

МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)

На правах рукописи



ТИХОНОВ ДЕНИС МАКСИМОВИЧ

СНИЖЕНИЕ РАЗМЕРНОСТИ
В ЗАДАЧАХ МУЛЬТИМОДЕЛИРОВАНИЯ ДИНАМИЧЕСКИХ СИСТЕМ

1.2.1 — Искусственный интеллект и машинное обучение

Диссертация на соискание ученой степени
кандидата физико-математических наук

Научный руководитель:
д.ф.-м.н. В. В. Стрижов

Москва — 2026

Оглавление

	Стр.
Введение	4
Глава 1. Постановка задачи восстановления пространства состояний дина- мической системы	12
1.1. Динамическая система, обыкновенные дифференциальные уравне- ния и наблюдаемые временные ряды	12
1.2. Линейные методы восстановления пространства состояний	16
1.3. Нелинейные методы восстановления пространства состояний	21
Глава 2. Полилинейные методы восстановления пространства состояний	24
2.1. Полилинейная динамическая система	25
2.2. Обобщение полилинейной модели на многомерные данные	32
2.3. Анализ полилинейных методов снижения размерности	41
Глава 3. Полилинейные методы объединения пространств состояний	49
3.1. Задача объединения пространств состояний	49
3.2. Нейросетевые модели построения пространства состояний	50
3.3. Анализ методов объединения пространств состояний	53
Глава 4. Разреженное восстановление системы обыкновенных дифференци- альных уравнений	56
4.1. Задача восстановления системы дифференциальных уравнений . . .	56
4.2. Методы восстановления вида системы дифференциальных уравнений	60
4.3. Полилинейный метод выбора функций кандидатов	62
4.4. Анализ восстановленных системы дифференциальных уравнений . .	66
Глава 5. Перенос пространства состояний внутри класса динамических си- стем	70

5.1. Полилинейная параметризация семейства проекций статическими параметрами	72
5.2. Параметризация перехода между пространствами состояний матрицей Якоби	76
5.3. Анализ полилинейной параметризации семейства проекций	78
5.4. Анализ перехода между пространствами состояний матрицей Якоби	81
Заключение	84
Список основных обозначений	86
Список иллюстраций	88
Список таблиц	91
Список литературы	92

Введение

Актуальность темы.

В работе исследуется проблема снижения размерности в задачах моделирования динамических систем по наблюдаемым временным рядам. Методы снижения размерности строят подпространство состояний исследуемой системы. В полученном подпространстве модели прогнозирования являются более устойчивыми в сравнении с моделями в исходных пространствах и имеют меньшее число параметров.

Пространство состояния системы восстанавливается с помощью векторизации временных рядов. Под векторизацией подразумевается *метод временных задержек*, который ставит в соответствие значению временного ряда его предысторию конечной длины. Векторное представление временных рядов и их синхронное наблюдение приводят к проблеме мультикорреляции. При высокой мультикорреляции модели прогнозирования оказываются неустойчивыми. Для построения простой, устойчивой и точной модели применяются методы снижения размерности [1–4].

Методы снижения размерности, применяемые при анализе динамических систем, строят низкоразмерные пространства состояний, сохраняющие основные топологические и динамические свойства исходной системы. Проекция системы на такое подпространство упрощает модели прогнозирования. Метод главных компонент и сингулярное разложение сохраняют дисперсию линейных признаков [1–3, 5]. Для моделей временных рядов аналогичным образом применяется собственное ортогональное разложение [6–8]. В задачах управления размерность пространств снижается выбором оптимального подмножества координат и проекцией на подпространство [4, 9–11]. Переход к подпространству упрощает интерпретацию и анализ моделей [12, 13].

Восстановление пространства состояний по наблюдаемым временным рядам опирается на теоремы о вложении. Теорема Уитни определяет достаточную раз-

мерность пространства для гладкого многообразия [14]. Теорема Такенса показывает, что для *почти всех функций* наблюдения метод задержек одномерного временного ряда строит пространство, диффеоморфное истинному пространству состояний [15, 16]. Для набора временных рядов результат обобщен в многомерном методе задержек [17, 18]. Параметры метода задержек, временная задержка и длина окна, выбираются методом минимума взаимной информации и методом ложных ближайших соседей [19–22]. Условия сохранения геометрических свойств и расстояний получены позднее в работе [23]. Эти результаты гарантируют, что восстановленное пространство состояний сохраняет топологические свойства истинной динамической системы.

Линейная проекция координат восстановленного пространства состояний не учитывает нелинейные связи компонент. Метод главных компонент, глобальное вложение по главным компонентам и конкатенация нескольких рядов строят линейные комбинации [3, 24, 25]. Синхронные временные ряды для восстановления пространства состояний при этом избыточны: каждый ряд повторяет схожие зависимости [18, 26]. Методы, использующие линейное объединение векторов задержек без учета внутренних зависимостей временных рядов, требуют большую размерность пространства для построения устойчивых модели прогнозирования. Выбор подмножества компонент вектора задержек есть частный случай снижения размерности, снижающий число параметров, но не учитывающий нелинейные зависимости [27].

Динамическую систему в получаемых пространствах описывают линейными и нелинейными моделями. Динамическая модовая декомпозиция и теория оператора Купмана [28–30] заменяют нелинейный *поток* линейным оператором в пространстве признаков большой размерности [31–33]. Нелинейная декомпозиция и нейросетевые автоэнкодеры подбирают это пространство по наблюдаемым временным рядам [34–37]. Нейросетевые модели дают точный прогноз, но не восстанавливают уравнения в исходных переменных [38–41].

