.2. Разделимость

Введем понятие разделимости, следуя [2].

Пусть \mathbf{X}_1 и \mathbf{X}_2 — временные ряды длины N и $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2$. Пусть выбрана длина окна L , тогда каждый из рядов порождает L -траекторную матрицу: \mathbf{X}_1 , \mathbf{X}_2 и \mathbf{X} . Обозначим \mathcal{L}_1^L и \mathcal{L}_2^L — линейные пространства, порожденные столбцами траекторных матриц \mathbf{X}_1 и \mathbf{X}_2 . Аналогично обозначим \mathcal{L}_1^K и \mathcal{L}_2^K — линейные пространства, порожденные строками траекторных матриц \mathbf{X}_1 и \mathbf{X}_2 , $K = N - L + 1$.

Определение 1. Ряды \mathbf{X}_1 и \mathbf{X}_2 называются слабо L -разделимыми, если $\mathcal{L}_1^L \perp \mathcal{L}_2^L$ и $\mathcal{L}_1^K \perp \mathcal{L}_2^K$.

В результате выполнения этапа разложения в алгоритме SSA получаем некоторое разложение траекторной матрицы ряда, которое не обязательно соответствует разделимости двух рядов. Поэтому необходимо усилить понятие разделимости.

Определение 2. Если ряды X_1 и X_2 слабо L -разделимы и множество собственных чисел сингулярного разложения траекторной матрицы одного ряда не пересекается с множеством собственных чисел разложения второго ряда ($\lambda_{1k} \neq \lambda_{2m} \forall k, m$), то ряды X_1 и X_2 сильно L -разделимы.

Утверждение 1. Пусть ряды X_1 и X_2 сильно L -разделимы. Тогда любое сингулярное разложение траекторной матрицы X ряда X можно разбить на две части, являющиеся сингулярными разложениями траекторных матриц рядов X_1 и X_2 .

Однако условия разделимости являются слишком жесткими и редко выполнены в реальных задачах. Поэтому в пособии [2] введено также понятие асимптотической разделимости.

Асимптотическая разделимость выполняется для более широкого класса рядов, чем точная разделимость. К примеру, $e^{\alpha n}$ и $\sin(2\pi\omega n)$, где $\alpha \neq 0$, $\omega \in (0, 0.5]$, асимптотически разделимы.

Для достижения лучшей разделимости необходимо выбирать большую длину окна ($L \sim N/2$). Большая длина окна позволяет выделить сигнал из зашумленного ряда, отделить тренд от периодических компонент. Не имеет смысла брать длину окна, большую чем половина длины ряда, а маленькая длина окна может привести к смешиванию компонент ряда.

1.3. Ранг ряда

Пусть $X_N = X_N^{(1)} + X_N^{(2)}$ и ряды $X_N^{(1)}$ и $X_N^{(2)}$ разделимы. Тогда в сингулярном разложении ряда X_N часть слагаемых относится к сингулярному разложению ряда $X_N^{(1)}$, а другая часть — к сингулярному разложению ряда $X_N^{(2)}$. Необходимо выяснить, сколько слагаемых относится к первому ряду и как их идентифицировать. На этапе группировки индексы слагаемых, относящихся к первому ряду, образуют подмножество I_1 , остальные — подмножество I_2 . Для того, чтобы понимать, сколько компонент относить к первому подмножеству, введем понятие ранга.

Рассмотрим ряд $X_N = (x_1, \dots, x_N)$, пусть L — длина окна.

Обозначим $\mathcal{L}^{(L)} = \text{span}(X_1, \dots, X_K)$ — траекторное пространство ряда X_N , где $X_i = (x_i, \dots, x_{i+L-1})^T$ — векторы вложения, $1 \leq i \leq K$.

Определение 3. Пусть $0 < d \leq \min(L, K)$. Будем говорить, что ряд X_N имеет L -ранг d , если $\dim \mathcal{L}^{(L)} = d$.

Например, в случае экспоненциального ряда $e^{\alpha n}$ для любых N и L ранг ряда равен 1, а ранг гармонического ряда $\sin(2\pi\omega n + \phi)$ равен 2 при $\omega < 1/2$ и 1 при $\omega = 1/2$, $\phi \in [0, 2\pi)$.

Модификации метода SSA с проекторами по некоторой норме

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 .

2.1. Схема методов

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

Выбирается длина окна L , и исходный временной ряд переводится в последовательность многомерных векторов длины L . В результате образуются $K = N - L + 1$ векторов вложения

$$X_i = (x_i, \dots, x_{i+L-1})^T, 1 \leq i \leq K.$$

Обозначим \mathcal{M} — пространство матриц $L \times K$,

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

\mathcal{M}_r — пространство матриц ранга, не превосходящего r .

Определим следующие операторы:

- Оператор вложения $\mathcal{T} : \mathbb{R}^N \rightarrow \mathcal{M}_{\mathcal{H}} : \mathcal{T}(\mathbf{X}) = \mathbf{X}$.
- $\Pi_r : \mathcal{M} \rightarrow \mathcal{M}_r$ — 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:

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

Это соответствует алгоритму SSA, описанному в разделе 1.1, для случая, когда восстановление производится по одной группе, состоящей из первых r компонент.

Проекторы Π_r и $\Pi_{\mathcal{H}}$ можно строить по различным нормам. С точки зрения вычислений, удобно выбирать \mathbb{L}_2 -норму для построения проекторов на пространство ганкелевых матриц и матриц ранга, не превосходящего r , поскольку целевая функция является гладкой и выпуклой, и решить задачу минимизации довольно просто, можно даже говорить о задании решения в явной форме. Однако при наличии выбросов норма Фробениуса оказывается недостаточно устойчивой. Норма в пространстве \mathbb{L}_1 является более устойчивой к выделяющимся наблюдениям, однако сложность в ее использовании состоит в негладкой и невыпуклой строго целевой функции, поэтому возникает проблема в применении стандартных методов оптимизации. Существует также вариант использования взвешенной нормы в \mathbb{L}_2 с присваиванием меньших весов точкам, содержащим выделяющиеся наблюдения.

В работе будут рассмотрены проекторы по нормам в пространствах \mathbb{L}_2 , \mathbb{L}_1 и взвешенной норме в пространстве \mathbb{L}_2 . Будем рассматривать следующие методы:

- Стандартный метод SSA, где оба проектора Π_r и $\Pi_{\mathcal{H}}$ строятся по норме в пространстве \mathbb{L}_2 ,
- Метод с проекторами Π_r и $\Pi_{\mathcal{H}}$ по норме в пространстве \mathbb{L}_1 ,
- Метод с проекторами Π_r и $\Pi_{\mathcal{H}}$ по взвешенной норме в пространстве \mathbb{L}_2 .

2.2. Вид проекторов на пространство ганкелевых матриц по различным нормам

Рассмотрим, как выглядят проекторы на пространство ганкелевых матриц по нормам в пространствах \mathbb{L}_2 , \mathbb{L}_1 и взвешенной норме в \mathbb{L}_2 .

Определение 4. Пусть \mathbf{A} — матрица $L \times K$.

Норма в пространстве \mathbb{L}_2 (норма Фробениуса): $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^L \sum_{j=1}^K a_{ij}^2}$.

Лемма 1 (Известный результат) $\operatorname{argmin}_a \mathbb{E}(\xi - a)^2 = \mathbb{E}\xi$.

Доказательство.

$$\begin{aligned} \mathbb{E}(\xi - a)^2 &= \mathbb{E}((\xi - \mathbb{E}\xi) + (\mathbb{E}\xi - a))^2 = \mathbb{E}((\xi - \mathbb{E}\xi)^2 + 2(\xi - \mathbb{E}\xi)(\mathbb{E}\xi - a) + (\mathbb{E}\xi - a)^2) = \\ &= \mathbb{E}(\xi - \mathbb{E}\xi)^2 + 2(\mathbb{E}\xi - a)\mathbb{E}(\xi - \mathbb{E}\xi) + (\mathbb{E}\xi - a)^2 = \mathbb{E}(\xi - \mathbb{E}\xi)^2 + (\mathbb{E}\xi - a)^2. \end{aligned}$$

Следовательно, $\operatorname{argmin}_a \mathbb{E}(\xi - a)^2 = \mathbb{E}\xi$. □

Необходимо построить проектор на множество ганкелевых матриц по норме Фробениуса. Заметим, что

$$\|\mathbf{X} - \mathbf{A}\|_F^2 = \sum_{i=1}^L \sum_{j=1}^K (x_{ij} - a_{ij})^2 = \sum_{k=1}^{L+K-1} \sum_{i+j=k+1} (x_{ij} - a_k)^2 \longrightarrow \min_{\mathbf{A} \in \mathcal{M}_{\mathcal{H}}}.$$

Тогда если взять в качестве ξ случайные величины, принимающие значения на побочной диагонали с равными вероятностями, то из Леммы 1 следует, что проектор на множество ганкелевых матриц по норме Фробениуса строится посредством усреднения элементов на диагоналях $i + j = \text{const}$.

Опишем теперь вид данного проектора по норме в пространстве \mathbb{L}_1 .

Определение 5. Пусть \mathbf{A} — матрица $L \times K$.

Норма в пространстве \mathbb{L}_1 : $\|\mathbf{A}\|_1 = \sum_{i,j} |a_{ij}|$.

Лемма 2 (Известный результат)

В случае непрерывного распределения $\operatorname{argmin}_a \mathbb{E}|\xi - a| = \text{med } \xi$.

Доказательство этой леммы приведено в работе [6].

Используя лемму 2, получаем, что $\Pi_{\mathcal{H}}$ в пространстве \mathbb{L}_1 строится посредством взятия выборочной медианы значений на побочных диагоналях $i + j = \text{const}$.

Определение 6. Пусть \mathbf{A} — матрица $L \times K$, \mathbf{W} — матрица весов $L \times K$.

Норма в пространстве \mathbb{L}_2 с весами \mathbf{W} : $\|\mathbf{A}\|_W = \sqrt{\sum_{i=1}^L \sum_{j=1}^K w_{ij} a_{ij}^2}$.

