

Санкт-Петербургский государственный университет
Прикладная математика и информатика

Отчет по научно-исследовательской работе

SSA для ВРЕМЕННЫХ РЯДОВ, ВАРИАНТЫ CIRCULANT SSA

Выполнил:

Погребников Николай Вадимович

группа 21.Б04-мм

Научный руководитель:

д. ф.-м. н., доц.

Голяндина Нина Эдуардовна

Кафедра Статистического Моделирования

Санкт-Петербург

2024

Оглавление

1.	Введение	3
2.	Постановка задачи	3
3.	Описание результатов	4
3.1.	Алгоритм метода SSA	4
3.2.	Разделимость временных рядов	5
3.3.	Пример работы алгоритма SSA на реальных данных	9
4.	Заключение	10
	Список литературы	11

1. Введение

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

Сингулярный спектральный анализ (**SSA**) — метод для разложения временных рядов, который выделяет частотные компоненты, тренды и шум. В данном исследовании рассматривается математическая составляющая базового алгоритма **SSA**, а также его применение с помощью языка R и библиотеки Rssa.

2. Постановка задачи

Перед началом исследования были поставлены следующие цели:

1. Ознакомиться с базовым алгоритмом **SSA**[1];
2. Научиться подбирать параметры метода для наилучшей разделимости;
3. Проверить полученные теоретические знания на практике с помощью языка R и библиотеки Rssa[2].

3. Описание результатов

Пусть $N > 2$. Рассмотрим вещественнозначный временной ряд $\mathbb{X} = (x_0, \dots, x_{N-1})$ длины N . Базовый алгоритм состоит из четырёх шагов.

3.1. Алгоритм метода SSA

Шаг 1. Вложение

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

$$\mathbf{X} = \begin{pmatrix} x_0 & x_1 & x_2 & \dots & x_{K-1} \\ x_1 & x_2 & x_3 & \dots & x_K \\ x_2 & x_3 & x_4 & \dots & x_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{L-1} & x_L & x_{L+1} & \dots & x_{N-1} \end{pmatrix}.$$

Полезным свойством является то, что матрица \mathbf{X} имеет одинаковые элементы на диагоналях. Таким образом, L -траекторная матрица является ганкелевой.

Шаг 2. Сингулярное разложение (SVD)

Результатом этого шага является сингулярное разложение (**SVD**= Singular Value Decomposition) траекторной матрицы ряда.

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

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

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

Набор $(\sqrt{\lambda_i}, U_i, V_i^T)$ называется i -й собственной тройкой разложения (1).

Шаг 3. Группировка

На основе разложения (1) производится процедура группировки, которая делит все множество индексов $\{1, \dots, d\}$ на m непересекающихся подмножеств I_1, \dots, I_d .

Пусть $I = \{i_1, \dots, i_p\}$, тогда $\mathbf{X}_I = \mathbf{X}_{i_1} + \dots + \mathbf{X}_{i_p}$. Такие матрицы вычисляются для каждого $I = I_1, \dots, I_m$. В результате получаются матрицы $\mathbf{X}_{I_1}, \dots, \mathbf{X}_{I_m}$. Тем самым разложение (1) может быть записано в сгруппированном виде:

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

Шаг 4. Диагональное усреднение

Пусть \mathbf{Y} — матрица размерности $L \times K$. Диагональное усреднение переводит матрицу \mathbf{Y} в ряд g_0, \dots, g_{N-1} :

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

Применяя данную операцию к матрицам $\mathbf{X}_{I_1}, \dots, \mathbf{X}_{I_m}$, получают m новых рядов: $\mathbb{X}_1, \dots, \mathbb{X}_m$. При этом, $\mathbb{X}_1 + \dots + \mathbb{X}_m = \mathbb{X}$.

