

Тензорная декомпозиция и прогноз для набора временных рядов

Сёмкин Кирилл¹

¹semkin.ki@phystech.edu

Вадим Стрижов²

²vadim.swifton@gmail.com

Аннотация

Декомпозиция временного ряда применяется для получения его структуры: разложения на простые и/или интерпретируемые составляющие, выявление периодичности, избавление от шума и т.д. В случае же набора нескольких рядов игнорирование взаимосвязей между ними может приводить к некачественным, ложным разложениям и предсказаниям. В данной работе предлагается метод декомпозиции, учитывающий фактор связанности и основанный на классическом методе SSA (Гусеница) и тензорных разложениях, а также способ прогноза будущих значений ряда. Предложенный подход сравнивается с методом mSSA, основанным на матричных разложениях, и реекуррентными нейросетями. Производится сравнение методов при обработке реальных данных (потребление электроэнергии, акселерометрия).

Ключевые слова: временные ряды, декомпозиция, прогноз, SVD, SSA, каноническое тензорное разложение.

1 Введение

Настоящая работа посвящена методам декомпозиции и построения прогноза для *набора временных рядов* — нескольких рассматриваемых вместе сигналов $\{x_i(t)\}_{i=1}^m$, где $t \in 1 \dots N$. Нас будет интересовать аддитивная декомпозиция — представление ряда в виде суммы нескольких составляющих.

Методы разложения [1, 2, 3] и прогнозирования [4, 1, 5] для одиночного сигнала не могут быть просто перенесены на наборы временных рядов в случае их взаимной зависимости друг от друга. Для иллюстрации достаточно взять модель хищник-жертва [6]. Размер популяции одних особей зависит от размера популяции других и наоборот. Мы будем предполагать некую связь рядов, строгое определение которой будет даваться в рамках рассматриваемых далее моделей. Выявление же этой связи — отдельная задача, см., например, [7, 8].

Рассмотрим несколько подходов к поставленному вопросу. Модель *реекуррентных нейронных сетей* (RNN) [9, 10] связывает ряды друг с другом и своими предыдущими реализациями композицией множества однотипных нелинейных преобразований. На каждом шаге происходит вычисление скрытого представления, инкапсулирующего в себе всю информацию о прошлом. С его помощью строится предсказания будущих значений рядов [11, 12].

Векторная авторегрессия (VAR) [13, 14] представляет собой линейную стохастическую модель динамики набора временных рядов. Если из значений сигналов в момент t составить вектор $\mathbf{x}_t = (x_1(t) \dots x_m(t))^T$, то динамика имеет вид:

$$\mathbf{x}_t = \boldsymbol{\mu} + \sum_{i=1}^p A_i \mathbf{x}_{t-i} + \mathbf{u}_t$$

Здесь μ - некоторый константный вектор, A_i матрицы $m \times m$, \mathbf{u}_t - случайный вектор (например, реализация белого шума $WN(t)$). Как мы видим, связь рядов задаётся линейными преобразованиями A_i , и каждый сигнал зависит как от собственных значений в прошлом, так и от прошлого других сигналов *линейным образом*.

Рассмотренные решения позволяют строить прогноз, но имеют большое количество подбираемых гиперпараметров и требуют затратных процедур обучения. Более того, структура этих моделей не содержит способа построения декомпозиции временных рядов.

В связи с этим мы разработали новый метод *tSSA*, имеющий всего одну настраиваемую переменную и основывающийся на каноническом тензорном разложении *CPD* [15]. Его вычисление — единственная необходимая процедура для получения декомпозиции и предсказания. Наш подход является расширением метода *SSA* [16] для многомерных временных рядов и опирается на теорию динамических систем: за основу взята модель *собственного пространства сигнала* [17]. Она опирается на построение фазового представления временных рядов, иллюстрация на рис. 1. Существует также другая модификация — *mSSA* [18], которое мы сравним с нашей в плане построения базиса в пространстве сигналов.

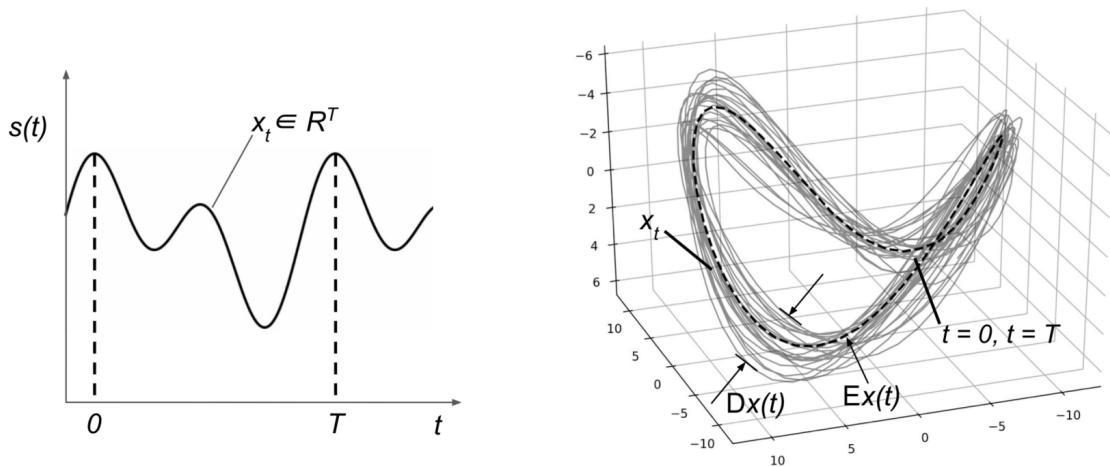


Рис. 1: Пример построения фазовой траектории временного ряда

Далее в работе будет формально введена математическая модель порождения временных рядов, поставлена задача поиска базиса в пространстве сигналов и предложено её решение. На его основе будет рассмотрен наш способ построения декомпозиции и прогноза для рядов, а также проанализированы его свойства и вычислительные особенности. Далее рассмотренные выше методы будут применены к реальным данным потребления электроэнергии и *TODO*, будут проанализированы их точность предсказания. С помощью *mSSA* и *tSSA* мы также построим декомпозиции сигналов и сравним их по специальной метрике — *невязки ганкелизации*, определение которой будет дано ниже.