Известно следующее утверждение о построении проектора на пространство ганкелевых матриц по данной норме.

Утверждение 2 ([9]) Для построения проекции $\hat{\mathbf{A}} = \Pi_{\mathcal{H}} \mathbf{A} = \{\hat{a}_{ij}\}_{i,j=1}^{L,K}$ необходимо суммировать элементы на диагоналях $i + j = \text{const}$ с весами и нормировать на сумму весов:

$$\hat{a}_{ij} = \frac{\sum_{l,k:l+k=i+j} w_{lk} a_{lk}}{\sum_{l,k:l+k=i+j} w_{lk}}.$$

Замечание 1. В случае ганкелевой матрицы весов \mathbf{W} проектор на пространство ганкелевых матриц по взвешенной норме в \mathbb{L}_2 совпадает с проектором на пространство ганкелевых матриц по норме в \mathbb{L}_2 .

2.3. Построение проектора по норме в \mathbb{L}_1 на множество матриц ранга, не превосходящего r . Последовательный метод

В отличие от проектора на множество матриц ранга r по норме Фробениуса, построение данного проектора в пространстве \mathbb{L}_1 является вычислительно сложной задачей. Рассмотрим один из методов построения проектора на множество матриц ранга, не превосходящего r , по норме в пространстве \mathbb{L}_1 .

Стоит задача проектирования матрицы \mathbf{X} на множество матриц ранга, не превосходящего r . Задачу оптимизации можно представить в виде

$$\min_{\mathbf{V}, \mathbf{U}} \|\mathbf{X}^T - \mathbf{V}\mathbf{U}^T\|_1 = \sum_{i=1}^L \|X_i - \mathbf{V}\mathbf{U}_i\|_1,$$

где \mathbf{V} — матрица $K \times r$, \mathbf{U} — матрица $L \times r$. Столбцы матрицы \mathbf{V} определяют главные компоненты. Матрица $\mathbf{E} = \mathbf{U}\mathbf{V}^T$ — проекция \mathbf{X} на множество матриц ранга, не превосходящего r , которую необходимо найти.

В пакете rsaL1 [10] имеется метод l1rsa, позволяющий вычислить проекцию в \mathbb{L}_1 на множество матриц ранга, не превосходящего r . Подробнее метод описан в статье [8].

Приведем алгоритм в наших обозначениях. Пусть $\mathbf{Y} \in \mathbb{R}^{L \times K}$. Задача выглядит следующим образом:

$$\|\mathbf{Y} - \mathbf{U}\mathbf{V}^T\|_1 \rightarrow \min_{\mathbf{U}, \mathbf{V}}.$$

Далее представлен алгоритм последовательного метода решения данной задачи.

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

Входные данные: $\mathbf{Y} \in \mathbb{R}^{L \times K}$ — траекторная матрица ряда, r — ранг сигнала; параметры критерия останова: $\varepsilon = 10^{-4}$, 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 $\mathbf{U}(0) \in \mathbb{R}^{L \times r}$, нормировка столбцов $\mathbf{U}(0)$;

2. $t := 0$;

3. **повторять**

- a. $t := t + 1$;
- b. $\mathbf{V}(t) = \underset{\mathbf{V} \in \mathbb{R}^{K \times r}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{U}(t-1)\mathbf{V}^T\|_1$;
- c. $\mathbf{U}(t) = \underset{\mathbf{U} \in \mathbb{R}^{L \times r}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{U}\mathbf{V}^T(t)\|_1$;
- d. Нормировка столбцов $\mathbf{U}(t)$;

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

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

Решаем задачу, меняя на каждой итерации \mathbf{U} и \mathbf{V} и разбивая исходную задачу на линейные подзадачи. $\mathbf{U}(0)$ можно инициализировать с помощью сингулярного разложения траекторной матрицы \mathbf{Y} в пространстве \mathbb{L}_2 .

Согласно [11], задача $\mathbf{V}(t) = \underset{\mathbf{V} \in \mathbb{R}^{K \times r}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{U}(t-1)\mathbf{V}^T\|_1$ может быть разбита на L линейных подзадач вида

$$\sum_{i=1}^K \delta_i \rightarrow \min_{\delta, v_i}$$

с ограничениями

$$-\delta \leq Y_i - \mathbf{U}(t-1)\mathbf{v}_i \leq \delta,$$

где $Y_i \in \mathbb{R}^L$ — столбцы \mathbf{Y} , $\mathbf{v}_i \in \mathbb{R}^r$ — строки \mathbf{V} . С решением каждой такой задачи получаем оценку \mathbf{v}_i , $i = 1, \dots, K$.

Задача $\mathbf{U}(t) = \underset{\mathbf{U} \in \mathbb{R}^{L \times r}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{U}\mathbf{V}^T(t)\|_1$ может быть записана как линейная задача вида

$$\sum_{i=1}^K \sum_{j=1}^L (\lambda_{ij}^+ + \lambda_{ij}^-) \rightarrow \min_{\mathbf{U}, \lambda^+, \lambda^-}$$

с ограничениями

$$\begin{aligned} Y_i - \mathbf{U} \mathbf{v}_i(t) + \lambda_i^+ - \lambda_i^- &= 0, \quad i = 1, \dots, K, \\ \lambda_i^+, \lambda_i^- &\geq 0, \quad i = 1, \dots, K, \end{aligned}$$

где $Y_i \in \mathbb{R}^L$ — столбцы \mathbf{Y} , $\mathbf{v}_i(t) \in \mathbb{R}^r$ — строки $\mathbf{V}(t)$.

2.4. Построение проектора по взвешенной норме в \mathbb{L}_2 на множество матриц ранга, не превосходящего r

В данном разделе опишем метод с итеративным обновлением весов из статьи [7], а также модификацию этого метода, подходящую для рядов с гетероскедастичным шумом.

2.4.1. Метод с итеративным обновлением весов

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

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

где \odot — поэлементное умножение, $\mathbf{W}^{1/2}$ — поэлементное взятие корня, веса $w_{ij} = w(\frac{y_{ij} - \hat{y}_{ij}}{\sigma_{ij}})$ вычисляются по формуле, как и в известном методе локальной регрессии loess,

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

Значения α и $\{\sigma_{ij}, i = 1, \dots, L, j = 1, \dots, K\}$ — параметры. График весовой функции $w(x)$ представлен на рисунке 2.1. Значение веса зависит от нормированных остатков: если остаток маленький по модулю, то вес максимальный, если остаток по модулю превосходит заданный параметр α , то вес обнуляется.

Алгоритм решения этой задачи представлен в статье [7]. Для начала опишем алгоритм решения задачи аппроксимации для фиксированной матрицы весов $\mathbf{W} \in \mathbb{R}^{L \times K}$. Алгоритм 2 решает задачу

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

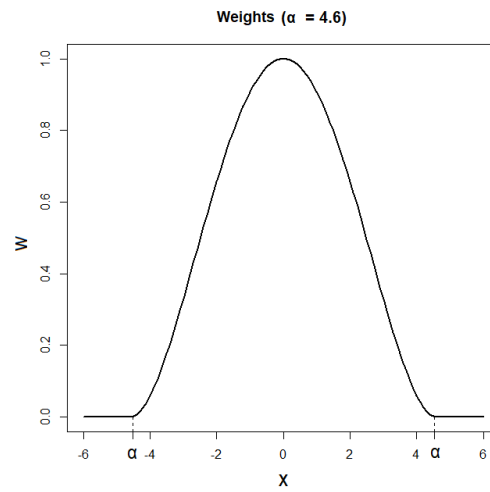


Рис. 2.1. График весовой функции $w(x)$.

при фиксированной \mathbf{W} .

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

Входные данные: $\mathbf{Y} \in \mathbb{R}^{L \times K}$ — траекторная матрица ряда, r — ранг сигнала, $\mathbf{W} \in \mathbb{R}^{L \times K}$ — матрица весов;
 параметры критерия останова: $\varepsilon = 10^{-4}$,
 максимальное число итераций $N_\alpha = 5$

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

1. $t := 0$;

2. **повторять**

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

$$(y_i - \mathbf{V}u_i)^T \mathbf{W}_i (y_i - \mathbf{V}u_i) \rightarrow \min_{u_i}, \quad i = 1, \dots, L, \quad (2.1)$$

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

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

$$(Y_j - \mathbf{U}v_j)^T \mathbf{W}^j (Y_j - \mathbf{U}v_j) \rightarrow \min_{v_j}, \quad j = 1, \dots, K, \quad (2.2)$$

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

в. $t := t + 1$.

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

Замечание 2. 1. Задача 2.1 решается с помощью QR -разложения матрицы $\mathbf{V}^T \mathbf{W}_i \mathbf{V}$.

2. Задача 2.2 решается с помощью QR -разложения матрицы $\mathbf{U}^T \mathbf{W}^j \mathbf{U}$.

Далее представлен алгоритм решения задачи построения проектора на множество матриц ранга, не превосходящего r , методом с итеративным обновлением весов. Алгоритм содержит вычисление вспомогательной матрицы $\Sigma = \{\sigma_{ij}\}_{i,j=1}^{L,K}$, которое обсудим далее в разделе 2.4.2.

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

Входные данные: $\mathbf{Y} \in \mathbb{R}^{L \times K}$ — траекторная матрица ряда, r — ранг сигнала; параметры критерия останова: $\varepsilon = 10^{-4}$, максимальное число итераций $N_{IRLS} = 10$

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

1. Инициализация $\mathbf{U} \in \mathbb{R}^{n \times r}$ и $\mathbf{V} \in \mathbb{R}^{p \times r}$ (например, с помощью сингулярного разложения матрицы \mathbf{Y});
2. Выбор параметра α ;
3. $t := 0$;
4. **повторять**

- a. Вычисление матрицы остатков $\mathbf{R} = \{r_{ij}\}_{i,j=1}^{n,p} = \mathbf{Y} - \mathbf{U}\mathbf{V}^T$;
- b. Обновление параметра σ_{ij} ;
- c. Вычисление матрицы весов $\mathbf{W} = \{w_{ij}\}_{i,j=1}^{L,K} = \{w(\frac{r_{ij}}{\sigma_{ij}})\}_{i,j=1}^{L,K}$, используя

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

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

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

с помощью алгоритма 2;

- e. $t := t + 1$.