3.2. Разделимость временных рядов

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

Точная разделимость

Определение 1. Будем говорить, что ряды $\mathbb{X}^{(1)}$ и $\mathbb{X}^{(2)}$ слабо L -разделимы, если пространства, порождаемые строками $\mathbf{X}^{(1)}$ и $\mathbf{X}^{(2)}$ соответственно, ортогональны. То же самое должно выполняться для столбцов [1].

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

Определение 2. Будем говорить, что ряды $\mathbb{X}^{(1)}, \mathbb{X}^{(2)}$ сильно L -разделимы, если они слабо L -разделимы и после процедуры **SVD** собственные числа рядов различны [1].

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

Примеры точной разделимости

Рассмотрим таблицу, в которой знаком + отмечены пары рядов, для которых существуют параметры функций и параметры метода L и $K = N - L + 1$, при которых они разделимы (точно разделимы). Данная таблица 1 и условия разделимости с доказательствами взяты из книги [1].

Таблица 1. Точная разделимость

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

Например, если взять за $\mathbb{X}^{(1)} = 2 \cos\left(\frac{2\pi}{24}x + \frac{3\pi}{11}\right)$, $\mathbb{X}^{(2)} = 2 \cos\left(\frac{2\pi}{12}x\right)$, $N = 143$, $L = 72$, получится рисунок 1.

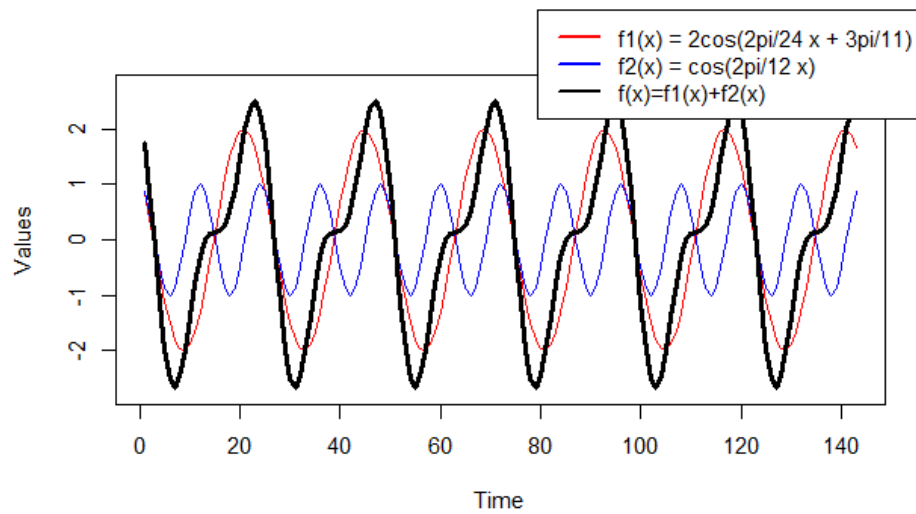


Рис. 1. $\mathbb{X} = \mathbb{X}^{(1)} + \mathbb{X}^{(2)} = 2 \cos\left(\frac{2\pi}{24}x + \frac{3\pi}{11}\right) + 2 \cos\left(\frac{2\pi}{12}x\right)$