2 Постановка задачи

Пусть имеем набор временных рядов $\{x_i(t)\}_{i=1}^m$. Гипотеза возникновения этих рядов — существование некой *порождающей динамической системы*. Под этим понимается модель эволюции во времени (для простоты, дискретном) обобщённой переменной $\mathbf{y} \in X$:

$$\begin{aligned}\mathbf{y}(t+1) &= f(\mathbf{y}(t)), \quad t \in \mathbb{N} \\ \mathbf{y}(0) &= \mathbf{y}_0\end{aligned}$$

В общем случае, X — гладкое многообразие большой размерности. Далее, конкретная траектория этой системы порождает наблюдаемые ряды через некое многомерное отображение $\varphi : X \rightarrow \mathbb{R}^m$:

$$\varphi(\mathbf{y}(t)) = \mathbf{x}_t \Leftrightarrow \begin{cases} \varphi_1(\mathbf{y}(t)) = x_1(t) \\ \dots \\ \varphi_m(\mathbf{y}(t)) = x_m(t) \end{cases}$$

Выдвигается предположение, что траектории $\mathbf{y}(t)$ лежат в многообразии $M \subset X$ размерности d_M , меньшей, чем у X . Т.о. ставится задача поиска вложения M в \mathbb{R}^L для некоторого L и нахождения базиса в образе этого вложения. С помощью него мы получаем описание исходной системы в терминах стандартного линейного пространства, а значит и описание наблюдаемых сигналов.

Решение для одного сигнала

Предлагаемое нами решение существенно опирается на методику SSA и теорему Таценса [19]. Кратко опишем их. Теорема формулируется для случая одного наблюдаемого сигнала и предоставляет явный вид вложения: произвольной точке $\mathbf{y}(t) \in M$ ставится в соответствие вектор $(\varphi \circ f^{t-L+1}(\mathbf{y}(t)), \dots, \varphi \circ f(\mathbf{y}(t)), \varphi \circ \mathbf{y}(t))^T = (\overset{\leftarrow}{x}(t-L+1) \dots \overset{\leftarrow}{x}(t))^T$. Он называется *вектором задержки* ряда в момент t и обозначается $\overset{\leftarrow}{\mathbf{x}}_t$. Размерность таких векторов должна удовлетворять условию $L > 2d_M$, а функция $\varphi(\cdot)$ некоторым условиям регулярности, которые мы считаем выполненными.

Т.о., имея временной ряд $x(t)$ длины N , получаем $N - L + 1$ векторов задержки. Пространство вложения $\text{Lin}(\{\overset{\leftarrow}{\mathbf{x}}_t\})$, в данном случае оно же *собственное пространство сигнала* $x(t)$, не должно иметь большую размерность. Соответственно ожидаем, что $\text{Lin}(\{\overset{\leftarrow}{\mathbf{x}}_t\}) \subset \mathbb{R}^L$. Ортонормированный базис в этом пространстве выбирается как U -компоненты SVD-разложения матрицы векторов задержек — так называемой *траекторной матрицы* $\mathbf{H}_x = [\overset{\leftarrow}{\mathbf{x}}_1 \dots \overset{\leftarrow}{\mathbf{x}}_{N-L+1}]$.

Взаимосвязь рядов и метод tSSA

Теперь обобщим данный подход на несколько временных рядов.

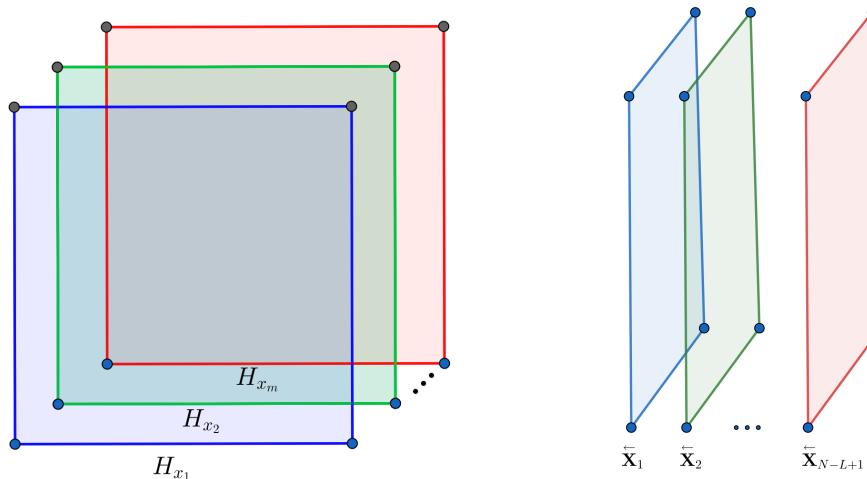


Рис. 2: Два вида на траекторный тензор. Слева — в виде набора траекторных матриц сигналов $\{x_i(t)\}_{i=1}^m$. Справа — в виде набора матриц задержки.

В данном случае $\varphi(\cdot)$ уже многомерное, а значит вместо вектора задержки одного сигнала необходимо рассматривать их набор для всех m сигналов в момент t . Т.е. образами вложения будут служить *матрицы задержки* $(\overleftarrow{\mathbf{x}}_{1_t} \dots \overleftarrow{\mathbf{x}}_{m_t}) := \overleftarrow{\mathbf{X}}_t$. Вместо же траекторной матрицы сигнала нужно рассматривать *траекторный тензор* \mathbf{T} набора сигналов. Он складывается из матриц задержки, выстроенных подряд по второму измерению тензора, см. рис.2. Из построения также следует, что \mathbf{T} является состыковкой траекторных матриц \mathbf{H}_{x_i} вдоль третьего измерения.

Продолжая аналогию подраздела выше, мы получаем *собственное пространство набора рядов* в виде линейной оболочки матриц задержек. Но нам интересно построить его для каждого сигнала отдельно.

Выше мы обещали дать определение взаимосвязи рядов в рамках нашей модели.

Определение. Мы называем набор временных рядов *связанным*, если они имеют одинаковое для всех собственное пространство с единым базисом.

