

# Робастные варианты метода анализа сингулярного спектра

Третьякова Александра Леонидовна

Санкт-Петербургский государственный университет  
Математико-механический факультет  
Кафедра статистического моделирования

Научный руководитель: к.ф.-м.н., доц. Голяндина Н.Э.  
Рецензент: к.ф.-м.н. Пепелышев А.Н.



Санкт-Петербург, 2020

Рассмотрим вещественнозначный **временной ряд**  $X = (x_1, \dots, x_N)$ , где  $N$  — длина ряда.

Предполагаем, что  $x_i = s_i + r_i$ ,  $i = 1, \dots, N$ , где  $r_i$  — шум.

## Задача

*Разложение временного ряда на интерпретируемые аддитивные составляющие:*

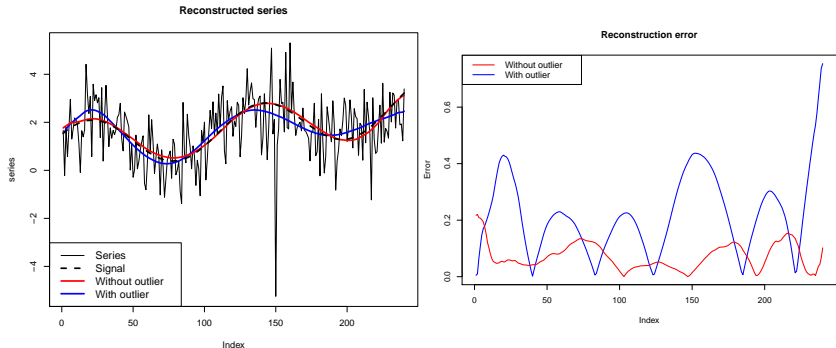
$$X = S + R,$$

$S$  — **сигнал**,

$R$  — **шум**.

**Метод:** «Гусеница»-SSA (Singular Spectrum Analysis) [Analysis of Time Series Structure: SSA and Related Techniques, Golyandina N., Nekrutkin V., Zhigljavsky A., 2001].

# Постановка задачи



**Рис.:** График ряда с выделяющимся наблюдением и модуль ошибок восстановления сигнала в присутствии выброса и без него.

**Задача:** предложить устойчивые к выбросам модификации метода анализа сингулярного спектра и сравнить их между собой.

Задачи:

- Обзор и структурирование устойчивых вариантов метода SSA,
- Предложение модификации для нестационарного шума,
- Сравнение рассмотренных методов по трудоемкости,
- Сравнение рассмотренных методов по точности на примерах.

# Метод SSA для выделения сигнала ранга, не превосходящего $r$

Ряд  $X = (x_1, \dots, x_N)$ .

Пусть  $0 < L < N$  — длина окна.  $K = N - L + 1$ .

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

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

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

Ряд  $\mapsto$  траекторная матрица  $\mathbf{X}$ :

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

- 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  $r$ .
- $\Pi_{\mathcal{H}} : \mathcal{M} \rightarrow \mathcal{M}_{\mathcal{H}}$  — проектор на пространство ганкелевых матриц.

Структура сигнала  $\mathbf{S}$ : ранг траекторной (ганкелевой) матрицы  $\mathcal{T}(\mathbf{S})$  равен  $r$ .

В результате получаем оценку сигнала:

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

где проекторы можно строить по различным нормам.

Будем рассматривать следующие варианты:

- Проекторы  $\Pi_r$  и  $\Pi_{\mathcal{H}}$  по норме в  $\mathbb{L}_2$  (стандартный L2-SSA),
- Проекторы  $\Pi_r$  и  $\Pi_{\mathcal{H}}$  по норме в  $\mathbb{L}_1$  (L1-SSA),
- Проекторы  $\Pi_r$  и  $\Pi_{\mathcal{H}}$  по взвешенной норме в  $\mathbb{L}_2$  (WL2-SSA).

