

САНКТ-ПЕТЕРБУРГСКИЙ УНИВЕРСИТЕТ

Математико-механический факультет

Кафедра статистического моделирования

Дипломная работа студента 55 группы
Александрова Фёдора Игоревича

ВЫДЕЛЕНИЕ
АДДИТИВНЫХ КОМПОНЕНТ ВРЕМЕННОГО РЯДА
НА ОСНОВЕ МЕТОДА “ГУСЕНИЦА”

Научный руководитель
Н.Э.Голяндина

Рецензент
В.В.Некруткин

“Допустить к защите”
Заведующий кафедрой

профессор С.М.Ермаков

Санкт-Петербург
2003

Содержание

1	Введение	3
2	Теоретические сведения о методе “Гусеница”	4
2.1	Алгоритм метода “Гусеница”	4
2.2	Разделимость временных рядов	6
2.3	Ряды конечного ранга	8
2.4	ЛРФ и корни характеристического полинома	10
3	Решение поставленных задач на основе метода “Гусеница”	11
3.1	Идентификация компонент	12
4	Описание методов автоматической идентификации	14
4.1	Идентификация компонент, соответствующих трендовым составляющим	14
4.1.1	Метод Кендалла	14
4.1.2	Метод нулей	16
4.1.3	Метод низких частот	16
4.2	Идентификация компонент, соответствующих э.-м. гармоническим составляющим	17
4.2.1	Метод Фурье	17
4.2.2	Усовершенствованный метод Фурье	19
5	Описание методов оценки частоты э.-м. гармонического ряда	23
5.1	Метод осреднения полярного угла	24
5.2	Метод характеристических корней	24
5.3	Периодограммный метод	25
6	Выбор параметров методов идентификации	25
6.1	Выбор параметров методов идентификации компонент, соответствующих трендовым составляющим	25
6.1.1	Метод Кендалла	25
6.1.2	Методы нулей и низких частот	26
6.2	Выбор параметров методов идентификации компонент, соответствующих э.-м. гармоническим составляющим	26
6.2.1	Метод Фурье и усовершенствованный метод Фурье	26
7	Сравнение методов идентификации	26
7.1	Методы идентификации тренда	27
7.2	Методы идентификации периодических составляющих	29
8	Описание программы	31
8.1	Работа с программой	31
	Литература	35

1 Введение

Рассмотрим вещественнозначный временной ряд $F_N = (f_0, \dots, f_{N-1})$ длины N . Одной из задач, связанных с временными рядами, является задача разбиения ряда на заданные аддитивные составляющие:

$$F_N = \sum_i F_N^{(i)}, \quad (1)$$

где $F_N^{(i)} = (f_0^{(i)}, \dots, f_{N-1}^{(i)})$, а (1) означает, что для $0 \leq n \leq N-1$ выполняется соотношение $f_n = \sum_i f_n^{(i)}$. Обычно интересует разбиение ряда на трендовую составляющую (или просто тренд), периодические составляющие (периодики) и шум.

В данной работе задача разложения ряда на аддитивные составляющие решается на основе метода исследования временных рядов “Гусеница” в зарубежной литературе этот метод также называется *Singular Spectrum Analysis (SSA)*. Он может применяться как к стационарным, так и нестационарным рядам и не предполагает знания параметрической модели ряда. Данному методу посвящены книги [1, 2]. Мы будем пользоваться книгой [1] как основным источником сведений о методе “Гусеница” (SSA) и придерживаться используемых там обозначений и понятий. При построении разложения вида (1) с помощью SSA оказывается заранее неизвестно, какие составляющие разложения (1) относятся к тренду, а какие составляют, например, периодику. Задачей дипломной работы является разработка, исследование и программная реализация алгоритмов идентификации составляющих разложения (1) временного ряда, полученного с помощью SSA. Часть методов идентификации основана на идеях, описанных в [3].

Опишем более подробно, составляющие какого вида мы будем идентифицировать. Начнём с задачи выделения тренда. Под трендом в общем случае понимается медленно меняющаяся составляющая ряда, но существует два подхода определения тренда. Первый заключается в параметрическом задании тренда, т. е. тренд задаётся, как некоторая функция f_n от номера n элемента ряда. При таком подходе часто различают *линейный, полиномиальный и экспоненциальный* тренды. Второй, частотный подход, определяет тренд как низкочастотную составляющую ряда. Для того, чтобы связать понятия низкочастотности и медленной изменчивости, введём *периодограмму ряда*.

Рассмотрим разложение Фурье вещественного временного ряда F_N :

$$f_n = c_0 + \sum_{1 \leq k \leq \frac{N-1}{2}} (c_k \cos(2\pi nk/N) + s_k \sin(2\pi nk/N)) + c_{N/2}, \quad 0 \leq n \leq N-1, \quad (2)$$

где $c_{N/2} = 0$, если N – нечётное, и $k \in \mathbb{Z}$. Тогда *периодограммой* $\Pi_f^N(\omega)$ ряда F_N , назовём функцию, определённую следующим образом при $\omega \in \{k/N\}_{k=0}^{\lfloor N/2 \rfloor}$:

$$\Pi_f^N(k/N) = \frac{N}{4} \begin{cases} 2c_0^2, & k = 0, \\ c_k^2 + s_k^2, & 1 \leq k \leq \frac{N-1}{2}, \\ 2c_{N/2}^2, & \text{если } N \text{ – чётное и } k = N/2. \end{cases} \quad (3)$$

Замечание 1.1. Существуют разные определения периодограммы, отличающиеся нормировкой. Мы будем рассматривать нормировку $N/4$, так как это будет удобно при дальнейшем использовании периодограммы в разделе 4.2.2.

Замечание 1.2. Исходя из данного определения периодограммы и вида разложения Фурье видно, что значение $\Pi_f^N(\omega)$, $\omega \in \{k/N\}_{k=0}^{\lfloor N/2 \rfloor}$, отражает вклад в вид ряда F_N гармоники с частотой ω и $\|F_N\|^2 = (F_N, F_N) = 2 \sum_{k=0}^{\lfloor N/2 \rfloor} \Pi_f^N(k/N)$.

Пользуясь понятием периодограммы, можно теперь формализовать понятие медленной изменчивости тренда. Будем считать, что составляющая ряда изменяется медленно, если гармоники в разложении (2) с низкими частотами дают большой вклад в вид ряда.

Применение метода “Гусеница” позволяет рассматривать понятие тренда, синтезирующее оба подхода (параметрический и частотный). Определим тренд, как медленно меняющуюся составляющую ряда, не являющуюся периодической. В разделе 3 рассмотрим понятие тренда более детально с учётом особенностей метода “Гусеница”.

Перейдём к задаче выделения периодики.

Определение 1.1. Ряд $F_N = (f_0, \dots, f_{N-1})$ называется *гармоническим* с частотой ω , если его элементы удовлетворяют следующему выражению:

$$f_n = A \cos(2\pi\omega n + \phi), \quad \omega \in (0, 0.5], \phi \in (0, 2\pi], 0 \leq n \leq N-1. \quad (4)$$

Периодом гармонического ряда с частотой ω , задаваемого (4), называется величина $T = 1/\omega$.

Определение 1.2. Под *периодической составляющей* (или просто *периодикой*) ряда с периодом T понимается сумма гармонических составляющих ряда, имеющих частоты ω_i : $\omega_i T \in \mathbb{Z}$.

Определение 1.3. Будем называть ряд F_N *экспоненциально-модулированным (э.-м.) гармоническим* с частотой ω и экспоненциальным показателем α , если его элементы удовлетворяют следующему выражению:

$$f_n = Ae^{\alpha n} \cos(2\pi\omega n + \phi), \quad A \in \mathbb{R}, \omega \in (0, 0.5], \phi \in (0, 2\pi], 0 \leq n \leq N-1. \quad (5)$$

Как и для гармонического ряда, *периодом* э.-м. гармонического ряда с частотой ω называется величина $T = 1/\omega$.

Определение 1.4. Составляющую ряда будем называть *обобщённой периодической* составляющей с периодом T , если она является суммой э.-м. гармоник с частотами ω_i : $\omega_i T \in \mathbb{Z}$ и необязательно равными экспоненциальными показателями α_i .

Обычно рассматриваются модулированные периодики, состоящие из гармонических составляющих с одинаковыми α_i , для метода же “Гусеница” равенство α_i не является обязательным требованием. Данная работа содержит следующие разделы. В разделе 2 приводится алгоритм метода “Гусеница”, а также теоретические сведения о разделимости рядов и ранге ряда, помогающие сформировать условия для решения задачи выделения составляющей. Также в нём вводятся понятия рядов, определяемых линейно-рекуррентной формулой и их характеристических полиномов, используемые в разделе 5. В разделе 3 формулируются задачи выделения составляющих ряда исходя из особенностей метода “Гусеница”, рассматриваются подходы к выделению тренда, периодик и обобщённых периодик ряда, приводится визуальный алгоритм решения поставленных задач и объясняется, почему для некоторых из поставленных задач необходимо уметь оценивать частоту э.-м. гармоники. В разделе 4 описываются алгоритмы автоматических методов выделения тренда, периодик и обобщённых периодик, а в следующем разделе приводятся методы оценки частоты э.-м. гармоники. В разделе 6 приводятся соображения о выборе параметров для этих методов при решении различных задач и в разделе 7 проводится сравнение характеристик работы методов с примерами работы на реальных временных рядах. В последний раздел помещено описание программы, решающей поставленные задачи на основе описанных методов.

2 Теоретические сведения о методе “Гусеница”

В данном разделе приведём алгоритм метода “Гусеница”, а также вкратце теоретические сведения, необходимые для его применения. Полную информацию о методе, а также доказательства приводимых фактов можно найти в [1].

Алгоритм метода состоит из четырёх шагов: вложение, сингулярное разложение, группировка и диагональное усреднение. Первые два в совокупности называются *разложением*, последние – *восстановлением*. Метод производит разбиение временного ряда на аддитивные компоненты и предоставляет информацию, на основе которой среди них можно найти те, которые формируют некоторую заданную составляющую ряда (например, тренд или периодику).

2.1 Алгоритм метода “Гусеница”

Первый этап: разложение

Шаг 1. Вложение. Рассмотрим одномерный вещественный временной ряд $F_N = (f_0, \dots, f_{N-1})$ длины $N \geq 3$. Выберем параметр вложения L ($L \in \mathbb{Z}$, $1 < L < N$), называемый длиной окна (или “длиной гусеницы”). Выбор L зависит как от вида ряда, так и от поставленной задачи.

Введём K , $K = N - L + 1$, и построим K L -мерных векторов $X_i \in \mathbb{R}^L$:

$$X_i = (f_{i-1}, \dots, f_{i+L-2})^T, \quad 1 \leq i \leq K.$$

Из них составим матрицу \mathbf{X} , называемую L -траекторной или просто траекторной матрицей L -вложения ряда F_N :

$$\mathbf{X} = [X_1 : \dots : X_K].$$

Так как $x_{ij} = f_{i+j-2}$, то матрица $\mathbf{X} = \{x_{ij}\}_{i,j=1}^{i=L, j=K}$ – ганкелева, т. е. для любых индексов i, j , $1 \leq i \leq L$, $1 \leq j \leq K$, таких, что $i + j = k$, $x_{ij} = c_k$.

Шаг 2. Сингулярное разложение. Пусть $\mathbf{S} = \mathbf{X}\mathbf{X}^T \in \mathbb{R}^{L \times L}$. Известно, что собственные числа $\{\lambda_i\}_{i=1}^L$ эрмитовой неотрицательно-определённой матрицы лежат в $\mathbb{R}_{\geq 0}$. Так как матрица \mathbf{S} является симметричной и неотрицательно-определённой, то её собственные числа $\{\lambda_i\}_{i=1}^L$ неотрицательны. Упорядочим их: $\lambda_1 \geq \dots \geq \lambda_L \geq 0$, и пусть U_1, \dots, U_L – соответствующие им ортонормированные собственные вектора. Будем называть $d = \max\{i : \lambda_i > 0\}$ порядком сингулярного разложения. Обозначим

$$V_i = \frac{1}{\sqrt{\lambda_i}} \mathbf{X}^T U_i, \quad i = 1, \dots, d,$$

V_i будем называть i -м факторным вектором.

Тогда сингулярное разложение матрицы \mathbf{X} имеет вид:

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_d, \quad \mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T, \quad (6)$$

U_i называют левым сингулярным вектором, V_i – правым сингулярным вектором, а \mathbf{X}_i – элементарной матрицей. Набор $(\sqrt{\lambda_i}, U_i, V_i)$ будем называть собственной тройкой.

Второй этап: восстановление

Шаг 3. Группировка. Разобьём множество индексов $\{1, \dots, d\}$ на непересекающиеся подмножества $I_j: \{1, \dots, d\} = \bigcup_{j=1}^m I_j$ и введём $\mathbf{X}_{I_j} = \sum_{k \in I_j} \mathbf{X}_k$. Таким образом, получаем разложение матрицы \mathbf{X} :

$$\mathbf{X} = \mathbf{X}_{I_1} + \mathbf{X}_{I_2} + \dots + \mathbf{X}_{I_m}.$$

Шаг 4. Диагональное усреднение. На этом шаге происходит построение временного ряда $F_N^{(j)}$ по матрице \mathbf{X}_{I_j} . Пусть есть матрица $\mathbf{Y} \in \mathbb{R}^{L \times K}$, $\mathbf{Y} = \{y_{i,j}\}_{i,j=1}^{i=L, j=K}$. Введём обозначения: $\tilde{L} = \min\{L, K\}$, $\tilde{K} = \max\{L, K\}$ и

$$\tilde{y}_{ij} = \begin{cases} y_{ij}, & \text{если } L \leq K, \\ y_{ji}, & \text{если } L > K. \end{cases}$$

Диагональным усреднением матрицы \mathbf{Y} называется получение ряда $G_N = (g_0, \dots, g_{N-1})$, $N = L + K - 1$, из матрицы \mathbf{Y} следующим образом:

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} \tilde{y}_{m, k-m+2}, & 0 \leq k < \tilde{L} - 1, \\ \frac{1}{\tilde{L}} \sum_{m=1}^{\tilde{L}} \tilde{y}_{m, k-m+2}, & \tilde{L} - 1 \leq k < \tilde{K}, \\ \frac{1}{N-k} \sum_{m=k-\tilde{K}+2}^{N-\tilde{K}+1} \tilde{y}_{m, k-m+2}, & \tilde{K} \leq k < N. \end{cases} \quad (7)$$

Применим диагональное усреднение к каждой из матриц \mathbf{X}_{I_j} , $j = 1 \dots m$, тем самым получив разложение ряда F_N в виде суммы рядов $F_N^{(j)}$:

$$F_N = F_N^{(1)} + F_N^{(2)} + \dots + F_N^{(m)}.$$

Определение 2.1. Будем называть ряд $F_N^{(j)}$, восстановленный по j -ой элементарной матрице, j -ым элементарным восстановленным рядом.

Замечание 2.1. Если рассмотреть матрицы $\mathbf{U}_d \in \mathbb{R}^{L \times d}$, $\mathbf{U}_d = [U_1 : \dots : U_d]$ и $\mathbf{V}_d \in \mathbb{R}^{K \times d}$, $\mathbf{V}_d = [V_1 : \dots : V_d]$, а также матрицу $\mathbf{\Lambda}_d \in \mathbb{R}^{d \times d}$, на диагонали которой стоят упорядоченные собственные числа $\lambda_1, \dots, \lambda_d$, то сингулярное разложение (6) может быть представлено в виде:

$$\mathbf{X} = \mathbf{U}_d \mathbf{\Lambda}_d^{\frac{1}{2}} \mathbf{V}_d^T \quad (8)$$

Это стандартная матричная форма сингулярного разложения матрицы \mathbf{X} .

Собственные числа имеют следующую интерпретацию.

Замечание 2.2. Отношение $\frac{\sum_{i \in I} \lambda_i}{\sum_{i=1}^d \lambda_i}$ отражает вклад составляющей, восстановленной по группе компонент $I \subset \{1, \dots, d\}$, в вид исходного ряда.

Так как собственные числа расположены по убыванию, собственные тройки, определяющие вид ряда, имеют меньшие номера, чем малозначачие. Например, в случае зашумления ряда тройки, соответствующие шуму, будут располагаться в конце набора всех собственных троек.

2.2 Разделимость временных рядов

Определение разделимости временных рядов Рассмотрим ряд $F_N = (f_0, \dots, f_{N-1})$, являющийся суммой рядов $F_N^{(1)}, F_N^{(2)}$, т. е. выполняется:

$$f_n = f_n^{(1)} + f_n^{(2)}, \quad 0 \leq n \leq N-1.$$

Для заданного L этим рядам соответствуют L -траекторные матрицы \mathbf{X} , $\mathbf{X}^{(1)}$ и $\mathbf{X}^{(2)}$.

Определение 2.2. L -траекторным пространством ряда F_N ($\mathcal{L}^{(L)}(F_N)$) будем называть пространство, являющееся линейной оболочкой набора векторов, стоящих в столбцах L -траекторной матрицы \mathbf{X} ряда F_N :

$$\mathcal{L}^{(L)} = \text{span}\langle X_1, \dots, X_K \rangle.$$

Возьмём $L: 1 < L < N$, $K = N - L + 1$.

Определение 2.3. $F_N^{(1)}$ и $F_N^{(2)}$ называются *слабо L -разделимыми*, если $\mathcal{L}^{(L)}(F_N^{(1)}) \perp \mathcal{L}^{(L)}(F_N^{(2)})$ и $\mathcal{L}^{(K)}(F_N^{(1)}) \perp \mathcal{L}^{(K)}(F_N^{(2)})$.

Замечание 2.3. Видно, что разделимость – это не что иное, как ортогональность всех отрезков ряда $F_N^{(1)}$ длины L и K всем отрезкам ряда $F_N^{(2)}$ длины L и K соответственно.

Пусть $\{\lambda_{1,i}\}_{i=1}^{d_1}$ и $\{\lambda_{2,i}\}_{i=1}^{d_2}$ – наборы собственных чисел, соответствующих рядам $F_N^{(1)}$ и $F_N^{(2)}$.