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

Авторы статьи [7], ссылаясь на проведенные численные эксперименты, предлагают $\forall i, j$ выбрать $\sigma_{ij} = \sigma = 1.4826 \text{ med } |\mathbf{R} - \text{med } |\mathbf{R}||$, где \mathbf{R} — это вектор, составленный из всех элементов матрицы остатков $\mathbf{R} = \{r_{ij}\}_{i,j=1}^{L,K}$, то есть

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

Параметр α предлагается взять равным 4.685. Также говорится, что максимальное ко-

личество итераций N_α и N_{IRLS} , необходимых для сходимости, достаточно взять 5 и 10 для достижения приемлемой точности.

2.4.2. Выбор параметра σ_{ij}

У выбора σ_{ij} не зависящими от i, j присутствуют существенные недостатки. Описанный алгоритм из статьи не подходит, к примеру, для рядов с гетероскедастичным шумом. По умолчанию остатки нормировались на σ_{ij} , которая задавалась константой. Нормировка остатков на константный параметр $\sigma_{ij} = \sigma \quad \forall i, j$ в случае шума с непостоянной дисперсией приводит к неправильной идентификации точек с выбросами. Если шум растет к концу ряда, то веса у всех значений на конце ряда некорректно занижаются, и точки, не содержащие выбросов, могут получить вес, меньший, чем у выбросов в начале ряда. Поэтому приходим к выводу, что нормирующий параметр необходимо задавать динамически.

Будем рассматривать матрицу $\Sigma = \{\sigma_{ij}\}_{i,j=1}^{L,K}$ ганкелевой, что соответствует приписыванию весов элементам ряда. Тогда элементы на диагоналях $i + j = \text{const}$ матрицы Σ равны между собой, и можем обозначить $\sigma_{i+j-1} = \sigma_{lk} \quad \forall l, k : l + k = i + j$, где $i = 1, \dots, L, j = 1, \dots, K$. Обозначим новый параметр $\sigma = (\sigma_1, \dots, \sigma_N)^T$. Будем задавать параметр σ как тренд (математическое ожидание) ряда, состоящего из модулей остатков. Для оценки математического ожидания как тренд из ряда модулей остатков будем предполагать медленную зависимость модуля остатков от индекса. Выделять тренд будем следующими способами: локальной регрессией, скользящей медианой или взвешенной локальной регрессией.

2.4.3. Выбор параметра α

У метода есть параметр α , который влияет на то, какие точки будем считать выбросами, а какие — нет. Для того, чтобы понять, какое значение следует взять в качестве α , выведем вероятностную формулу для этого параметра. Будем задавать вероятность γ и подбирать параметр α так, чтобы нормированные модули остатков попадали в промежуток $(0; \alpha)$ с вероятностью γ . Для вывода вероятностной формулы для параметра α введем следующее определение.

Определение 7. Если $r \sim N(0, \sigma^2)$, то $|r| \sim N_H(\sigma^2)$ — полунормальное распределение

с параметром σ^2 , функция распределения:

$$F_H(x; \sigma^2) = \frac{2}{\sqrt{\pi}} \int_0^{x/\sqrt{2}\sigma^2} e^{-z^2} dz = \operatorname{erf}\left(\frac{x}{\sqrt{2}\sigma^2}\right). \quad (2.3)$$

Среднее и дисперсия: $\mathbb{E}|r| = \sigma\sqrt{\frac{2}{\pi}}$, $\mathbb{D}|r| = \sigma^2(1 - \frac{2}{\pi})$.

После ганкелизации матрицы остатков $\mathbf{R} = \mathbf{Y} - \mathbf{U}\mathbf{V}^T$, получаем ряд из остатков $\mathbf{R} = \{r_i\}_{i=1}^N$. Напомним, что мы рассматриваем ряд вида

$$x_i = s_i + \varepsilon_i, \quad i = 1, \dots, N,$$

где s_i — сигнал, ε_i — шум. Если предположить отделимость сигнала от шума, то матрица \mathbf{R} соответствует траекторной матрице шума $\{\varepsilon_i\}_{i=1}^N$. Для случайного шума точное совпадение матрицы остатков и траекторной матрицы шума невозможно, однако будем это предполагать в дальнейших рассуждениях. Тогда в предположении точной отделимости сигнала от шума ряд из остатков соответствует шуму, то есть $\mathbf{R} = \{r_i\}_{i=1}^N = \{\varepsilon_i\}_{i=1}^N$. Обозначим $r^* = \frac{|\varepsilon|}{\sigma}$, где $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_N)$ — тренд из ряда $|\mathbf{R}|$. В предположении точной отделимости сигнала от шума это означает, что $\sigma_i = \mathbb{E}|\varepsilon_i|$.

Зададим вероятность γ : $P(r^* \in (0, \alpha)) = \gamma$. Распишем:

$$P(r^* \in (0, \alpha)) = P(0 \leq \frac{|\varepsilon|}{\sigma} \leq \alpha) = P(\frac{|\varepsilon|}{\sigma} \leq \alpha) = \gamma.$$

Для того, чтобы получить выражение для α , нам необходимо знать распределение $\frac{|\varepsilon|}{\sigma}$.

Утверждение 3. Пусть $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. Пусть $\sigma = \mathbb{E}|\varepsilon|$. Тогда $r^* = \frac{|\varepsilon|}{\sigma}$ имеет полунормальное распределение $N_h(\frac{\pi}{2})$, среднее $\mathbb{E}r^* = 1$, дисперсия $\mathbb{D}r^* = \frac{\pi}{2} - 1$.

Доказательство. Если $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$, то

$$\frac{\varepsilon}{\sigma} = \frac{\varepsilon}{\mathbb{E}|\varepsilon|} \sim N(0, \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 \frac{2}{\pi}}) = N(0, \frac{\pi}{2}).$$

Тогда $\frac{|\varepsilon|}{\sigma} \sim N_h(\frac{\pi}{2})$ по определению полунормального распределения.

Посчитаем среднее и дисперсию r^* .

$$r^* = \frac{|\varepsilon|}{\sigma} = \frac{|\varepsilon|}{\mathbb{E}|\varepsilon|} = \frac{|\varepsilon|}{\sigma_\varepsilon \sqrt{\frac{2}{\pi}}}.$$

Тогда

$$\mathbb{E}r^* = \frac{\mathbb{E}|\varepsilon|}{\sigma_\varepsilon \sqrt{\frac{2}{\pi}}} = 1,$$

$$\mathbb{D}r^* = \frac{\mathbb{D}|\varepsilon|}{\sigma_\varepsilon^2 \frac{2}{\pi}} = \frac{\sigma_\varepsilon^2 (1 - \frac{2}{\pi})}{\sigma_\varepsilon^2 \frac{2}{\pi}} = \frac{\pi}{2} - 1.$$

□

Из утверждения 3 следует, что уравнение $P(\frac{|\varepsilon|}{\sigma} \leq \alpha) = \gamma$ переписывается в виде $F_h(\alpha) = \gamma$, где F_h — функция распределения $N_h(\frac{\pi}{2})$.

Используя формулу (2.3) для функции распределения полунормального распределения, получаем уравнение

$$\text{erf}\left(\frac{\alpha}{\sqrt{2}\frac{\pi}{2}}\right) = \gamma,$$

где erf — функция ошибок, которая имеется во многих пакетах в R. Отсюда получаем вероятностную формулу для параметра α

$$\alpha = \frac{\sqrt{2}\pi}{2} \text{erf}^{-1}(\gamma). \quad (2.4)$$

К примеру, для $\gamma = 0.95$ получим $\alpha \approx 3.079$. Для $\gamma = 0.99$ получим $\alpha \approx 4.046$.

Замечание 3. На основе утверждения 3 можно провести следующие рассуждения для временных рядов:

1. Пусть дисперсия шума постоянна, $x_i = s_i + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$, сигнал s_i точно отделим от шума. Тогда $r^* = \frac{|\varepsilon|}{\sigma}$ имеет полунормальное распределение $N_h(\frac{\pi}{2})$ среднее $\mathbb{E}r^* = 1$, дисперсия $\mathbb{D}r^* = \frac{\pi}{2} - 1$.
2. Пусть шум гетероскедастичный, $x_i = s_i + t_i \varepsilon_i$, $t_i > 0$, $\varepsilon_i \sim N(0, 1)$, сигнал s_i точно отделим от шума. Пусть $\mathbf{R} = (r_1, \dots, r_N)^T$ — вектор из остатков, $\mathbf{R}_+ = (|r_1|, \dots, |r_N|)^T$ — вектор из модулей остатков, $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_N)^T = \mathbb{E}(\mathbf{R}_+)$. Тогда каждая компонента вектора $\mathbf{r}^* = (\frac{|r_1|}{\sigma_1}, \dots, \frac{|r_N|}{\sigma_N})^T$ имеет полунормальное распределение: $r_i^* = \frac{|r_i|}{\sigma_i} \sim N_h(\frac{\pi}{2})$. В таком случае $\mathbb{E}r_i^* = 1$, $\mathbb{D}r_i^* = \frac{\pi}{2} - 1$. Другими словами, получаем, что в случае гетероскедастичного шума дисперсия \mathbf{r}^* постоянна, и тогда все рассуждения остаются верными в предположениях нормальности шума.

2.4.4. Модификация метода с итеративным обновлением весов

Далее представлена модификация алгоритма, подходящая для рядов с гетероскедастичным шумом.