Существенной задачей остаётся построение интерпретируемых моделей по

наблюдаемым данным при слабых априорных предположениях о структуре системы. Разреженное восстановление нелинейной динамической системы восстанавливает явный интерпретируемый вид системы дифференциальных уравнений, описывающих динамическую систему [42–45]. Нелинейные зависимости учитываются в модели заранее заданным набором нелинейных преобразований. Сложность модели растёт экспоненциально при увеличении длины окна метода задержек и числа исследуемых преобразований. Тензорные разложения [46, 47] записывают те же произведения меньшим числом параметров [48–50].

В диссертации решается задача снижения размерности пространства состояний по набору рядов. Предлагается полилинейное отображение векторов задержек, дополненных константой. Линейные задержки сохраняются и дополняются произведения компонент набора временных рядов. Число параметров снижается тензорным разложением. Для восстановления пространств состояний используются линейные и нелинейные методы. Лемма о полилинейном расширении показывает, что предлагаемое преобразование векторов задержек является вложением ровно тогда, когда вложением является многомерный метод задержек. Теорема о многомерном полилинейном вложении устанавливает достаточные условия на ранг канонического разложения и на размерность восстановленного пространства. Доказаны включение линейной векторной авторегрессии в класс полилинейных моделей и липшицевость проекции, параметризованной статическими параметрами системы.

Цели работы.

1. Исследовать способы решения задачи восстановления пространства состояний динамических систем по набору временных рядов с использованием инструментов тензорной алгебры.
2. Теоретически обосновать применимость полилинейных отображений, снижающих размерность пространств состояний, для многомерных временных рядов.
3. Предложить методы снижения размерности пространств состояний, обобщающие метод анализа сингулярного спектра и учитывающие внутренние зависимости компонент временных рядов.
4. Предложить способ построения интерпретируемой модели, восстанавливающей неизвестные зависимости динамических систем с малым числом параметров.

Основные положения, выносимые на защиту.

1. Исследована проблема восстановления пространств состояний динамической системы по многомерным наблюдаемым временным рядам. Предложен полилинейный метод снижения размерности пространств состояний.
2. Исследованы свойства предлагаемых методов. Доказаны теоремы о полилинейном вложении метода задержек для многомерных временных рядов.
3. Предложены полилинейные методы снижения размерности набора пространств состояний. Предлагаемые методы сохраняют свойства исходных систем и учитывают нелинейные связи наблюдаемых временных рядов.
4. Предложен метод построения интерпретируемой нелинейной модели, восстанавливающий приближённый вид дифференциальных уравнений, описывающих динамическую систему.
5. Разработан алгоритм построения полилинейной модели с малым числом параметров, оптимизированной с помощью тензорных разложений. Про-

ведены вычислительные эксперименты, подтверждающие эффективность предложенных методов.

Методы исследования. Используются линейные и нелинейные методы восстановления пространств состояний динамической системы. Для восстановления пространства состояний используются методы, основанные на теореме Такенса. Для отбора интерпретируемых нелинейных признаков используются методы тензорной алгебры. Для снижения размерности пространств состояний используются метод главных компонент и нейросетевые методы. Для оптимизации моделей прогнозирования в пространстве состояний используются составные функции потерь.

Научная новизна. Разработан полилинейный метод восстановления пространства состояний динамической системы. Предложен метод объединения подпространств состояний нескольких временных рядов на основе полилинейного отображения, учитывающий нелинейные зависимости между рядами. Разработаны методы снижения числа параметров модели на основе тензорных разложений, адаптированные к применению в нейросетевых моделях. Разработано полилинейное расширение метода разреженного восстановления дифференциальных уравнений динамической системы, обеспечивающее линейный рост числа параметров и интерпретируемость модели.

Теоретическая значимость. Доказана теорема о полилинейном вложении метода задержек для многомерного временного ряда в случае канонического разложения. Доказана теорема о включении линейной векторной авторегрессии в класс полилинейных моделей. Доказана теорема о липшицевой непрерывности параметризованного полилинейного отображения.

Практическая значимость. Предложенные в работе методы предназначены для восстановления пространств состояний динамических систем по временным рядам в условиях ограниченных данных и зашумлённых измерений. Разработанные методы используются в задачах прогнозирования состояний динамических систем и строят интерпретируемые модели с малым числом параметров.

Степень достоверности и апробация работы. Достоверность результатов подтверждена математическими доказательствами, экспериментальной проверкой полученных методов на задачах восстановления динамических систем по временным рядам, а также публикациями результатов исследования в рецензируемых научных изданиях, в том числе рекомендованных ВАК. Результаты работы докладывались и обсуждались на следующих научных конференциях:

1. Объединение пространств состояний динамической системы через полилинейное отображение, Международная конференция «Иванниковские чтения», Москва, 2026.
2. Разреженное восстановление функции эволюции связанных динамических систем, 68-я Всероссийская научная конференция МФТИ, Москва, 2026.
3. Dynamic system reconstruction from multivariate time series via multilinear map and time delay embedding, International Conference ICOMP, Иннополис, 2024.
4. Восстановление зашумленных значений многомерных временных рядов с помощью тензорного метода задержек и тензорного разложения, 66-я Всероссийская научная конференция МФТИ, Москва, 2024.
5. Классификация квазипериодических временных сигналов с помощью сферических гармоник, Международная научная конференция студентов, аспирантов и молодых учёных «Ломоносов-2022», Москва, 2022.