Определение 2.4. $F_N^{(1)}$ и $F_N^{(2)}$ называются *сильно L -разделимыми*, если

1. $\mathcal{L}^{(L)}(F_N^{(1)}) \perp \mathcal{L}^{(L)}(F_N^{(2)})$ и $\mathcal{L}^{(K)}(F_N^{(1)}) \perp \mathcal{L}^{(K)}(F_N^{(2)})$,
2. $\lambda_{1,i} \neq \lambda_{2,j}$, $1 \leq i \leq d_1$, $1 \leq j \leq d_2$.

Предложение 2.1. Пусть $F_N^{(1)}$ и $F_N^{(2)}$ – слабо L -разделимы. Тогда существует такое сингулярное разложение траекторной матрицы \mathbf{X} , что его можно разбить на два сингулярных разложения траекторных матриц $\mathbf{X}^{(1)}$ и $\mathbf{X}^{(2)}$.

Предложение 2.2. Пусть $F_N^{(1)}$ и $F_N^{(2)}$ – сильно L -разделимы. Тогда любое сингулярное разложение траекторной матрицы \mathbf{X} можно разбить на два сингулярных разложения траекторных матриц $\mathbf{X}^{(1)}$ и $\mathbf{X}^{(2)}$.

Предложения (2.1) и (2.2) имеют существенное практическое применение, так как они дают условия, при котором мы можем разбить сингулярное разложение исходного временного ряда на сумму сингулярных разложений рядов, его составляющих. Это разбиение положим в основу формирования непересекающихся множеств компонент на шаге группировки.

Замечание 2.4. При реализации алгоритмов, основанных на методе “Гусеница”, надо принимать во внимание, что на сингулярного разложения метода “Гусеница” используется некоторое произвольное разложение, а, следовательно, условием того, что его можно разбить на два сингулярных разложения соответствующих рядов, будет сильная разделимость этих рядов.

Далее при употреблении термина разделимость по умолчанию подразумевается слабая разделимость.

Примеры разделимых рядов

1. Отделимость от константного ряда.

Пусть ряд $F_N^{(1)}$ задаётся следующим образом: $f_n^{(1)} = c \neq 0$. Тогда гармонический ряд $F_N^{(2)}$ с частотой ω разделим с $F_N^{(1)}$ при $N, L: L\omega, K\omega \in \mathbb{Z}$.

2. Отделимость от экспоненциального ряда.

Пусть ряд $F_N^{(1)}$: $f_n^{(1)} = e^{\alpha n}$. Тогда э.-м. гармонический ряд $F_N^{(2)}$: $f_n^{(2)} = e^{-\alpha n} \sin(2\pi\omega n + \phi)$ разделим с $F_N^{(1)}$ при $N, L: L\omega, K\omega \in \mathbb{Z}$.

3. Отделимость от гармонического ряда.

Пусть $F_N^{(1)}$: $f_n^{(1)} = \cos(2\pi\omega_1 n + \phi_1)$. Тогда гармонический ряд $F_N^{(2)}$: $f_n^{(2)} = \cos(2\pi\omega_2 n + \phi_2)$, разделим с $F_N^{(1)}$ при $N, L: L\omega_1, L\omega_2, K\omega_1, K\omega_2 \in \mathbb{Z}$ и $\omega_1 \neq \omega_2$.

Естественно, в силу п. 1, гармонический ряд $F_N^{(1)}$ отделим также от константы при выполнении соответствующих условий.

4. Отделимость от э.-м. гармонического ряда.

Пусть $F_N^{(1)}$: $f_n^{(1)} = e^{\alpha n} \cos(2\pi\omega_1 n + \phi_1)$. Тогда экспоненциально-модулированный гармонический ряд $F_N^{(2)}$: $f_n^{(2)} = e^{-\alpha n} \cos(2\pi\omega_2 n + \phi_2)$, разделим с $F_N^{(1)}$ при $N, L: L\omega_1, L\omega_2, K\omega_1, K\omega_2 \in \mathbb{Z}$ и $\omega_1 \neq \omega_2$.

Приближённая и асимптотическая разделимость Требования разделимости являются очень жесткими и редко выполняются в реальных рядах. Так как разделимость – это ортогональность отрезков рядов длины L и K (см. замечание (2.3)), то введём понятие приближённой разделимости через меру ортогональности и через приближённую разделимость введём понятие асимптотической разделимости.

Пусть есть два ряда $F_N^{(1)}$ и $F_N^{(2)}$ длины N . Обозначим

$$\rho_{i,j}^{(M)} = \frac{\sum_{l=0}^{M-1} f_{i+l}^{(1)} f_{j+l}^{(2)}}{\sqrt{\sum_{l=0}^{M-1} (f_{i+l}^{(1)})^2} \sqrt{\sum_{l=0}^{M-1} (f_{j+l}^{(2)})^2}}$$

и введём меру ортогональности отрезков рядов

$$\rho^{(L,K)} = \max\left(\max_{1 \leq i,j \leq K} \rho_{i,j}^{(L)}, \max_{1 \leq i,j \leq L} \rho_{i,j}^{(K)}\right). \quad (9)$$

Определение 2.5. Будем называть ряды $F_N^{(1)}$ и $F_N^{(2)}$ ε -разделимыми при длине окна L , если $|\rho^{(L,K)}| < \varepsilon$.

В случае, когда ε мало, будем называть ε -разделимость *приближённой разделимостью*.

Приближённая разделимость может возникнуть в следующих случаях. Если ряды $F_N^{(1)}$ и $F_N^{(2)}$ разделимы, то при малых искажениях $\Delta^{(1)}$ и $\Delta^{(2)}$ ряды $F_N^{(1)} + \Delta^{(1)}$ и $F_N^{(2)} + \Delta^{(2)}$ будут приближённо разделимы. Если $F_N^{(1)}$ и $F_N^{(2)}$ L -разделимы, то при $L' \approx L$ они также будут приближённо разделимы. Рассмотрим теперь бесконечные временные ряды. Выберем последовательность $L(N)$ длин окон, зависящих от N , $1 < L(N) < N$, каждому $L(N)$ соответствует своё значение $\rho^{(L(N),K(N))}$.

Определение 2.6. Два бесконечных ряда $F^{(1)}$ и $F^{(2)}$ называются $L(N)$ -асимптотически разделимыми, если

$$\rho^{(L(N),K(N))} \xrightarrow{N \rightarrow \infty} 0.$$

Примеры асимптотической разделимости Приведём таблицу 1, которая показывает, существуют ли ограничения на параметры функций, задающих элементы рядов, а также на параметр разложения L и длину ряда N , при которых рассматриваемые ряды будут асимптотически разделимы. Более подробно про условия и скорость асимптотической сходимости можно узнать в [1, раздел 6.1.2].

Таблица 1: Наличие асимптотической разделимости между различными рядами

	конст.	гарм-ка	эксп-та	э.-м. гарм-ка	полином
конст.	–	+	+	+	–
гарм-ка	+	+	+	+	+
эксп-та	+	+	+	+	+
э.-м.гарм-ка	+	+	+	+	+
полином	–	+	+	+	–

Сравнивая таблицу с примерами точной разделимости, видно, что требование асимптотической разделимости гораздо более слабое, чем требование точной разделимости.

Можно ввести понятие стохастической разделимости [1, раздел 6.1.2], тогда справедливо также следующее утверждение [1, следствие 6.1]:

Предложение 2.3. *Любой бесконечный детерминированный ряд $F^{(1)}$ стохастически разделим с гауссовским белым шумом $F^{(2)}$, если $L, K \rightarrow \infty$ и $L/K \rightarrow a > 0$.*

2.3 Ряды конечного ранга

Рассмотрим ряд $F_N = (f_0, \dots, f_{N-1})$, $N \geq 3$, $1 < L < N$, $K = N - L + 1$, и $\mathcal{L}^{(L)}(F_N)$ – его L -траекторное пространство.

Определение 2.7. Ряд F_N имеет L -ранг, равный d ($\text{rank}_L F_N = d$), $d \leq \min(L, K)$, если $d = \dim \mathcal{L}^{(L)}(F_N)$.

Определение 2.8. Если $d \leq \min(L, N - L + 1)$, то тройку чисел (d, L, N) будем называть *допустимой*.

Определение 2.9. Ряд F_N называется рядом конечного ранга d ($\text{rank} F_N = d$), если для любого L , такого что тройка (d, L, N) – допустимая, $\text{rank}_L F_N = d$.

Предложение 2.4. *Справедливы следующие утверждения:*

1. $\text{rank}_L F_N = \text{rank} \mathbf{X} = \text{rank} \mathbf{X}\mathbf{X}^T = \text{rank} \mathbf{X}^T \mathbf{X}$, где \mathbf{X} – L -траекторная матрица ряда F_N .
2. Если $\text{rank}_L F_N = d$, тогда $\mathbf{X}\mathbf{X}^T$ имеет d ненулевых собственных чисел и соответствующие собственные вектора образуют ортонормированный базис траекторного пространства $\mathcal{L}^{(L)}(F_N)$.
3. Сингулярное разложение L -траекторной матрицы \mathbf{X} имеет порядок d .

Рассмотрим примеры рядов конечного ранга.

Экспоненциально-модулированные гармонические ряды Рассмотрим э.-м. гармонический ряд $F_N = (f_0, \dots, f_{N-1})$ с частотой ω и экспоненциальным показателем α . Напомним, его n -ая компонента задаётся следующим образом:

$$f_n = Ae^{\alpha n} \cos(2\pi\omega n + \phi), \quad A \in \mathbb{R}, \quad \omega \in (0, 0.5], \quad \phi \in (0, 2\pi]. \quad (10)$$

Ранг э.-м. гармонического ряда и вид сингулярных векторов, ему соответствующих, зависит от ω и α . Сначала рассмотрим общий случай, когда $\omega \in (0, 0.5)$. Для любых L, N ряд (10) при $\omega \in (0, 0.5)$ имеет ранг 2, то есть ему соответствуют две компоненты сингулярного разложения.

Лемма 2.1. *Если $\omega \in (0, 0.5)$, $L\omega \in \mathbb{Z}$, то элементы собственных векторов U_1, U_2 , соответствующих ряду (10), имеют вид*

$$u_n^{(1)} = B_1 e^{\alpha n} \cos(2\pi\omega n + \tilde{\phi}_1), \quad u_n^{(2)} = B_2 e^{\alpha n} \cos(2\pi\omega n + \tilde{\phi}_2) \quad (11)$$

при $1 \leq n \leq L$, $\tilde{\phi}_j \in [0, 2\pi)$. Причём если $L \rightarrow \infty$, $\alpha \rightarrow 0$, $L\alpha \rightarrow \gamma \in \mathbb{R}$, то

$$|\tilde{\phi}_1 - \tilde{\phi}_2| \rightarrow \pi/2. \quad (12)$$

Аналогичное утверждение верно для элементов факторных векторов, только требования накладываются на K , а не на L .

Доказательство. Из того, что U_1, U_2 образуют ортонормированный базис L -тракторного пространства $\mathcal{L}^{(L)}(F_N)$, следует, что их элементы представляются в виде (11). Докажем теперь, что верно (12). Исходя из вида векторов U_1, U_2 , запишем условие их ортогональности в следующем виде

$$\sum_{n=0}^{L-1} e^{2\alpha n} \cos(2\pi\omega n + \psi_1) \cos(2\pi\omega n + \psi_2) = 0,$$

где $\psi_j = \tilde{\phi}_j + 2\pi\omega$, $j = 1, 2$, заменив нумерацию при суммировании для удобства доказательства. Будем доказывать теперь, что для $\psi_1 = \psi_1(\alpha, L)$ и $\psi_2 = \psi_2(\alpha, L)$ выполняется $|\psi_1 - \psi_2| \rightarrow \pi/2$. Очевидно, что это равносильно утверждению (12). Очевидным образом получаем

$$\sum_{n=0}^{L-1} e^{2\alpha n} \cos(4\pi\omega n + \psi_1 + \psi_2) + \cos(\psi_1 - \psi_2) \sum_{n=0}^{L-1} e^{2\alpha n} = 0, \quad (13)$$

Для любых $\psi \in [0, 2\pi)$, β верно:

$$\sum_{n=0}^{L-1} e^{\beta n} \cos(4\pi\omega n + \psi) = \frac{2e^\beta \cos^2(2\pi\omega) \cos \psi - \cos \psi - e^\beta \cos \psi + e^\beta \sin \psi \sin(4\pi\omega)}{4e^\beta \cos^2(2\pi\omega) - 1 - 2e^\beta - e^{2\beta}} (1 - e^{\beta L})$$

и, так как $\omega < 0.5$, а, следовательно, $\cos^2(2\pi\omega) \neq 1$, то

$$\sum_{n=0}^{L-1} e^{\beta n} \cos(4\pi\omega n + \psi) \xrightarrow[L\beta \rightarrow \gamma']{L \rightarrow \infty} \frac{\cos \psi \sin^2(2\pi\omega) - \sin \psi \cos(2\pi\omega) \sin(2\pi\omega)}{2 \sin^2(2\pi\omega)} (1 - e^{\gamma'}).$$

Поэтому первое слагаемое суммы (13) имеет конечный предел:

$$\sum_{n=0}^{L-1} e^{2\alpha n} \cos(4\pi\omega n + \psi_1 + \psi_2) \xrightarrow[L\alpha \rightarrow \gamma]{L \rightarrow \infty} C,$$

$$C = \frac{\cos(\psi_1 + \psi_2) \sin^2(2\pi\omega) - \sin(\psi_1 + \psi_2) \cos(2\pi\omega) \sin(2\pi\omega)}{2 \sin^2(2\pi\omega)} (1 - e^{2\gamma}).$$

Так как

$$\sum_{n=0}^{L-1} e^{2\alpha n} = \frac{e^{2\alpha L} - 1}{e^{2\alpha} - 1} \xrightarrow[L\alpha \rightarrow \gamma]{L \rightarrow \infty} \infty,$$

то для того, чтобы выполнялось равенство (13), необходимо, чтобы $\cos(\psi_1 - \psi_2) \rightarrow 0$, что и доказывает (12). \square

Рассмотрим теперь два особых случая, когда $\omega = 0.5$ и когда $\omega = 0$.

Если $\omega = 0.5$ и $\cos(\phi) \neq 0$, тогда ряд $F_N = (f_0, \dots, f_{N-1})$ имеет вид: $f_n = (-e^\alpha)^n$. При любых N и L ранг такого ряда равен 1, причём собственный и факторный вектора, ему соответствующие, имеют такой же вид, как у исходного ряда.

Если $\omega = 0$ и $\cos(\phi) \neq 0$, тогда F_N имеет вид: $f_n = e^{\alpha n}$. При любых N и L такой ряд имеет ранг 1. Вид собственного и факторного вектора совпадает с видом исходного ряда.

Замечание 2.5. Если $\alpha = 0$ и $\omega \in (0, 0.5)$, то F_N принимает вид гармоник с частотой ω . При $L\omega \in \mathbb{Z}$ ($K\omega \in \mathbb{Z}$) элементы собственных (факторных) векторов имеют вид $B \cos(2\pi\omega n + \tilde{\phi})$, $B \sin(2\pi\omega n + \tilde{\phi})$ и оба собственных числа совпадают (собственные числа э.-м. гармонического ряда с экспоненциальным показателем $\alpha \leq 0$ асимптотически совпадают при $L, K \rightarrow \infty$).

Замечание 2.6. Вид элементов собственных (факторных) векторов, соответствующих гармонике с частотой ω , будет $B \cos(2\pi\omega n + \tilde{\phi})$, $B \sin(2\pi\omega n + \tilde{\phi})$ и при более слабом требовании на ω , когда $2L\omega \in \mathbb{Z}$ ($2K\omega \in \mathbb{Z}$).

Попробуем узнать, что представляют собой элементарные восстановленные ряды, соответствующие гармонической составляющей. Пусть F_N – гармонический ряд с частотой ω . Пусть $L\omega \in \mathbb{Z}$, $K\omega \in \mathbb{Z}$. Обозначим, как и при описании алгоритма “Тусеницы” в разделе 2.1, $\tilde{L} = \min(L, K)$,

$\tilde{K} = \max(L, K)$. Ряду F_N соответствуют две собственных тройки, пусть $F_N^{(1)} = (f_0^{(1)}, \dots, f_{N-1}^{(1)})$ и $F_N^{(2)} = (f_0^{(2)}, \dots, f_{N-1}^{(2)})$ – элементарные восстановленные ряды для каждой из этих троек. Покажем, что ряды $(f_{\tilde{L}-1}^{(1)}, \dots, f_{\tilde{K}-1}^{(1)})$ и $(f_{\tilde{L}-1}^{(2)}, \dots, f_{\tilde{K}-1}^{(2)})$ имеют такой же вид, какой имеет исходный ряд. Известен вид сингулярных векторов U_1, U_2, V_1, V_2 :

$$u_k^{(1)} = A \cos(2\pi\omega k + \phi), \quad u_k^{(2)} = A \sin(2\pi\omega k + \phi), \quad 1 \leq k \leq L,$$

$$v_k^{(1)} = B \cos(2\pi\omega k + \psi), \quad v_k^{(2)} = B \sin(2\pi\omega k + \psi), \quad 1 \leq k \leq K.$$

По определению элементарного восстановленного ряда получаем, что для $\tilde{L} - 1 \leq n \leq \tilde{K} - 1$

$$f_n^{(1)} = AB \frac{1}{\tilde{L}} \sum_{k=0}^{\tilde{L}-1} \cos(2\pi\omega(\tilde{L} - k - 1) + \phi) \cos(2\pi\omega(n - \tilde{L} + 1 + k) + \psi).$$

Так как

$$\begin{aligned} & \cos(2\pi\omega(\tilde{L} - k - 1) + \phi) \cos(2\pi\omega(n - \tilde{L} + 1 + k) + \psi) = \\ & \frac{1}{2} (\cos(2\pi\omega(2\tilde{L} - 2 - n - 2k) + \phi + \psi) + \cos(2\pi\omega n + \phi + \psi)), \end{aligned}$$

и для любого ξ справедливо, что $\sum_{k=0}^{\tilde{L}-1} \cos(2\pi\omega k + \xi) = 0$, то

$$f_n^{(1)} = \frac{1}{2} \cos(2\pi\omega n + \phi + \psi), \quad \tilde{L} - 1 \leq n \leq \tilde{K} - 1. \quad (14)$$

Аналогичным образом доказывается, что для элементов второго элементарного восстановленного ряда $F_N^{(2)}$ также выполняется (14) при $\tilde{L} - 1 \leq n \leq \tilde{K} - 1$.