Предъявим метод построения такого пространства и базиса. Заметим, что для этого достаточно разложить матрицы задержки каждого сигнала по общему набору векторов. Теперь применим *CPD-разложение* к траекторному тензору и рассмотрим его вид для каждого сечения по третьему измерению:

$$\mathbf{T} = \sum_{i=1}^r \mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{c}_i \Leftrightarrow \begin{cases} \mathbf{H}_{x_1} = \sum_{i=1}^r \boldsymbol{\sigma}_{x_1}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top \\ \mathbf{H}_{x_2} = \sum_{i=1}^r \boldsymbol{\sigma}_{x_2}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top \\ \dots \\ \mathbf{H}_{x_m} = \sum_{i=1}^r \boldsymbol{\sigma}_{x_m}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top \end{cases} \quad (1)$$

Проанализируем полученное. Разложение определяется каноническим тензорным рангом r и набором векторов, которые удобно упаковать в матрицы: $A = [\mathbf{a}_1 \dots \mathbf{a}_r]; B = [\mathbf{b}_1 \dots \mathbf{b}_r]; C = [\mathbf{c}_1 \dots \mathbf{c}_r]$, k -ую строку матрицы C мы обозначаем $\boldsymbol{\sigma}_{x_k}$. Она имеет смысл набора сингулярных чисел для \mathbf{H}_{x_k} , только в данном случае они могут быть отрицательными. Из определения CPD как разложения вида (1) с минимальным r следует, что матрицы A, B, C имеют полный ранг. Общим собственным пространством сигналов является $\text{Lin}(\{\mathbf{a}_i\})$ с единым базисом $\{\mathbf{a}_i\}_{i=1}^r$, в общем случае не ортогональным. В формулировке задачи мы предполагали, что размерность вложения не должна быть большой, т.е. $r \ll L$. Наконец заметим, что строки матриц \mathbf{H}_{x_k} также являются векторами задержек размерности $N - L + 1$, и в их терминах общим пространством сигналов будет $\text{Lin}(\{\mathbf{b}_i\})$. Т.о. наш подход обладает согласованностью относительно выбора $\overleftarrow{\mathbf{x}}_{k_t}$, чего, например, нет у метода mSSA.

В данном разложении связь рядов определяют $\boldsymbol{\sigma}_{x_k}$: если для разных сигналов они имеют нулевые элементы в разных позициях, то собственные пространства будут различаться (т.к. в качестве базисных векторов будут участвовать разные наборы \mathbf{a}_i).

Декомпозиция набора рядов

Теперь мы можем представить каждый сигнал $x_k(t)$ суммой компонент по следующему принципу: *декомпозиция траекторной матрицы* \mathbf{H}_{x_k} определяет *декомпозицию временного ряда*. Эта методика лежит в основе метода SSA. Т.к. разложение у каждой матрицы своё, то и сигналы можно обрабатывать независимо.

В данном подходе основным является свойство *ганкелевости* траекторных матриц — равенства элементов на одной антидиагонали. Каждомуциальному ряду длины N взаимнооднозначно сопоставляется ганкелева матрица $L \times (N - L + 1)$.

Для получения декомпозиции предполагается, что факторы разложения \mathbf{H}_{x_k} можно сгруппировать таким образом, что, сложив их в каждой группе, мы получим ганкелевы матрицы C_1, \dots, C_s . Это бы в точности соответствовало представлению $x_k(t)$ в виде суммы s сигналов. На практике такая ситуация почти невозможна (см. [16]), поэтому все C_i дополнительно *ганкелизируют* — усредняют каждую антидиагональ. Будем обозначать эту операцию $\text{Hankel}(\cdot)$. Описанная процедура представлена выражением (2).

$$\begin{aligned}\mathbf{H}_{x_k} &= \sum_{i=1}^r \boldsymbol{\sigma}_{x_k}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top = \sum_{i \in \mathbb{I}_1} \boldsymbol{\sigma}_{x_k}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top + \dots + \sum_{i \in \mathbb{I}_s} \boldsymbol{\sigma}_{x_k}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top = \\ &= C_1 + \dots + C_s = \text{Hankel}(C_1) + \dots + \text{Hankel}(C_s) \Leftrightarrow x_k(t) = c_1(t) + \dots + c_s(t)\end{aligned}\quad (2)$$

Здесь $\mathbb{I}_1 \sqcup \dots \sqcup \mathbb{I}_s = \{1, \dots, r\}$ — выбранные группы индексов. Четвёрте равенство в (2) нуждается в обосновании, для этого докажем простое свойство оператора ганкелизации:

Лемма 1. *Оператор $\text{Hankel}(\cdot)$ линейный.*

Доказательство. Возьмём две матрицы A и B одного размера и рассмотрим их элементы с одной антидиагонали a_1, \dots, a_n и b_1, \dots, b_n . Тогда соответствующая антидиагональ в $\text{Hankel}(A + B)$ равна $\frac{1}{n} \sum_{i=1}^n (a_i + b_i) = \frac{1}{n} \sum_{i=1}^n a_i + \frac{1}{n} \sum_{i=1}^n b_i$, т.е. равна сумме антидиагоналей матриц $\text{Hankel}(A)$ и $\text{Hankel}(B)$.

Для произвольного скаляра α также очевидно $\frac{1}{n} \sum_{i=1}^n \alpha \cdot a_i = \alpha \frac{1}{n} \sum_{i=1}^n a_i$, из чего следует $\text{Hankel}(\alpha A) = \alpha \text{Hankel}(A)$. \square

Учитывая, что \mathbf{H}_{x_k} — ганкелева, мы полностью обосновали корректность процедуры построения декомпозиции временного ряда.

Вычислительная сложность декомпозиции

Несмотря на полученный результат, хочется группировать факторы так, что каждая матрица C_i была как можно более "ганкелевой". Это соответствует изначальной идеи о связи разложений матриц и сигналов, а также повышает интерпретируемость компонент. Опишем задачу более формально и исследуем её.