Публикации по теме диссертации. Основные результаты по теме диссертации изложены в 6 изданиях, 3 из которых изданы в журналах, рекомендованных ВАК, 3 — в тезисах докладов.

Личный вклад. Все приведенные результаты, кроме отдельно оговоренных случаев, получены диссертантом лично при научном руководстве д.ф.-м.н. В. В. Стрижова. В работах [51–53] Д. М. Тихоновым разработаны полилинейные методы восстановления пространства состояний по многомерным временным рядам, в том числе по зашумлённым наблюдениям, и метод объединения таких пространств, а также проведён анализ полученных результатов. В работах [54–56] Д. М. Тихоновым теоретически обоснована применимость предлагаемых отображений к восстановлению дифференциальных уравнений и к переносу пространства состояний внутри класса систем, а также развиты и исследованы математические модели, методы и вычислительные алгоритмы с реализацией в виде программных решений и их приложениями к модельным динамическим системам и рядам акселерометра.

Структура и объем работы. Диссертация состоит из оглавления, введения, пяти разделов, заключения, списка иллюстраций, списка таблиц, перечня основных обозначений и списка литературы из 87 наименований. Основной текст занимает 101 страницу.

Краткое содержание работы по главам. Глава 1 посвящена постановке задачи восстановления пространства состояний динамической системы по наблюдаемым временным рядам. Приведён обзор метода задержек, метода главных компонент, глобального вложения по главным компонентам и нелинейных подходов: нейросетевых методов, оператора Купмана и нейронных обыкновенных дифференциальных уравнений.

В главе 2 представлены полилинейные методы восстановления пространства состояний. Вводятся полилинейная динамическая система и тензорное произве-

ление, построено обобщение на многомерные временные ряды. Число параметров снижается каноническим разложением. Лемма о полилинейном расширении сводит полилинейное отображение к многомерному методу задержек. Теорема о многомерном полилинейном вложении ограничивает ранг канонического разложения и размерность восстановленного пространства. Доказано включение линейной векторной авторегрессии в класс полилинейных моделей.

В главе 3 полилинейное отображение объединяет нейросетевые представления нескольких временных рядов. Каноническое разложение снижает число параметров тензора связей. Приведены результаты восстановления рядов акселерометра при ходьбе и спуске по лестнице.

В главе 4 предложен тензорный метод для восстановления аналитического вида дифференциальных уравнений. Тензорное расширение использует каноническое разложение: число параметров растёт линейно, модель остаётся интерпретируемой. Слагаемые отбираются разреживанием фактор-матриц, уравнения записываются в исходных переменных.

В главе 5 пространство состояний системы класса восстанавливается по короткому ряду и статическим параметрам: семейство проекций параметризуется полилинейным отображением, а переход между пространствами состояний двух систем задаётся матрицей Якоби. Доказана липшицевость параметризованной проекции по статическим параметрам. На рядах акселерометра новых испытуемых пространство состояний строится по короткому ряду.

Глава 1

Постановка задачи восстановления пространства состояний динамической системы

В данной главе ставится задача восстановления пространства состояний динамической системы по наблюдаемым временным рядам. Вводятся динамическая система и метод временных задержек, обоснованный теоремой Такенса. Приводится обзор линейных методов: метод задержек, метод главных компонент и их сочетание. Затем рассматриваются нелинейные методы: оператор Купмана, нейросетевые модели и разреженное восстановление вида дифференциальных уравнений.

1.1. Динамическая система, обыкновенные дифференциальные уравнения и наблюдаемые временные ряды

Определение 1. *Классическая динамическая система* [57–60] $(\mathcal{M}, \mu, \phi_t)$ — набор из гладкого многообразия $\mathcal{M}$, меры μ на $\mathcal{M}$ с непрерывной положительной плотностью и однопараметрической группы ϕ_t диффеоморфизмов многообразия $\mathcal{M}$, сохраняющих меру $\mu(A) = \mu(\phi_t A)$ при всех t и всех измеримых $A \subset \mathcal{M}$. Параметр t принимает значения в $\mathbb{R}$ (непрерывное время) или в $\mathbb{Z}$ (дискретное время).

В локальных координатах при $t \in \mathbb{R}$ действие ϕ_t задается гладкими функциями

$$\frac{d}{dt}\mathbf{v}(t) = \mathbf{f}(\mathbf{v}(t), t; \mathbf{c}), \quad \mathbf{v} \in U \subseteq \mathbb{R}^{d_s}, \quad \mathbf{f} : U \rightarrow \mathbb{R}^{d_s}, \quad (1.1)$$

где $\mathbf{v}$ — состояние системы, $\mathbf{f}$ — векторное поле, зависящее от состояния, времени t и вектора параметров $\mathbf{c}$.

Определение 2. *Фазовое пространство или пространство состояний* — это пространство U всех возможных состояний системы $\mathbf{v}$. Каждая точка фазового пространства полностью описывает состояние динамической системы в момент времени t .

Определение 3. *Фазовая траектория* — кривая $\{\mathbf{v}(t, \mathbf{v}_0)\}$ в фазовом пространстве, то есть последовательность состояний $\{\mathbf{v}(t) : t \in \mathcal{T}\}$ на интервале времени $\mathcal{T}$ при начальном условии $\mathbf{v}(0) = \mathbf{v}_0$.