Полиномиальные ряды Рассмотрим временной ряд $F_N = (f_0, \dots, f_{N-1})$ следующего вида:

$$f_n = \sum_{k=0}^m a_k n^k, \quad a_m \neq 0. \quad (15)$$

При достаточно больших N и любых L такой ряд имеет ранг $m + 1$ и элементы соответствующих сингулярных векторов удовлетворяют (15).

Справедливо следующее утверждение:

Предложение 2.5. *Любой ряд, являющийся линейной комбинацией произведений полиномов, экспонент и гармоник, является рядом конечного ранга.*

2.4 ЛРФ и корни характеристического полинома

Определение 2.10. Ненулевой ряд $G_N = (g_0, \dots, g_{N-1})$ определяется линейно-рекуррентной формулой (ЛРФ) размерности не более d , $1 \leq d \leq N - 1$, и это обозначается, как $\text{fdim}(G_N) \leq d$, если существуют такие $a_1, \dots, a_d \in \mathbb{R}$, $a_d \neq 0$, что верно следующее равенство:

$$g_{n+d} = \sum_{k=1}^d a_k g_{n+d-k}, \quad 0 \leq n \leq N - d - 1. \quad (16)$$

Определение 2.11. Число $d = \min\{k : \text{fdim}(G_N) \leq k\}$, называется *конечно-разностной размерностью ряда G_N* и это обозначается, как $\text{fdim}(G_N) = d$. ЛРФ вида (16) с $d = \text{fdim}(G_N)$ называется *минимальной ЛРФ*.

Предложение 2.6. *Э.-м. гармонический ряд $G_N = (g_0, \dots, g_{N-1})$ с частотой ω и экспоненциальным показателем α имеет конечно-разностную размерность 2 ($\text{fdim}(G_N) = 2$) и определяется ЛРФ следующего вида:*

$$g_{n+2} = 2e^\alpha \cos(2\pi\omega)g_{n+1} - e^{2\alpha}g_n, \quad 0 \leq n \leq N - 3. \quad (17)$$

Определение 2.12. Если ряд G_N имеет конечно-разностную размерность d ($\text{fdim}(G_N) = d$) и определяется ЛРФ (16), то полином $\mathcal{X}_d(x)$ вида (18) называется *характеристическим полиномом* этой ЛРФ.

$$\mathcal{X}_d(x) = x^d - \sum_{k=1}^d a_k x^{d-k} \quad (18)$$

Пусть полином имеет p различных корней. Обозначим за k_m , $1 \leq m \leq p$, кратность корня x_m . Тогда верна следующая теорема, [1, теорема 5.3]:

Теорема 2.1. *Вещественный ряд G_N , удовлетворяет ЛРФ (16) тогда и только тогда, когда*

$$g_n = \sum_{m=1}^p \sum_{j=0}^{k_m-1} c_{mj} n^j x_m^n,$$

где c_{mj} - комплексные коэффициенты, определенные по первым d значениям ряда g_0, \dots, g_{d-1} .

Рассмотрим теперь, как можно найти ЛРФ для некоторого ряда G_N ранга d .

Обозначим для вектора $X \in \mathbb{R}^L$ за $X^\nabla \in \mathbb{R}^{L-1}$ вектор, сформированный из первых $L-1$ компонент вектора X . Рассмотрим теперь $\mathcal{L}_d = \mathcal{L}^{(L)}$ - L -траекторное пространство ряда G_N . Обозначим за P_1, \dots, P_d ортонормированный базис \mathcal{L}_d , а за π_j , $1 \leq j \leq d$, - L -ый элемент вектора P_j . Рассмотрим линейное векторное пространство $\mathcal{L}_d^\nabla = \text{span}\langle P_1^\nabla, \dots, P_d^\nabla \rangle$.

Введём ν^2 - коэффициент вертикальности пространства \mathcal{L}_d , $\nu^2 = \pi_1^2 + \dots + \pi_d^2$. Предположим, что $e_L = (0, \dots, 0, 1)^T \in \mathbb{R}^L$ не принадлежит \mathcal{L}_d , тогда $\nu^2 < 1$. Заметим, что, так как ν есть косинус угла между e_L и \mathcal{L}_d , то он не зависит от выбора базиса P_1, \dots, P_d .

Следующая теорема, [1, теорема 5.2], предоставляет способ вычисления коэффициентов ЛРФ, определяющей ряд G_N .

Теорема 2.2. *Рассмотрим ряд G_N , $\text{rank}_L(G_N) = d$. Если $1 \leq d < L$ и e_L не принадлежит \mathcal{L}_d , тогда ряд G_N удовлетворяет следующей ЛРФ:*

$$g_{n+L-1} = \sum_{k=1}^{L-1} a_k g_{n+L-1-k}, \quad 0 \leq n \leq K-1,$$

где коэффициенты a_k определяются через следующее равенство:

$$(a_{L-1}, \dots, a_1)^T = \frac{1}{1-\nu^2} \sum_{j=1}^d \pi_j P_j^\nabla. \quad (19)$$

3 Решение поставленных задач на основе метода “Гусеница”

Метод “Гусеница”, применённый к ряду F_N , разбивает ряд на аддитивные компоненты (элементарные восстановленные ряды) и при этом он основан на сингулярном разложении траекторной матрицы (8), т. е. предоставляет информацию о компонентах в виде матриц собственных и сингулярных векторов и набора собственных чисел, причём каждой компоненте соответствует собственная тройка $(\sqrt{\lambda_i}, U_i, V_i)$.

Задача выделения любой аддитивной составляющей ряда длины N с помощью метода “Гусеница” опирается на следующее предположение: существует такое L , что возможно получить искомую составляющую (может быть, приближенно) как сумму элементарных восстановленных рядов после сингулярного разложения ряда с длиной гусеницы L . Тогда задача выделения сводится к задаче поиска компонент, соответствующих выделяемой составляющей, их группировке и восстановлению по ним аддитивной составляющей ряда, как это было описано в разделе 2.1.

Список решаемых задач Во введении тренд был определён, как медленно меняющаяся составляющая ряда, не являющаяся периодической. Метод “Гусеница” приводит к разложению тренда на элементарные восстановленные ряды, которые можно условно классифицировать, как монотонные, константные или медленно-осциллирующие составляющие тренда. Поэтому задачу выделения тренда можно разбить соответственно на более частные задачи выделения составляющих тренда таких типов.

Что касается обобщённых периодик, то метод “Гусеница” приводит к разложению ряда на э.-м. гармоники. Поэтому идентификация периодики сводится к обнаружению э.-м. гармоник с соответствующими частотами. Только после этого, собрав э.-м. гармоники с нужными частотами, мы получаем обобщённую периодику. Отметим, что с точки зрения метода “Гусеница”, э.-м. гармоника с периодом 2 ближе к экспоненциальному ряду, чем к гармоническому. Поэтому способ обнаружения такой э.-м. гармоники отличается от общего случая.

Сформируем список задач, решения которых на основе метода “Гусеница” будут приведены далее.

1. Выделение тренда ряда:
 - выделение монотонной составляющей тренда,
 - выделение константной составляющей тренда,
 - выделение медленно осциллирующей составляющей тренда.
2. Выделение периодик и обобщённых периодик ряда (задачи этого типа можно ставить также с ограничением на период, например, выделение периодики с периодом из заданного интервала):
 - выделение э.-м. гармоник ряда:
 - 1) с частотой $\omega < 0.5$,
 - 2) с частотой $\omega = 0.5$,
 - выделение периодик (может быть, обобщённых) ряда,
3. Разбиение временного ряда на тренд, периодики (может быть, обобщённые) и остаток. Остаток может быть как шумом, так и составляющей, не вписывающейся в модели тренда и периодик. Эту задачу можно рассматривать, как состоящую из задач, сформулированных выше и её решение будет состоять в последовательном выделении тренда и обобщённых периодик ряда.

Условия разрешимости Сформулируем условия разрешимости задачи выделения произвольной аддитивной составляющей ряда на основе метода “Гусеница”. Для того, чтобы её можно было выделить, она должна быть *сильно разделима с остатком ряда*. Тогда она будет рядом конечного ранга и в любом сингулярном разложении ей будет соответствовать набор собственных троек, размером равный её рангу. Выделение такой составляющей будет заключаться в поиске этих компонент и восстановлении по ним искомой части ряда.

Но при исследовании реальных рядов требование разделимости является чересчур сильным и практически никогда не выполняется. Условие приближённой разделимости в таком случае является более адекватным. Напомним, что согласно определениям 2.5, 2.6, если два бесконечных ряда $F^{(1)}$, $F^{(2)}$ асимптотически разделимы, то конечные ряды $F_N^{(1)}$, $F_N^{(2)}$ будут приближённо разделимы при достаточно большом N .

Будем предполагать, что выделяемая составляющая приближённо разделима с остатком при выбранной длине окна L . Например, если составляющая будет асимптотически отделима от остатка, то при достаточно большом N она будет приближённо с ним разделима. Тогда для достаточно больших N в сингулярном разложении ей будет соответствовать набор компонент разложения, размером равный её рангу и в этом случае достаточно будет найти эти компоненты и по ним приближённо восстановить искомую часть ряда.

3.1 Идентификация компонент

Как уже было ранее сказано, выделение составляющей заключается в поиске собственных троек сингулярного разложения, которые ей соответствуют. Процедуру их нахождения будем называть идентификацией. В качестве базы для идентификации будем использовать вид сингулярных векторов. В тех случаях, когда ранг d выделяемой составляющей известен, будем искать d собственных троек.

Тренд Рассмотрим, как в общем случае проводится идентификация собственных троек, соответствующих тренду ряда. Для тренда, как он определялся ранее, во многих случаях неизвестен его ранг. При классификации трендовых составляющих мы отдельно рассматривали монотонные и константные

составляющие тренда. Ранг константных составляющих известен, в отличие от ранга монотонной составляющей, которая может быть приближена несколькими способами, либо экспонентами, либо гармониками с большим периодом. Ранг трендовых составляющих, не являющихся ни монотонными, ни константными, тоже в общем случае неизвестен. Эти замечания не позволяют ставить обязательным условие известности ранга составляющей при решении задачи выделения тренда.

Сформулируем общий подход к выделению тренда как медленно меняющейся составляющей. Такой тренд может быть приближен либо экспонентами, либо низкочастотными гармониками. В обоих случаях сингулярные вектора ведут себя подобно самому тренду, т. е. тоже изменяются медленно. Поэтому при идентификации компонент, соответствующих тренду, будем искать компоненты, сингулярные вектора которых медленно меняются.

Периодическая составляющая Сведения, на основании которых проводится идентификация периодических и обобщённых периодических составляющих, были приведены в разделе 2.3. Для гармонической или э.-м. гармонической составляющей знаем её ранг, он равен 2 в случае, если её частота $\omega < 0.5$ и 1 в случае, когда $\omega = 0.5$. Также известен вид сингулярных векторов из собственных троек, соответствующих такой составляющей: они имеют вид, сходный с видом самой составляющей. Поэтому при идентификации компонент, соответствующих гармонике или э.-м. гармонике, будем искать две собственные тройки (или одну, но с частотой $\omega = 0.5$), сингулярные вектора которых имеют вид, сходный с видом самой выделяемой составляющей. Так как собственные числа, порожденные э.-м. гармоникой, асимптотически равны при $\alpha \leq 0$ и $L, K \rightarrow \infty$, то компоненты, соответствующие гармонике, будут иметь соседние номера, хотя при достаточно сильной зашумлённости ряда и недостаточно больших L и K это свойство может нарушиться.

Замечание 3.1. Можно пользоваться также тем, что для гармонического ряда известен вид участков элементарных восстановленных рядов от $\tilde{L} - 1$ до $\tilde{K} - 1$. Как было показано, они ведут себя, как и сингулярные вектора, подобно исходному ряду. Соответственно, методы идентификации, применяемые к сингулярным векторам, можно также применять и к соответствующим участкам элементарных восстановленных рядов.

Визуальная идентификация Задачи, сформулированные в разделе 3, можно решать следующим визуальным способом.

При выделении тренда рассматривают графики сингулярных векторов и отбирают компоненты с медленно изменяющимися векторами. При поиске же пар компонент, соответствующих гармонике и э.-м. гармонике, используется следующий приём. Изобразим в виде парной диаграммы два сингулярных вектора: значения одного будет откладывать по оси X , а другого по оси Y . Если эта пара сингулярных векторов порождена гармоникой и период гармонике T – целый, то построенный график выглядит, как правильный T -угольник. Даже если $T \notin \mathbb{Z}$, всё равно график имеет распознаваемую регулярную структуру. В случае, если пара сингулярных векторов соответствует э.-м. гармонической составляющей, график будет близок по виду к спирали.

Для примера приведём графики, построенные так, как это описано выше, для собственных векторов, соответствующих гармонике с периодом 12 ($L = 36$) и э.-м. гармонике с периодом 6 и экспоненциальным показателем $\ln(1.05)$ ($L = 48$). Амплитуда такой э.-м. гармонике возрастёт через 50 точек приблизительно в 10 раз. Точки графика для гармонике совпадают через каждые $T = 12$ шагов.

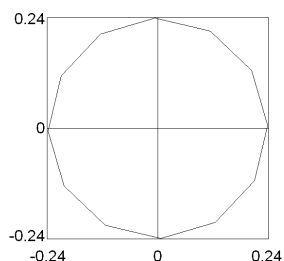


Рис. 1: “Кружок” для гармоники

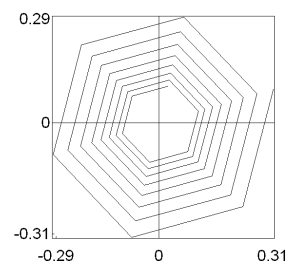


Рис. 2: “Спираль” для э.-м. гармонике

В сложных случаях, для решения которых подобных графиков недостаточно, также используется периодограмма, так как периодограмма гармонике длины M с частотой ω , $M\omega \in \mathbb{Z}$, принимает

единственное ненулевое значение в ω . Если $M\omega \notin \mathbb{Z}$, то периодограмма всё равно имеет единственный пик, который легко увидеть.

Задачи, требующие оценки частоты гармоник и э.-м. гармоник Среди поставленных задач есть задачи выделения периодической или обобщённой периодической составляющей с ограничениями на её период. Так как периодика и обобщённая периодика есть суммы гармоник или, соответственно, э.-м. гармоник, то для решения таких задач необходимо уметь оценивать частоту гармонической и э.-м. гармонической составляющей.

Опишем, как ограничение на период периодики формулируется для гармоник, её составляющих. Если поставлена задача выделения периодики с периодом из интервала $[a, b]$, тогда частота каждой её гармоники должна лежать в одном из интервалов вида $[k/a, k/b] \subset [0, 0.5]$, $k \in \mathbb{Z}$. Аналогичным образом формулируются условия для обобщённой периодики.

4 Описание методов автоматической идентификации

В качестве основы при разработке описанных в этом разделе методов использовалась статья [3]. Методы автоматической идентификации можно рассматривать, во-первых, как возможность автоматизировать процесс поиска и группировки компонент, соответствующих искомой составляющей (а, значит, автоматизировать её выделение), а во-вторых, как дополнительную информацию при выделении из ряда нужной составляющей, которую можно использовать при “ручном” формировании группы компонент для восстановления.

4.1 Идентификация компонент, соответствующих трендовым составляющим

Основанием работы этих методов служат рассуждения, описанные в разделе 3.1.

Вкратце их можно свести к тому, что сингулярные вектора компонент, соответствующих тренду, ведут себя подобно самому тренду. Поэтому методы идентификации должны определять по заданной последовательности (например, по элементам сингулярного вектора), близка ли она к тренду. Вследствие этого методы идентификации тренда сформулированы в применении к произвольному ряду, а не только к сингулярным векторам.

4.1.1 Метод Кендалла

Пусть есть два ряда длины M , $G_M = (g_0, \dots, g_{M-1})$ и $H_M = (h_0, \dots, h_{M-1})$. Введём $\mathcal{K}(G_M, H_M)$:

$$\mathcal{K}(G_M, H_M) = \# \{(i, j) : g_i < g_j, h_i < h_j, \quad 0 \leq i, j \leq M-1\}, \quad (20)$$

т. е. $\mathcal{K}(G_M, H_M)$ – это количество всех пар номеров элементов обоих рядов (i, j) , таких что выполняется $g_i < g_j, h_i < h_j$. Тогда *коэффициент корреляции Кендалла* между последовательностями g_0, \dots, g_{M-1} и h_0, \dots, h_{M-1} , называется $\tau(G_M, H_M)$ (см. [4]):

$$\tau(G_M, H_M) = \frac{4\mathcal{K}(G_M, H_M)}{M(M-1)} - 1. \quad (21)$$

Коэффициент корреляции Кендалла можно рассматривать, как меру монотонной зависимости между последовательностями g_0, \dots, g_{M-1} и h_0, \dots, h_{M-1} .

Возьмём теперь в качестве H_M ряд $H_M^{(0)} = (0, 1, \dots, M-1)$ и обозначим $\tau(G_M) = \tau(G_M, H_M^{(0)})$. Тогда

$$\tau(G_M) = \frac{4\mathcal{K}(G_M)}{M(M-1)} - 1, \quad (22)$$

где

$$\mathcal{K}(G_M) = \# \{(i, j) : g_i < g_j, i < j, \quad 0 \leq i, j \leq M-1\}.$$

Видно, что если ряд G_M – монотонный, то $|\tau(G_M)| = 1$.

