

# Тензорная декомпозиция и прогноз для набора временных рядов

Сёмкин Кирилл<sup>1</sup>

<sup>1</sup>semkin.ki@phystech.edu

Вадим Стрижов<sup>2</sup>

<sup>2</sup>vadim.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 рекуррентными 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 *рекуррентных 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 строятся 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  $\mu$  - некоторый константный вектор,  $A_i$  матрицы  $m \times m$ ,  $u_t$  - 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.

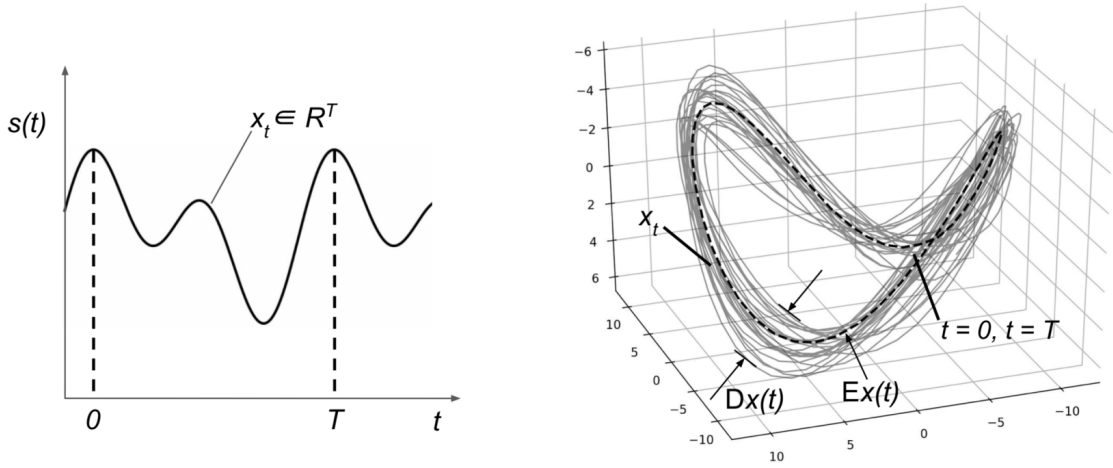


Рис. 1: Пример построения фазовой траектории временного ряда

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  $y \in X$ :

$$\begin{aligned} y(t+1) &= f(y(t)), \quad t \in \mathbb{N} \\ y(0) &= y_0 \end{aligned}$$

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 Такенса [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  $(\varphi \circ f^{t-L+1}(\mathbf{y}(t)), \dots, \varphi \circ f(\mathbf{y}(t)), \varphi \circ \mathbf{y}(t))^\top = (x(t-L+1) \dots x(t))^\top$ . Он называется *вектором задержки* ряда в момент  $t$  и обозначается  $\overleftarrow{\mathbf{x}}_t$ . 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  $\text{Lin}(\{\overleftarrow{\mathbf{x}}_t\})$ , в данном случае оно же *собственное пространство сигнала*  $x(t)$ , не должно иметь большую размерность. Соответственно ожидаем, что  $\text{Lin}(\{\overleftarrow{\mathbf{x}}_t\}) \subset \mathbb{R}^L$ . Ортонормированный базис в этом пространстве выбирается как  $U$ -компонента SVD-разложения матрицы векторов задержек — так называемой *траекторной матрицы*  $\mathbf{H}_x = [\overleftarrow{\mathbf{x}}_1 \dots \overleftarrow{\mathbf{x}}_{N-L+1}]$ .

## Взаимосвязь рядов и метод tSSA

Теперь обобщим данный подход на несколько временных рядов.

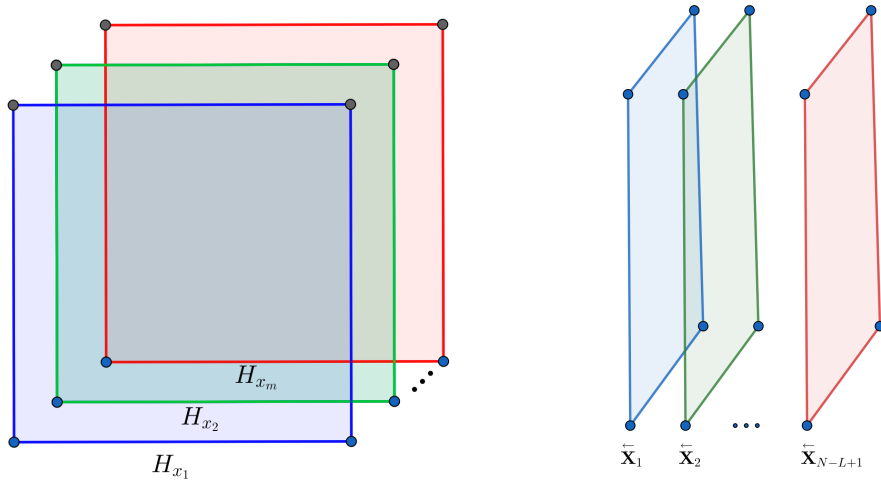


Рис. 2: Два вида на траекторный тензор. Слева — в виде набора траекторных матриц сигналов  $\{x_i(t)\}_{i=1}^m$ . Справа — в виде набора матриц задержки.