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

Входные данные: $\mathbf{Y} \in \mathbb{R}^{L \times K}$ — траекторная матрица ряда, r — ранг сигнала; параметры критерия остановки: $\varepsilon = 10^{-4}$, максимальное число итераций $N_{IRLS} = 10$

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

1. Инициализация $\mathbf{U} \in \mathbb{R}^{n \times r}$ и $\mathbf{V} \in \mathbb{R}^{p \times r}$ (например, с помощью сингулярного разложения матрицы \mathbf{Y});
2. Выбор параметра α ;
3. $t := 0$;
4. **повторять**

- a. Вычисление матрицы остатков $\mathbf{R} = \{r_{ij}\}_{i,j=1}^{n,p} = \mathbf{Y} - \mathbf{U}\mathbf{V}^T$;
- b. Ганкелизация матрицы \mathbf{R} и получение ряда длины N из остатков:
 $\mathbf{R} = \mathcal{T}^{-1}\Pi_{\mathcal{H}}(\mathbf{R}) = (r_1, \dots, r_N)^T$;
- c. Пусть $\mathbf{R}_+ = (|r_1|, \dots, |r_N|)^T$ — вектор из модулей остатков. Вычисление $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_N)^T$ как оценки мат. ожидания $\mathbb{E}(\mathbf{R}_+)$ некоторым методом: локальной регрессией loess, скользящей медианой или взвешенной локальной регрессией lowess (подробнее оценка мат. ожидания обсуждалась в пункте 2.4.2);
- d. Вычисление ряда $|\boldsymbol{\sigma}^{-1}\mathbf{R}| = (\frac{|r_1|}{\sigma_1}, \dots, \frac{|r_N|}{\sigma_N})^T$ и получение матрицы
 $\mathbf{R}^* = \{r_{ij}^*\}_{i,j=1}^{L,K} = \mathcal{T}(|\boldsymbol{\sigma}^{-1}\mathbf{R}|)$;
- e. Вычисление матрицы весов $\mathbf{W} = \{w_{ij}\}_{i,j=1}^{L,K} = \{w(r_{ij}^*)\}_{i,j=1}^{L,K}$, 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})

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

- g. $t := t + 1$.

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

Посмотрим на график весов, чтобы убедиться, что только точки, содержащие выбросы, получили маленькие веса. На рисунках 2.2 и 2.3 изображен график ряда с 5% выбросов и веса, получившиеся в результате применения модификации метода. На графике весов видно, что точки, в которых содержались выделяющиеся наблюдения, получили нулевые веса. В остальных точках веса колеблются от 0.8 до 1.

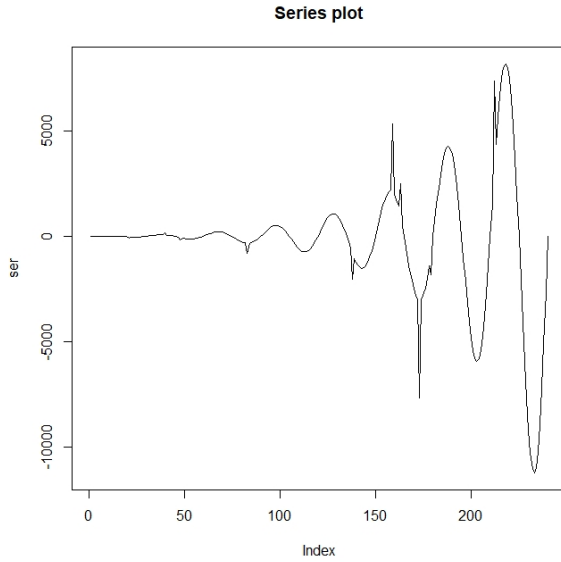


Рис. 2.2. График ряда с 5% выбросов.

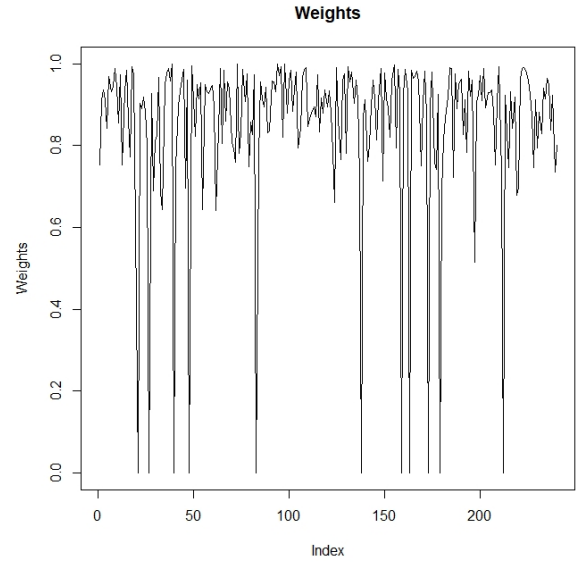


Рис. 2.3. Веса.

2.4.5. Модификация метода с итеративным обновлением весов с использованием метода Гаусса-Ньютона

Заметим, что использование ганкелевой матрицы $\{\sigma_{ij}\}_{i,j=1}^{L,K}$ и ее замена на параметр $\sigma = (\sigma_1, \dots, \sigma_N)^T$ соответствует приписыванию весов элементам ряда. Однако в рассмотренной модификации после нормировки модуля остатков на параметр σ и вычисления матрицы весов, мы приписываем веса элементам траекторной матрицы ряда и вычисляем проекцию на множество матриц ранга, не превосходящего r . Возникает идея приписывать веса самому ряду, а не матрице, и находить оценку сигнала с этими весами.

Для нахождения проекции на множество рядов ранга, не превосходящего r , по взвешенной норме можно использовать модифицированный метод Гаусса-Ньютона, опи-

санный в статье [12], который итеративно решает задачу

$$\mathbf{X}^* = \operatorname{argmin}_{\mathbf{X} \in \bar{\mathcal{D}}_r} \|\mathbf{Y} - \mathbf{X}\|_{\mathbf{W}},$$

где \mathbf{Y} — ряд длины N , $\bar{\mathcal{D}}_r$ — замыкание множества рядов ранга, не превосходящего r , $\mathbf{W} \in \mathbb{R}^{N \times N}$ — матрица весов, $\|\mathbf{Y}\|_{\mathbf{W}} = \mathbf{Y}^T \mathbf{W} \mathbf{Y}$.

Ниже представлен алгоритм модификации метода с итеративным обновлением весов с использованием метода Гаусса-Ньютона.

Алгоритм 5: Модификация метода с итеративным обновлением весов с использованием метода Гаусса-Ньютона

Входные данные: \mathbf{Y} — ряд, $\mathbf{Y} \in \mathbb{R}^{L \times K}$ — траекторная матрица ряда \mathbf{Y} , r — ранг сигнала; параметры критерия останковки: $\varepsilon = 10^{-4}$, максимальное число итераций $N_{IRLS} = 10$

Выходные данные: $\hat{\mathbf{Y}}$ — проекция ряда \mathbf{Y} на множество рядов ранга, не превосходящего r

1. Инициализация $\mathbf{U} \in \mathbb{R}^{n \times r}$ и $\mathbf{V} \in \mathbb{R}^{p \times r}$ (например, с помощью сингулярного разложения матрицы \mathbf{Y});
2. Выбор параметра α ;
3. Вычисление матрицы остатков $\mathbf{R} = \{r_{ij}\}_{i,j=1}^{n,p} = \mathbf{Y} - \mathbf{UV}^T$;
4. $t := 0$;

5. повторять

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

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

- b. Пусть $\mathbf{R}_+ = (|r_1|, \dots, |r_N|)^T$ — вектор из модулей остатков. Вычисление $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_N)^T$ как оценки мат. ожидания $\mathbb{E}(\mathbf{R}_+)$ некоторым выбранным методом (подробнее оценка мат. ожидания обсуждалась в пункте 2.4.2);

- c. Вычисление ряда $|\boldsymbol{\sigma}^{-1}\mathbf{R}| = (\frac{|r_1|}{\sigma_1}, \dots, \frac{|r_N|}{\sigma_N})^T = (r_1^*, \dots, r_N^*)^T$;

- d. Вычисление вектора весов $\mathbf{W} = (w_1, \dots, w_N)^T = (w(r_1^*), \dots, w(r_N^*))^T$,

используя

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

- e. Решение задачи взвешенной аппроксимации с помощью модифицированного метода Гаусса-Ньютона с матрицей весов

$$\mathbf{W} = \text{diag}(\mathbf{W}) \in \mathbb{R}^{N \times N};$$

$$\mathbf{X}^* = \underset{\mathbf{X} \in \mathcal{D}_r}{\text{argmin}} \|\mathbf{Y} - \mathbf{X}\|_{\mathbf{W}}.$$

- f. Пусть $\hat{\mathbf{Y}}$ — траекторная матрица ряда \mathbf{X}^* . Обновление матрицы остатков

$$\mathbf{R} = \{r_{ij}\}_{i,j=1}^{n,p} = \mathbf{Y} - \hat{\mathbf{Y}} \text{ и ряда } \mathbf{Y} = \mathbf{X}^* ;$$

- g. $t := t + 1$.

до тех пор, пока $\|\mathbf{Y} - \hat{\mathbf{Y}}\|_{\mathbf{W}} > \varepsilon$ и $t < N_{IRLS}$;

2.5. Оценка трудоемкости методов

Вопросу трудоемкости L1-SSA уделялось большое внимание во многих работах, посвященных построению \mathbb{L}_1 -проекции на множество матриц ранга, не превосходящего r . В статье [13] приведен алгоритм, решающий точно эту задачу. Пусть имеется вещественная матрица $\mathbf{X} \in \mathbb{R}^{L \times K}$, необходимо вычислить проекцию на множество матриц ранга, не превосходящего r . Трудоемкость алгоритма, решающего задачу в явном виде, составляет $O(L^{\text{rank}(\mathbf{X})r-r+1})$. В данной работе мы рассматривали методы, решающие эту задачу приближенно, но более эффективно.

Сравним теоретические трудоемкости описанных алгоритмов.