Примером служит аттрактор Лоренца [61], заданный нелинейной системой обыкновенных дифференциальных уравнений

$$\begin{cases} \dot{x} = \sigma(y - x) \\ \dot{y} = x(\rho - z) - y, \\ \dot{z} = xy - \beta z \end{cases}$$

где параметры не зависят от времени. В классическом случае $\sigma = 10$, $\rho = 28$ и $\beta = 8/3$. Вектор состояния имеет вид $\mathbf{v} = [x, y, z]^\top$, вектор параметров — $\mathbf{c} = [\sigma, \rho, \beta]^\top$. Система Лоренца — одна из простейших изученных динамических систем с хаотическим поведением: малые изменения начальных условий меняют траекторию [62, 63]. При малом отличии начальных данных фазовые траектории быстро расходятся, что ограничивает точность прогноза.

Частным случаем (1.1) являются *автономные системы*, не зависящие в явном виде от времени:

$$\frac{d}{dt}\mathbf{v} = \mathbf{f}(\mathbf{v}) \quad (1.2)$$

Решение — гладкое отображение $\phi : T \rightarrow U$ оси времени в фазовое пространство:

$$\frac{d}{dt}\phi(t) = \mathbf{f}(\phi(t)).$$

В прикладных задачах встречаются дискретные системы, порождённые непрерывными. Такие системы (1.1) и (1.2) сводятся к дискретному виду путём дискретизации траектории с шагом по времени Δt , так что $\mathbf{v}_k = \mathbf{v}(k\Delta t)$:

$$\mathbf{v}_{k+1} = \mathbf{F}_{\Delta t}(\mathbf{v}_k).$$

Дискретный *оператор эволюции* $\mathbf{F}_{\Delta t}$ определяется выбранным шагом Δt . Для произвольного момента времени t_k и Δt с заданными начальными услови-

ями $\mathbf{v}(t_0) = \mathbf{v}_0 = \text{const}$ отображение $\mathbf{F}_{\Delta t}$ задаётся выражением:

$$\mathbf{v}_{k+1} = \mathbf{F}_{\Delta t}(\mathbf{v}_k) = \mathbf{v}_k + \int_{t_k}^{t_k+\Delta t} \mathbf{f}(\mathbf{v}(s)) ds.$$

Дискретное представление согласовано с экспериментальными данными: показания устройств, датчиков и камер снимаются с заданной частотой. Далее $\mathbf{F}_{\Delta t}$ обозначается $\mathbf{F}$: шаг Δt фиксирован экспериментом.

На практике $\mathbf{v}_t$ редко доступно или даже известно. Исходное фазовое пространство $\mathbf{v}_t$ связано с временным рядом $\{\mathbf{v}_{\text{obs } t}\}_{t=1}^N$, полученным в эксперименте, посредством гладкой функции наблюдения $\mathbf{h} : \mathcal{M} \rightarrow \mathbb{R}^{d_o}$, как

$$\mathbf{v}_{\text{obs } t} = [x_{1,t}, \dots, x_{d_o,t}]^\top = [h_1(\mathbf{v}_t), \dots, h_{d_o}(\mathbf{v}_t)] = \mathbf{h}(\mathbf{v}_t),$$

где $x_{i,t}$ — скалярные значения временного ряда в момент времени t , $i = 1, \dots, d_o$.

Число наблюдаемых временных рядов $d_o = 1$ для одномерного случая, а функция наблюдения обозначается как $h : \mathcal{M} \rightarrow \mathbb{R}$, $d_o \geq 2$ для многомерного случая.

Определение 4. *Функция лаговых координат* $\Phi_{h,F,\tau} : \mathcal{M} \rightarrow \mathbb{R}^n$ в одномерном случае сопоставляет значению наблюдаемого временного ряда его предыдущие во времени и определяется как

$$\begin{aligned} \Phi_{h,F,\tau}(\mathbf{v}_k) &= [h(\mathbf{F}^{\circ(-(n-1)\tau)}(\mathbf{v}_k)), \dots, h(\mathbf{F}^{-\tau}(\mathbf{v}_k)), h(\mathbf{v}_k)]^\top \\ &= [x_{k-(n-1)\tau}, \dots, x_{k-\tau}, x_k]^\top, \end{aligned} \tag{1.3}$$

с параметрами τ — задержка (лаг) и n — размерность вложения, операция $\mathbf{F}^{\circ i}$ обозначает i -ю итерацию отображения, например, $\mathbf{F}^{\circ 3}(\mathbf{v}_k) = \mathbf{F}(\mathbf{F}(\mathbf{F}(\mathbf{v}_k))) = \mathbf{v}_{k+3}$.

Метод временных задержек отображает скалярный временной ряд $\{x_t\}_{t=1}^N$ в пространство более высокой размерности функцией лаговых координат $\Phi_{h,F,\tau}$ и формирует вектор временных задержек

$$\mathbf{x}_k = [x_{k-(n-1)\tau}, \dots, x_{k-\tau}, x_k]^\top.$$

Матрица, образованная векторами задержек, называется *траекторной матрицей* временного ряда $\mathbf{x}$ и определяется как

$$\mathbf{X} = \begin{bmatrix} x_1 & \dots & x_n \\ x_2 & \dots & x_{n+1} \\ \vdots & \ddots & \vdots \\ x_m & \dots & x_N \end{bmatrix}^T = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m], \quad m = N - n + 1, \quad (1.4)$$

где n — ширина окна, N — длина временного ряда, а τ равен 1.