## Определение

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

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

- $\Pi_{\mathcal{H}}$  — проектор на множество ганкелевых матриц по норме Фробениуса посредством усреднения элементов на диагоналях  $i + j = \text{const}$ :  $\|\mathbf{X} - \mathbf{Y}\|_F^2 \longrightarrow \min_{\mathbf{Y} \in \mathcal{M}_{\mathcal{H}}} .$
- $\Pi_r$  — проектор на множество матриц ранга  $r$  по норме Фробениуса:  $\|\mathbf{X} - \mathbf{Y}\|_F^2 \longrightarrow \min_{\mathbf{Y} \in \mathcal{M}_r} , \mathbf{Y} = \sum_{i=1}^r \sqrt{\lambda_i} U_i V_i^T .$

## Определение

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

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

## Замечание

Так как  $\operatorname{argmin}_a \mathbb{E}|\xi - a| = \operatorname{med}\xi$ , то  $\Pi_{\mathcal{H}}$  строится посредством выбора медианы значений на диагоналях  $i + j = \text{const}$ .

Для построения проектора на множество матриц ранга  $r$  в  $\mathbb{L}_1$  будем рассматривать **последовательный метод**.



# L1-SSA. Реализация. Последовательный метод

В R-пакете `rcaL1` [Jot et al., 2017] имеется реализация последовательного метода решения задачи  $\|\mathbf{Y} - \mathbf{U}\mathbf{V}^T\|_1 \rightarrow \min_{\mathbf{U}, \mathbf{V}}$ .

Алгоритм `l1pca` [Brooks J. P., Jot S., 2013] :

❶ Инициализация  $\mathbf{U}(0) \in \mathbb{R}^{L \times r}$ , нормировка столбцов  $\mathbf{U}(0)$ ,

❷  $t := t + 1$ ,

❸  $\mathbf{V}(t) = \underset{\mathbf{V} \in \mathbb{R}^{K \times r}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{U}(t-1)\mathbf{V}^T\|_1$

Задача разбивается на  $K$  независимых подзадач вида

$\mathbf{v}_i = \underset{\mathbf{x}}{\operatorname{argmin}} \|Y_i - \mathbf{U}(t-1)\mathbf{x}\|_1$ , где  $Y_i \in \mathbb{R}^L$  — столбцы  $\mathbf{Y}$ ,  $\mathbf{v}_i \in \mathbb{R}^r$  — строки  $\mathbf{V}$ ,  $i = 1, \dots, K$ ,

❹  $\mathbf{U}(t) = \underset{\mathbf{U} \in \mathbb{R}^{L \times r}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{U}\mathbf{V}^T(t)\|_1$  (решается аналогично п.3)

❺ Нормировка столбцов  $\mathbf{U}(t)$ ,

❻ if  $\mathbf{U}(t) \neq \mathbf{U}(t-1)$  (по крит. остановки) then Go to Step 2  
else  $\mathbf{U} := \mathbf{U}(t)$ ;  $\mathbf{V} := \mathbf{V}(t)$ .

Крит. остановки:  $\max_{i=1, \dots, L, j=1, \dots, r} |u_{ij}(t) - u_{ij}(t-1)| > \varepsilon$  или  $t > N_{\text{iter}}$ .

Решаем задачу, меняя на каждой итерации  $\mathbf{U}$  и  $\mathbf{V}$  и разбивая исходную задачу на линейные подзадачи.

## Определение

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

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

## Утверждение (Zvonarev, Golyandina, 2015)

Для построения проекции  $\Pi_{\mathcal{H}} \mathbf{Y} = \hat{\mathbf{Y}} = \{\hat{y}_{ij}\}_{i,j=1}^{L,K}$  необходимо суммировать элементы на диагоналях  $i + j = \text{const}$  с весами и нормировать на сумму весов: 
$$\hat{y}_{ij} = \frac{\sum_{l,k:l+k=i+j} w_{lk} y_{lk}}{\sum_{l,k:l+k=i+j} w_{lk}}.$$

**Замечание:** 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  $\mathbb{L}_2$ .

# WL2-SSA. Метод с итеративным обновлением весов

Пусть  $\mathbf{Y} \in \mathbb{R}^{L \times K}$  — траекторная матрица ряда,  $\hat{\mathbf{Y}} = \mathbf{U}\mathbf{V}^T = \{\hat{y}_{ij}\}_{i,j=1}^{L,K}$ .  
Решаем задачу