Рассмотрим теперь двумерный случайный вектор $\begin{pmatrix} \xi \\ \eta \end{pmatrix}$. Пусть

$$\begin{pmatrix} g_0 \\ h_0 \end{pmatrix}, \begin{pmatrix} g_1 \\ h_1 \end{pmatrix}, \dots, \begin{pmatrix} g_{M-1} \\ h_{M-1} \end{pmatrix}$$

выборка объёма M из генерального закона распределения $\binom{\xi}{\eta}$, $G_M = (g_0, \dots, g_{M-1})$, и $H_M = (h_0, \dots, h_{M-1})$, тогда $\tau(G_M, H_M)$ – выборочный коэффициент корреляции Кендалла. Воспользуемся следующим утверждением ([4, разделы 4.8, 5.5, 5.6]):

Лемма 4.1. Пусть ξ и η – независимы. Тогда

$$P\left(\frac{|\tau(G_M, H_M) - a_M|}{\sqrt{\sigma_M}} > t\right) \xrightarrow{M \rightarrow \infty} 2 - 2\Phi(t), \quad t > 0,$$

где $\Phi(x)$ – функция нормального распределения с нулевым средним и единичной дисперсией, $\sigma_M = \frac{2(2M+5)}{9M(M-1)}$, $a_M = \frac{2}{M(M-1)}$.

Вот как этот результат можно применить к временным рядам. Пусть $\xi_0, \dots, \xi_M, \dots$ – бесконечная последовательность случайных величин и g_0, \dots, g_{M-1} – выборка объёма M , то есть последовательность реализаций M первых случайных величин.

Лемма 4.2. Если верна гипотеза \mathcal{H}_0 о том, что ξ_i – независимые и одинаково распределённые случайные величины, то для $\tau(G_M)$, $G_M = (g_0, \dots, g_{M-1})$, посчитанного по формуле (22), верно:

$$P\left(\frac{|\tau(G_M) - a_M|}{\sqrt{\sigma_M}} > t\right) \xrightarrow{M \rightarrow \infty} 2 - 2\Phi(t), \quad t > 0.$$

Известно, что критерий независимости, основанный на статистике τ , является мощным относительно альтернативы $\mathcal{H}_1 : \xi_i = f(i) + \varepsilon_i$, где $f(i)$ – монотонная функция, а ε_i – независимые и одинаково распределённые случайные величины, т. е. мощным против альтернатив с монотонным трендом $f(i)$. В то же время такой критерий маломощный против альтернативы со “стационарным” трендом вроде синуса. Поэтому можно перейти к следующей мере, отражающей наличие монотонного тренда в G_M :

$$\alpha(G_M) = 2 - 2\Phi\left(\frac{|\tau(G_M) - a_M|}{\sqrt{\sigma_M}}\right).$$

Для заранее заданного порогового уровня α_0 , $\alpha_0 \in (0, 1)$, будем считать, что если для ряда $G_M = (g_0, \dots, g_{M-1})$, выполняется неравенство $\alpha(G_M) < \alpha_0$, то он содержит монотонный тренд. В качестве примера применения данного метода приведём алгоритм идентификации собственных троек для решения задачи выделения тренда ряда на основе рассмотрения сингулярных векторов.

Алгоритм 4.1.

Исходные данные:

- матрица факторных или собственных векторов сингулярного разложения исходного временного ряда $\mathbf{Y} \in \mathbb{R}^{M \times L}$,
- пороговое значение для α_0 , $0 < \alpha_0 < 1$,
- количество рассматриваемых компонент сингулярного разложения A , $1 \leq A \leq L$ (для уменьшения времени работы алгоритма можно рассматривать только первые A компонент).

Результат работы: группа собственных троек, соответствующих трендовой составляющей исходного временного ряда.

Алгоритм:

1. $k := 1$;
2. пусть $Y_k = (y_0^{(k)}, \dots, y_{M-1}^{(k)})^T$ – k -ый столбец матрицы \mathbf{Y} , $G_M = (y_0^{(k)}, \dots, y_{M-1}^{(k)})$;
3. сосчитаем $\tau(G_M)$ по (22);
4. сосчитаем $\alpha(G_M) = 2 - 2\Phi\left(\frac{3}{2}\sqrt{\frac{M(M-1)}{M+2.5}}\left(|\tau(G_M)| - \frac{2}{M(M-1)}\right)\right)$;
5. если $\alpha(G_M) < \alpha_0$, то считаем k -ую компоненту соответствующей трендовой составляющей исходного ряда;
6. если $k = A$, то заканчиваем работу;
7. $k := k + 1$;
8. goto 2.

4.1.2 Метод нулей

Метод основан на представлении о тренде, как о медленно меняющейся, необязательно монотонной, составляющей ряда.

Для рассматриваемого ряда $G_M = (g_0, \dots, g_{M-1})$ считается

$$\mathcal{Z}(G_M) = \#\{i : g_i g_{i+1} \leq 0, |g_i - g_{i+1}| > \varepsilon, 0 \leq i \leq M-2\}, \quad (23)$$

где ε – заданный допуск, сравнимый с точностью вычислений. Для очень маленького ε величину $\mathcal{Z}(G_M)$ можно воспринимать как количество нулей кусочно-линейной функции, образованной рядом G_M следующим образом: в точках i , $0 \leq i \leq M-1$, она принимает значения g_i . Отсюда и название метода.

Для ряда G_M , являющегося трендом, значение $\mathcal{Z}(G_M)$ не должно быть большим, поэтому введём \mathcal{Z}_0 – ограничение сверху на $\mathcal{Z}(G_M)$ и будем считать, что G_M является трендом, если $\mathcal{Z}(G_M) \leq \mathcal{Z}_0$. Как и для метода Кендалла, приведём алгоритм идентификации собственных троек, соответствующих тренду, на основе сингулярных векторов.

Алгоритм 4.2.

Исходные данные:

- матрица факторных или собственных векторов сингулярного разложения исходного временного ряда $\mathbf{Y} \in \mathbb{R}^{M \times L}$,
- максимально возможное для тренда количество “нулей” \mathcal{Z}_0 , $\mathcal{Z}_0 \geq 0$,
- количество рассматриваемых компонент сингулярного разложения A , $1 \leq A \leq L$.

Результат работы: группа собственных троек, соответствующих трендовой составляющей исходного временного ряда.

Алгоритм:

1. $k := 1$;
2. пусть $Y_k = (y_0^{(k)}, \dots, y_{M-1}^{(k)})^T$ – k -ый столбец матрицы \mathbf{Y} , $G_M = (y_0^{(k)}, \dots, y_{M-1}^{(k)})$;
3. считаем $\mathcal{Z}(G_M)$ согласно определению (23);
4. если $\mathcal{Z}(G_M) \leq \mathcal{Z}_0$, то считаем k -ую компоненту соответствующей трендовой составляющей исходного ряда;
5. если $k = A$, то заканчиваем работу;
6. $k := k + 1$;
7. goto 2.

4.1.3 Метод низких частот

Метод нулей основан на представлении о тренде, как о медленно меняющейся составляющей, но для определения того, насколько медленно меняется ряд, использует искусственную характеристику – число “нулей”. Рассмотрим более естественный подход, основанный на периодограмме (3). Будем считать, что ряд является трендом, если в его разложении Фурье гармонические составляющие с низкими частотами дают большой вклад. Данный метод основывается на таком подходе и поэтому называется методом низких частот.

Пусть есть ряд $G_M = (g_0, \dots, g_{M-1})$. Будем считать областью низких частот интервал $[0, \omega_0]$, $0 < \omega_0 < 0.5$. Посчитаем для ряда G_M отношение

$$\mathcal{C}(G_M) = \frac{\Pi_f^M(0) + \sum_{k: \frac{k}{M} \leq \omega_0} \Pi_f^M(k/M)}{\Pi_f^M(0) + \sum_{1 \leq k \leq \frac{M-1}{2}} \Pi_f^M(k/M) + \Pi_f^M(M/2)}, \quad \Pi_f^M(M/2) = 0, \text{ если } M\text{-нечётное.} \quad (24)$$

$\mathcal{C}(G_M)$ можно интерпретировать, как вклад гармоник с низкими частотами в разложение Фурье последовательности g_0, \dots, g_{M-1} .

Для тренда значение $\mathcal{C}(G_M)$ должно быть большим, но так как требование большого вклада в вид ряда низкочастотных гармоник равносильно требованию маленького вклада оставшихся

(высокочастотных) гармоник, будем считать, что ряд G_M содержит трендовую составляющую, если $1 - \mathcal{C}(G_M) \leq \mathcal{C}_0$.

Приведём алгоритм для идентификации собственных троек, которые соответствуют трендовой составляющей ряда, на основе рассмотрения сингулярных векторов.

Алгоритм 4.3.

Исходные данные:

- матрица факторных или собственных векторов сингулярного разложения исходного временного ряда $\mathbf{Y} \in \mathbb{R}^{M \times L}$,
- верхняя граница области низких частот ω_0 , $0 < \omega_0 < 0.5$
- максимально вклад высокочастотных гармоник \mathcal{C}_0 , $0 < \mathcal{C}_0 < 1$,
- количество рассматриваемых компонент сингулярного разложения A , $1 \leq A \leq L$.

Результат работы: группа компонент, соответствующих трендовой составляющей исходного временного ряда.

Алгоритм:

1. $k := 1$;
2. пусть $Y_k = (y_0^{(k)}, \dots, y_{M-1}^{(k)})^T$ – k -ый столбец матрицы \mathbf{Y} , $G_M = (y_0^{(k)}, \dots, y_{M-1}^{(k)})$;
3. считаем $\mathcal{C}(G_M)$ согласно определению (24);
4. если $1 - \mathcal{C}(G_M) \leq \mathcal{C}_0$, то считаем k -ую компоненту соответствующей трендовой составляющей исходного ряда;
5. если $k = A$, то заканчиваем работу;
6. $k := k + 1$;
7. goto 2.

4.2 Идентификация компонент, соответствующих э.-м. гармоническим составляющим

Методы, используемые для идентификации компонент, соответствующих э.-м. гармоническим составляющим, основаны на периодограмме (3).

Для гармонического ряда длины M с частотой $\omega = k_0/M$, $k_0 \in \mathbb{Z}$, периодограмма равна нулю во всех точках, где она определена, кроме k_0/M :

$$P_f^M(k/M) = 0, \quad 0 \leq k \leq M/2, k \neq k_0. \quad (25)$$

Так как соответствие между периодограммой и рядом однозначно, то вышеописанный вид может иметь только периодограмма гармонического ряда. Периодограмма э.-м. гармонического ряда длины M с частотой k_0/M имеет похожий на (25), но более сложный вид. Она тоже имеет “пик” в точке k_0/M , но при этом в близких к k_0/M точках не принимает нулевого значения. Этот эффект называется “растеканием” периодограммы. Подобное растекание можно наблюдать для периодограммы гармонического ряда длины M , частота которого не попадает на решётку $\{k/M\}_{k=0}^{M/2}$, а также для периодограммы ряда, являющегося суммой двух гармоник с близкими частотами.

4.2.1 Метод Фурье

Метод основывается на том, что пик периодограммы для гармонического и э.-м. гармонического рядов длины M с частотой k_0/M находится в точке k_0/M . Будем понимать под периодограммой вектора $Y = (y_0, \dots, y_{M-1})^T$ периодограмму ряда (y_0, \dots, y_{M-1}) . Так как гармонике с частотой ω , $\omega < 0.5$, соответствуют два сингулярных вектора, являющихся э.-м. гармониками с таким же периодом, то аргументы максимумов их периодограмм будут совпадать.

Для данного метода нет различия между идентификацией гармоники и э.-м. гармоники, так как он основан только на аргументе “пика” периодограммы, а не на её форме и “растекание” периодограммы здесь не вредит.

Рассмотрим матрицу сингулярных векторов $\mathbf{Y} \in \mathbb{R}^{M \times L}$ и пусть $Y_i = (y_0^{(i)}, \dots, y_{M-1}^{(i)})^T$ – i -ый сингулярный вектор. Будем искать пары собственных троек, соответствующих э.-м. гармонике. Достаточно рассматривать только пары вида $i, i+1$, $0 \leq i \leq L-2$, так как собственные числа упорядочены по убыванию, а собственные соответствующие э.-м. гармонике, асимптотически равны при $L, K \rightarrow \infty$. Для каждого сингулярного вектора $Y_j \in \mathbb{R}^M$, $j = i, i+1$, найдём частоту θ_j , на которой достигается максимум периодограммы $\Pi_{Y_j}^M$:

$$\max_{0 \leq k \leq M/2} \{\Pi_{Y_j}^M(k/M)\} = \Pi_{Y_j}^M(\theta_j). \quad (26)$$

Сравним θ_i, θ_{i+1} . Если

$$M|\theta_i - \theta_{i+1}| \leq s_0, \quad (27)$$

(где s_0 – параметр метода, $s_0 \in \mathbb{Z}$), то будем считать, что пара компонент с номерами $i, i+1$ соответствуют э.-м. гармонической составляющей исходного ряда.

Замечание 4.1. Частоты, на которых производится сравнение значений периодограмм, принадлежат множеству $\{k/M\}_{k=0}^{M/2}$. Поэтому в качестве расстояния между θ_i и θ_{i+1} мы используем количество интервалов длины $1/M$, содержащихся между θ_i и θ_{i+1} .

Замечание 4.2. Пары номеров собственных троек, идентифицированных с помощью метода Фурье, могут пересекаться.

Так как э.-м. гармоническому ряду с частотой $\omega = 0.5$ соответствует один сингулярный вектор и восстановленный ряд с тем же периодом, то при идентификации такие составляющие надо рассматривать отдельно и идентифицировать i -ую собственную тройку, если $M|\theta_i - 0.5| \leq s_0$, где θ_i – частота, на которой достигается максимум периодограммы $\Pi_{Y_i}^M$ i -ого сингулярного вектора.

Алгоритм 4.4.

Исходные данные:

- матрица факторных или собственных векторов сингулярного разложения исходного временного ряда $\mathbf{Y} \in \mathbb{R}^{M \times L}$,
- максимально возможное расстояние s_0 , $s_0 \in \mathbb{Z}$, между аргументами максимумов периодограмм сингулярных векторов, соответствующих э.-м. гармонической составляющей или максимальное расстояние от 0.5 до максимума периодограммы сингулярного вектора, достаточное для того, чтобы считать вектор соответствующей э.-м. гармонике с частотой $\omega = 0.5$,
- количество рассматриваемых компонент сингулярного разложения A , $1 \leq A \leq L$.

Результат работы: группа компонент, соответствующих э.-м. гармоническим составляющим исходного временного ряда.

Алгоритм:

1. $k := 1$;
2. сосчитаем θ_k по (26);
3. если $M|\theta_k - 0.5| \leq s_0$, то
 - 1) считаем k -ую компоненту соответствующей э.-м. гармонике с частотой $\omega = 0.5$;
 - 2) goto 6;
4. сосчитаем θ_{k+1} по (26);
5. если $M|\theta_k - \theta_{k+1}| \leq s_0$, то считаем компоненты с номерами k и $k+1$ соответствующими э.-м. гармонической составляющей с частотой $\omega < 0.5$;
6. если $k = A - 1$, то заканчиваем работу;
7. $k := k + 1$;
8. goto 4.

4.2.2 Усовершенствованный метод Фурье

Метод Фурье плох тем, что не принимает во внимание форму периодограмм сингулярных векторов, а сравнивает только аргументы их максимумов. Но максимумы периодограмм для двух вовсе не соответствующих гармонике сингулярных векторов могут случайно совпасть. Поэтому хотелось бы учитывать форму периодограммы, которая должна быть близка к форме периодограммы гармонического ряда. Исходя из этого требования, усовершенствуем метод Фурье, пользуясь [3, раздел 4.2].

Заметим, что “растекание” периодограммы э.-м. гармонического ряда может ухудшить качество работы метода, поэтому сформулируем вначале метод в применении к задаче выделения гармоник, а потом рассмотрим, что можно сделать для того, чтобы он работал и при выделении э.-м. гармоник ряда.

Рассмотрим ряд G_N , пусть ранг его равен d . Выберем, как и ранее, тип сингулярных векторов, к которым будем применять метод и обозначим их длину за M . Для множества $\mathcal{A} \subset \{1, \dots, d\}$ номеров компонент сингулярного разложения введём функцию $\rho_{\mathcal{A}}(f)$, определённую на множестве $\{k/M\}_{k=0}^{\lfloor M/2 \rfloor}$:

$$\rho_{\mathcal{A}}(f) = \sum_{k \in \mathcal{A}} \Pi_{Y_k}^M(f), \quad \text{где } Y_k \in \mathbb{R}^M \text{ — } k\text{-ый сингулярный вектор.} \quad (28)$$

Лемма 4.3. *Рассмотрим гармонический ряд $G_N = (g_0, \dots, g_{N-1})$:*

$$g_n = A \cos(2\pi\omega n + \phi), \quad 0 \leq n \leq N-1, \quad \omega \in (0, 0.5),$$

причем длина ряда N и длина окна L , при котором проводится сингулярное разложение, такие, что $L\omega \in \mathbb{Z}$, $L \geq 2$, $K\omega \in \mathbb{Z}$, $K \geq 2$, $K = N - L + 1$. Тогда

$$\rho_{\{1,2\}}(f) = \begin{cases} 1, & f = \omega, \\ 0, & f \neq \omega. \end{cases}$$

Доказательство. В силу условий леммы элементы $Y_1, Y_2 \in \mathbb{R}^M$ — двух первых сингулярных векторов рассматриваемого типа разложения ряда G_N имеют вид:

$$y_n^{(1)} = A \cos(2\pi\omega n + \phi), \quad y_n^{(2)} = A \sin(2\pi\omega n + \phi), \quad n = 0 \dots M-1.$$

Воспользуемся тем, что $\|Y_1\| = \|Y_2\| = 1$ и найдём A_1 :

$$A^2 = \left(\sum_{n=0}^{M-1} (\cos(2\pi\omega n + \phi))^2 \right)^{-1} = \frac{2}{M}.$$

Пользуясь определением периодограммы (3), находим, что для любого $f = \frac{k}{M}$, $k \in \mathbb{Z}$, $0 < k < \frac{M}{2}$,

$$\rho_{\{1,2\}}(f) = \Pi_{Y_1}^M(f) + \Pi_{Y_2}^M(f) = \frac{M}{4}(2A^2) = 1, \quad \text{если } f = \omega,$$

$$\rho_{\{1,2\}}(f) = \Pi_{Y_1}^M(f) + \Pi_{Y_2}^M(f) = 0, \quad \text{если } f \neq \omega.$$

□

Следствие 4.1. *Пусть $g_n = g_n^{(1)} + g_n^{(2)}$, $g_n^{(1)} = A \cos(2\pi\omega n + \phi)$, $0 \leq n \leq N-1$, $\omega \in (0, 0.5)$. Тогда, если для длины ряда N и длины окна L , с которой проводится разложение ряда G_N , выполняется $L\omega \in \mathbb{Z}$, $L \geq 2$, $K\omega \in \mathbb{Z}$, $K \geq 2$, $K = N - L + 1$ и $G^{(1)}$ отделим от $G_N^{(2)}$, то*

$$\rho_{\{a,b\}}(f) = \begin{cases} 1, & f = \omega, \\ 0, & f \neq \omega, \end{cases}$$

где a и b — номера компонент сингулярного разложения, соответствующих $G_N^{(1)}$.