Рассмотрим первое равенство в (2), а также применим к нему оператор ганкелизации:

$$\begin{cases} \mathbf{H}_{x_k} = \sum_{i=1}^r \boldsymbol{\sigma}_{x_k}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top \\ \mathbf{H}_{x_k} = \sum_{i=1}^r \text{Hankel}(\boldsymbol{\sigma}_{x_k}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top) \end{cases}$$

Вычтем из первого втрое и обозначим $H_i = \boldsymbol{\sigma}_{x_k}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top - \text{Hankel}(\boldsymbol{\sigma}_{x_k}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top)$, что соответствует невязке между исходной матрицей и её ганкелизацией. Т.о. получим:

$$H_1 + \dots + H_r = 0 \Leftrightarrow H_r = - \sum_{j=1}^{r-1} H_j$$

Теперь группировка факторов и ганкелевость итоговых матриц C_i становится равносильна $\sum_{k \in \mathbb{I}_i} H_k = 0, \forall i \in 1, \dots, s$, т.е. суммированию в ноль матриц невязок внутри каждой группы. Несложно заметить, что если существует разложение на s групп, то к нему можно прийти, раскладывая последовательно факторы на две группы, так что остановимся на поиске этого простого разделения.

Мы видим, что матрицы невязок все вместе уже суммируются в 0, т.е. группу для H_r можно определить, сгруппировав оставшиеся невязки. Введём для каждой из них переменную-индикатор $\beta_j \in \{0, 1\}$, показывающей, к какой из двух групп относится j -ая. Невязки в одной группе суммируются в 0; также считаем, что все они ненулевые (иначе мы уже имеем искомую группировку), т.е. в группе должно быть хотя бы две невязки. Т.о. задача поиска группировки свелась к следующей:

$$\begin{cases} \sum_{j=1}^{r-1} \beta_j H_j = 0 \\ \beta_j \in \{0, 1\}, \forall j \in 1, \dots, r \\ \sum_{i=1}^{r-1} \beta_i \geq 2 \end{cases} \quad (3)$$

Для придания задаче более привычного вида, во-первых, векторизуем каждую H_i , и соберём все эти векторы в одну матрицу Λ , а также введём $\beta = (\beta_1 \dots \beta_{r-1})$. Также вспомним, что мы не ждём идеальной ситуации суммирования в 0 матриц невязок (ганкелевости C_i), нам достаточно сделать сумму как можно более близкой к нулю. Это приводит нас к конечной формулировке:

$$\begin{cases} \|\Lambda\beta\| \rightarrow \min_{\beta} \\ \beta_j \in \{0, 1\}, \forall j \in 1, \dots, r \\ \sum_{i=1}^{r-1} \beta_i \geq 2 \end{cases} \quad (4)$$

Назовём решение этой системы *оптимальной декомпозицией ряда*. Она представляет из себя задачу наименьших квадратов с целочисленными ограничениями (ILS), которая, как показано в [20], является NP-сложной. Все проведённые рассуждения приводят к:

Теорема 1. *Оптимальная декомпозиция ряда, в смысле поиска группировки факторов разложения траекторной матрицы ряда, является NP-сложной задачей.*

Несмотря на данный результат существуют методы и эвристики для сведения ILS к эффективно вычислим задачам [21].

Прогноз для набора рядов

Получение прогноза в нашей модели сводится к восстановлению компонент векторов задержек каждого входного ряда. Т.к. мы построили общий для всех сигналов базис $\{\mathbf{a}_i\}_{i=1}^r$, то каждый сигнал может обрабатываться независимо.

Если бы $\{\mathbf{a}_i\}_{i=1}^r$ был ортогональным, то мы бы могли применить технику прогнозирования SSA [16]. Наша модель не даёт такой гарантии, поэтому метод SSA необходимо модифицировать.

Рассмотрим произвольный входной ряд $x(t)$ и построим для него прогноз на один шаг вперёд. Вектор задержки $\overleftarrow{\mathbf{x}}_{N+1} = (x(N-L) \dots x(N+1))^T$, в котором последняя компонента неизвестна, лежит в пространстве $\text{Lin}(\{\mathbf{a}_i\})$. В терминах вводимой ранее матрицы $A \in \mathbb{R}^{L \times r}$ полного ранга это записывается:

$$\begin{aligned}\overleftarrow{\mathbf{x}}_{N+1} = A\boldsymbol{\lambda} \Leftrightarrow & \begin{cases} \mathbf{x}_{kn} = A_{kn}\boldsymbol{\lambda} \\ x(N+1) = \mathbf{a}_{pr}^\top \boldsymbol{\lambda} \end{cases}, \text{ где} \\ & A = \left(\frac{A_{kn}}{\mathbf{a}_{pr}^\top} \right) \\ \overleftarrow{\mathbf{x}}_{N+1} = & (\mathbf{x}_{kn} \ x(N+1))^\top\end{aligned}\quad (5)$$

Здесь $\boldsymbol{\lambda} \in \mathbb{R}^r$, вектор задержки и матрица представлены в блочном виде для отделения известной части сигнала и его будущего значения. Т.о. для прогноза необходимо найти $\boldsymbol{\lambda}$ из первой части системы (5). Матрица данной СЛАУ $A_{kn} \in \mathbb{R}^{(L-1) \times r}$ имеет ранг не менее $r - 1$, но даже пусть ранг будет полный. Сама система в силу предположения модели $r \ll L$ является переопределённой. Значит, решение здраво искать только в смысле метода наименьших квадратов: $\boldsymbol{\lambda} = (A_{kn}^T A_{kn})^{-1} A_{kn}^T \mathbf{x}_{kn}$. В итоге прогноз модели даётся:

$$x(N+1) = \mathbf{a}_{pr}^\top (A_{kn}^T A_{kn})^{-1} A_{kn}^T \mathbf{x}_{kn} \quad (6)$$

Здесь $\mathbf{a}_{pr}^\top (A_{kn}^T A_{kn})^{-1} A_{kn}^T = \mathbf{d}^\top$ — вектор-строка, и, вычислив её один раз, можно быстро строить предсказания на несколько шагов, последовательно получая $x(N+1), x(N+2)$, и т.д.

Распишем (6) в терминах $x(t)$:

$$x(t) = \sum_{i=1}^{L-1} d_i \cdot x(t-i)$$

Данное выражение приводит нас к результату:

Теорема 2. *Модель прогнозируемого методом tSSA ряда — авторегрессионная AR(L-1). Поведение прогноза полностью определяется коэффициентами авторегрессии d_i , найденными в (6).*

3 Эксперимент

Цели и постановка

В данном разделе мы рассмотрим применение метода tSSA для декомпозиции и прогноза нескольких многомерных временных рядов, а также сравним его с другими моделями, упомянутыми во введении: mSSA, VAR, RNN. Упомянем, что mSSA практически идентична нашему подходу, за исключением способа построения разложения траекторных матриц, а две другие модели способны только к предсказанию.