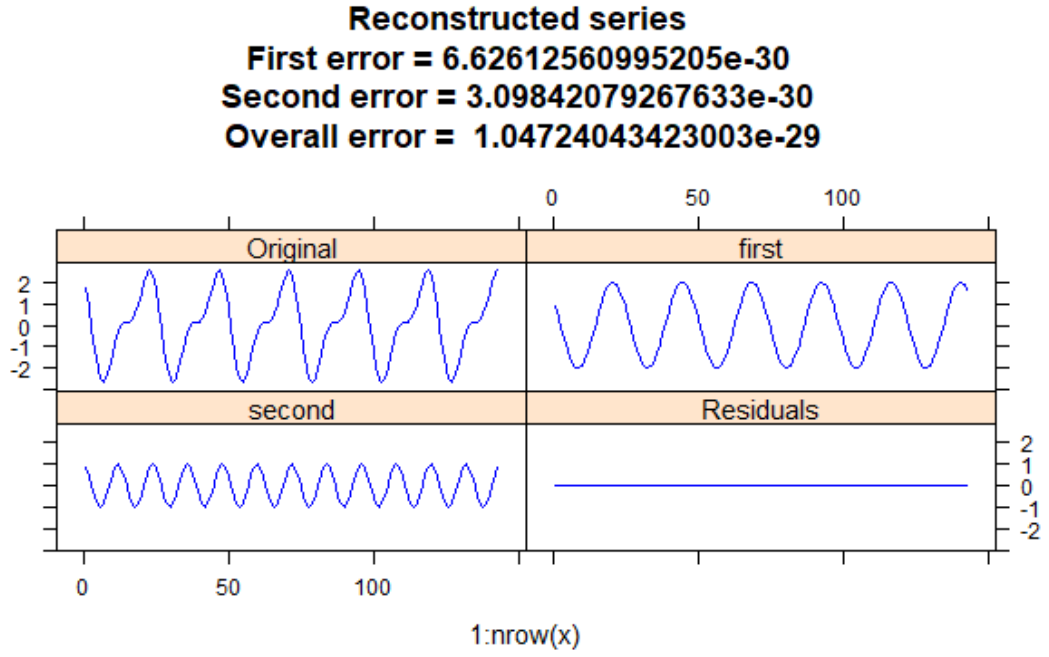


Рис. 2. $\mathbb{X} = \mathbb{X}^{(1)} + \mathbb{X}^{(2)} = 2 \cos\left(\frac{2\pi}{24}x + \frac{3\pi}{11}\right) + 2 \cos\left(\frac{2\pi}{12}x\right)$ после применения **SSA**

После применения **SSA** ряды были сгруппированы на ряды с названиями **first** и **second**. Как можно заметить на рисунке 2, ошибка разделения равна нулю (машинному нулю). Значит, ряды точно разделились.

Однако, по таблице 1 видно, что условия точной разделимости достаточно жесткие и вряд ли выполнимы в реальных задачах. Тогда появляется такое понятие, как асимптотическая разделимость.

Асимптотическая разделимость

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

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

Определение 3. Ряды $F^{(1)}, F^{(2)}$ называются ε -разделимыми при длине окна L , если

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

Определение 4. Если $\rho^{(\mathbb{X}(N), K(N))} \rightarrow 0$ при некоторой последовательности $L = L(N)$, $N \rightarrow \infty$, то ряды $\mathbb{X}^{(1)}, \mathbb{X}^{(2)}$ называются асимптотически $L(N)$ -разделимыми [1].

Примеры асимптотической разделимости

Как можно заметить по таблице 2, для гораздо большего класса функций асимптотическая разделимость имеет место.

Таблица 2. Асимптотическая разделимость

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

Так, для $\mathbb{X}^{(1)} = 2 \cos\left(\frac{2\pi}{24}x + \frac{3\pi}{11}\right)$, $\mathbb{X}^{(2)} = \frac{x}{20} - 4$, $N = 12 * 8$, получится рисунок 3.