Теорема 1 (Такенса для диффеоморфизмов [16]). Пусть $\mathcal{M}$ — компактное многообразие размерности d_s , g — дважды непрерывно дифференцируемый диффеоморфизм в себя и h — дважды непрерывно дифференцируемая функция, определенная на многообразии $\mathcal{M}$ и принимающая значения $\mathbb{R}$ и $n \geq 2d_s + 1$. Тогда для почти всех функций $h \in C^2(\mathcal{M}, \mathbb{R})$ и $g \in \text{Diff}^2(\mathcal{M})$ отображение $\Phi_{h,g} : \mathcal{M} \rightarrow \mathbb{R}^{2d_s+1}$, определяемое как

$$\Phi_{h,g}(\mathbf{v}_k) = [h(\mathbf{g}^{\circ n}(\mathbf{v}_k)), \dots, h(\mathbf{v}_k)]^T,$$

является биекцией и вложением $\mathcal{M}$ в $\mathbb{R}^{2d_s+1}$.

В случае дискретной динамической системы как исходной, так и порожденной из непрерывной требуется ввести дополнительные условия для сохранения необходимых свойств.

Теорема 2 (Такенса для дискретного случая [16, 17]). Пусть $\mathcal{M}$ — компактное многообразие размерности d_s , $\mathbf{F} : \mathcal{M} \rightarrow \mathcal{M}$ — диффеоморфизм, полученный дискретизацией потока, со свойствами:

- множество неподвижных точек конечно;
- отсутствуют периодические точки с периодом τ и 2τ ;
- для любого $k \geq 3$ множество периодических точек с периодом $k\tau$ конечно и в каждой такой точке собственные значения дифференциала $D\mathbf{F}^{\circ k}$ различны и не равны 1.

Тогда для почти всех функций $h \in C^2(\mathcal{M}, \mathbb{R})$ отображение (1.3) при $n \geq 2d_s + 1$

$$\Phi_{h, \mathbf{F}, \tau}(\mathbf{v}_k) = [h(\mathbf{F}^{\circ(-(n-1)\tau)}(\mathbf{v}_k)), \dots, h(\mathbf{F}^{\circ(-\tau)}(\mathbf{v}_k)), h(\mathbf{v}_k)]^\top,$$

является биекцией и вложением $\mathcal{M}$ в $\mathbb{R}^{2d_s+1}$.

Определение 5. Почти все функции — это такие функции $h_\alpha = h_0 + \sum_{i=1}^{n_\alpha} \alpha_i h_i$, где $\{h_i\}_{i=1}^{n_\alpha}$ — базис многочленов степени меньше или равной $2n$ и h_0 — произвольная гладкая функция, для которых для почти всех $\alpha \in \mathbb{R}^{n_\alpha}$ по мере Лебега выполняются условия теорем Такенса.

Иными словами, наборы параметров, для которых теорема неверна, имеют меру Лебега нуль, что соответствует вероятностной интерпретации: при случайном выборе коэффициентов с любой абсолютно непрерывной плотностью вероятность получить неподходящую функцию h_α равна нулю.

Важное свойство вложения — сохранение характеристик исходной системы: корреляционной размерности, характеристических показателей, описывающих чувствительность к начальным условиям, а также периодичности, квазипериодичности, затухания и петель траектории.

Теорема Такенса связывает измеряемые временные ряды с истинным фазовым пространством. При этом выбор n нетривиален, а получаемое фазовое пространство (1.4) имеет высокую размерность. Необходимо снизить размерность получаемых пространства.

1.2. Линейные методы восстановления пространства состояний

Прямой выбор параметров вложения оптимизирует критерии для множества лагов τ и повторяется для увеличивающихся n . Используются метод ложных ближайших соседей и нелинейный прогноз по ближайшим соседям [26, 64].

Основная идея заключается в определении критерия для нахождения *некорректной проекции* в итеративном алгоритме:

1. Найти ближайшего соседа $\mathbf{x}_{\text{nn } t}^d \in \mathbb{R}^d$ для каждой точки $\mathbf{x}_t^d \in \mathbb{R}^d$,

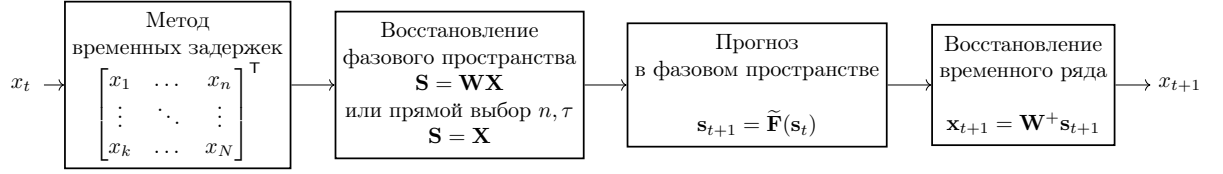


Рис. 1.1: Упрощённая схема получения фазового пространства в одномерном случае. Показаны три подхода: одномерный метод временных задержек, прямой выбор n и Глобальное вложение по главным значениям.

2. Вычислить расстояние в фазовом пространстве $\|\mathbf{x}_t^d - \mathbf{x}_{nn\ t}^d\|$,
3. Увеличить размерность вектора временных задержек на 1: $\mathbf{x}_t^{d+1}, \mathbf{x}_{nn\ t}^{d+1} \in \mathbb{R}^{d+1}$,
4. Вычислить расстояние в фазовом пространстве $\|\mathbf{x}_t^{d+1} - \mathbf{x}_{nn\ t}^{d+1}\|$,
5. Проверить критерий для истинного/ложного ближайшего соседа,
6. Повторять, пока все ближайшие соседи не станут истинными.