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

где  $\odot$  — поэлементное умножение,  $\mathbf{W}^{1/2}$  — поэлементное взятие корня, веса  $w_{ij} = w\left(\frac{y_{ij} - \hat{y}_{ij}}{\sigma_{ij}}\right)$  вычисляются по формуле

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

Значения  $\alpha$  и  $\{\sigma_{ij}, i = 1, \dots, L, j = 1, \dots, K\}$  — параметры.

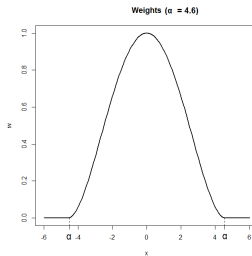


Рис.: График функции  $w(x)$ .

# WL2-SSA. Метод с итеративным обновлением весов.

## Реализация

Задача:

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

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

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

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

где  $\mathbf{W}_i = \text{diag}(w_i) \in \mathbb{R}^{K \times K}$  составлена из  $i$ -ой строки  $\mathbf{W}$ .

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

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

$$(Y_j - \mathbf{U}v_j)^T \mathbf{W}^j (Y_j - \mathbf{U}v_j) \rightarrow \min_{v_j}, \quad j = 1, \dots, K, \quad (2)$$

где  $\mathbf{W}^j = \text{diag}(W_j) \in \mathbb{R}^{L \times L}$  составлена из  $j$ -го столбца  $\mathbf{W}$ .

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

- ❸ Повторяем шаги 1–2, пока не выполнен критерий сходимости

$$\left\| \mathbf{W}^{1/2} \odot (\mathbf{Y} - \mathbf{UV}^T) \right\|_F^2 \leq \varepsilon$$

или не достигнуто максимальное число итераций  $N_\alpha$ .

# WL2-SSA. Метод с итеративным обновлением весов.

## Реализация

**Алгоритм IRLS** (параметры  $\alpha$  и  $\sigma$ ) [Chen K., Sacchi M., 2015]:

- 1 Инициализация  $\mathbf{U} \in \mathbb{R}^{L \times r}$  и  $\mathbf{V} \in \mathbb{R}^{K \times r}$  (например, с помощью сингулярного разложения матрицы  $\mathbf{Y}$ ),
- 2 Выбор параметра  $\alpha$  (величина, начиная с которой точку ряда считать выбросом),
- 3 Вычисление матрицы остатков  $\mathbf{R} = \{r_{ij}\}_{i,j=1}^{L,K} = \mathbf{Y} - \mathbf{U}\mathbf{V}^T$ ,
- 4 Обновление параметра  $\sigma_{ij}$  (нормировка для остатков),
- 5 Вычисление матрицы весов  $\mathbf{W} = \{w_{ij}\}_{i,j=1}^{L,K} = \{w(\frac{r_{ij}}{\sigma_{ij}})\}_{i,j=1}^{L,K}$ , используя

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

- 6 Решение задачи взвешенной аппроксимации (обновление матриц  $\mathbf{U}$  и  $\mathbf{V}$ )

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

- 7 Повторяем шаги 3–6, пока не выполнен критерий сходимости

$$\left\| \mathbf{W}^{1/2} \odot (\mathbf{Y} - \mathbf{U}\mathbf{V}^T) \right\|_F^2 \leq \varepsilon$$

или не достигнуто максимальное число итераций  $N_{IRLS}$ .

# WL2-SSA. Метод с итеративным обновлением весов.

## Выбор параметров. Параметр $\sigma$

**Проблема 1:** Нормировка остатков на константный параметр  $\sigma_{ij} = \sigma \ \forall i, j$  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.

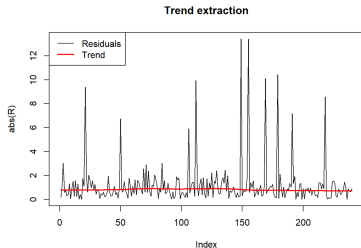


Рис.: График модуля остатков.  
Постоянная дисперсия шума.