Для измерения **качества декомпозиции** введём метрики:

Определение. *Абсолютной ошибкой ганкелизации* матрицы M назовём

$$\text{AHE} = \|M - \text{Hankel}(M)\|_2$$

Определение. *Относительной ошибкой ганкелизации* матрицы M назовём

$$\text{RHE} = \frac{\text{AHE}}{\|M\|_2}$$

AHE можно представлять как суммарное стандартное отклонение антидиагоналей матрицы. RHE её более интерпретируемая модификация, которая и будет использоваться далее.

Напомним, что разложение ряда на компоненты эквивалентно группировке факторов разложения траекторной матрицы сигнала, т.е. получению матриц C_i (см. раздел Декомпозиция набора рядов). Именно для них мы и будем рассчитывать RHE, а также среднее RHE по одному ряду — \bar{RHE}_{ts} , и среднее RHE всех рядов датасета — \bar{RHE} .

Для измерения **качества прогноза** ряда будем использовать стандартные метрики: среднеквадратическую ошибку MSE_{ts} и среднюю абсолютную процентную ошибку $MAPE_{ts}$, а также их усреднения по всему датасету с обозначениями \bar{MSE} и \bar{MAPE} .

В качестве сигналов будут рассмотрены два набора данных: потребление электроэнергии, её цена за единицу мощности (рис. 3); а также три ряда, характеризующие погоду г. Берлин в период с 1980 по 1990 год: средние за день температура, осадки и атмосферное давление (рис. 4). Мы предполагаем наличие связи рядов в каждом наборе (на основе экономических/физических причин), т.к. это важное условие для применения нашего метода, см. раздел Взаимосвязь рядов и метод tSSA. Каждый ряд имеет длину $\approx 3 \cdot 10^3$, где около 20% — отложенная выборка для оценки качества прогноза.

Рис. 3: Многомерный ряд потребления электричества и его цены за единицу мощности

Рис. 4: Многомерный ряд физических величин: температуры, осадков и давления, измеренных на метеостанции г. Берлин

Опишем нюансы метода tSSA и других моделей. Во-первых, вычисление канонического ранга тензора — NP-трудная задача [22], поэтому каждый раз ранг разложения будет указываться явно. Во-вторых, используемый алгоритм СР-разложения — ALS [15], имеет ошибку аппроксимации, из-за чего полученные разложения траекторных матриц будут неточными. Относительные величины этих невязок будут указаны ниже. В-третьих, для поиска декомпозиции рядов будет решаться задача ILS (4), в которой матрица имеет высокую строковую размерность. Для ускорения мы сократим её до нескольких сотен, что всё ещё будет много больше размерности искомого вектора параметров. Солвер для ILS — SCIP [23].

Наконец, мы зафиксируем единый для всех моделей параметр, отвечающий за длину предуславливания методов на историю рядов, который обозначался за L . Для tSSA и mSSA он был описан в начале раздела 2; в RNN это можно считать длиной энкодера; в VAR это максимальная длина истории ряда, на которую производится авторегрессия. Везде ниже $L = 500$.

Результаты и обсуждение

Рассмотрим решение задачи прогнозирования. На рис. 5 и рис. 6 отображена зависимость введённых метрик качества прогноза tSSA для разных рангов разложения. Точка минимума и характер рассматриваемой зависимости MSE и MAPE совпадают. Также наблюдается эффект переобучения при повышении значения ранга, особенно резкое ухудшение метрик у данных погоды. При этом ошибка СР-разложения монотонно убывает, см. рис. 7.

Рис. 5: Значения метрик \bar{MSE} и \bar{MAPE} предсказания tSSA для разных рангов СР-разложения. Отдельно выделен оптимальный ранг. Данные электроэнергии.

Рис. 6: Значения метрик $\overline{\text{MSE}}$ и $\overline{\text{MAPE}}$ предсказания tSSA для разных рангов СР-разложения. Отдельно выделен оптимальный ранг. Данные погодных условий.

Рис. 7: Относительные ошибки CPD-аппроксимации траекторных тензоров в зависимости от ранга разложения. Слева — электроэнергии, справа — погоды.

Графики наилучшего прогноза tSSA приведены на рис. 8 и 9. Хорошо виден характер предсказания в зависимости от величины ранга: в первом случае модель авторегрессии большого порядка (см. теорему раздела Прогноз для набора рядов), что позволяет точно аппроксимировать ряд. В случае же погоды большой порядок приводил к нестабильности и неограниченному росту прогноза, но когда порядок небольшой возможно аппроксимировать только тренд сигналов.

Итоговые метрики качества всех моделей — в табл. 1 и 2. Наш метод показал наилучшие результаты, хотя mSSA был близок по точности. Метод VAR оказался нестабилен на выбранном горизонте прогнозирования, а RNN обучался в константную функцию.

Рис. 8: Предсказание данных электроэнергии методом tSSA. CPD-ранг = 30

Рис. 9: Предсказание данных погоды методом tSSA. CPD-ранг = 5

Таблица 1: Метрики моделей на прогнозировании данных электроэнергии

	<i>tSSA</i>	<i>mSSA</i>	<i>VAR</i>	<i>RNN</i>
MSE(Production), 10^6	1.24	1.51	7.81	2.70
MSE(Price), 10^3	0.88	1.03	4.85	30.0
MSE, 10^6	0.62	0.75	3.91	135.00
MAPE(Production)	0.054	0.060	0.137	0.999
MAPE(Price)	0.164	0.170	0.360	1.004
MAPE	0.109	0.115	0.249	1.002

Таблица 2: Метрики моделей на прогнозировании данных погоды

	<i>tSSA</i>	<i>mSSA</i>	<i>VAR</i>	<i>RNN</i>
MSE(Temp)	14.16	20.12	40.67	-
MSE(Prec)	11.02	11.23	22.84	-
MSE(Pres)	98.02	101.23	156.75	-
MSE	41.067	44.193	73.420	-
MAPE(Temp)	0.850	0.626	0.468	-
MAPE(Prec)	3.151	2.834	4.637	-
MAPE(Pres)	0.007	1.254	3.983	-
MAPE	1.336	1.571	3.029	-

Теперь рассмотрим задачу декомпозиции. На рис. 10 представлены зависимости качества разложения от значения СР-ранга. В обоих случаях $\overline{\text{RHE}}$ достаточно быстро достигает минимума, после чего не меняется или растёт. Вспоминая аналогичный результат для прогнозирования, мы получаем, что метод tSSA достигает наибольшей обобщающей способности на малом каноническом ранге траекторных тензоров данных. В свою очередь,

этот ранг пропорционален размерности собственного пространства сигналов, см. начало раздела 2. Т.о. можно заключить о выполнении главной теоретической предпосылки применения нашего метода.