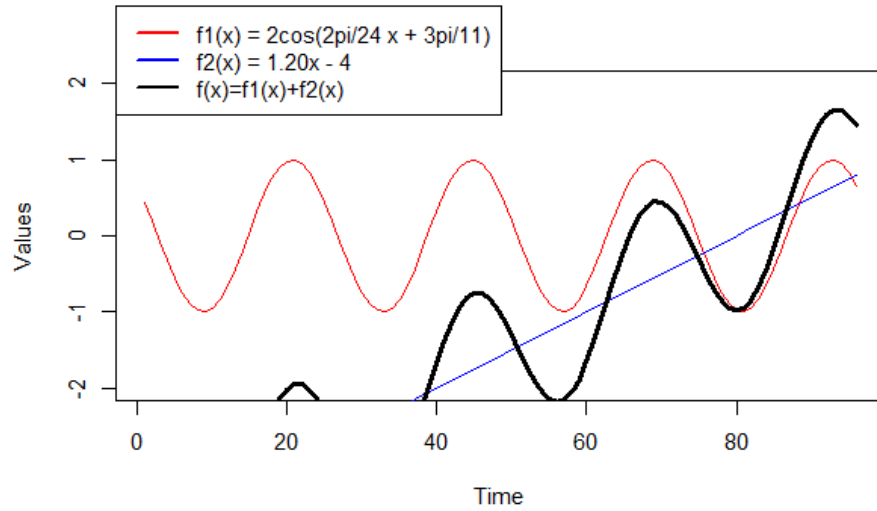


Рис. 3. $\mathbb{X} = \mathbb{X}^{(1)} + \mathbb{X}^{(2)} = 2 \cos\left(\frac{2\pi}{24}x + \frac{3\pi}{11}\right) + \frac{x}{20} - 4$

Если взять $N = 12 \times 8$, $L = 48$, то с увеличением N ошибки для разложения по первой и второй компонентах будут стремиться к нулю, как на рисунках 4, 5.

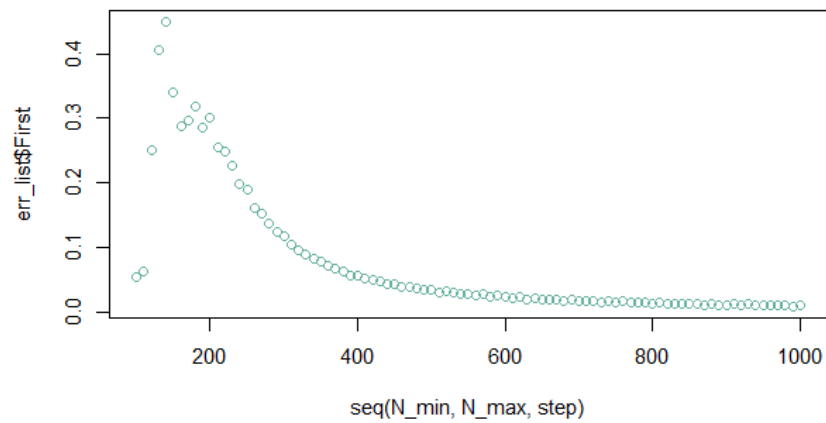


Рис. 4. Ошибка для $2 \cos\left(\frac{2\pi}{24}x + \frac{3\pi}{11}\right)$

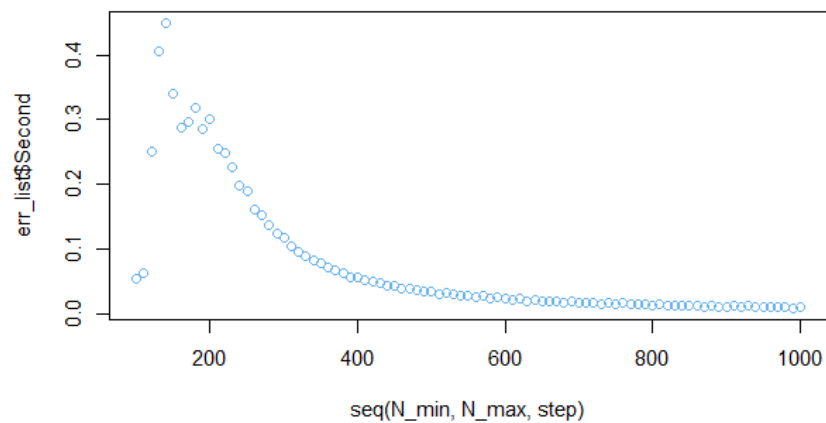


Рис. 5. Ошибка для $\frac{x}{20} - 4$

3.3. Пример работы алгоритма SSA на реальных данных

Есть набор данных, состоящий из объема месячных продаж крепленых вин в Австралии с января 1984 года по июнь 1994 года. Этот пример также подробно разобран в книге [2].