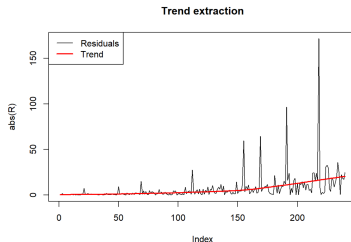


Рис.: График модуля остатков.  
Гетероскедастичный шум.

**Решение:** Будем рассматривать матрицу  $\Sigma = \{\sigma_{ij}\}_{i,j=1}^{L,K}$  ганкелевой, что соответствует приписыванию весов элементам ряда. Обозначим параметр  $\sigma = (\sigma_1, \dots, \sigma_N)^T$ . Будем задавать параметр  $\sigma$  как тренд (мат. ожидание) ряда, состоящего из модулей остатков.

# WL2-SSA. Выбор параметров. Параметр $\alpha$

**Проблема 2:** Непонятно, как задавать параметр  $\alpha$ , влияющий на то, какие точки считать выбросами, а какие — нет.

**Решение:** Выведем вероятностную формулу для параметра  $\alpha$ .

Модель ряда:  $x_i = s_i + \varepsilon_i$ ,  $i = 1, \dots, N$ . Ганкелизуем матрицу остатков

$\mathbf{R} = \mathbf{Y} - \mathbf{UV}^T$ , получим ряд  $R = \{r_i\}_{i=1}^N$ . В предположении точной отделимости сигнала от шума, ряд  $R = \{\varepsilon_i\}_{i=1}^N$ . Будем задавать вероятность  $\gamma$ :

$P(r^* \in (0, \alpha)) = \gamma$ , где  $r^* = \frac{|\varepsilon|}{\sigma}$ ,  $\sigma = (\sigma_1, \dots, \sigma_N)$  — тренд из ряда  $|R|$ .

## Определение

Если  $r \sim N(0, \sigma^2)$ , то  $|r| \sim N_H(\sigma^2)$  — *полунормальное распределение* с параметром  $\sigma^2$ , ф.р.  $F_H(x; \sigma^2) = \frac{2}{\sqrt{\pi}} \int_0^{x/\sqrt{2}\sigma^2} e^{-z^2} dz = \text{erf}(\frac{x}{\sqrt{2}\sigma^2})$ .

## Утверждение

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

Получаем выражение для  $\alpha$ :

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

**Замечание:** В предположении нормальности шума и точной отделимости сигнала от шума формула верна и для нестационарного шума.

# WL2-SSA. Метод с итеративным обновлением весов.

## Реализация

**Алгоритм IRLS** (параметры  $\alpha$  и  $\sigma$ ) [Chen K., Sacchi M., 2015]:

- 1 Инициализация  $\mathbf{U} \in \mathbb{R}^{L \times r}$  и  $\mathbf{V} \in \mathbb{R}^{K \times r}$  (например, с помощью сингулярного разложения матрицы  $\mathbf{Y}$ ),
- 2 Выбор параметра  $\alpha$  (величина, начиная с которой точку ряда считать выбросом),
- 3 Вычисление матрицы остатков  $\mathbf{R} = \{r_{ij}\}_{i,j=1}^{L,K} = \mathbf{Y} - \mathbf{U}\mathbf{V}^T$ ,
- 4 Обновление параметра  $\sigma_{ij}$  (нормировка для остатков),
- 5 Вычисление матрицы весов  $\mathbf{W} = \{w_{ij}\}_{i,j=1}^{L,K} = \{w(\frac{r_{ij}}{\sigma_{ij}})\}_{i,j=1}^{L,K}$ , используя

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

- 6 Решение задачи взвешенной аппроксимации (обновление матриц  $\mathbf{U}$  и  $\mathbf{V}$ )

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

- 7 Повторяем шаги 3–6, пока не выполнен критерий сходимости

$$\left\| \mathbf{W}^{1/2} \odot (\mathbf{Y} - \mathbf{U}\mathbf{V}^T) \right\|_F^2 \leq \varepsilon$$

или не достигнуто максимальное число итераций  $N_{IRLS}$ .