Точки $\mathbf{x}_t^{d+1}, \mathbf{x}_{nn\ t}^{d+1}$ — ложные ближайшие соседи, если отношение расстояний больше некоторого экспертно заданного порога $r_{\text{threshold}}$

$$r(\mathbf{x}_t^d, \mathbf{x}_t^{d+1}) = \frac{\|\mathbf{x}_t^d - \mathbf{x}_{nn\ t}^d\|}{\|\mathbf{x}_t^{d+1} - \mathbf{x}_{nn\ t}^{d+1}\|} > r_{\text{threshold}}.$$

Для снижения размерности применяется метод главных компонент [1–3, 5, 65]. Для траекторной матрицы та же проекция совпадает с разложением по собственным ортогональным модам [6–8]. Пусть $\mathbf{X} \in \mathbb{R}^{n \times m}$ — центрированная траекторная матрица векторов задержек из (1.4). Тогда ее сингулярное разложение определяется как

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \quad (1.5)$$

где $\mathbf{U}$ — матрица ортонормированных собственных векторов $\mathbf{X}\mathbf{X}^T$, $\mathbf{V}$ — матрица ортонормированных собственных векторов $\mathbf{X}^T\mathbf{X}$, $\mathbf{\Sigma}$ — диагональная матрица сингулярных чисел. С помощью p главных сингулярных чисел на основе матрицы $\mathbf{U}$, строится проекция в подпространство:

$$\mathbf{S} = \mathbf{W}\mathbf{X} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m], \quad \mathbf{s}_i \in \mathbb{R}^p, \quad p \ll n. \quad (1.6)$$

Параметр p выбирается по доле объяснённой дисперсии при последовательном увеличении размерности подпространства. Для периодического ряда p увеличивают до исчезновения *самопересечений*: одной области фазового пространства не должны соответствовать фрагменты ряда с различной фазой. Объединение PCA (1.5) с методом задержек (1.4) даёт *вложение глобальных главных компонент*.

В многомерном случае вектор временных задержек содержит каждый из рядов $\{x_{i,t}\}_{t=1}^N$, $i = 1, \dots, d_o$, и объединяет их в одно фазовое пространство:

$$\mathbf{x}_t = [\mathbf{x}_{1,t}, \dots, \mathbf{x}_{d_o,t}]^T = [x_{1,t}, \dots, x_{1,t-(n_1-1)\tau_1}, \dots, x_{d_o,t}, \dots, x_{d_o,t-(n_{d_o}-1)\tau_{d_o}}]^T.$$

Размерность n_{tot} равна сумме размерностей вложения отдельных рядов: $n_{\text{tot}} = \sum_{i=1}^{d_o} n_i$. Таким образом, исходное многомерное фазовое пространство получается прямым объединением фазовых пространств каждого временного ряда.

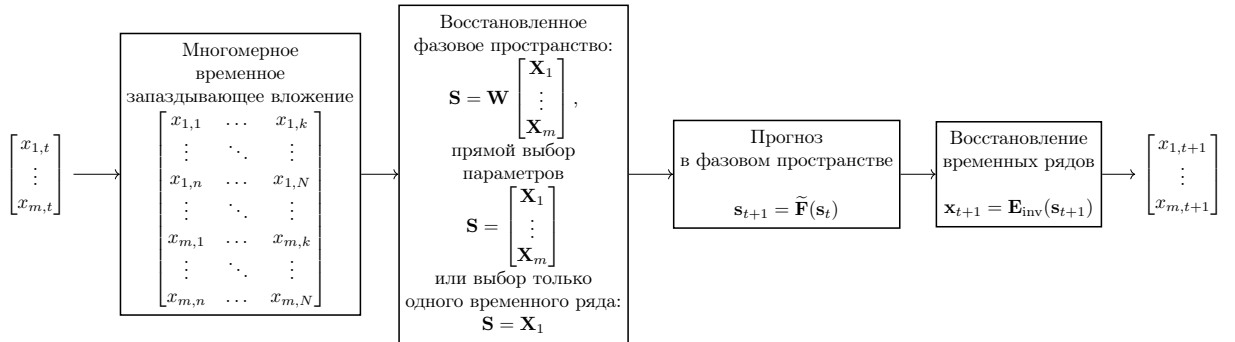


Рис. 1.2: Упрощённая схема получения фазового пространства в многомерном случае. Показаны четыре подхода: метод временных задержек, прямой выбор n , главные глобальные компоненты или реконструкция с использованием одного выбранного временного ряда. **Примечание:** на этой схеме матрица траектории не транспонирована, в отличие от рисунка 1.1.

Простейший подход — выбор одного временного ряда, который раскрывает фазовое пространство динамической системы лучше других. Альтернативно, процесс сводится к одномерному случаю как есть. Подходы, основанные на ближайших соседях, итеративно увеличивают размерность вложения каждого

временного ряда, получая наилучшее приближение. Поиск размерности искомого подпространства ведётся независимо для каждого временного ряда либо одновременно по объединённому вектору [25]. Упрощённая схема реконструкции фазового пространства для многомерного случая показана на рисунке 1.2.

Вышеописанные подходы сталкиваются с рядом проблем:

1. *проблема нахождения параметров*,
2. *проблема избыточной информации* для многомерного случая,
3. *несинхронизированность временных рядов* в многомерном случае.

Перечисленные задачи далее решаются полилинейной моделью.