После применения алгоритма с выбором параметров $L = 84$ и определенной группировкой, получается разложение на трендовую, сезонную и шумовую составляющие.

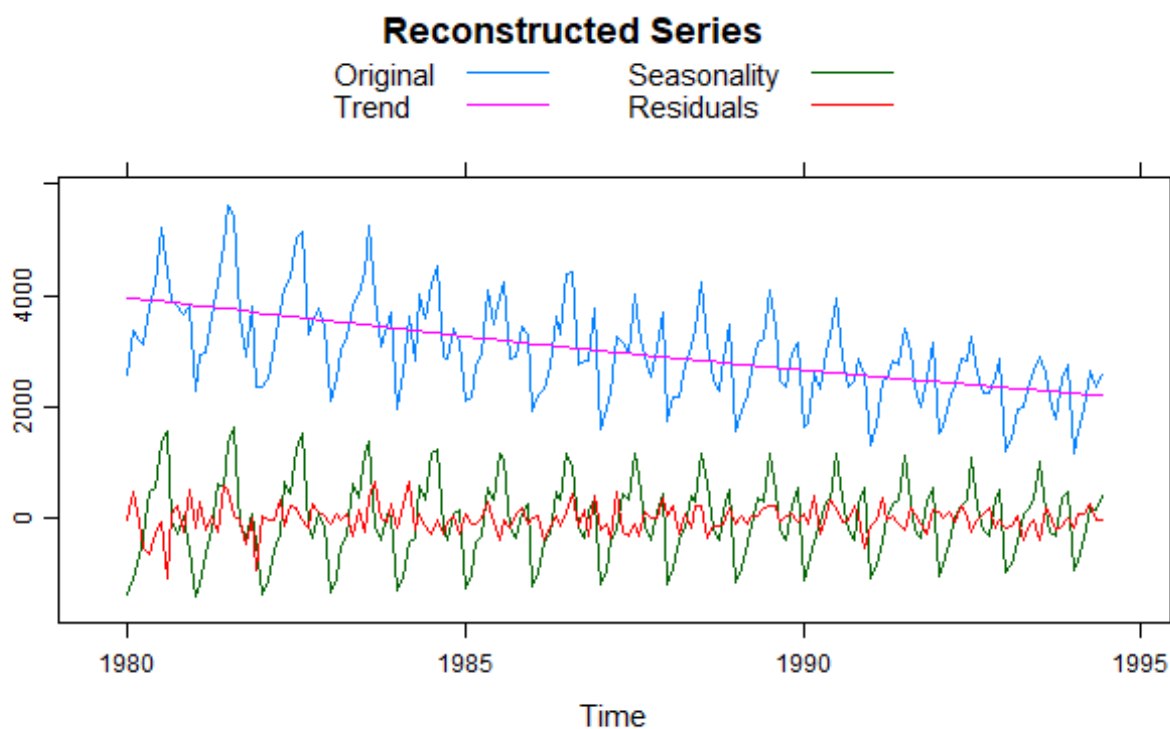


Рис. 6. График продаж вин с 1980 по 1995 года

4. Заключение

В данной работе был исследован базовый алгоритм метода **SSA** и полученные знания были проверены на реальных и смоделированных примерах. Однако данная научно-исследовательская работа на этом не ограничивается.

Дальнейшими действиями является: научиться прогнозировать значения на основе предыдущих, рассмотреть различные модификации метода **SSA**.

Список литературы

1. Golyandina Nina, Nekrutkin Vladimir, Zhigljavsky Anatoly. Analysis of Time Series Structure: SSA and Related Techniques. — Chapman and Hall/CRC, 2001.
2. Golyandina Nina, Korobeynikov Anton, Zhigljavsky Anatoly. Singular Spectrum Analysis with R. — Springer, 2018.