# WL2-SSA. Метод с итеративным обновлением весов.

## Модификация

Модификация 5-ого пункта алгоритма IRLS:

- 5.a Ганкелизация матрицы  $\mathbf{R}$  и получение ряда длины  $N$  из остатков:  $\mathbf{R} = \mathcal{T}^{-1}\Pi_{\mathcal{H}}(\mathbf{R}) = (r_1, \dots, r_N)^T$ ,
- 5.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,
- 5.c Вычисление ряда  $|\boldsymbol{\sigma}^{-1}\mathbf{R}| = (\frac{|r_1|}{\sigma_1}, \dots, \frac{|r_N|}{\sigma_N})^T$  и получение матрицы  $\mathbf{R}^* = \{r_{ij}^*\}_{i,j=1}^{L,K} = \mathcal{T}(|\boldsymbol{\sigma}^{-1}\mathbf{R}|)$ ,
- 5.d Вычисление матрицы весов  $\mathbf{W} = \{w_{ij}\}_{i,j=1}^{L,K} = \{w(r_{ij}^*)\}_{i,j=1}^{L,K}$ , используя

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

# Сравнение теоретических трудоемкостей

Ряд  $X = (x_1, \dots, x_N)$  длины  $N$ , матрица  $Y \in \mathbb{R}^{L \times K}$  — траекторная матрица ряда  $X$ . Ранг траекторной матрицы сигнала равен  $r$ .

- Трудоемкость последовательного метода:

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

где  $N_{\text{iter}}$  — общее кол-во итераций для сходимости метода (по выбранному критерию сходимости).

- Трудоемкость метода с обновлением весов:

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

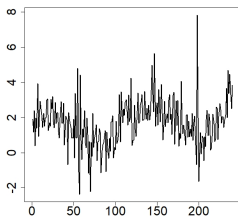
где  $N_{\alpha}$  и  $N_{\text{IRLS}}$  — общее кол-во итераций для решения задач (1), (2) и сходимости метода (по выбранному критерию сходимости).

Число итераций в статьях предполагается фиксированным. Однако предположение о достаточности фиксированного числа итераций не верно. В предположении, что число итераций не растет с увеличением длины ряда, так как зависит от разделимости, метод с итеративным обновлением весов оказывается менее трудоемким.

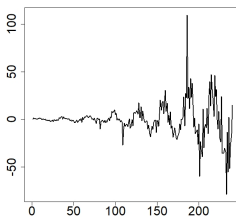
Пусть длина ряда  $N = 240$ . Рассмотрим следующие примеры:

- ❶  $x_n = e^{n/N} + \sin(2\pi n/120 + \pi/6) + \varepsilon_n, \quad \varepsilon_n \sim N(0, 1), r = 3,$
- ❷  $x_n = e^{4n/N} \sin(2\pi n/30) + Ae^{4n/N} \varepsilon_n, \quad \varepsilon_n \sim N(0, 1), r = 2,$
- ❸  $x_n = ne^{4n/N} \sin(2\pi n/30) + \varepsilon_n, \quad \varepsilon_n \sim N(0, 1), r = 4.$

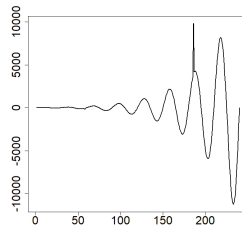
Выбросы: в случайно выбранных точках ряда  $x_i$  значение заменяется на  $x_i + \delta x_i$ , где  $\delta$  — заданная константа.



Пр. 1: 1% выбросов,  
выброс размера  $5x_i$ .



Пр. 2: 1% выбросов,  
выброс размера  $5x_i$ .



Пр. 3: 1% выбросов,  
выброс размера  $1.5x_i$ .

Временной ряд  $X = (x_1, \dots, x_N)$ ,  $x_i = s_i + \varepsilon_i$ ,  $i = 1, \dots, N$ .  
Обозначим  $S = (s_1, \dots, s_N)^T$  — сигнал.