Последовательный метод

Вычислим трудоемкость последовательного алгоритма из раздела 2.3. Трудоемкость составляет $O((KP_1 + P_2)N_{iter})$, где P_1 и P_2 — трудоемкость решения задач линейного программирования, а N_{iter} — общее количество итераций для сходимости метода, которое также считаем не зависящим от L, K, r . Согласно [14], сложность вычисления задачи линейного программирования с v переменными и s ограничениями составляет $O(c \log v)$. В статье [8], содержащей описание алгоритма последовательного метода, вычислено количество переменных и ограничений в решаемых нами задачах, и получено, что трудоемкость может быть оценена как

$$T_{\text{lpca}} = O(LK \log(2LK + Lr)N_{iter}), \quad (2.5)$$

Метод с итеративным обновлением весов

Теоретическая трудоемкость метода из раздела 2.4 составляет, согласно статье [7],

$$T_{\text{IRLS}} = O(LKr^2N_\alpha N_{\text{IRLS}}), \quad (2.6)$$

где N_α и N_{IRLS} — общее количество итераций для решения задач (2.1), (2.2) и сходимости взвешенного метода наименьших квадратов с обновлением весов. Количество итераций мы брали постоянными и не зависящими от L, K и r . При подсчете трудоемкости авторы статьи [7] используют книгу [15], в которой приводятся эффективные алгоритмы QR-разложения матрицы.

Сравнение теоретических трудоемкостей

Сравним теоретические трудоемкости последовательного метода и взвешенного метода наименьших квадратов. Необходимо сравнить (2.5) и (2.6). Задача сводится к сравнению $\log(L(2K + r))N_{iter}$ и $r^2 N_\alpha N_{IRLS}$. Авторы статьи [7] утверждают, что максимальное число итераций N_α и N_{IRLS} для метода с обновлением весов достаточно взять 5 и 10 соответственно. Проведенные нами вычислительные эксперименты, описанные в разделе 3.5, показали, что максимальное число итераций для последовательного метода N_{iter} достаточно взять равным 5 (эксперименты проводились на рядах ранга 3 и 2 при $L = 120$, $K = 121$). Отличия в ошибках восстановления сигнала при увеличении числа итераций незначительны — ошибка уменьшается не более чем на 0.2%. Рассмотрим 2 случая.

Пусть L фиксировано, маленькое, а $K = N - L + 1 \sim N$. Это соответствует случаю, когда длина окна маленькая, и траекторная матрица вытянута. Асимптотически трудоемкость метода с обновлением весов оказывается меньше.

Пусть траекторная матрица ряда близка к квадратной, то есть $L \sim N/2$, $K = N - L + 1 \sim N/2$. Тогда, если поводить сравнение, то получаем, что надо сравнить $\log(\frac{N}{2}(N + r))$ и r^2 . Получаем, что асимптотически трудоемкость метода с обновлением весов снова оказывается меньше трудоемкости последовательного метода.

Для того, чтобы иметь возможность сравнить время работы методов, необходимо критерии остановки сделать такими, чтобы методы выдавали примерно одинаковые по точности результаты. Однако это оказывается нетривиальной задачей, поэтому поставим максимальное количество итераций для каждого метода такие, чтобы точность оказывалась приемлемой для каждого из методов. Сравним время работы, учитывая количество итераций.

В таблице 2.1 приведено время работы нескольких методов с 1% выбросов для 30 реализаций ряда длины $N = 240$

$$f_n = e^{4n/N} \sin(2\pi n/30) + A e^{4n/N} \varepsilon_n, \quad \varepsilon_n \sim N(0, 1).$$

Ранг такого ряда равен 2, длина окна выбрана $L = 120$. Различные модификации метода IRLS по времени работают примерно одинаково. Поэтому приведем результат только для одной из них.

На первый взгляд, результаты противоречат теоретическому выводу о том, что

Таблица 2.1. Время работы программы и число итераций для различных методов для $M = 30$ реализаций ряда.

	l1pca	IRLS (orig.)	IRLS (loess)
time	281 sec	860 sec	871 sec
N_{iter}	5	5*10	5*10

трудоемкость метода IRLS меньше. Это можно объяснить тем, что теоретические трудоемкости мы сравнивали асимптотически, а при выбранных нами для вычислительных экспериментов N, L, r результат получился обратный.

Вычислим, при каких N для такого примера трудоемкость последовательного метода начнет превосходить трудоемкость IRLS.

Траекторная матрица ряда близка к квадратной, $L \sim \frac{N}{2}, K \sim \frac{N}{2}$. Используя формулы 2.5 и 2.6, получаем, что надо сравнить $T_{l1pca} = \log(\frac{N^2}{2} + N)$ и $T_{IRLS} = 2^2 * 10$. Лишь при N порядка $6 * 10^8$ трудоемкость последовательного метода становится больше.

Вычислительные эксперименты

Сравним результаты работы описанных методов на нескольких примерах. Для начала возьмем ряд с экспоненциальным трендом и гауссовским шумом, который уже был рассмотрен в работе [6]. Затем рассмотрим ряд с растущей амплитудой и два случая: случай с гетероскедастичным шумом и с шумом с постоянной дисперсией, проведем сравнения для таких рядов.

3.1. Пример 1

Для начала рассмотрим пример из работы [6], но добавим большее количество выбросов (1% и 5%) в случайных точках ряда. Проверим, какой из приведенных алгоритмов окажется наиболее устойчивым.

Длина ряда $N = 240$. Рассмотрим временной ряд

$$f_n = e^{n/N} + \sin(2\pi n/120 + \pi/6) + \varepsilon_n, \quad \varepsilon_n \sim N(0, 1).$$

На рис. 3.1 изображен график ряда при 1% выбросов с величиной выброса $5f_i$. В случайно выбранных точках ряда f_i значение заменяется на $f_i + 5f_i$.

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

$$\text{MSE} = \mathbb{E} \left(\frac{1}{N} \sum_{i=1}^N (s_i - \hat{s}_i)^2 \right), \quad (3.1)$$

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

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

а также будем сравнивать методы по величине ошибки, согласованной с \mathbb{L}_1 , которая имеет вид

$$\text{MAD} = \mathbb{E} \left(\frac{1}{N} \sum_{i=1}^N |s_i - \hat{s}_i| \right). \quad (3.2)$$

Возьмем количество реализаций ряда $M = 30$. Будем находить оценки математических ожиданий (3.1) и (3.2), а далее из оценки MSE будем извлекать корень, получая RMSE.

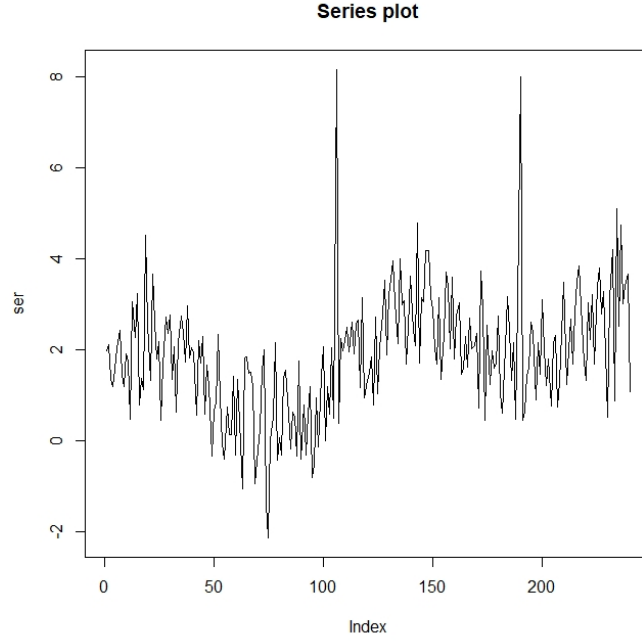


Рис. 3.1. График ряда при 1% выбросов с величиной выброса $5f_i$.

Ранг ряда равен 3. Во всех методах берется длина окна $L = 120$, равная половине длины ряда, и восстановление сигнала ведется по 3 компонентам.

Для метода с итеративным обновлением весов и его модификации выберем следующие параметры:

1. **Инициализация.** Возьмем в качестве начальных значений для обоих методов $\mathbf{U} = \mathbf{U}_r \mathbf{\Lambda}_r^{1/2}$, $\mathbf{V} = \mathbf{V}_r$, где $\mathbf{U}_r = [U_1 : \dots : U_r]$, $\mathbf{V}_r = [V_1 : \dots : V_r]$, $\mathbf{\Lambda}_r = \text{diag}(\lambda_1, \dots, \lambda_r)$ — первые r компонент сингулярного разложения траекторной матрицы.
2. **Параметр α .** Авторы статьи [7], содержащей описание исходного метода, предлагают выбрать $\alpha = 4.685$. Исходя из полученной формулы (2.4), связывающей параметр α и вероятность γ , получим для $\gamma = 0.99$ значение параметра $\alpha = 4.046$. Его и будем брать. Однако стоит заметить, что такая вероятностная интерпретация параметра α верна только для модификации метода.

В таблице 3.1 представлены результаты сравнения для четырех методов. Выброс добавлялся в случайных точках ряда заменой значения f_i на $f_i + 5f_i$.

При сравнении методов со стандартным использовался пакет Rssa [16]. Реализация модифицированного метода Гаусса-Ньютона находится в репозитории [17].

Таблица 3.1. Оценки RMSE и MAD для различных методов для $M = 30$ реализаций ряда.