Воспользуемся следствием 4.1 для усовершенствования метода Фурье. Сначала проведём идентификацию методом Фурье и будем проверять, является ли каждая из найденных пар компонент парой, соответствующей гармонике. Будем считать, что пара компонент с номерами a и b соответствует гармонике, если выполняется:

$$\max_{0 \leq k \leq M/2} \rho_{\{a,b\}}(k/M) \geq \rho_0, \quad (29)$$

где $\rho_0 \in (0, 1)$ – заранее заданное пороговое значение. Ясно, что чем больше ρ_0 , тем строже условие.

Переформулируем лемму 4.3 в применении к гармонике с частотой $\omega = 0.5$.

Лемма 4.4. *Рассмотрим гармонический ряд $G_N = (g_0, \dots, g_{N-1})$ с периодом 2:*

$$g_n = A(-1)^n, \quad 0 \leq n \leq N-1,$$

причем длина ряда N и длина окна L при котором проводится сингулярное разложение L такие, что $L/2 \in \mathbb{Z}, L \geq 2, K/2 \in \mathbb{Z}, K \geq 2, K = N - L + 1$. Тогда

$$\rho_{\{1\}}(f) = \begin{cases} \frac{1}{2}, & f = \frac{1}{2}, \\ 0, & f \neq \frac{1}{2}. \end{cases}$$

Доказательство. В силу условий леммы элементы первого сингулярного вектора $Y_1 \in \mathbb{R}^M$ разложения ряда G_N имеют вид:

$$y_n^{(1)} = A_1(-1)^n, \quad 0 \leq n \leq N-1.$$

Из того, что $\|Y_1\| = 1$, очевидно, что $A_1^2 = 1/M$. Пользуясь определением периодограммы (3), находим, что для любого $f = \frac{k}{M}, k \in \mathbb{Z}, 0 < k < \frac{M}{2}$,

$$\rho_{\{1\}}(f) = \Pi_{Y_1}^M(f) = \frac{M}{2}(A_1^2) = 1/2, \quad \text{если } f = 1/2,$$

$$\rho_{\{1\}}(f) = \Pi_{Y_1}^M(f) = 0, \quad \text{если } f \neq 1/2.$$

□

Приведём алгоритм, реализующий усовершенствованный метод Фурье для идентификации собственных троек, соответствующих гармоническим составляющим как с частотой $\omega < 0.5$, так и с частотой $\omega = 0.5$.

Алгоритм 4.5.

Исходные данные:

- матрица факторных или собственных векторов сингулярного разложения исходного временного ряда $\mathbf{Y} \in \mathbb{R}^{M \times L}$,
- множество размером Q , состоящее из
 - пар номеров собственных троек, предварительно идентифицированных с помощью метода Фурье как соответствующих гармонике с частотой $\omega < 0.5$,
 - номеров собственных троек, идентифицированных методом Фурье как соответствующих гармонике с частотой $\omega = 0.5$,
- минимальный пороговый уровень ρ_0 .

Результат работы: группа собственных троек, соответствующих гармоническим составляющим исходного временного ряда.

Алгоритм:

1. $j := 1$;
2. если j -ый элемент множества – пара чисел, goto 7;

3. иначе элемент множества – номер собственной тройки a , возможно соответствующей гармонике с частотой $\omega = 0.5$.
4. считаем $\max_{0 \leq k \leq M/2} \rho_{\{a\}}(k/M)$ для вектора Y_a по (28);
5. если $2 \max_{0 \leq k \leq M/2} \rho_{\{a\}}(k/M) \geq \rho_0$, то считаем тройку с номером a соответствующей гармонической составляющей с частотой $\omega = 0.5$;
6. goto 10;
7. пусть b – это первый номер, а c – второй номер из j -ой пары;
8. считаем $\max_{0 \leq k \leq M/2} \rho_{\{b,c\}}(k/M)$ для векторов Y_b, Y_c по (28);
9. если $\max_{0 \leq k \leq M/2} \rho_{\{b,c\}}(k/M) \geq \rho_0$, то считаем тройки с номерами b и c соответствующими гармонической составляющей с частотой $\omega < 0.5$;
10. если $j = Q$, то заканчиваем работу;
11. $j := j + 1$;
12. goto 2.

Рассмотрим теперь, чем вредит “растекание” периодограммы э.-м. гармонического ряда. Так как сингулярные вектора – нормированы и, следовательно, периодограмма – тоже, “растекание” приведёт к тому, что максимум периодограммы будет меньше, чем для немодулированной гармоники с той же частотой. Поэтому меньше будет и значение $\max_{0 \leq k \leq M/2} \rho_{\{a,b\}}(k/M)$. Попробуем сосчитать, насколько меньше будет это значение для э.-м. гармонического ряда длины M с экспоненциальным показателем α .

Рассмотрим нормированный э.-м. гармонический ряд $G_N = (g_0, \dots, g_{N-1})$ ($\|G_N\| = 1$) длины N с экспоненциальным показателем α . Если длина ряда N и длина окна, при котором проводится сингулярное разложение L такие, что $L\omega \in \mathbb{Z}, L \geq 2, K\omega \in \mathbb{Z}, K \geq 2, K = N - L + 1$, то элементы двух первых сингулярных векторов $Y_1, Y_2 \in \mathbb{R}^M$ разложения ряда G_N имеют вид:

$$y_n^{(1)} = A_1 e^{\alpha n} \cos(2\pi\omega n + \phi), \quad y_n^{(2)} = A_2 e^{\alpha n} \sin(2\pi\omega n + \phi), \quad n = 0 \dots M-1. \quad (30)$$

Примем для простоты $\phi = 0$. Воспользуемся тем, что $\|Y_1\| = \|Y_2\| = 1$. Тогда для A_1 и A_2

$$A_1(\alpha, \omega, M) = \left(\sum_{n=0}^{M-1} e^{2\alpha n} \cos^2(2\pi\omega n) \right)^{-1}, \quad A_2(\alpha, \omega, M) = \left(\sum_{n=0}^{M-1} e^{2\alpha n} \sin^2(2\pi\omega n) \right)^{-1}. \quad (31)$$

Согласно определению (28), $\rho_{\{1,2\}}(\omega) = \Pi_{Y_1}^M(\omega) + \Pi_{Y_2}^M(\omega)$. Сосчитаем теперь значения периодограмм векторов Y_1 и Y_2 в точке ω .

Обозначим за $H_0 = (h_0^{(0)}, \dots, h_{M-1}^{(0)})$ и $H_1 = (h_0^{(1)}, \dots, h_{M-1}^{(1)})$ ряды, задаваемые следующими функциями:

$$h_n^{(0)} = \cos(2\pi n\omega), \quad h_n^{(1)} = \sin(2\pi n\omega), \quad 0 \leq n \leq M-1.$$

Для вектора $Y \in \mathbb{R}^M, Y = (y_0, \dots, y_{M-1})^T$ и ряда H длины M под (Y, H) будем понимать скалярное произведение ряда H и ряда (y_0, \dots, y_{M-1}) . Из вида разложения Фурье (2) и, так как $\sum_{n=0}^{M-1} g_n \cos(2\pi k/Mn) = c_k \frac{M}{2}$, видно, что при $\omega: \omega M \in \mathbb{Z}$ значение периодограммы вектора Y_j , в ω выражается следующим образом

$$\Pi_{Y_j}^M(\omega) = \frac{1}{M} \left\{ (Y_j, H_0)^2 + (Y_j, H_1)^2 \right\}, \quad j = 1, 2.$$

Тогда для $\rho_{\{1,2\}}(\omega)$ верно:

$$\rho_{\{1,2\}}(\omega) = \frac{1}{M} \left\{ (Y_1, H_0)^2 + (Y_1, H_1)^2 + (Y_2, H_0)^2 + (Y_2, H_1)^2 \right\}. \quad (32)$$

Введём следующие обозначения:

$$\Sigma_1(\alpha, M) = \sum_{n=0}^{M-1} e^{\alpha n} \cos^2(2\pi\omega n), \quad \Sigma_2(\alpha, M) = \sum_{n=0}^{M-1} e^{\alpha n} \cos(2\pi\omega n) \sin(2\pi\omega n), \quad \Sigma_3(\alpha, M) = \Sigma_1(2\alpha, M),$$

$$\Sigma_4(\alpha, M) = \sum_{n=0}^{M-1} e^{\alpha n} \sin^2(2\pi\omega n), \quad \Sigma_5(\alpha, M) = \Sigma_4(2\alpha, M).$$

Тогда (32) можно записать в следующем виде:

$$\rho_{\{1,2\}}(\omega) = \frac{1}{M} \left\{ \frac{(\Sigma_1(\alpha, M))^2 + (\Sigma_2(\alpha, M))^2}{\Sigma_3(\alpha, M)} + \frac{(\Sigma_2(\alpha, M))^2 + (\Sigma_4(\alpha, M))^2}{\Sigma_5(\alpha, M)} \right\}.$$

Обозначим за $I_j(\alpha, M)$, $j = 1, \dots, 5$ следующие интегралы:

$$I_1(\alpha, M) = \int_0^M e^{\alpha t} \cos^2(2\pi t\omega) dt, \quad I_2(\alpha, M) = \int_0^M e^{\alpha t} \cos(2\pi t\omega) \sin(2\pi t\omega) dt, \quad I_3(\alpha, M) = I_1(2\alpha, M),$$

$$I_4(\alpha, M) = \int_0^M e^{\alpha t} \sin^2(2\pi t\omega) dt, \quad I_5(\alpha, M) = I_4(2\alpha, M).$$

Введём

$$I(\alpha, M) = \frac{1}{M} \left\{ \frac{(I_1(\alpha, M))^2 + (I_2(\alpha, M))^2}{I_3(\alpha, M)} + \frac{(I_2(\alpha, M))^2 + (I_4(\alpha, M))^2}{I_5(\alpha, M)} \right\} \quad (33)$$

Предложение 4.1. Пусть $\omega \in (0, 0.5)$, $M \in \mathbb{Z}$: $M\omega \in \mathbb{Z}$, $\alpha \in \mathbb{R}$. Если $M \rightarrow \infty$, $\alpha \rightarrow 0$: $M\alpha \rightarrow \gamma \in \mathbb{R}$, тогда

$$I(\alpha, M) \rightarrow \frac{2(e^\gamma - 1)}{\gamma(e^\gamma + 1)}. \quad (34)$$

Доказательство. Так как

$$\int e^{\beta t} \cos^2(ct) dt = \frac{1}{2} \frac{e^{\beta t}}{(\beta^2 + 4c^2)\beta} (\beta^2 \cos(2ct) + 2\beta c \sin(2ct) + \beta^2 + 4c^2), \quad (35)$$

то получаем для $c = 2\pi\omega$:

$$I_1(\alpha, \omega, M) = \frac{\alpha^2 + 2c^2}{\alpha(\alpha^2 + 4c^2)} (e^{\alpha M} - 1), \quad I_3(\alpha, M) = \frac{2\alpha^2 + c^2}{4\alpha(\alpha^2 + c^2)} (e^{2\alpha M} - 1),$$

$$I_4(\alpha, M) = \frac{2c^2}{\alpha(\alpha^2 + 4c^2)} (e^{\alpha M} - 1), \quad I_5(\alpha, M) = \frac{c^2}{4\alpha(\alpha^2 + c^2)} (e^{2\alpha M} - 1).$$

Из того, что $I_2(\alpha, M)$ и $I_4(\alpha, M)$ связаны следующим соотношением:

$$I_2(\alpha, M) = -\frac{\alpha}{2c} I_4(\alpha, M),$$

находим, что

$$I_3(\alpha, M) = \frac{-c^2}{c(\alpha^2 + 4c^2)} (e^{\alpha M} - 1).$$

Подставив полученные значения для $I_j(\alpha, M)$, $j = 1, \dots, 5$, в (33) и устремив $M \rightarrow \infty$, $\alpha \rightarrow 0$, $M\alpha \rightarrow \gamma$, получаем искомый предел:

$$I(\alpha, M) \rightarrow \frac{2(e^\gamma - 1)}{\gamma(e^\gamma + 1)}.$$

□

Сформулируем усовершенствованный метод Фурье для идентификации э.-м. гармонических составляющих ряда. Пусть

$$\tilde{\rho}(\alpha, M) = \frac{2}{\alpha M} \frac{(e^{\alpha M} - 1)}{(e^{\alpha M} + 1)}.$$

Будем считать, что собственные тройки с номерами a, b соответствуют э.-м. гармонической составляющей, если

$$\frac{\max_{0 \leq k \leq M/2} \rho_{\{a,b\}}(k/M)}{\tilde{\rho}(\alpha, M)} \geq \rho_0, \quad \rho_0 \in (0, 1). \quad (36)$$

Оценка экспоненциального показателя в данной работе не рассматривается. Эта задача может быть решена как с использованием свойств метода “Гусеница”, так и различными сторонними методами.

Построим алгоритм для идентификации собственных троек, соответствующих э.-м. гармонике с частотой ω и экспоненциальным показателем α на основании вида собственных или факторных векторов.

Алгоритм 4.6.

Исходные данные:

- матрица факторных или собственных векторов сингулярного разложения исходного временного ряда $\mathbf{Y} \in \mathbb{R}^{M \times L}$,
- множество размером Q пар номеров собственных троек, предварительно идентифицированных (как соответствующих э.-м. гармонике) с помощью метода Фурье,
- минимальный пороговый уровень ρ_0 ;

Результат работы: группа собственных троек, соответствующих гармоническим составляющим исходного временного ряда.

Алгоритм:

1. $j := 1$;
2. пусть a это первый номер, а b – второй номер из j -ой пары;
3. оценим по векторам Y_a, Y_b экспоненциальный показатель э.-м. гармонической составляющей ряда, обозначим за $\tilde{\alpha}$ полученную оценку;
4. считаем $\max_{0 \leq k \leq M/2} \rho_{\{a,b\}}(k/M)$ для векторов Y_a, Y_b согласно (28);
5. пусть $\tilde{\rho} = \frac{2}{\tilde{\alpha}M} \frac{e^{\tilde{\alpha}M} - 1}{e^{\tilde{\alpha}M} + 1}$;
6. если $\max_{0 \leq k \leq M/2} \rho_{\{a,b\}}(k/M) \geq \tilde{\rho}\rho_0$, то считаем компоненты с номерами a и b соответствующими э.-м. гармонической составляющей;
7. если $j = Q$, то заканчиваем работу;
8. $j := j + 1$;
9. goto 2.

5 Описание методов оценки частоты э.-м. гармонического ряда

Как было сказано в разделе 3.1, решение некоторых задач требует оценки частоты э.-м. гармоники. В данном разделе приведём методы оценки частоты на основе соответствующих собственных троек. Предположим, что частота э.-м. гармоники $\omega < 0.5$, то есть ей соответствуют две собственных тройки.

5.1 Метод осреднения полярного угла

Этот метод основан на идее, используемой при визуальной идентификации э.-м. гармонических составляющих с помощью парных диаграмм сингулярных векторов, и описанной в разделе 3.1. Пусть $\mathbf{Y} \in \mathbb{R}^{M \times L}$ – матрица собственных или факторных векторов разложения. Рассмотрим сингулярные вектора $Y_a = (y_1^{(a)}, \dots, y_M^{(a)})^T$, $Y_b = (y_1^{(b)}, \dots, y_M^{(b)})^T$, соответствующие э.-м. гармонике. Построим по ним набор векторов A_n , $1 \leq n \leq M$, следующим образом:

$$A_n = (y_n^{(1)}, y_n^{(2)})^T. \quad (37)$$

Рассмотрим случай оценки частоты гармонике. При условии точной её отделимости (см. раздел 2.3) известен асимптотический вид векторов Y_1, Y_2 и A_n имеет следующий асимптотический вид:

$$\tilde{A}_n = (B \cos(2\pi\omega n + \phi), B \sin(2\pi\omega n + \phi))^T.$$

Пусть $\tilde{\Delta}_n$, $1 \leq n \leq M - 1$, – угол между \tilde{A}_n и \tilde{A}_{n+1} , а Δ_n , $1 \leq n \leq M - 1$, – угол между A_n и A_{n+1} . Тогда $\omega = \tilde{\Delta}_n/2\pi$ для любого n .

Так как нам нужно найти оценку частоты на основе сингулярных векторов в неасимптотическом случае и при условии приближённой разделимости, то будем искать оценку частоты $\hat{\omega}$ через среднее значение для Δ_n , $1 \leq n \leq M - 1$:

$$\bar{\Delta} = \frac{1}{M-1} \sum_{n=1}^{M-1} \Delta_n, \quad \hat{\omega} = \frac{\bar{\Delta}}{2\pi}. \quad (38)$$

Алгоритм 5.1.