На рис. 11 и 12 представлены результаты наилучшей найденной декомпозиции рядов на две компоненты методом tSSA. В силу вычислительной сложности задачи ILS разложение на большее количество компонент не производится. Из графиков видно, как сильно влияет общий базис сигналов на их декомпозицию: для генерации электроэнергии и её цены компоненты получились почти идентичными с точностью до смещения и масштаба. Аналогичная ситуация для температуры и осадков в данных погоды, а для атмосферного давления разложение получилось своё. Далее, на рис. 12 метод выделил для каждого сигнала по две периодические составляющие: одна из них большой амплитуды, другая много меньшей. Небольшой шум остался в каждой компоненте. На рис. 11 также выделил по два высокочастотных колебания разной амплитуды и смещения.

Для сравнения на рис. 13 и 14 приведены результаты декомпозиции методом mSSA на пять компонент. Для их получения не было необходимости в решении ILS или подбора СР-ранка, но они подбирались вручную на основе близости сингулярных чисел некоторой матрицы, см. [16]. Тем не менее компоненты получились более интерпретируемыми и подробными: метод выделял несколько трендов и гармоник отдельно. Также, как видно из табл. 3 и 4, он оказался более точным. Т.о. вычислительная сложность декомпозиции tSSA является главной проблемой метода и требует разработки собственных эвристик и методов, которые у mSSA уже имеются.

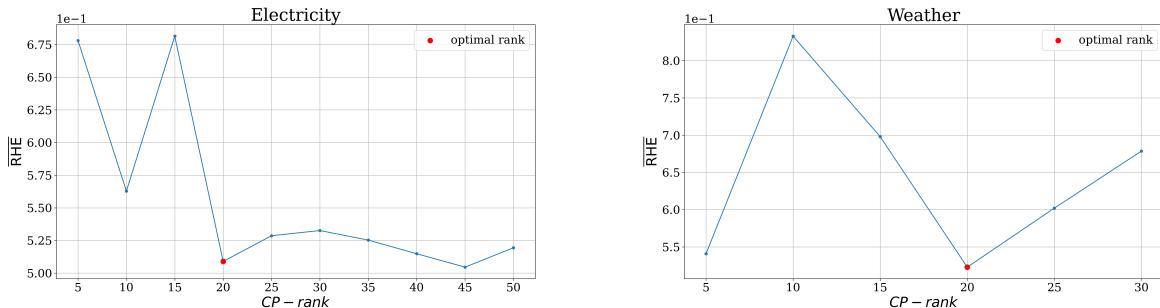


Рис. 10: Зависимость метрики \overline{RHE} от ранга СР-разложения. Слева — для данных электроэнергии, справа — погоды. Отдельно выделен оптимальный ранг.

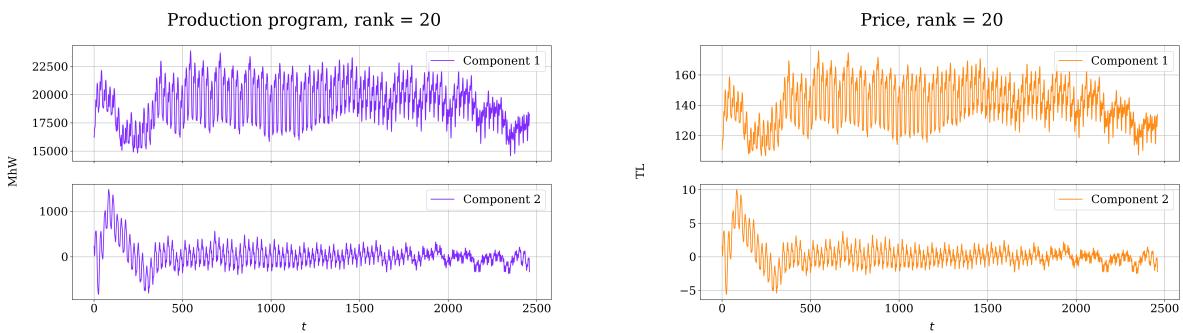


Рис. 11: Разложение рядов на две компоненты методом tSSA. Данные электроэнергии. СРД-ранг = 20