**Выбросы:** в случайно выбранных точках ряда  $x_i$  значение заменяется на  $x_i + \delta x_i$ , где  $\delta$  — заданная константа. Будем сравнивать результаты при отсутствии (0%) выбросов, при 1% и 5% выделяющихся наблюдений.

На каждой реализации ряда случайными являются шум и местоположения выбросов.

Сравнения проводятся по величине ошибки, согласованной с  $\mathbb{L}_2$  (MSE) и ошибки, согласованной с  $\mathbb{L}_1$  (MAD):

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

где  $S$  — сигнал,  $\tilde{S}$  — его оценка. Будем вычислять  $\text{RMSE} = \sqrt{\text{MSE}}$ .

Таблица: Оценки RMSE для трех примеров для  $M = 30$  реализаций ряда.

	Пример 1		Пример 2		Пример 3	
Method	0%	5%	0%	5%	0%	5%
Basic SSA	<b>0.184</b>	0.653	<b>2.16</b>	5.96	<b>0.215</b>	459.6
l1pca	0.217	0.250	<b>2.45</b>	2.87	0.256	21.11
IRLS (orig.)	<b>0.184</b>	<b>0.206</b>	3.52	3.61	<b>0.216</b>	398.2
IRLS (loess)	0.196	<b>0.204</b>	<b>2.31</b>	<b>2.39</b>	<b>0.227</b>	303.2
IRLS (median)	0.210	0.223	2.84	2.86	0.256	38.21
IRLS (lowess)	0.206	<b>0.211</b>	2.59	2.63	0.243	<b>0.301</b>

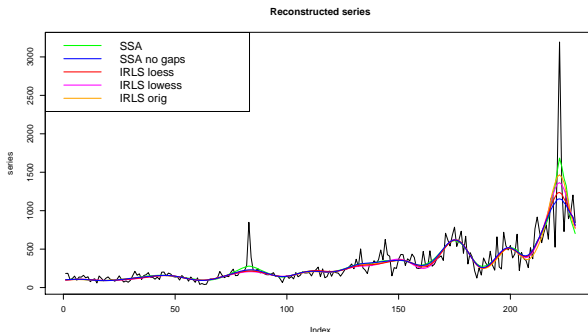
В каждом столбце выделен наилучший метод (**красным**) и незначимо отличающиеся от него (**синим**) при уровне значимости  $\alpha = 0.05$ . Выводы:

- Для первого примера наиболее устойчивыми являются оригинальный метод IRLS и его модификации с loess и lowess.
- Для ряда с гетероскедастичным шумом наиболее устойчивый метод — модификация IRLS с использованием локальной регрессии.
- В случае быстрорастущей амплитуды ряда модификация с использованием взвешенной локальной регрессии оказывается наиболее устойчивой.

# Реальный пример

Рассмотрим ряд — импорт товаров в США из Кувейта с ноября 1993 г. по ноябрь 2012 г.. Имеются данные за каждый месяц. Длина ряда  $N = 229$ . Возьмем длину окна  $L = 60$ , будем восстанавливать сигнал по 5 компонентам.

Выбросы находятся в точках  $x_{83}$  и  $x_{222}$ . В качестве истинного сигнала будем брать результат восстановления сигнала стандартным SSA для ряда с поставленными на места выбросов и впоследствии заполненными пропусками.



## Результаты:

- Структурированы устойчивые модификации по используемым в проекторах, входящих в алгоритм, нормам.
- Предложена новая модификация метода для рядов с нестационарным шумом, в которой нормировка для определения весов не константа, а определяется на основе тренда модуля ряда остатков.
- Рассматриваемые устойчивые модификации SSA были реализованы на R.
- Исследованы теоретические трудоемкости рассмотренных методов, проведено сравнение по трудоемкости. Метод с итеративным обновлением весов оказался менее трудоемким.
- Проведено сравнение методов по точности на модельных примерах и на реальном ряде. Сравнение методов по точности на примерах подтвердило, что предложенная модификация оказывается лучше в случае нестационарного шума. Также сравнение позволило дать рекомендации по использованию методов.