Исходные данные:

- сингулярные вектора $Y_a, Y_b \in \mathbb{R}^M$, соответствующие э.-м. гармонике;

Результат работы: оценка частоты $\hat{\omega}$.

Алгоритм:

1. Построим последовательность векторов A_n , $1 \leq n \leq M$, на основе векторов Y_a, Y_b по (37);
2. Определим направление “вращения” вокруг нуля при перемещении от вектора A_n к A_{n+1} , $1 \leq n \leq M - 1$;
3. Посчитаем Δ_n , $1 \leq n \leq M - 1$, – угол между векторами A_n и A_{n+1} ;
4. Посчитаем для Δ_n , $1 \leq n \leq M - 1$, среднее значение $\bar{\Delta}$;
5. Выразим $\hat{\omega}$ через $\bar{\Delta}$ по (38).

5.2 Метод характеристических корней

Теорема 2.1 из раздела 2.4 дает связь между корнями характеристического полинома и аддитивными составляющими исходного ряда, и помимо этого, из неё видно, что два сопряженных корня $x_1 = e^{\alpha+i2\pi\omega}$, $x_2 = e^{\alpha-i2\pi\omega}$ соответствуют э.-м. гармонике с частотой ω и экспоненциальным показателем α . Таким образом для того, чтобы найти частоту гармонике, достаточно найти соответствующие ей корни x_1, x_2 характеристического полинома и выразить её частоту через x_1 или x_2 :

$$\hat{\omega} = \frac{1}{2\pi} \arccos(\operatorname{Re}(x_1)/|x_1|). \quad (39)$$

В соответствии с теоремой 2.2 можно построить следующий алгоритм оценки периода выделенной э.-м. гармонической составляющей с использованием корней характеристического полинома. Будем в качестве ортонормированного базиса пространства \mathcal{L}_2 , соответствующего э.-м. гармонике брать собственные вектора U_a, U_b соответствующие э.-м. гармонической составляющей.

Алгоритм 5.2.

Исходные данные:

- собственные вектора U_a, U_b .

Результат работы: оценка частоты $\hat{\omega}$.

Алгоритм:

1. Найдём ЛРФ, посчитав её коэффициенты по (19), используя в качестве базиса P_1, P_2 набор U_1, U_2 ;
2. Посчитаем корни характеристического полинома этой ЛРФ;
3. Найдём среди корней два корня x_1, x_2 с максимальными модулями из таких корней, что $\text{Im } x_i \neq 0, \text{Re } x_1 = \text{Re } x_2$;
4. Через x_1, x_2 выразим $\hat{\omega}$ по (39);

5.3 Периодограммный метод

Данный метод опирается на вид периодограммы э.-м. гармонической последовательности. Будем применять его к ряду G_N , восстановленному по собственным тройкам, соответствующим э.-м. гармонике и оценивать её частоту по аргументу максимума θ_{ab} периодограммы ряда G_N . Тогда оценка частоты $\hat{\omega} = \theta_{ab}$.

Так как периодограмма ряда длины N задана только на решётке $\{k/N\}_{k=0}^{\lfloor N/2 \rfloor}$, то при $N\omega \notin \mathbb{Z}$ оценка $\hat{\omega}$ будет найдено заведомо с ошибкой, сравнимой по порядку с $1/N$.

Приведём алгоритм оценки частоты э.-м. гармонической составляющей ряда с помощью периодограммы.

Алгоритм 5.3.

Исходные данные:

- собственные тройки с номерами a и b , соответствующие э.-м. гармонической составляющей.

Результат работы: оценка частоты $\hat{\omega}$.

Алгоритм:

1. Восстановим по собственным тройкам с номерами a и b ряд G_N ;
2. Найдём аргумент максимума θ_{ab} периодограммы ряда G_N ;
3. Возьмём в качестве оценки частоты $\hat{\omega} = \theta_{ab}$.

6 Выбор параметров методов идентификации

В данном разделе для методов идентификации даётся более подробная, чем при их описании в разделе 4, интерпретация параметров, а также приведены соображения по выбору параметров при решении различных задач.

6.1 Выбор параметров методов идентификации компонент, соответствующих трендовым составляющим

6.1.1 Метод Кендалла

Параметром метода Кендалла является пороговое значение $\alpha_0 \in (0, 1)$. Сравнимое с ним значение $\alpha(G_M)$, вычисляемое по ряду G_M , хорошо интерпретируется в случае, когда G_M – монотонный ряд, так как в этом случае $\alpha(G_M) = 0$. В остальных случаях можно сказать, что если ряд G_M содержит монотонный тренд, то $\alpha(G_M)$ не слишком велико.

Стандартным пороговым значением для этого метода можно считать $\alpha_0 = 0.05$.

6.1.2 Методы нулей и низких частот

Покажем согласованность параметра метода нулей Z_0 и параметра метода низких частот ω_0 . Так как мы рассматриваем тренд как медленно меняющуюся составляющую ряда, его (и соответствующие тренду сингулярные вектора) можно аппроксимировать суммой гармоник с низкими частотами (включая нулевую частоту). Поэтому продемонстрируем связь параметров методов на примере гармонического ряда G_M с частотой ω . Если к нему применить метод нулей с параметром Z_0 , то он будет считаться трендом, если для ω выполняется $[2M\omega] \leq Z_0$. Мы получили, что методы согласованы при $\omega_0 = Z_0/(2M)$.

Таким образом, оба метода по существу сводятся к заданию максимально допустимой частоты ω_0 . Однако в терминах частот может быть сложно работать. Предложим два других, эквивалентных, способа задания частоты ω_0 . Введем $T_0 = 1/\omega_0$ — период гармонического ряда с частотой ω_0 , $\nu_0 = N\omega_0$ — число периодов такого ряда, укладывающихся в длину исходного ряда. Тогда вместо ω_0 можно задать:

1. T_0 - минимальный период. Такое задание может быть удобно, если, например, в ежемесячных данных мы хотим выделить составляющие с периодом, большим, чем 12, считая их относящимися к тренду. Тогда можно установить T_0 , большим, чем, скажем, 14 (с небольшим запасом).
2. ν_0 - максимальное число периодов, укладывающихся в длину исходного ряда. Естественно, это число не обязательно должно быть целым. Как правило, имеет смысл выбирать ν_0 не больше двух-трёх.

6.2 Выбор параметров методов идентификации компонент, соответствующих э.-м. гармоническим составляющим

6.2.1 Метод Фурье и усовершенствованный метод Фурье

Выбор порогового значения метода Фурье не нуждается в особых комментариях, следует лишь добавить, что для гармонического ряда длины M с частотой ω в случае, когда $M\omega \notin \mathbb{Z}$, максимальное значение периодограммы не обязательно находится в ближайшей к ω точке вида $k/M, k \in \mathbb{Z}$, поэтому в качестве выбираемого по умолчанию порогового значения стоит взять число больше 0, например, 1.

Вычисляемое в усовершенствованном методе Фурье значение $\max_{0 \leq k \leq M/2} \rho_{\{a,b\}}(k/M)$ не интерпретируется простым образом. Ясно, что чем пороговое значение ближе к 1, тем метод строже. В качестве значения по умолчанию выбрано 0.65.

7 Сравнение методов идентификации

Сравним методы, применяя их для выделения описанных в разделе 3 составляющих. Будем считать, что имеет место приближённая делимость выделяемой составляющей (тренда или периодики) с остатком. Поэтому сингулярные вектора, соответствующие выделяемой составляющей, в общем разложении будут отличаться от сингулярных векторов, полученных в результате непосредственного разложения рассмотренной составляющей. Искажения векторов могут быть как регулярными (например, в случае приближённой делимости выделяемой составляющей с гармоникой, искажения будут гармонического вида), так и нерегулярными. Например, если остаток является высокочастотной нерегулярной составляющей, то сингулярные вектора будут выглядеть зашумлёнными. Поэтому будем сравнивать методы идентификации также в применении к искажённым сингулярным векторам.

Сравнивать методы идентификации будем согласно двум подходам. Сформулируем данные подходы для методов идентификации трендовых компонент, аналогично формулируются подходы и для методов идентификации периодических составляющих. Будем называть метод *хорошим согласно первому подходу*, если он идентифицирует все компоненты, сингулярные вектора которых соответствуют тренду ряда. *Согласно второму подходу метод будем называть хорошим*, если он не идентифицирует компоненты, сингулярные вектора для которых не соответствуют тренду.

Если компонента не была идентифицирована методом идентификации какой-то составляющей, хорошим по первому подходу, значит она не соответствует этой составляющей. Если же она была идентифицирована методом, хорошим по второму подходу, значит она точно соответствует этой составляющей.

7.1 Методы идентификации тренда

Приведём таблицу 2, изображающую характеристики качества работы методов, применяя их к собственным тройкам, соответствующим различным составляющим тренда. Элемент таблицы описывает, идентифицирует ли метод собственные тройки, соответствующие составляющей. Знак “+” в таблице означает, что метод идентифицирует нужные собственные тройки, “±” – что на результат идентификации влияет степень зашумлённости сингулярных векторов, “undef” означает, что результат применения метода непредсказуем. Хороший по первому подходу метод

Таблица 2: Сводная таблица качества работы методов выделения тренда

Типы составляющих	монот. тренд	конст. тренд	медл. осцилл. тренд	гарм-ка	э.-м. гарм-ка
Кендалл	+	<i>undef</i>	<i>undef</i>	<i>undef</i>	<i>undef</i>
Мет.нулей	±	±	±	–	–
Мет.НЧ	+	+	+	–	–

должен иметь знаки “+” во всех столбцах, соответствующих тренду. Хороший согласно второму подходу должен иметь знаки “–” во всех столбцах, не соответствующих тренду. Из таблицы видно, что с точки зрения первого подхода метод Кендалла – это наихудший метод, а метод нулей хуже метода низких частот. С точки зрения второго подхода метод Кендалла снова наихудший, а метод нулей и метод низких частот одинаково хороши.

Рассмотрим теперь подробно свойства методов идентификации тренда. Идеи методов Кендалла и нулей были взяты из [3], тогда как метод низких частот был придуман, как аналог метода нулей, но с тем, чтобы устранить его недостаток при работе с зашумлёнными сингулярными векторами.

Метод Кендалла Метод Кендалла идеален при выделении монотонной составляющей тренда. Он идентифицирует только собственные тройки, сингулярные вектора которых содержат монотонную составляющую. Собственные тройки, соответствующие монотонному тренду, даже при искажении сингулярных векторов, будут им идентифицированы.

Результат применения метода Кендалла при выделении константного тренда непредсказуем, так как в реальных исследованиях константному тренду соответствуют сингулярные вектора по значениям близкие к константе, но слегка отклоняющиеся от неё произвольным образом. Эти отклонения появляются, если тренд приближённо отделим от остатка. Так как метод Кендалла зависит не от масштаба, а только от формы ряда, к которому применяется, из-за случайности отклонений его результат будет непредсказуемым.

При выделении медленно-осциллирующего тренда результат метода Кендалла также непредсказуем, так как сингулярные вектора, соответствующие такому тренду, могут не содержать монотонную составляющую. Объясним это на примере гармонического ряда с чётным периодом $T = N$. Рассмотрим сингулярный вектор, $L = T/2$, его n -ый элемент имеет вид $A \cos(2\pi n/N + \phi)$, $1 \leq n \leq L$, но значение ϕ неизвестно. В случае, если $\phi = \pi/2$, сингулярный вектор является монотонным монотонному и метод Кендалла идентифицирует компоненту, которой такой вектор соответствует, а если $\phi = 0$, то метод Кендалла никогда такой сингулярный вектор не идентифицирует. Так как значение ϕ заранее неизвестно, результат метода Кендалла непредсказуем. Такая же неопределённость проявляется при применении метода Кендалла к сингулярным векторам, соответствующим гармоникам с меньшими периодами.

Несмотря на неопределённость результатов, метод Кендалла лучше остальных методов, если необходимо выделить только монотонную составляющую тренда или при выделении тренда известно, что он состоит только из монотонной составляющей, так как его результат мало зависит от искажений сингулярных векторов.

Метод нулей В отличие от метода Кендалла метод нулей позволяет выделять не обязательно монотонные составляющие тренда. Но у него есть недостаток, его результат зависит от искажений сингулярных векторов, особенно от степени их зашумлённости, так как искажение вектора может увеличить количество его “нулей”. В целом можно сказать, что метод нулей не очень хорош, зато

он универсален, так как не зависит от формы выделяемой составляющей, а опирается только на медленную изменяемость тренда.

Метод низких частот Так же, как и метод нулей, метод низких частот можно при рассмотрении аппроксимации ряда гармониками свести к выбору наибольшей возможной частоты для гармоник, приближающих тренд (см. раздел 6.1.2), поэтому свойства этих методов очень похожи при работе с неискажёнными сингулярными векторами. Но метод низких частот, в отличие от метода нулей, позволяет с помощью выбора верхнего порогового уровня C_0 вклада в вид векторов гармоник с высокими частотами идентифицировать компоненты, сингулярные вектора которых зашумлены. Продемонстрируем работу методов идентификации тренда на примере исследования реального временного ряда.

Пример 7.1. Ряд *traffat*, выделение тренда.

Рассмотрим ряд *traffat*, содержащий ежемесячные данные о дорожных авариях с 1960 по 1974 годы в Онтарио. Ряд изображён на рис. 3, длина ряда равна 180. Проведём сингулярное разложение

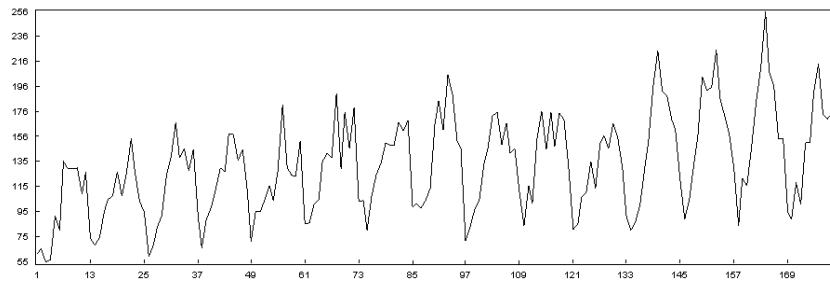


Рис. 3: *traffat*: исходный ряд

с длиной окна $L = 60$. Рассмотрим первые двадцать собственных троек (будем использовать аббревиатуру ЕТ, *eigentriples*) и применим к первым двадцати собственным векторам методы идентификации компонент тренда со следующими параметрами. Для метода Кендалла возьмём параметр по умолчанию $\alpha_0 = 0.05$, параметры для методов нулей и низких частот выберем, установив $\nu_0 = 2$ (см. раздел 6.1.2): $Z_0 = 2$, $\omega_0 = 0.02$. В таблице 3 приводятся результаты работы

Таблица 3: *traffat*: результаты идентификации тренда

ЕТ	1	2	3	4	5	6	7	8	9	10
Кендалл	0.00	0.24	0.93	0.00	0.39	0.90	0.79	0.49	0.96	0.86
Мет.нулей	0(0)	1(9)	1(10)	0(1)	0(2)	1(20)	1(20)	1(59)	1(12)	1(12)
Мет.НЧ	0.00	1.00	1.00	0.10	0.20	1.00	0.95	1.00	1.00	0.93
ЕТ	11	12	13	14	15	16	17	18	19	20
Кендалл	0.85	0.91	0.78	0.92	0.83	0.62	0.99	0.34	0.69	0.93
Мет.нулей	1(29)	1(30)	1(49)	1(49)	1(25)	1(15)	1(41)	1(33)	1(23)	1(24)
Мет.НЧ	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00

методов: $\alpha(U_i)$ для метода Кендалла, $1 - \mathcal{C}(U_i)$ для метода низких частот. Для метода нулей приводится “0”, если компонента была идентифицирована, “1” в противном случае и в скобках указывается количество нулей собственного вектора $Z(U_i)$. Выделенные жирным шрифтом результаты соответствуют идентифицированным собственным тройкам. На рис. 4 изображены первые десять собственных векторов. Из таблицы 3 видно, что ЕТ1,4 были идентифицированы всеми методами, тогда как ЕТ5 не была идентифицирована методом Кендалла. Из изображений собственных векторов U_4, U_5 видно, что метод Кендалла не идентифицировал ЕТ5, так как вектора U_4, U_5 соответствуют медленно-осциллирующей составляющей тренда, а результат метода Кендалла при работе с такими рядами непредсказуем.