Method	0%	1%	5%
Оценки RMSE			
Basic SSA	0.184	0.256	0.653
l1pca	0.217	0.223	0.250
IRLS (orig.)	0.184	0.189	0.206
IRLS (loess)	0.196	0.199	0.204
IRLS (median)	0.210	0.221	0.223
IRLS (lowess)	0.206	0.207	0.211
IRLS (MGN)	0.398	0.222	0.798
Оценки MAD			
Basic SSA	0.143	0.197	0.505
l1pca	0.175	0.179	0.203
IRLS (orig.)	0.145	0.147	0.161
IRLS (loess)	0.155	0.158	0.160
IRLS (median)	0.170	0.178	0.178
IRLS (lowess)	0.165	0.166	0.168
IRLS (MGN)	0.259	0.178	0.601

Первая строка таблиц соответствует стандартному методу SSA с большой длиной окна ($L = 120$). Вторая строка — метод `l1rsa` из пакета `rcaL1` [10] (соответствует последовательному методу из раздела 2.3). Третья строка соответствует стандартному методу с итеративным обновлением весов из раздела 2.4. Четвертая, пятая и шестая строки соответствуют модификации взвешенного метода наименьших квадратов с различными вариантами выделения тренда (локальная регрессия с параметром сглаживания 0.35, скользящая медиана с длиной окна 80 и взвешенная локальная регрессия с параметром сглаживания 0.35). Была использована реализация `lowess` в R из статьи [18]. Параметр сглаживания выбран 0.35, остальные параметры оставлены по умолчанию: число итераций равно 3, параметр δ , требующийся для ускорения вычисления, равен $0.01(\max_i f_i - \min_i f_i)$. Седьмая строка посвящена модификации метода с обновлением весов с использованием метода Гаусса-Ньютона.

Из-за того, что дисперсия шума постоянная, можем предположить, что модификации IRLS с различными вариантами выделения тренда будут работать примерно одинаково. Поэтому подробнее об отличиях модификаций с разными вариантами выделения тренда пока что говорить не будем.

Жирным шрифтом в каждом столбце выделено лучшее значение и значение, которое незначимо отличается от лучшего. Проверка значимости сравнений представлена далее в таблицах 3.3 и 3.4.

Можно заметить, что при 5% выбросов модификация с использованием метода Гаусса-Ньютона оказывается неустойчивой. Можно предположить, что это происходит из-за того, что выбросы оказываются близко друг к другу, либо близко к началу или концу ряда. Исследуем, как ведут себя методы в зависимости от положения выброса. Результаты представлены в таблице 3.2.

Действительно, если выброс попадает близко к концу ряда (в точку x_{235}), то модификация с использованием метода Гаусса-Ньютона дает большую ошибку восстановления сигнала. Однако при наличии выброса в середине ряда ошибка восстановления сигнала с использованием этой модификации маленькая. Можно сделать вывод, что этот метод лучше использовать, если выбросов небольшое количество, либо при наличии информации, что выбросы не содержатся близко к началу или концу ряда. В наших исследованиях выбросы находятся в случайно выбранных точках ряда, чтобы максимально обобщить все варианты местоположения выделяющихся наблюдений. По-

Таблица 3.2. RMSE в зависимости от положения выброса.

Method	x_{50}	x_{100}	x_{150}	x_{200}	x_{235}
Basic SSA	0.106	0.239	0.137	0.170	0.259
l1pca	0.213	0.251	0.195	0.165	0.278
IRLS (orig.)	0.111	0.201	0.121	0.163	0.222
IRLS (loess)	0.135	0.217	0.163	0.178	0.243
IRLS (median)	0.223	0.255	0.227	0.179	0.283
IRLS (lowess)	0.168	0.230	0.188	0.187	0.255
IRLS (MGN)	0.092	0.220	0.134	0.156	0.655

этому в дальнейших сравнениях не будем рассматривать эту модификацию.

При отсутствии выбросов наиболее точным все так же остается классический метод SSA, а также метод с итеративным обновлением весов. В присутствии выделяющихся наблюдений наиболее устойчивым являются метод обновлением весов и его модификация с использованием локальной регрессии.

Проверка значимости сравнения

Опишем подробнее, как происходит сравнение метода, выдающего наименьшую ошибку, с остальными методами. Проверим значимость сравнения по критерию для зависимых выборок. Проверим 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. Критерий является двухсторонним.

Проверим, является ли отличие между этими методами значимым. В таблице 3.3 приведены p-value для сравнения среднеквадратичных ошибок для стандартного SSA и остальных методов без выделяющихся наблюдений. Все сравнения оказываются значимыми при уровне значимости 0.05. В таблице 3.4 приведены p-value для сравнения ошибок для метода взвешенных наименьших квадратов и остальных методов при 5%

Таблица 3.3. P-value для сравнения различных методов с наилучшим без выбросов.

0%	l1pca	IRLS (orig.)	IRLS (loess)	IRLS (median)	IRLS (lowess)
Basic SSA	4.7e-5	0.08	0.022	0.001	0.009

Таблица 3.4. P-value для сравнения различных методов с наилучшим в присутствии выбросов.

5%	Basic SSA	l1pca	IRLS (orig.)	IRLS (median)	IRLS (lowess)
IRLS (loess)	3.1e-13	1.1e-5	0.589	0.019	0.185

выбросов. При уровне значимости 0.05 при без выбросов сравнение стандартного SSA с оригинальным методом с обновлением весов оказывается незначимым. При 5% выбросов сравнения модификации IRLS (loess) с оригинальным IRLS и IRLS (lowess) оказываются незначимыми.

3.2. Пример 2

Попробуем рассмотреть не похожий на предыдущий пример ряд, добавив растущую амплитуду и шум непостоянной дисперсии, и исследуем устойчивость методов.

Длина ряда $N = 240$. Рассмотрим ряд с гетероскедастичным шумом

$$f_n = e^{4n/N} \sin(2\pi n/30) + Ae^{4n/N} \varepsilon_n, \quad \varepsilon_n \sim N(0, 1).$$

График ряда представлен на рисунке 3.2. Ранг ряда равен 2.

Из-за наличия шума непостоянной дисперсии можно предположить, что модификация метода с итеративным обновлением весов даст хороший результат. Тогда необходимо выбрать наиболее подходящий метод оценки математического ожидания ряда из модулей остатков для задания параметра σ . На рисунке 3.3 представлено выделение тренда из модуля остатков различными способами: локальной регрессией с параметром сглаживания 0.35, скользящей медианой с длиной окна 80 и взвешенной локальной регрессией с параметром сглаживания 0.35. Вычислим реальный тренд из ряда из модулей остатков:

$$\mathbb{E}|\mathbf{R}| = Ae^{4n/N} \mathbb{E}|\varepsilon| = Ae^{4n/N} \sqrt{\frac{2}{\pi}}.$$

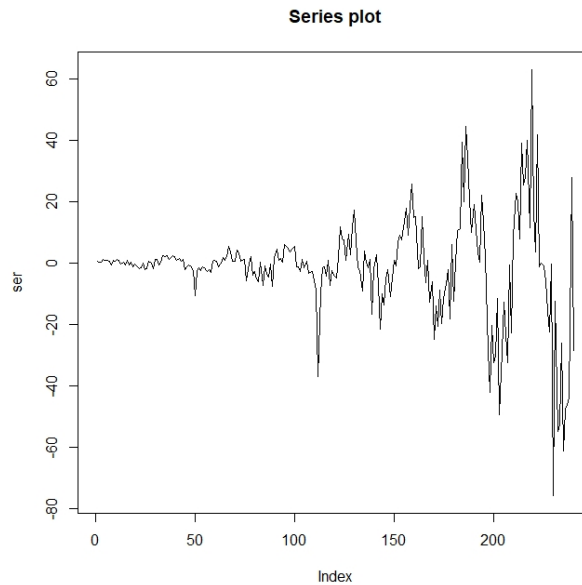


Рис. 3.2. График ряда при 1% выбросов с величиной выброса $5f_i$, $A = 0.5$.

Можно сделать следующие выводы. Прежде всего, локальная регрессия сильнее реагирует на выбросы, тренд чуть завышен. Скользящая медиана плохо работает на конце ряда, даже если брать длину окна в скользящей медиане большой. Взвешенная локальная регрессия хорошо справляется с выбросами и показывает результат, близкий к реальному тренду.

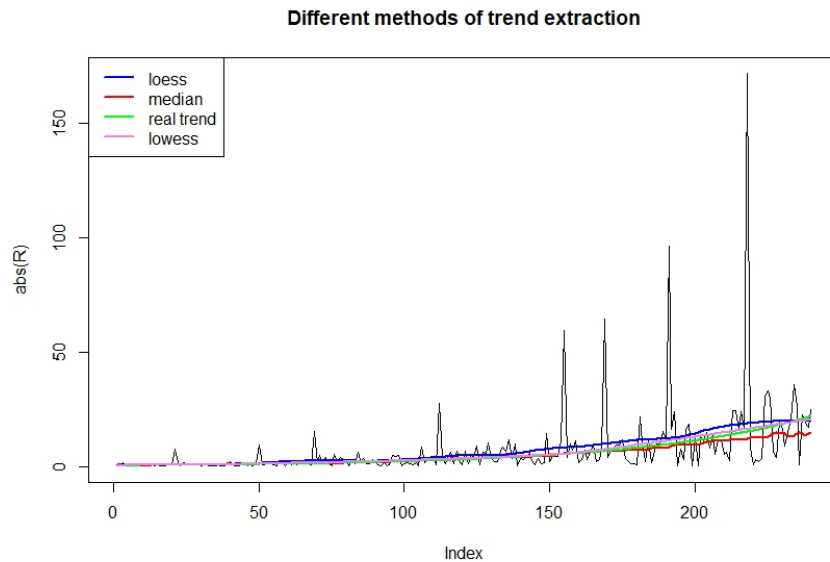


Рис. 3.3. Выделение тренда из ряда $|R|$ несколькими способами.

Если сравнить точки, получившие нулевые веса различными методами, можно сде-

Таблица 3.5. Оценки RMSE для различных методов для $M = 30$ реализаций ряда.