В одномерном и многомерном случаях основная трудность — *выбор параметров*. Два параметра метода временных задержек: τ — задержка (лаг), и n — размер вложения, которые определяют фазовое пространство. Размерность результирующего фазового пространства часто очень высока, а реальная размерность вложения часто значительно меньше. Отображение Хенона [66] задаётся системой

$$\begin{cases} x_{t+1} = a + by_t + cx_t^2, \\ y_{t+1} = dx_t, \end{cases}$$

где $a = 1, b = 1, c = -1.4, d = 0.3$ — параметры. Подходящие параметры вложения для реконструкции фазового пространства — это $n = 2$ и $\tau = 1$. Рисунок 1.3 показывает три примера различных подходов к реконструкции. Фазовое пространство строится с использованием каждого временного ряда x_t и y_t при построении вложении размерности два:

$$\begin{aligned} x_t : \{x_t, x_{t-1}\} &\rightarrow \left\{x_t, \frac{1}{d}y_t\right\}, \\ y_t : \{y_t, y_{t-1}\} &\rightarrow \{dx_{t-1}, y_{t-1}\}, \end{aligned}$$

Однако теорема Такенса даёт верхнюю границу $n = 2d_s + 1 = 5$.

Для многомерного случая это пример *избыточной информации*. Добавление вектора задержек y_t к вектору x_t не улучшает качество восстановленного фазового пространства из-за линейной зависимости между x_{t-1} и y_t .

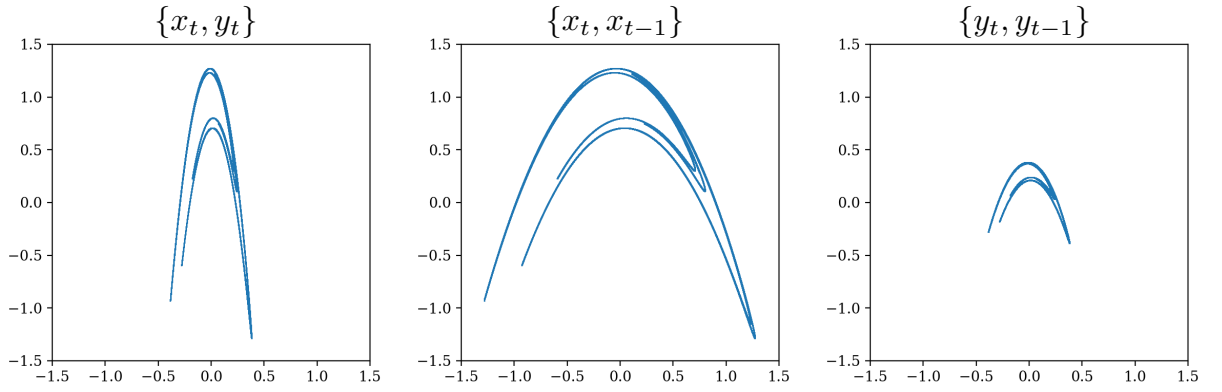


Рис. 1.3: Пример восстановленного пространства состояний из различных вложений: исходные координаты (слева), метод задержек с временным рядом x_t (в центре), метод задержек с временным рядом y_t (справа). Чтобы подчеркнуть линейную зависимость между результирующими многообразиями, все рисунки показаны в одинаковых масштабах по осям.

Проблема избыточной информации возникает в многомерных временных рядах из-за синхронности наблюдений. Это приводит к высокой корреляции и избыточности информативных компонент.

Несинхронизированные ряды с разными шагами дискретизации требуют снижения размерности фазового пространства для рядов с более высокой частотой за счёт пропуска значений либо повышения размерности для рядов с более низкой частотой за счёт расширения окна. Это критично для методов, основанных на расстояниях в фазовых пространствах, из-за проклятия размерности.

1.3. Нелинейные методы восстановления пространства состояний

Рекуррентные нейронные сети и их модификации с долгой краткосрочной памятью (LSTM) и управляемыми рекуррентными блоками (GRU) решают прямую задачу прогноза следующих значений временного ряда. Рекуррентная сеть отображает вектор задержек $\mathbf{x}_t$ в скрытое состояние $\mathbf{h}_t$:

$$\mathbf{h}_t = \sigma(\mathbf{W}_h \mathbf{h}_{t-1} + \mathbf{W}_x \mathbf{x}_t + \mathbf{b}),$$

где σ — нелинейная функция активации, $\mathbf{W}_h, \mathbf{W}_x$ — матрицы весов, а $\mathbf{b}$ — вектор смещения. Такие модели являются *чёрными ящиками* из-за сложной интерпретации внутреннего состояния. Внутреннее состояние $\mathbf{h}_t$ получается в результате оптимизации модели, минимизирующие ошибки прогноза, и не учитывает гладкость, непрерывность и устойчивость, требуемые для восстановленных пространств состояний.