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:

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

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  $\sigma_{x_k}$ . 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  $\sigma_{x_k}$ : 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. Каждому временному 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).

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

Здесь  $\mathbb{I}_1 \sqcup \dots \sqcup \mathbb{I}_s = \{1, \dots, r\}$  — выбранные группы индексов. Четвёртое 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  $\text{Hankel}(B)$ .  $\square$

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

Несмотря на это, хочется группировать факторы так, что каждая матрица  $C_i$  была как можно более "ганкелевой". Это соответствует изначальной идее о связи разложений матриц и сигналов, а также повышает интерпретируемость компонент. Для дальнейшей оптимизации группировки введём меру "ганкелевости".

**Определение.** *Невязкой ганкелизации* матрицы  $\mathbf{H}$  назовём  $r(\mathbf{H}) = \|\mathbf{H} - \text{Hankel}(\mathbf{H})\|_F^2$

В статистическом плане, эта невязка соответствует суммарной дисперсии антидиагоналей матрицы.

Теперь для поиска декомпозиции ряда будем искать разбиение факторов, наилучшее в плане средней ганкелевой невязки:

$$\begin{aligned} &\frac{1}{s} \sum_{j=1}^s r(C_j) \rightarrow \min \\ &\begin{cases} C_j = \sum_{i \in \mathbb{I}_j} \sigma_{x_k}(i) \cdot \mathbf{a}_i \mathbf{b}_i^\top, \quad \forall j = 1 \dots s \\ \mathbb{I}_1 \sqcup \dots \sqcup \mathbb{I}_s = \{1, \dots, r\} \\ s \in \{2 \dots r\} \end{cases} \end{aligned}$$

В итоге имеем задачу дискретной оптимизации по множеству всех разбиений и, скорей всего, для неё нет быстрого алгоритма поиска решения [TODO NP-hard]. Поэтому предлагается следующая эвристика: имея оптимальное разбиение на данное число групп, продолжать разбивать каждую из них на две оптимальным внутри этой группы способом. Так можно продолжать до получения достаточного числа компонент. Данный принцип проиллюстрирован на рис.3. С точки зрения декомпозиции исходного ряда мы получаем его оптимальное разбиение на две компоненты, а затем продолжаем декомпозировать их независимо друг от друга. Получаем рекурсивную процедуру *DichotomyPartition*, которую можно адаптивно перезапускать для получения нужного количества компонент:

```

1      def DichotomyPartition(num_components):
2          """
3          Args:
4          num_components: how many components from initial signal
5                          is desired
6
7          Returns:
8          factor groups related to best signal decomposition
9          """
10         # at the start we have one group of factors,
11         # which forms hankel trajectory matrix
12         partition_list = [[1, 2, 3, ..., r]]
13         mean_hankel_resid = 0
14
15         # proceed until we obtain enough groups
16         while len(partition_list) < num_components:
17             new_partition_list = []
18
19             # dissect each group into two optimally within this group
20             # FullSearch test all possible partitions into 2 groups
21             # and find best in terms of Mean Hankel Residual
22             for partition in partition_list:
23                 sub_partition_1, sub_partition_2 = FullSearch(partition)
24
25                 new_partition_list.extend(sub_partition_1, sub_partition_2)
26
27             # calculate Mean Hankel Residual of new partition
28             mean_hankel_resid = CalculateResidual(new_partition_list)
29
30             yield new_partition_list, mean_hankel_resid
31
32             # save new partition
33             partition_list = new_partition_list
34

```

Листинг 1: Псевдокод процедуры рекурсивного получения декомпозиции временного ряда

## 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:

$$\begin{aligned}
 \overleftarrow{\mathbf{x}}_{N+1} = A\boldsymbol{\lambda} &\Leftrightarrow \begin{cases} \mathbf{x}_{kn} = A_{kn}\boldsymbol{\lambda} \\ x(N+1) = \mathbf{a}_{pr}^T \boldsymbol{\lambda} \end{cases}, \text{ где} \\
 A &= \begin{pmatrix} A_{kn} \\ \mathbf{a}_{pr}^T \end{pmatrix} \\
 \overleftarrow{\mathbf{x}}_{N+1} &= (\mathbf{x}_{kn} \ x(N+1))^T
 \end{aligned} \tag{3}$$