4 Заключение

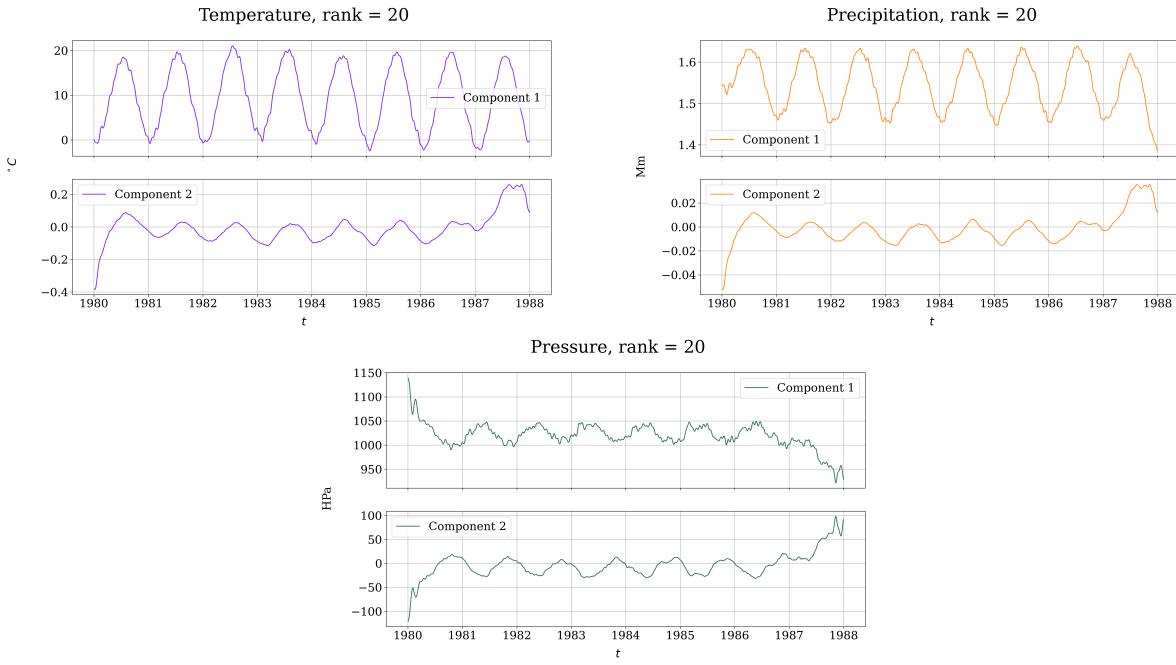


Рис. 12: Разложение рядов на две компоненты методом tSSA. Данные погоды. CPD-ранг = 20

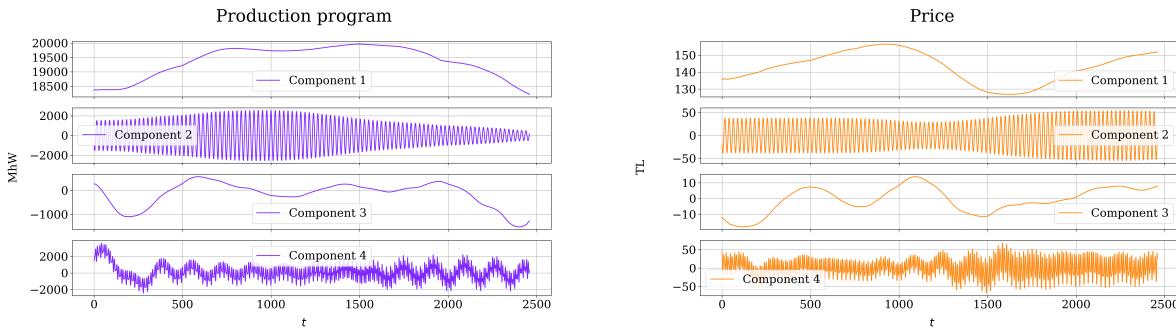


Рис. 13: Разложение рядов на компоненты методом mSSA. Данные электроэнергии.

Таблица 3: Метрики моделей на задаче декомпозиции данных электроэнергии

	tSSA	mSSA
RHE(Production)	0.507	0.308
RHE(Price)	0.511	0.31
RHE	0.509	0.309

Таблица 4: Метрики моделей на задаче декомпозиции данных погоды

	tSSA	mSSA
RHE(Temp)	0.512	0.467
RHE(Precip)	0.508	0.538
RHE(Pres)	0.542	0.502
RHE	0.521	0.502

Список литературы

- [1] W. Enders и John Wiley & Sons. *Applied Econometric Time Series*. Wiley series in probability and statistics. Wiley, 2010. ISBN: 9788126543915. URL: <https://books.google.ru/books?id=vzJ0CgAAQBAJ>.

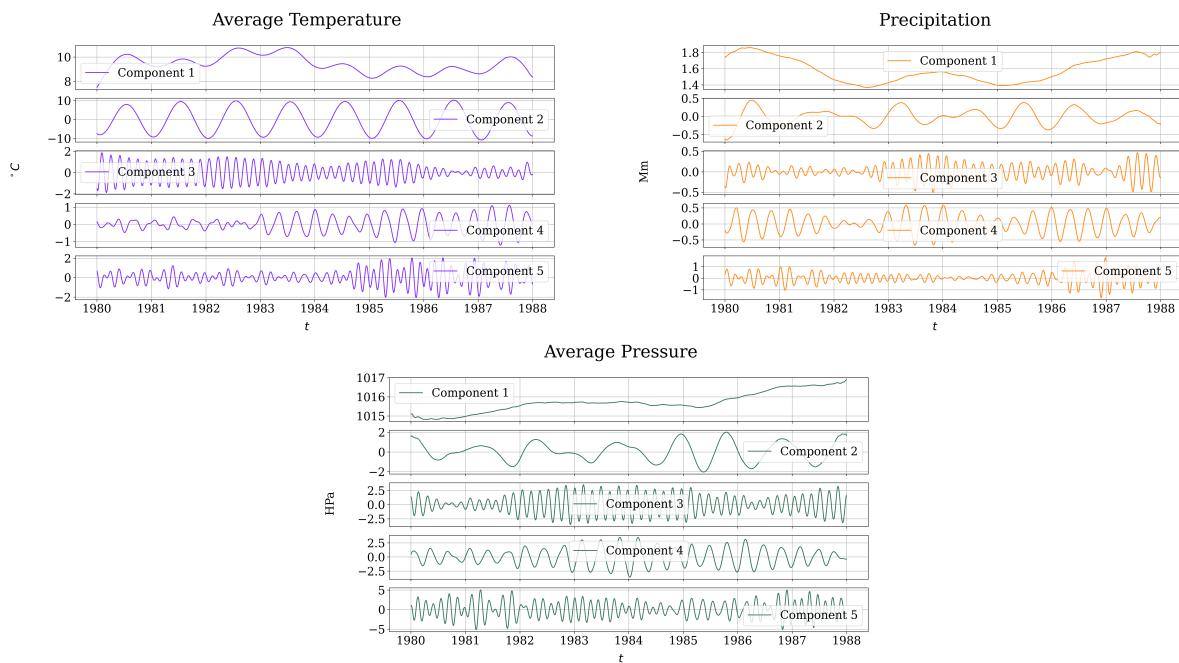


Рис. 14: Разложение рядов на компоненты методом mSSA. Данные погоды.