Method	0%	1%	5%
Оценки RMSE			
Basic SSA	2.16	2.78	5.96
l1pca	2.45	2.47	2.87
IRLS (orig.)	3.52	3.59	3.61
IRLS (loess)	2.31	2.36	2.39
IRLS (median)	2.84	2.84	2.86
IRLS (lowess)	2.59	2.60	2.63
Оценки MAD			
Basic SSA	1.08	1.33	2.91
l1pca	1.18	1.26	1.36
IRLS (orig.)	1.63	1.65	1.66
IRLS (loess)	1.16	1.19	1.21
IRLS (median)	1.37	1.38	1.38
IRLS (lowess)	1.26	1.28	1.29

лать следующие выводы: наибольшие выбросы все методы идентифицировали одинаково хорошо, скользящая медиана обнуляет наибольшее количество точек, нулевые веса получили также точки, не являющиеся выбросами. Loess и lowess присваивают нулевые веса одинаковым точки и не обнуляют ничего лишнего. Есть предположение, что с использованием этих двух методов ошибка восстановления сигнала будет наименьшая. Отличие в этих методах состоит в том, что lowess чуть сильнее занижает веса в точках, не являющихся выбросами, чем loess. Возможно, из-за этого loess окажется лучше.

Результаты сравнения методов представлены в таблице 3.5. Сразу можно отметить, что для ряда с гетероскедастичным шумом метод с обновлением весов дает большую ошибку даже в отсутствии выбросов. Без выбросов модификация метода IRLS с выделением тренда с помощью локальной регрессии, а также последовательный метод дают маленькую ошибку, как и стандартный SSA. В присутствии выбросов (1%) оба эти метода показывают хорошие результаты, а при 5% выбросов наилучшим оказывается метод с обновлением весов с использованием локальной регрессии.

В таблицах 3.6 и 3.7 представлены p-value для проверки значимости сравнения

Таблица 3.6. P-value для сравнения различных методов с наилучшим без выбросов.

0%	l1pca	IRLS (orig.)	IRLS (loess)	IRLS (median)	IRLS (lowess)
Basic SSA	0.57	6.7e-9	0.69	0.001	0.021

Таблица 3.7. P-value для сравнения различных методов с наилучшим в присутствии выбросов.

5%	Basic SSA	l1pca	IRLS (orig.)	IRLS (median)	IRLS (lowess)
IRLS (loess)	1.7e-5	0.020	8.7e-4	0.018	0.026

наилучших методов с остальными при 0% и 5% выбросов.

3.3. Пример 3

Возьмем похожий ряд, но с шумом, имеющим постоянную дисперсию. Рассмотрим пример, предложенный в статье [19], и проведем для этого примера вычислительный эксперимент.

Пусть длина ряда $N = 240$. Рассмотрим временной ряд

$$f_n = ne^{4n/N} \sin(2\pi n/30) + \varepsilon_n, \quad \varepsilon_n \sim N(0, 1).$$

Ранг ряда равен 4. У такого ряда разброс собственных значений очень велик. Это может приводить к тому, что некоторые компоненты сигнала могут смешиваться с шумом. Однако шум рассматриваемого размера не портит делимость сигнала от шума. Выбросы будут находиться в случайно выбранных точках ряда. Сравнение будем проводить при 1% и 5% выбросов, а также без выделяющихся наблюдений. В случайно выбранных точках f_i значение будет заменяться на $f_i + 1.5f_i$.

На рис. 3.4 изображен график ряда при 1% выбросов с величиной выброса $1.5f_i$.

Результаты сравнения методов при различном проценте выделяющихся наблюдений представлены в таблице 3.8.

Без выделяющихся наблюдений оригинальный IRLS и его модификация с использованием loess незначимо отличаются от стандартного SSA. В присутствии выделяющихся наблюдений только модификация IRLS с использованием взвешенной локальной

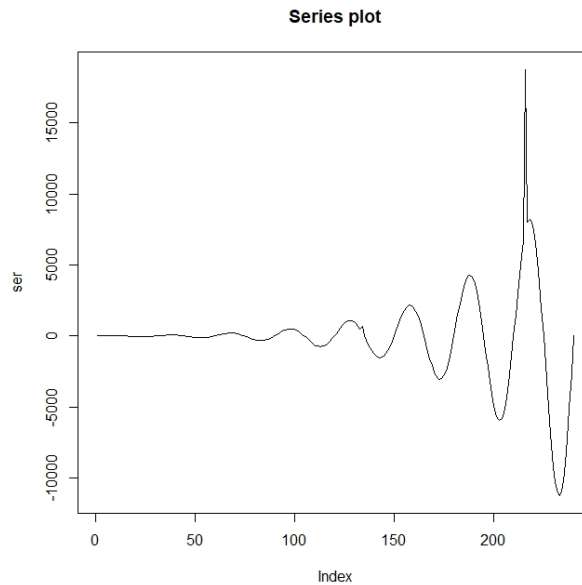


Рис. 3.4. График ряда при 1% выбросов с величиной выброса $1.5f_i$.

регрессии работает хорошо.

Проверка значимости сравнения наилучшего метода с остальными при отсутствии выбросов представлена в таблице 3.9.

3.4. Выводы

Результаты исследования для трех рассмотренных примеров представлены в таблице 3.10. На основе проведенного исследования можно сделать следующие выводы. Для первого примера без растущей амплитуды ряда в случае гауссовского шума наиболее точным методом без выбросов оказывается стандартный SSA. Стандартный метод с обновлением весов также является точной при отсутствии выбросов. Самыми устойчивыми к выбросам являются стандартный IRLS, а также его модификации IRLS (loess) и IRLS (lowess).

В случае шума с непостоянной дисперсией преимущество стандартного метода с итеративным обновлением весов пропадает, однако его модификация с использованием локальной регрессии при выделении тренда из остатков оказывается устойчивой к выбросам и точной без выделяющихся наблюдений.

Если же разброс значений ряда очень велик, то при отсутствии выбросов точными являются стандартный SSA, оригинальный IRLS и модификация IRLS (loess). Однако устойчивой является IRLS (lowess).

Таблица 3.8. Оценки RMSE и MAD для различных методов для $M = 30$ реализаций ряда.

Method	0%	1%	5%
Оценки RMSE			
Basic SSA	0.215	123.7	459.6
l1pca	0.256	8.65	21.11
IRLS (orig.)	0.216	220.4	398.2
IRLS (loess)	0.227	115.25	303.2
IRLS (median)	0.256	20.21	38.21
IRLS (lowess)	0.243	0.260	0.301
Оценки MAD			
Basic SSA	0.164	33.32	168.2
l1pca	0.197	1.117	3.145
IRLS (orig.)	0.165	27.311	88.54
IRLS (loess)	0.174	22.18	31.180
IRLS (median)	0.187	8.182	12.20
IRLS (lowess)	0.180	0.179	0.187

Таблица 3.9. P-value для сравнения различных методов с наилучшим без выбросов.

0%	l1pca	IRLS (orig.)	IRLS (loess)	IRLS (median)	IRLS (lowess)
Basic SSA	4.1e-4	0.962	0.107	0.004	0.024

Таблица 3.10. Оценки RMSE и MAD для трех рассмотренных примеров для $M = 10$ реализаций ряда.

Оценки RMSE						
Method	0%	5%	0%	5%	0%	5%
Basic SSA	0.184	0.653	2.16	5.96	0.215	459.6
l1pca	0.217	0.250	2.45	2.87	0.256	21.11
IRLS (orig.)	0.184	0.206	3.52	3.61	0.216	398.2
IRLS (loess)	0.196	0.204	2.31	2.39	0.227	303.2
IRLS (median)	0.210	0.223	2.84	2.86	0.256	38.21
IRLS (lowess)	0.206	0.211	2.59	2.63	0.243	0.301
Оценки MAD						
Method	0%	5%	0%	5%	0%	5%
Basic SSA	0.143	0.505	1.08	2.91	0.164	168.2
l1pca	0.175	0.203	1.18	1.36	0.197	3.145
IRLS (orig.)	0.145	0.161	1.63	1.66	0.165	88.54
IRLS (loess)	0.155	0.160	1.16	1.21	0.174	31.180
IRLS (median)	0.170	0.178	1.37	1.38	0.187	12.20
IRLS (lowess)	0.165	0.188	1.26	1.29	0.180	0.187

Исходя из проведенного исследования можно сказать, что если у ряда нет растущей амплитуды и разброс значений небольшой, то можно использовать метод с обновлением весов. Он достаточно точный без выбросов и устойчивый к выделяющимся наблюдениям. В случае появления растущей амплитуды ряда и шума с непостоянной дисперсией, преимущество метода с обновлением весов пропадает. В таком случае следует использовать его модификацию с использованием локальной регрессии. Если же разброс значений у ряда большой, то следует использовать модификацию с выделением тренда с помощью взвешенной локальной регрессии, которая хорошо справляется с выбросами.

3.5. Исследование числа итераций

В последовательном методе l1pca и методе IRLS есть дополнительный параметр: максимальное число итераций в цикле. Во взвешенном методе наименьших квадратов задается максимальное число итераций для внешнего цикла и для внутреннего. Для того, чтобы иметь возможность сравнить трудоемкости методов, необходимо исследовать, какое количество итераций требуется для сходимости этих методов. На рисунках 3.5 и 3.6 показана зависимость ошибки RMSE от числа итераций для двух примеров (первый пример с экспоненциальным трендом, второй пример с быстрорастущей амплитудой ряда). На рисунках 3.7 и 3.8 представлена зависимость RMSE от числа итераций во внешнем цикле для метода IRLS. Сравнения для различного числа итераций проводились на одинаковых реализациях ряда.

Из графиков видно, что отличия в ошибках при увеличении итераций совсем незначительные, и в последовательном методе можно было бы ограничиться и 5 итерациями.

Для метода IRLS достаточно брать порядка 7 итераций внешнего цикла для второго примера. Для первого примера ошибка практически перестает убывать при 20 итерациях. Число итераций во внутреннем цикле можно оставить по умолчанию равным 5, так как отличия в таком случае совсем незначительные.