В завершение приведём ряд, восстановленный по группе собственных троек ЕТ1,4,5, изображённый на фоне исходного ряда (рис. 5). Если в программе, описанной в разделе 8, решать задачу

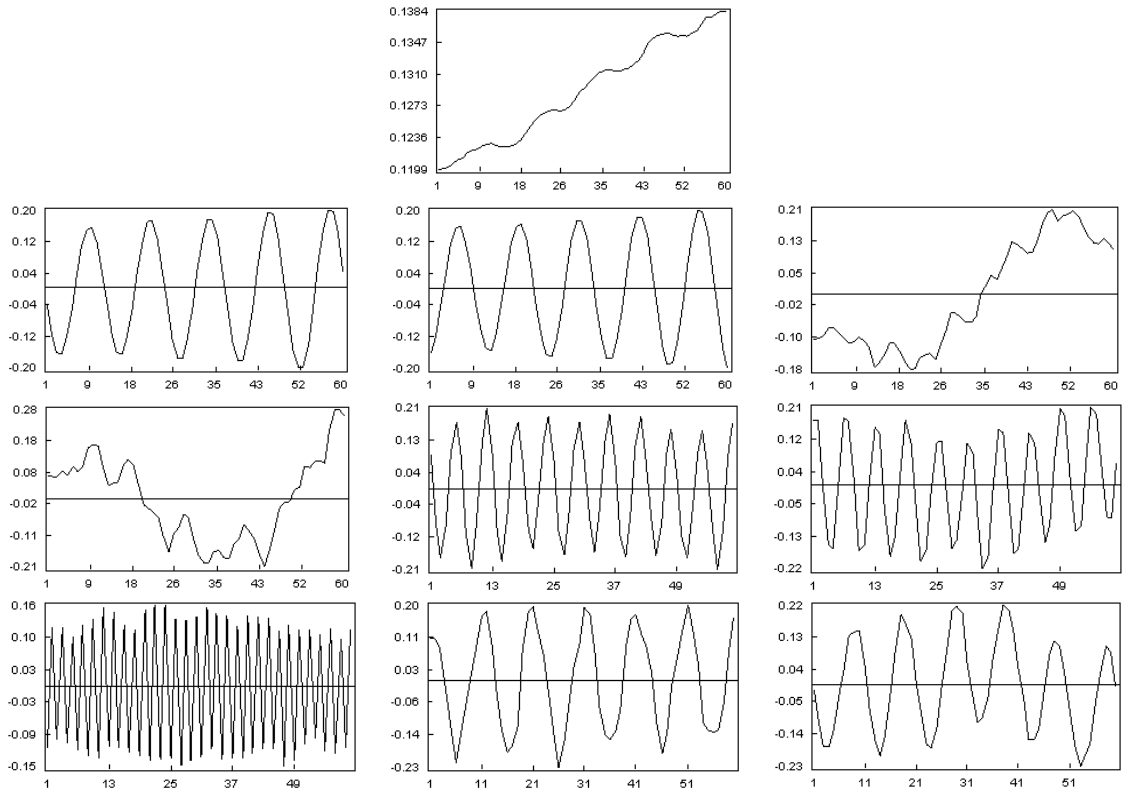


Рис. 4: traffat: собственные векторов первых десяти ET (слева направо)

выделения тренда, указав следующие параметры: $L = 60$, $\alpha_0 = 0.05$, $Z_0 = 2$, $\omega_0 = 0.02$, $C_0 = 0.5$ и проводить идентификацию на основании вида собственных векторов, то результат работы программы будет совпадать с результатом, приведённым на рис. 5.

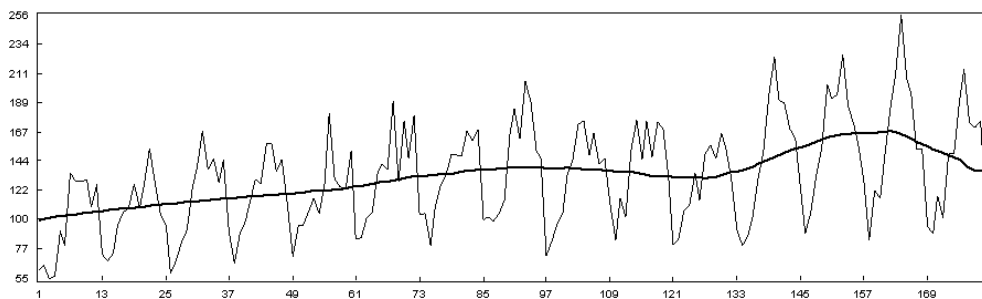


Рис. 5: traffat: исходный и восстановленный по ET с номерами 1,4,5 ряды

Приведём теперь несколько примеров, демонстрирующих недостатки методов.

Пример 7.2. *Особенность метода Кендалла при выделении константной составляющей тренда.* Возьмём ряд egg, отражающий ежемесячное среднее число яиц на одну несушку в США с 1938 по 1940 годы. Проведём сингулярное разложение с $L = 18$. Первый собственный вектор, изображённый на рис. 6, соответствует константной составляющей тренда, но он искажён из-за того, что не выполнены условия точной разделимости с гармоникой ряда. Это видно на рис. 7, на котором изображён тот же вектор, но в более крупном масштабе. По форме вектор похож на такой участок гармонического ряда, что метод Кендалла ошибочно не идентифицирует ET1 ($\alpha(U_1) \approx 0.57$). Возможно, если рассмотреть другое сингулярное разложения, тогда этот сингулярный вектор будет искажён таким образом, что метод Кендалла идентифицирует ET1, но предугадать это нельзя.

Пример 7.3. *Особенность метода нулей при работе с зашумлёнными сингулярными векторам.*

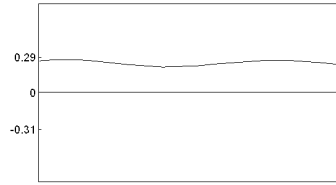


Рис. 6: egg: U_1

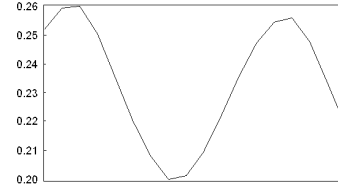


Рис. 7: egg: U_1 , другой масштаб

Второй пример демонстрирует недостаток метода нулей при работе с зашумлёнными собственными векторами. Рассмотрим ряд `airpass` длины $N = 144$, содержащий ежемесячные данные о количестве пассажиров на международных авиалиниях. Проведём сингулярное разложение с $L = 74$. Применим методы нулей и низких частот к собственному вектору ET12. Параметры для них выберем, установив $\nu_0 = 2$ (см. раздел 6.1.2): $\omega_0 = 0.03$, $Z_0 = 2$, в качестве C_0 возьмём стандартное $C_0 = 0.5$. Метод низких частот идентифицировал ET12 ($1 - C(U_{12}) \approx 0.47 \leq 0.5$), в отличие от метода нулей ($Z(U_{12}) = 17 > 2$). Собственный вектор показан на рис. 8 и видно, что именно

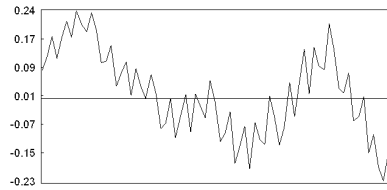


Рис. 8: airpass: U_{12}

из-за зашумления график вектора U_{12} имеет большое количество пересечений с нулевой осью.

7.2 Методы идентификации периодических составляющих

Таблица 4: Сводная таблица качества работы методов выделения гармоник и э.-м. гармоник

Типы составляющих	гарм-ка	э.-м. гарм-ка	шум
Мет.Фурье	+	+	<i>undef</i>
Усов.мет.Фурье	+	\pm	—

Приведём сравнительные характеристики методов в таблице 4, аналогичной таблице 2. Знаки “+”, “*undef*” используются в том же смысле, что и в таблице 2. Под шумом понимается нерегулярная высокочастотная составляющая.

Объясним, что означает знак “ \pm ”, характеризующий качество идентификации э.-м. гармоники с помощью усовершенствованного метода Фурье. При идентификации э.-м. гармоники с помощью усовершенствованного метода Фурье необходимо оценить её экспоненциальный показатель. Ошибка, появившаяся при оценке, может привести к тому, что для э.-м. гармоники будет неправильно посчитано $\tilde{\rho}$ и она не будет идентифицирована.

Продemonстрируем работу методов на ряде `traffat`, выделение тренда которого было проведено в разделе 7.1, а также приведём пример, демонстрирующий неопределённость результата простого метода Фурье при работе с шумом.

Пример 7.4. *Ряд `traffat`, выделение сезонной составляющей.*

Проведём разложение ряда `traffat` с тем же $L = 60$ и применим к собственным векторам простой и усовершенствованный методы Фурье с параметрами по умолчанию $s_0 = 1$, $\rho_0 = 0.65$. В таблице 5 приводятся результаты работы методов (для простого метода Фурье вычисляемые согласно (27), для усовершенствованного – согласно (29)). Выделенные жирным шрифтом результаты соответствуют идентифицированным компонентам. Одномерные графики первых десяти собственных

Таблица 5: traffat: результаты идентификации периодик

ЕТ	1,2	2,3	3,4	4,5	5,6	6,7	7,8	8	8,9	9,10
Мет.Фурье	5	0	4	0	9	0	19	0	23	0
Усов.мет.Фурье	.	0.99	.	0.85	.	0.96	.	0.98	.	0.91
ЕТ	10,11	11,12	12,13	13,14	14,15	15,16	16,17	17,18	18,19	19,20
Мет.Фурье	8	0	10	0	3	19	19	0	9	0
Усов.мет.Фурье	.	0.93	.	0.85	.	.	.	0.61	.	0.82

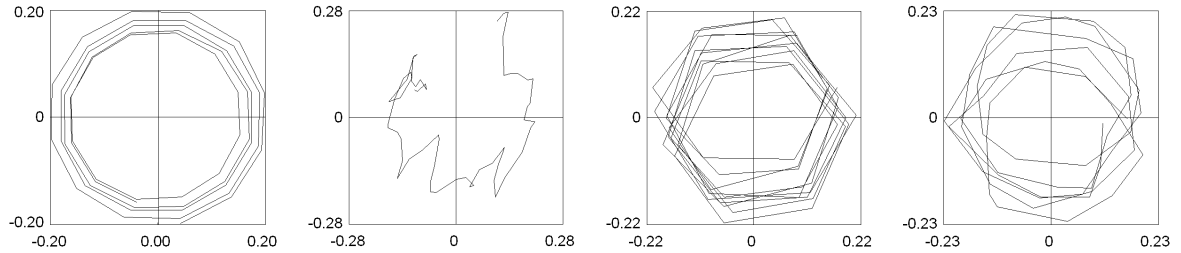


Рис. 9: traffat: парные диаграммы для собственных векторов (слева направо) ЕТ2,3 ЕТ4,5, ЕТ6,7, ЕТ9,10

векторов были приведены на рис. 4. На рис. 9 изображены парные диаграммы для идентифицированных пар из первых десяти ЕТ, построенные, как это было описано в разделе 3.1.

ЕТ4,5 идентифицированы, как периодические, но помним, что они были также идентифицированы с помощью методов нулей и низких частот. Задав $\nu_0 = 2$, мы установили считать составляющие, которым соответствуют ЕТ4,5, медленно-осциллирующими составляющими тренда, поэтому будем считать ЕТ4,5 соответствующими тренду. На парной диаграмме для них (рис. 9, вторая картинка слева) регулярности не видно, так как период слишком велик ($\nu_0 = 2$ соответствует тому, что период $T > N/2$) и вектора сильно искажены. Собственная тройка ЕТ8 была идентифицирована, как соответствующая гармонике с периодом 2 и для неё в таблице приводится удвоенный результат (29), как это было описано в алгоритме 4.5.

Так как задача состоит в выделении сезонной составляющей (то есть периодики с периодом $T = 12$), то необходимо оценить периоды выделенных гармоник. Для оценки периодов идентифицированных гармоник применим методы, описанные в разделе 5, результаты которых приведены в табл. 6. Выделенные жирным шрифтом результаты соответствуют сезонной составляющей, так как эти периоды являются делителями $T = 12$ (с точностью до погрешности, вносимой искажениями собственных векторов). В завершение приведём на рис. 10 результат восстановления сезонной

Таблица 6: traffat: результаты оценки периода гармоник и э.-м. гармоник

ЕТ	2,3	6,7	8	9,10	11,12	13,14	19,20
Мет.поляр.угла	11.97	5.97	.	9.86	4.00	2.40	5.02
Мет.хар.корней	11.95	5.95	2.00	9.65	3.98	2.40	5.03
Периодогр.мет.	12.00	6.00	2.00	10.00	4.00	2.40	5.14

составляющей ряда по группе собственных троек ЕТ2,3,6-8,11-14. Это результат, который выдаст программа, описанная в разделе 8 при выделении периодики с периодом, оценка которого должна лежать в $[11, 13]$ со следующими параметрами: $L = 60$, $s_0 = 1$, $\rho_0 = 0.65$, при проведении идентификации на основании вида собственных векторов.

Пример 7.5. Приведём теперь пример, демонстрирующий преимущество усовершенствованного метода Фурье перед простым. Возьмём начальный участок wheat100 ряда wheat длины 100 ($N = 100$), полная длина которого составляет 733, содержащий ежегодные данные о ценах на пшеницу

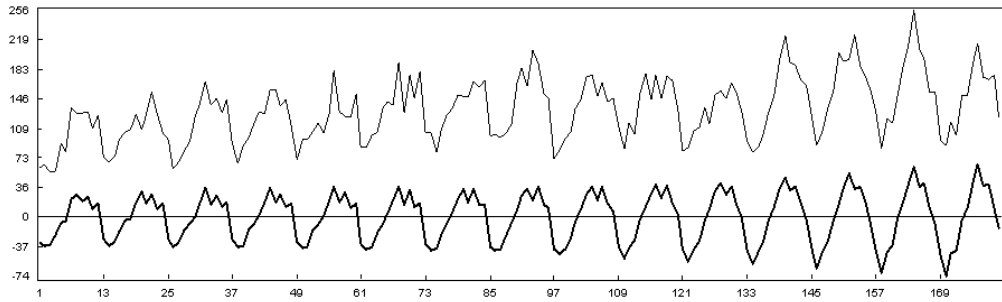


Рис. 10: traffat: исходный и восстановленный по ЕТ с номерами 2,3,6-8,11-14 ряды

с 1264 по 1996 годы, и проведём сингулярное разложение с длиной окна $L = 50$. Рассмотрим 6-ой и 7-ой собственные вектора, изображённые на рис. 11,12. Аргументы пиков периодограмм этих

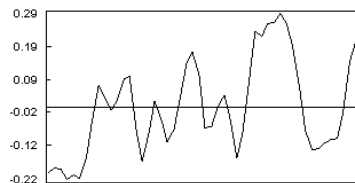


Рис. 11: wheat100: U_6

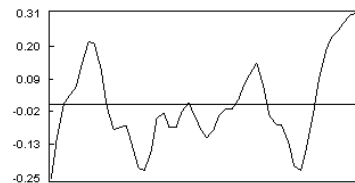


Рис. 12: wheat100: U_7

векторов совпадают (они равны $4/50$), поэтому простой метод Фурье ошибочно идентифицирует их. Но периодограммы этих векторов, изображённые на рис. 13,14, сильно отличаются от периодограммы

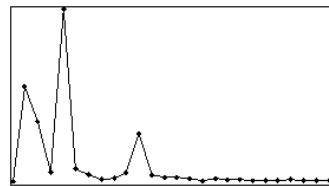


Рис. 13: wheat100: $\Pi_{U_6}^{50}$

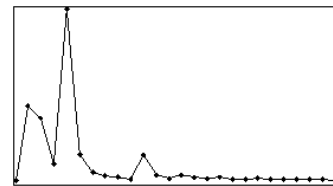


Рис. 14: wheat100: $\Pi_{U_7}^{50}$

гармоники и усовершенствованный метод Фурье правильно не идентифицирует ЕТ6,7 ($\max_{0 \leq k \leq 50/2} \rho_{\{6,7\}}(k/50) = 0.3891$).

8 Описание программы

Программа WinSsaAuto позволяет решать все задачи из поставленных в разделе 3 с помощью описанных в разделе 4 методов автоматической идентификации и приведённых в разделе 5 методов оценки периода.

Программа представляет возможность провести три типа исследования ряда. Для любого исследования вначале проводится сингулярное разложение ряда с заданной длиной окна L . Первое исследование – *автоматическое выделение тренда*, во время него производится поиск соответствующих тренду собственных троек с помощью одного из методов, Кендалла, нулей или низких частот, и восстановление по найденным компонентам. Второе исследование – *автоматическое выделение периодики* (в том числе и обобщённой) с помощью простого и усовершенствованного методов Фурье. В рамках этого же исследования можно решить задачу выделения периодики (может быть, обобщённой) с заданным периодом или гармоник с периодом из заданного интервала с помощью методов оценки частоты гармоник и э.-м. гармоник, описанных в разделе 5. Третье исследование выдаёт результаты работы и методов идентификации тренда и методов идентификации периодик (обобщённых периодик) для всех собственных троек разложения подобно тому, как это было приведено в таблицах 3, 5.

8.1 Работа с программой

Управление программой осуществляется с помощью кнопок инструментальной панели, расположенной в верхней части окна. Опишем их по порядку (слева направо):

1. открытие файла,
2. сохранение ряда в файл,
3. запуск одного из исследований,
4. просмотр сингулярных векторов или собственных чисел разложения,
5. вывод дополнительной информации о том, как прошло исследование,
6. для результата исследования периодик: разбиение восстановленного ряда на гармоники с оценкой их периодов,
7. сглаживание/детализация ряда, являющегося результатом исследования,
8. просмотр двух рядов в разных масштабах (удобно для результата исследования периодик),
9. закрытие всех открытых окон (эта функция доступна при хотя бы при одном открытом окне, далее не будем её упоминать).

Кнопка находится в активном состоянии (её можно нажать), если применение её функции в данный момент осмысленно.

Работу с программой необходимо начинать с открытия файла ряда. Файл должен представлять собой одну или две колонки чисел с пробельными символами между ними.

После открытия файла нужно вызвать форму запуска исследования с помощью кнопки 3 инструментальной панели и выбрать вид проводимого исследования. В форме "Start investigation" можно сделать активными кнопки "L" и "Amount of components to investigate" и указать значения соответствующих параметров (длина окна и число рассматриваемых собственных троек). Значение L можно будет поменять позднее, уже при проведении конкретного исследования.