- [2] Estela Dagum и Silvia Bianconcini. *Seasonal Adjustment Methods and Real Time Trend-Cycle Estimation*. Авг. 2016. ISBN: 978-3-319-31820-2.
- [3] Robert B. Cleveland и др. «STL: A Seasonal-Trend Decomposition Procedure Based on Loess (with Discussion)». В: *Journal of Official Statistics* 6 (1990), с. 3–73.
- [4] Robin John Hyndman и George Athanasopoulos. *Forecasting: Principles and Practice*. English. 2nd. Australia: OTexts, 2018.
- [5] Granville Tunnicliffe Wilson. «Time Series Analysis: Forecasting and Control, 5th Edition, by George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung, 2015. Published by John Wiley and Sons Inc., Hoboken, New Jersey, pp. 712. ISBN: 978-1-118-67502-1». В: *Journal of Time Series Analysis* 37 (март 2016), n/a—n/a. DOI: [10.1111/jtsa.12194](https://doi.org/10.1111/jtsa.12194).
- [6] Б. Вольтерра. «Математическая теория борьбы за существование». В: *Усп. Физ. Нук* 8.1 (1928), с. 13—34. DOI: [10.3367/UFNr.0008.192801b.0013](https://doi.org/10.3367/UFNr.0008.192801b.0013). URL: <https://ufn.ru/ru/articles/1928/1/b/>.
- [7] C. W. J. Granger. «Investigating Causal Relations by Econometric Models and Cross-spectral Methods». В: *Econometrica* 37.3 (1969), с. 424—438. ISSN: 00129682, 14680262. URL: <http://www.jstor.org/stable/1912791> (дата обр. 24.01.2024).
- [8] George Sugihara и др. «Detecting Causality in Complex Ecosystems». В: *Science* 338.6106 (окт. 2012), с. 496. DOI: [10.1126/science.1227079](https://doi.org/10.1126/science.1227079).
- [9] Sepp Hochreiter и Jürgen Schmidhuber. «Long Short-Term Memory». В: *Neural Computation* 9.8 (нояб. 1997), с. 1735—1780. ISSN: 0899-7667. DOI: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735). eprint: <https://direct.mit.edu/neco/article-pdf/9/8/1735/813796/neco.1997.9.8.1735.pdf>. URL: <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [10] Ahmed Tealab. «Time series forecasting using artificial neural networks methodologies: A systematic review». В: *Future Computing and Informatics Journal* 3.2 (2018), с. 334—340. ISSN: 2314-7288. DOI: <https://doi.org/10.1016/j.fcij.2018.10.003>. URL: <https://www.sciencedirect.com/science/article/pii/S2314728817300715>.

- [11] Xueli Zhang и др. «Robust recurrent neural networks for time series forecasting». B: *Neurocomputing* 526 (2023), c. 143—157. ISSN: 0925-2312. DOI: <https://doi.org/10.1016/j.neucom.2023.01.037>. URL: <https://www.sciencedirect.com/science/article/pii/S0925231223000462>.
- [12] Hansika Hewamalage, Christoph Bergmeir и Kasun Bandara. «Recurrent Neural Networks for Time Series Forecasting: Current status and future directions». B: *International Journal of Forecasting* 37.1 (2021), c. 388—427. ISSN: 0169-2070. DOI: <https://doi.org/10.1016/j.ijforecast.2020.06.008>. URL: <https://www.sciencedirect.com/science/article/pii/S0169207020300996>.
- [13] Dimitrios Asteriou и Stephen Hall. «Vector Autoregressive (VAR) Models and Causality Tests». B: янв. 2016, c. 333—346. ISBN: 978-1-137-41546-2. DOI: 10.1057/978-1-137-41547-9_15.
- [14] Arnold Zellner. «An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias». B: *Journal of the American Statistical Association* 57.298 (1962), c. 348—368. DOI: <10.1080/01621459.1962.10480664>. eprint: <https://www.tandfonline.com/doi/pdf/10.1080/01621459.1962.10480664>. URL: <https://www.tandfonline.com/doi/abs/10.1080/01621459.1962.10480664>.
- [15] Tamara Kolda и Brett Bader. «Tensor Decompositions and Applications». B: *SIAM Review* 51 (авг. 2009), c. 455—500. DOI: <10.1137/07070111X>.
- [16] D. Stepanov и N. Golyandina. «SSA-based approaches to analysis and forecast of multidimensional time series». B: *Proceedings of the 5th St.Petersburg Workshop on Simulation*. 2005, c. 293—298.
- [17] Hamid KRIM. «Two Decades of Array Signal Processing Research». B: *IEEE Signal Processing Magazine* 0 (1994), c. 67—94. URL: <https://cir.nii.ac.jp/crid/1572261550523548160>.
- [18] Hossein Hassani и Rahim Mahmoudvand. «Multivariate singular spectrum analysis: A general view and new vector forecasting approach». B: *International Journal of Energy and Statistics* 01 (апр. 2013). DOI: <10.1142/S2335680413500051>.
- [19] Floris Takens. «Detecting Strange Attractors in Turbulence». B: *Dynamical Systems and Turbulence, Warwick 1980*. Под ред. David Rand и Lai-Sang Young. Т. 898. Lecture Notes in Mathematics. Berlin: Springer, 1981. Гл. 21, с. 366—381. ISBN: 978-3-540-11171-9. DOI: <10.1007/bfb0091924>. URL: <http://dx.doi.org/10.1007/bfb0091924>.
- [20] P. van Emde-Boas. *Another NP-complete partition problem and the complexity of computing short vectors in a lattice*. Report. Department of Mathematics. University of Amsterdam. Department, Univ., 1981.
- [21] Erik Grafarend, Silvelyn Zwanzig и Joseph Awange. «Integer Least Squares». B: *Applications of Linear and Nonlinear Models: Fixed Effects, Random Effects, and Total Least Squares*. Cham: Springer International Publishing, 2022, c. 579—603. ISBN: 978-3-030-94598-5. DOI: 10.1007/978-3-030-94598-5_16. URL: https://doi.org/10.1007/978-3-030-94598-5_16.
- [22] Johan Håstad. «Tensor rank is NP-complete». B: *Journal of Algorithms* 11.4 (1990), c. 644—654. ISSN: 0196-6774. DOI: [https://doi.org/10.1016/0196-6774\(90\)90014-6](https://doi.org/10.1016/0196-6774(90)90014-6). URL: <https://www.sciencedirect.com/science/article/pii/0196677490900146>.
- [23] Suresh Bolusani и др. *The SCIP Optimization Suite 9.0*. ZIB-Report 24-02-29. Zuse Institute Berlin, февр. 2024. URL: <https://nbn-resolving.org/urn:nbn:de:0297-zib-95528>.