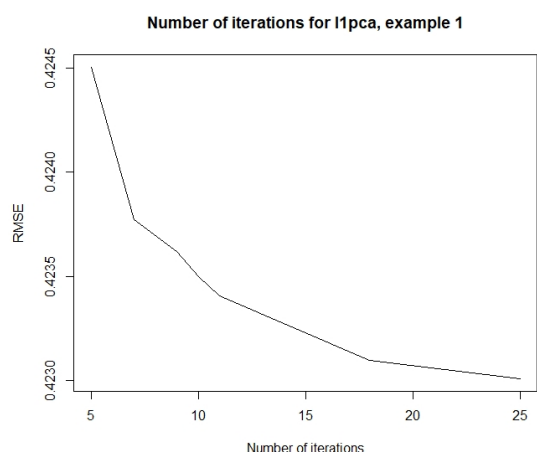


Рис. 3.5. Зависимость RMSE от числа итераций для последовательного метода l1pca , пример 1.

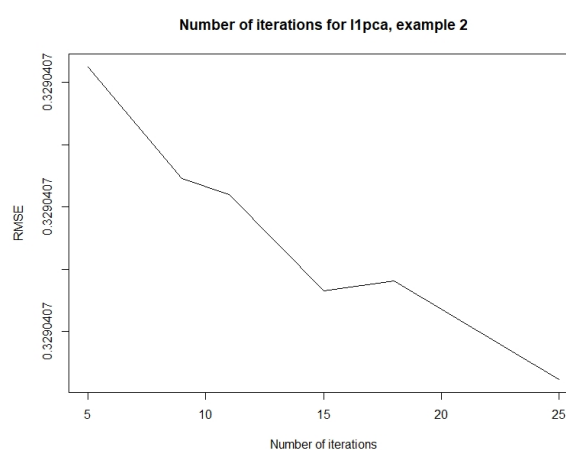


Рис. 3.6. Зависимость RMSE от числа итераций для последовательного метода l1pca , пример 2.

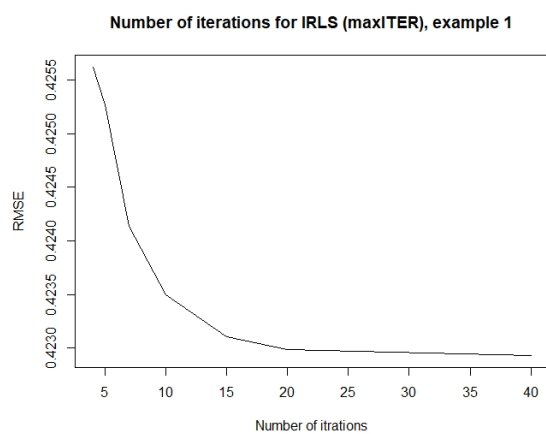


Рис. 3.7. Зависимость RMSE от числа итераций для метода IRLS, пример 1.

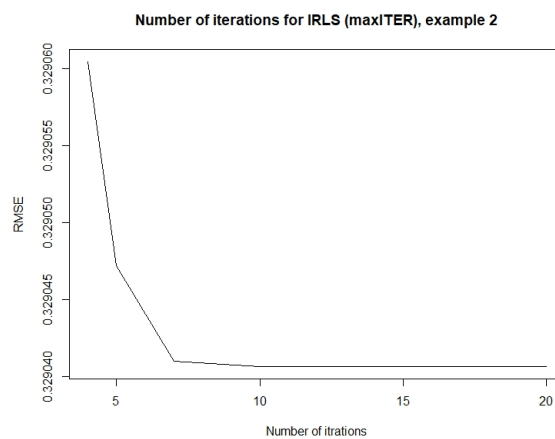


Рис. 3.8. Зависимость RMSE от числа итераций для метода IRLS, пример 2.

Заключение

В работе были приведены и исследованы некоторые варианты модификации метода SSA с целью повышения устойчивости к выбросам.

В главе 1 были введены основные понятия и обозначения, описан алгоритм базового метода.

В главе 2 была введена общая схема метода с проекторами на пространство ганкелевых матриц и множество матриц ранга, не превосходящего r , без указания конкретной нормы. Далее описано построение проектора на пространство ганкелевых матриц по нормам в \mathbb{L}_2 , \mathbb{L}_1 и взвешенной норме в \mathbb{L}_2 . Для построения \mathbb{L}_1 -проектора на множество матриц ранга, не превосходящего r , был предложен последовательный метод, который уже реализован в R-пакете [10]. Также была рассмотрена другая идея построения устойчивого метода — присвоение точкам, содержащим выбросы, меньший вес. Этой идее соответствует метод с итеративным обновлением весов из статьи [7]. Но так как было обнаружено, что этот метод не подходит для рядов с гетероскедастичным шумом, было разработано несколько модификаций данного метода. Были рассмотрены несколько вариантов выделения тренда: скользящая медиана, локальная регрессия loess, а также взвешенная локальная регрессия lowess. Выведена формула для параметра α , зависящая от вероятности γ . Благодаря этой формуле стала понятна интерпретация параметра, а также даны рекомендации по выбору α . Изначально шум предполагался нормальный, с постоянной дисперсией. Было показано, что для нормального шума с меняющейся дисперсией формула для параметра α остается верной.

Бало проведено сравнение теоретических трудоемкостей методов. Рассмотрено два случая: случай с квадратной и вытянутой траекторной матрицей. В обоих случаях асимптотически теоретическая трудоемкость метода с обновлением весов оказалась наименьшей. Однако при выбранных нами для вычислительных экспериментов рядах трудоемкость последовательного метода оказывается меньше трудоемкости метода с обновлением весов.

Глава 3 была посвящена численным экспериментам. Классический метод SSA сравнивался с последовательным методом, методом с обновлением весов и его модификациями. В случайных точках ряда f_i добавлялись выделяющиеся наблюдения, равные $5f_i$ для первого и третьего примеров и $1.5f_i$ для второго. Для начала проводилось сравне-

ние методов без выделяющихся наблюдений, а затем при 1% и 5% выбросов. Сравнение проводилось по величине оценок ошибок RMSE и MAD.

Сначала рассматривался пример из работы [6], но с добавлением большего количества выбросов. Было показано, что в таком случае метод IRLS оказался довольно точным и устойчивым.

Далее ряд был заменен на ряд с шумом непостоянной дисперсии. На таком примере преимущество метода IRLS пропадает. Однако его модификация с выделением тренда с помощью локальной регрессии оказывается устойчивой к выбросам и точной при отсутствии выбросов. В случае шума с постоянной дисперсией, но с большим разбросом значений ряда устойчивым методом оказалась модификация IRLS с использованием взвешенной локальной регрессии.

Список литературы

1. Golyandina N., Nekrutkin V., Zhigljavsky A. Analysis of Time Series Structure: SSA and Related Techniques. — Boca Raton, Fla. : Chapman & Hall/CRC, 2001.
2. Голяндина Н. Э. Метод «Гусеница»-SSA: анализ временных рядов: Учеб. пособие. — Санкт-Петербург : BBM, 2004.
3. Advanced spectral methods for climatic time series / M. Ghil, R. M. Allen, M. D. Dettinger et al. // Reviews of Geophysics. — 2002. — Vol. 40, no. 1. — P. 1–41.
4. Broomhead D., King G. Extracting qualitative dynamics from experimental data // Physica D. — 1986. — Vol. 20. — P. 217–236.
5. Vautard M., Ghil M. Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series // Physica D. — 1989. — Vol. 35. — P. 395–424.
6. Третьякова А. Л. Устойчивые варианты метода SSA для анализа временных рядов: выпускная квалификационная работа, науч.рук. к.ф.-м.н., доцент Голяндина Н.Э. — 2018.
7. Chen K., Sacchi M. Robust reduced-rank filtering for erratic seismic noise attenuation // GEOPHYSICS. — 2015. — 01. — Vol. 80. — P. V1–V11.
8. Brooks J. P., Jot S. pcaL1: An implementation in R of three methods for L1-norm principal component analysis. — 2013.
9. Zvonarev N., Golyandina N. Iterative algorithms for weighted and unweighted finite-rank time-series approximations // arXiv: Methodology. — 2015. — P. 5–18.
10. Jot S., Brooks J. P., Visentin A., Park Y. W. — pcaL1: L1-Norm PCA Methods, 2017. — R package version 1.5.2.
11. Ke Q., Kanade T. 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.
12. Zvonarev N., Golyandina N. Image space projection for low-rank signal estimation: Modified gauss-newton method // arXiv preprint arXiv:1803.01419. — 2018.
13. Markopoulos P., Karystinos G., Pados D. Optimal algorithms for L1-subspace signal processing // IEEE Transactions on Signal Processing. — 2014. — 10. — Vol. 62. — P. 5046–5058.
14. Chvátal V. Linear Programming. Series of books in the mathematical sciences. — New

- York : W.H. Freeman Company, 1983. — ISBN: 9780716711957.
15. Golub G. H., Van Loan C. F. Matrix Computations. — Third edition. — The Johns Hopkins University Press, 1996.
 16. Korobeynikov A., Shlemov A., Usevich K., Golyandina N. — Rssa: A collection of methods for singular spectrum analysis, 2016. — R package version 1.0. Access mode: <http://CRAN.R-project.org/package=Rssa>.
 17. Zvonarev N. R code for Modified Gauss-Newton algorithm. — <https://github.com/neg99/MGN>. — 2019.
 18. Cleveland W. S. Robust locally weighted regression and smoothing scatterplots // Journal of the American Statistical Association. — 1979. — Vol. 74, no. 368. — P. 829–836.
 19. Rodrigues P., Lourenco V., Mahmoudvand R. A robust approach to singular spectrum analysis // Quality and Reliability Engineering International. — 2018. — 06. — Vol. 34.
 20. pcaMethods – a Bioconductor package providing PCA methods for incomplete data / Wolfram Stacklies, Henning Redestig, Matthias Scholz et al. // Bioinformatics. — 2007. — Vol. 23. — P. 1164–1167.
 21. Stacklies W., Redestig H., Scholz M. et al. — pcaMethods – a Bioconductor package providing PCA methods for incomplete data, 2007. — R package version 1.72.0. Access mode: <https://bioconductor.org/biocLite.R>.
 22. Hawkins D. M., Liu L., Young S. S. Robust singular value decomposition technical report number 122, National Institute of Statistical Sciences 19. — 2002.