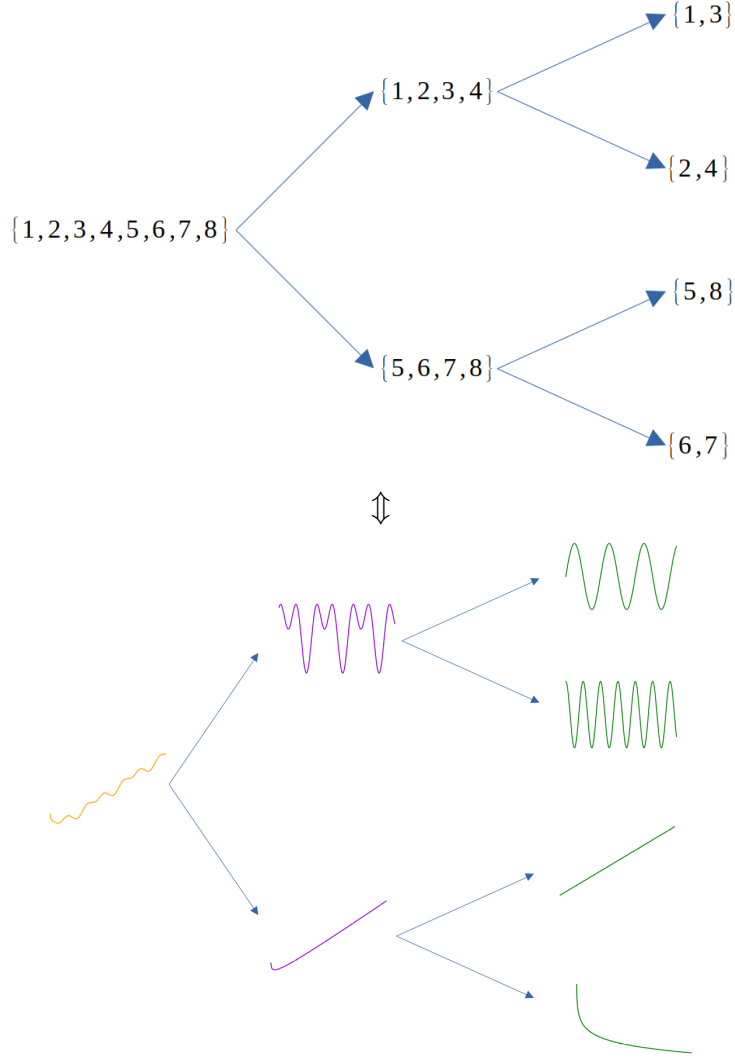


Рис. 3: Пример рекурсивного разбиения факторов (в виде индексов) на 4 группы. На каждом уровне ищется оптимальное разбиение внутри уже построенных группировок. С точки зрения разложения временных рядов мы рекурсивно раскладываем каждую полученную на данном уровне компоненту на две подкомпоненты.

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 (3). 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 формулой:

$$x(N+1) = \mathbf{a}_{pr}^T (A_{kn}^T A_{kn})^{-1} A_{kn}^T \mathbf{x}_{kn} \quad (4)$$

Здесь  $\mathbf{a}_{pr}^T (A_{kn}^T A_{kn})^{-1} A_{kn}^T = \mathbf{d}^T$  — 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.

Если расписать (4) в терминах  $x(t)$ :

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

Получаем, что предсказание методом tSSA, по-сути, сводится к *авторегрессионной*

модели ряда  $AR(L - 1)$ . Соответственно, поведение прогноза полностью определяется коэффициентами  $d_i$ .

### 3 Эксперимент



## 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=vzJOCgAAQBAJ>.
- [2] Estela Dagum и Silvia Bianconcini. *Seasonal Adjustment Methods and Real Time Trend-Cycle Estimation*. Авт. 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] В. Вольтерра. «Matematicheskaya teoriya bor by za sushchestvovanie». В: *Usp. Fiz. Nauk* 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». В: *Neurocomputing* 526 (2023), с. 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». В: *International Journal of Forecasting* 37.1 (2021), с. 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». В: янв. 2016, с. 333—346. ISBN: 978-1-137-41546-2. DOI: [10.1057/978-1-137-41547-9\\_15](https://doi.org/10.1057/978-1-137-41547-9_15).

- [14] Arnold Zellner. «An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias». В: *Journal of the American Statistical Association* 57.298 (1962), с. 348—368. DOI: [10.1080/01621459.1962.10480664](https://doi.org/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». В: *SIAM Review* 51 (авг. 2009), с. 455—500. DOI: [10.1137/07070111X](https://doi.org/10.1137/07070111X).
- [16] D. Stepanov и N. Golyandina. «SSA-based approaches to analysis and forecast of multidimensional time series». не определен. В: *Proceedings of the 5th St.Petersburg Workshop on Simulation*. 2005, с. 293—298.
- [17] Hamid KRIM. «Two Decades of Array Signal Processing Research». В: *IEEE Signal Processing Magazine* 0 (1994), с. 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». В: *International Journal of Energy and Statistics* 01 (апр. 2013). DOI: [10.1142/S2335680413500051](https://doi.org/10.1142/S2335680413500051).
- [19] Floris Takens. «Detecting Strange Attractors in Turbulence». В: *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](https://doi.org/10.1007/bfb0091924). URL: <http://dx.doi.org/10.1007/bfb0091924>.