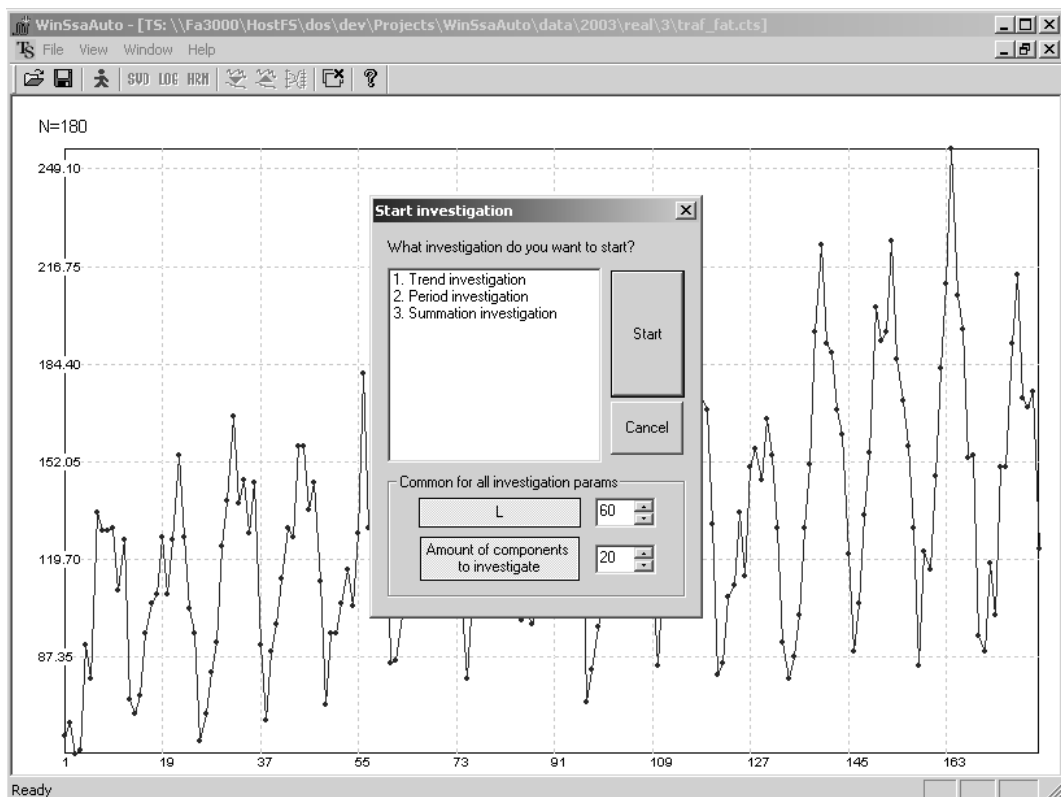


Рис. 15: WinSsaAuto: выбор исследования

Программа позволяет работать сразу с несколькими рядами. Можно загружать новые ряды, тогда они будут открыты в новых окнах. Работа производится с активным окном. Изменение активного окна меняет (если это нужно) статус кнопок в инструментальной панели. Для того, чтобы после проведённого исследования ряда провести над ним ещё одно, можно воспользоваться окном результата исследования и сделав его активным, вызвать форму запуска исследования. В таком случае будет проведено исследование исходного ряда, а не ряда, являющегося результатом. Опишем виды исследований подробно. При любом исследовании необходимо установить длину окна L (если это не было сделано при выборе исследования) и выбрать тип сингулярных векторов (или элементарные восстановленные ряды), на основе которых будет проводиться идентификация.

Автоматическое выделение тренда При выборе “Trend investigation” необходимо установить параметры методов идентификации. Результатом работы является восстановленный по идентифицированным компонентам ряд.

Соответственно активным кнопкам инструментальной панели, для результата можно:

1. начать новое исследование, при этом в качестве исследуемого ряда будет взят исходный ряд, а не результат исследования,
2. посмотреть сингулярные вектора или собственные числа (см. рис. 17),
3. посмотреть промежуточные результаты исследования (см. рис. 17),
4. сгладить/детализировать результат, при сглаживании отбрасываются последние (по номеру) идентифицированные компоненты и ряд восстанавливается, при детализации – добавляются отброшенные,
5. посмотреть ряды в разных масштабах.

Автоматическое выделение периодик При выделении периодики (см. рис. 18) нужно выбрать тип решаемой задачи, установить параметры методов идентификации, и, если задача ставится с ограничением на период, указать допустимый интервал.

Для результата исследования периодик по сравнению с результатом выделения тренда доступна ещё одна функция – разложение ряда, являющегося результатом исследования, на составляющие гармоники с оценкой их периода (см. рис. 18) с помощью кнопки 6 инструментальной панели.

Сводная таблица результатов Во время исследования “Summation investigation” производится применение к собственным тройкам всех методов и вывод результатов в виде единой таблицы. Идентифицированные тройки выделяются с помощью цвета.

В форме “Choose summation investigation parameters” необходимо установить нужные значения параметров. В блоке “Methods parameters” нужно указать параметры, необходимые для получения результатов, это ε для метода нулей (см. раздел 4.1.2) и ω_0 для метода низких частот (см. разделы 4.1.3 и 6.1.2). В блоке “Thresholds” нужно указать пороговые значения методов. В соответствии с ними, среди элементов таблицы будут выделены цветом те, которые позволяют идентифицироваться собственную тройку.

Таблица организована следующим образом. Первый столбец содержит номера собственных троек, следующие столбцы соответствуют методам идентификации, за исключением столбцов, выводящих дополнительную информацию (корни собственных чисел). Так как методы идентификации гармоник и э.-м. гармоник применяются к парам собственных троек, то для них результаты указываются на строке, соответствующей тройке с большим номером. Если метод идентифицировал собственную тройку, то соответствующая ячейка таблицы будет выделена цветом: бежевым для трендовых методов и светло-зелёным – для методов Фурье. Если тройка была идентифицирована хотя бы одним методом, ячейка первого столбца выделяется одним из следующих цветов:

- розовым – если ЕТ была идентифицирована всеми методами идентификации тренда и не была идентифицирована, как соответствующая гармонике или э.-м. гармонике (то есть она точно соответствует тренду),
- синим – если ЕТ идентифицирована усовершенствованным методом Фурье, но не идентифицирована ни одним трендовым методом (точно соответствует гармонике или э.-м. гармонике),
- светло-фиолетовым во всех остальных случаях.

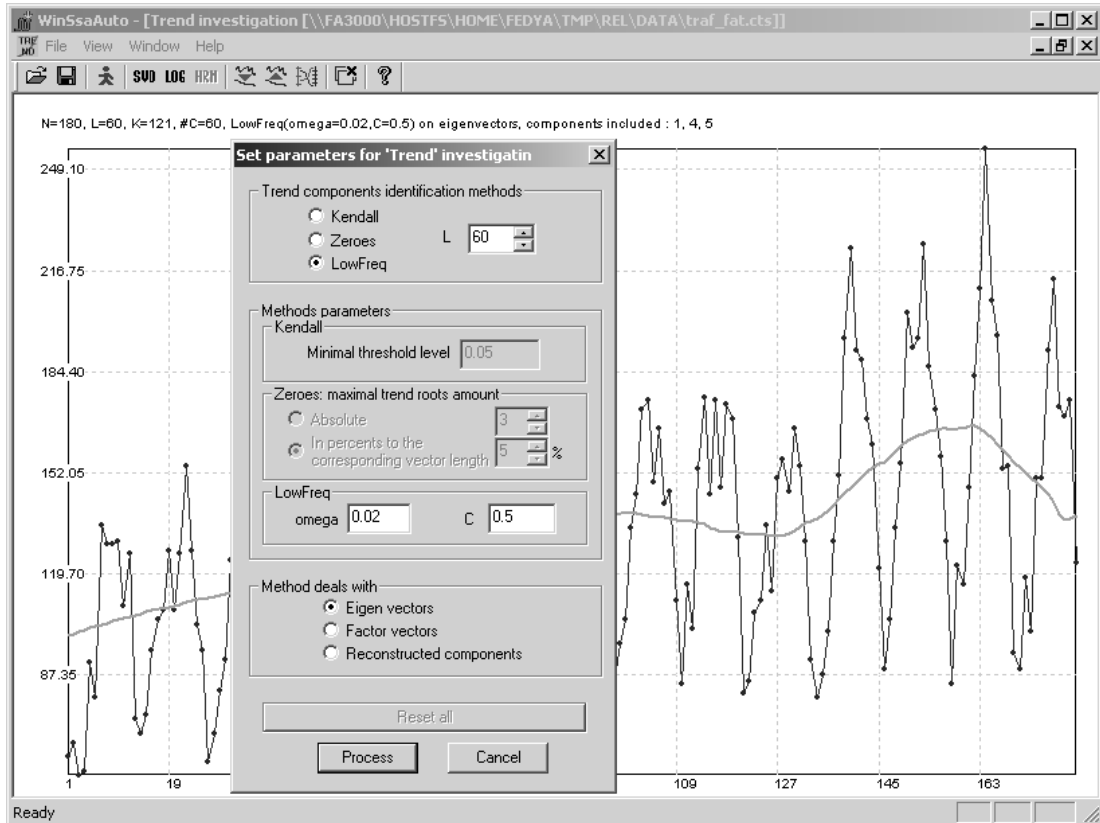


Рис. 16: WinSsaAuto: выбор параметров и результат выделения тренда

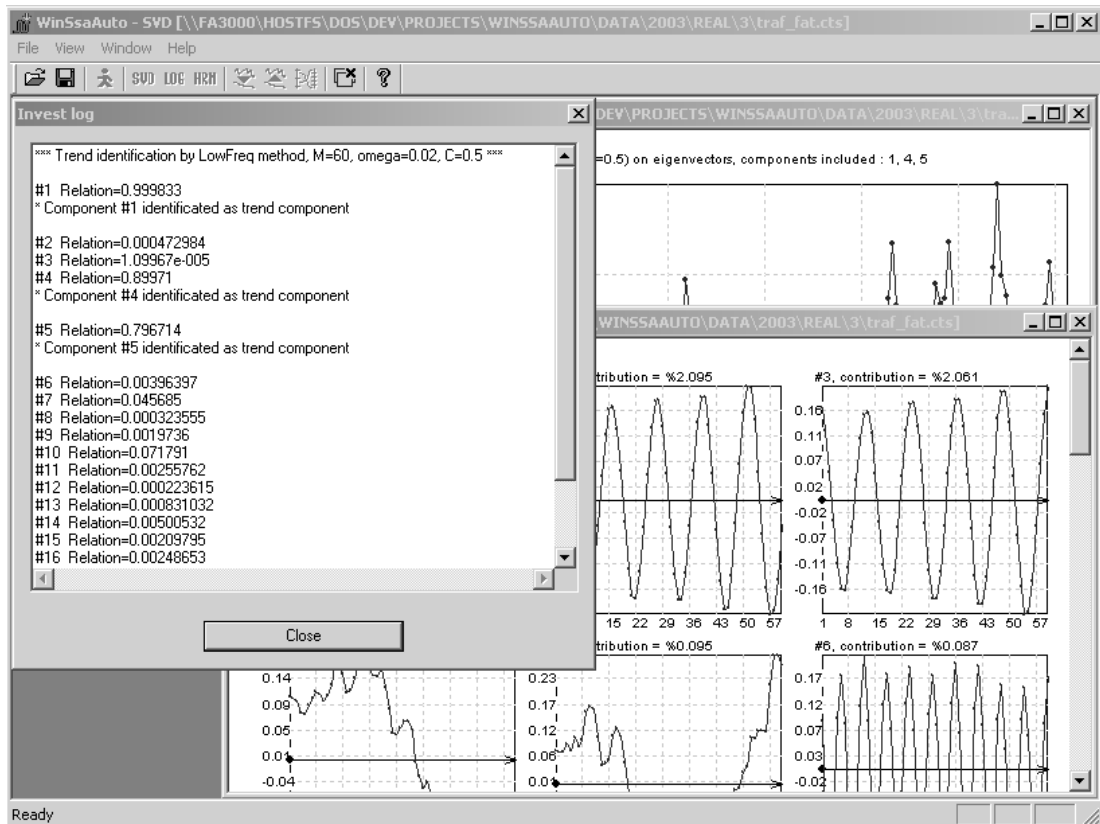


Рис. 17: WinSsaAuto: промежуточные результаты исследования и собственные вектора

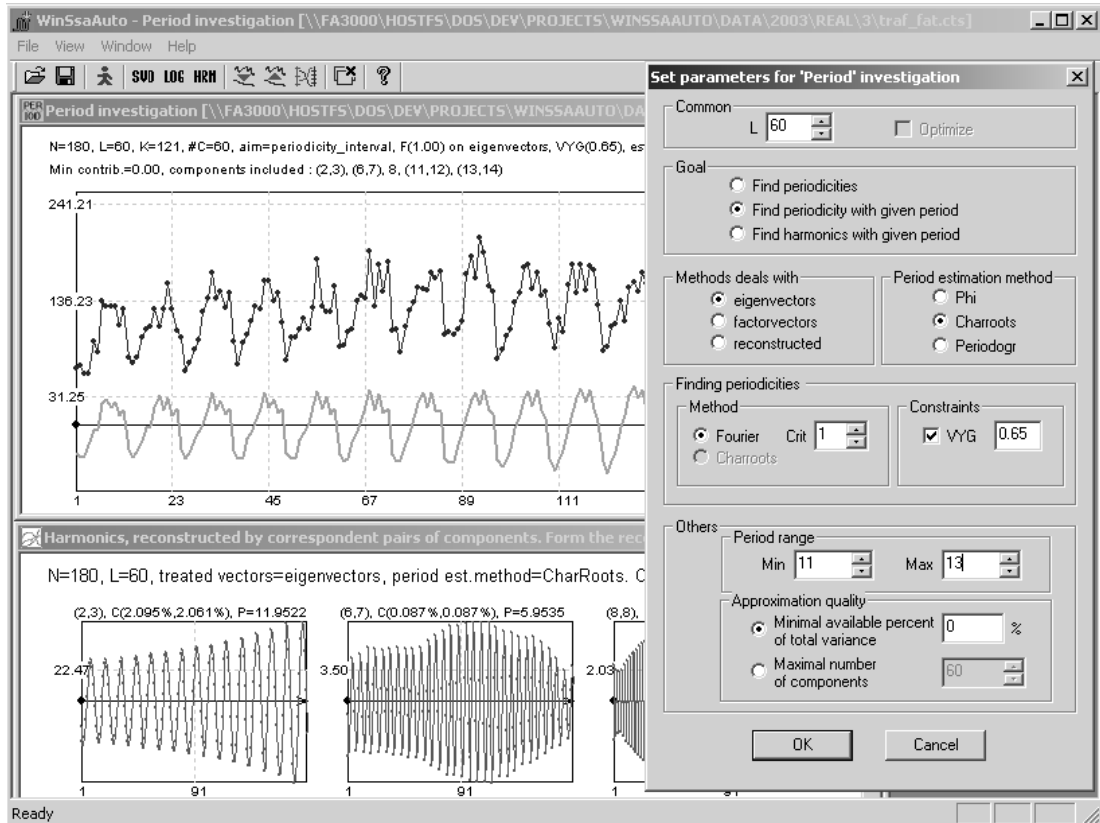


Рис. 18: WinSsaAuto: выбор параметров и результат выделения периодики, составляющие периодику гармоники

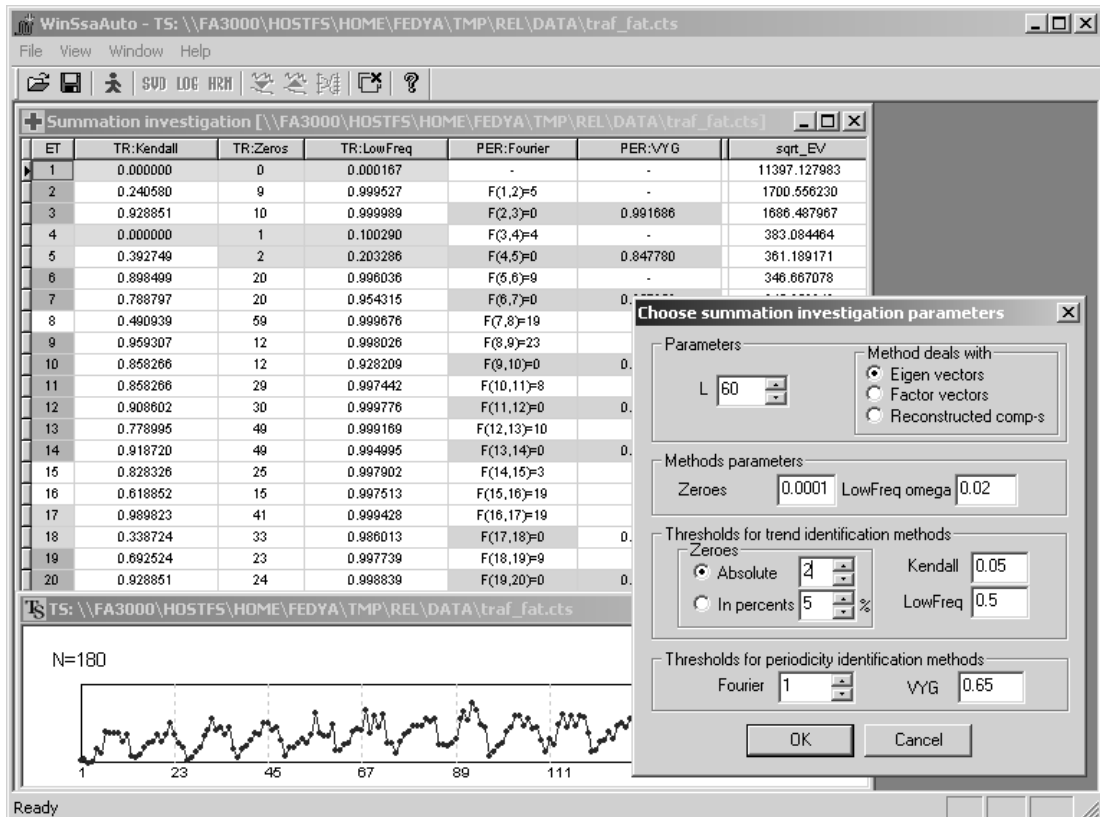


Рис. 19: WinSsaAuto: Сводная таблица результатов

Список литературы

- [1] Golyandina, N.E., Nekrutkin, V.V., Zhigljavsky, A.A., (2001), *Analysis of Time Series Structure. SSA and Related Techniques*, Chapman & Hall/CRC.
- [2] Elsner, J.B. and Tsonis, A.A., (1996), *Singular Spectral Analysis. A New Tool in Time Series Analysis*, Plenum Press, New York and London, 164 pp.
- [3] Vautard, R., Yiou, P., Chil, M., (1992), *Singular-spectrum analysis: A toolkit for short, noisy chaotic signals*, Physica D 58, 95-126.
- [4] Кендэлл, М., (1975), *Ранговые корреляции*, М., Статистика.