Нейросетевые методы, учитывающие специфику задачи моделирование динамических систем, разделяют задачу на три части: *энкодер*, отображающий наблюдаемый ряд в пространство состояний $\mathbf{s}_t$, *модель динамики*, аппроксимирующую динамическую систему, и *декодер*, восстанавливающий исходный ряд из пространства состояний. Математически схема записывается так:

$$\begin{aligned}\mathbf{s}_t &= \text{Encoder}(\mathbf{x}_{\leq t}), \\ \mathbf{s}_{t+1} &= \text{Dynamics}(\mathbf{s}_t), \\ \hat{\mathbf{x}}_t &= \text{Decoder}(\mathbf{s}_t).\end{aligned}$$

Другие методы моделирования динамических систем используют переход в пространства большей размерности, например, оператор Купмана [28–32, 35]. Нелинейная система в конечномерном пространстве состояний рассматривается как линейная в бесконечномерном пространстве. Пусть $\mathbf{x}_{t+1} = \mathbf{F}(\mathbf{x}_t)$ — нелинейное отображение. Оператор Купмана $\mathcal{K}$ действует на временные ряды

и функции наблюдения h следующим образом:

$$\mathcal{K}h = h \circ \mathbf{F}.$$

Нейронные сети строят конечномерное приближение этого оператора и описывают нелинейную динамику линейной моделью в подпространстве малой размерности. Архитектуры класса *Deep Koopman* используют автоэнкодеры для построения линейных зависимостей в скрытом пространстве [36]:

$$\mathbf{s}_{t+1} = \mathbf{K}\mathbf{s}_t,$$

где $\mathbf{K}$ — матрица, аппроксимирующая конечномерное представление оператора Купмана. Та же схема автоэнкодера переносится на системы с явной зависимостью от времени [67]. Линейное представление в скрытом пространстве применяется и к выделению причинных связей по наблюдаемым рядам [68].

Интерпретируемые методы восстановления пространства состояний рассматривают систему как набор автономных дифференциальных уравнений. К ним относятся модели с явным восстановлением систем дифференциальных уравнений и *нейронные обыкновенные дифференциальные уравнения* [38]. В отличие от дискретного оператора эволюции, Neural ODE восстанавливает функциональную зависимость непрерывной динамической системы в подпространстве состояний как решение задачи Коши:

$$\frac{d}{dt}\mathbf{s}(t) = \mathbf{f}_\theta(\mathbf{s}(t), t), \quad \mathbf{s}(t_0) = \mathbf{s}_0,$$

где векторное поле $\mathbf{f}_\theta$ параметризуется весами нейронной сети θ . Постановка задаёт модель в непрерывном времени и снимает ограничение общего шага дискретизации Δt для несинхронизированных рядов.

Непрерывные модели продолжены *латентными Neural ODE* и стохастическими *SDE-сетями*. Такие подходы остаются устойчивыми при появлении шума в функциях наблюдения. Подход отделяет детерминированную компоненту динамической системы от случайной компоненты функции наблюдения.

Для зашумлённых и неполных данных требуется оценка неопределённости. Байесовские нейросетевые подходы вводят априорные распределения на параметры модели или состояния, после чего апостериорные распределения вычисляются по наблюдаемым временным рядам. В частности, динамические вариационные автоэнкодеры, например, Deep Variational Bayes Filters, рассматривают последовательность состояний как марковский процесс и строят аппроксимацию их апостериорного распределения:

$$\begin{aligned}\mathbf{z}_t &\sim p_\theta(\mathbf{z}_t \mid \mathbf{z}_{t-1}), \\ \mathbf{x}_t &\sim p_\theta(\mathbf{x}_t \mid \mathbf{z}_t),\end{aligned}$$

с вариационным приближением $q_\phi(\mathbf{z}_{1:T} \mid \mathbf{x}_{1:T})$. В таком методе энкодер параметризует распределение скрытого состояния, а модель динамической системы предсказывает распределения будущих состояний.

Отдельная задача — это восстановление аналитического вида дифференциальных уравнений нелинейной системы. Нелинейные нейросетевые модели параметризует правые части уравнений, а применение L_1 -регуляризации минимизирует число параметров. Метод разреженной идентификации нелинейной динамики строит правую часть как линейную комбинацию функций из заданной библиотеки [42]:

$$\frac{d}{dt}\mathbf{x} = \boldsymbol{\Theta}(\mathbf{x})\boldsymbol{\Xi}.$$

Комбинированные подходы формируют или расширяют базис функций в системах высокой размерности при помощи нейросетей. Анализ системы в полученном подпространстве проводится с помощью инструментов линейной алгебры. То есть нейронная сеть аппроксимирует экспериментальные данные, а разреженная нелинейная регрессия выделяет комбинацию простых базисных функций.

Глава 2

Полилинейные методы восстановления пространства состояний

В данной главе пространство состояний восстанавливается полилинейным отображением векторов задержек. Лемма о полилинейном расширении связывает предлагаемое отображение с многомерным методом задержек. Теорема о многомерном полилинейном вложении ограничивает ранг канонического разложения и размерность восстановленного пространства. Вводятся каноническое разложение, разложение Таккера и тензорный поезд [46–48, 50, 69]. Приводятся вычислительные эксперименты на аттракторе Лоренца и на рядах акселерометра.

Для введения полилинейного отображения определяется тензорное произведение.

Определение 6. *Тензорное произведение по n -ой моде* тензора $\mathbf{A} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ и матрицы $\mathbf{M} \in \mathbb{R}^{J \times I_n}$ определяется формулой $\mathbf{C} \in \mathbb{R}^{I_1 \times \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N}$, где $J, I_1, \dots, I_N$ — соответствующие индексы, с элементами:

$$\mathbf{C} = \mathbf{A} \times_n^2 \mathbf{M} = \mathbf{A} \times_n \mathbf{M}, \quad c_{i_1, \dots, i_{n-1}, j, i_{n+1}, \dots, i_N} = \sum_{i_n=1}^{I_n} a_{i_1, \dots, i_n, \dots, i_N} m_{j, i_n}. \quad (2.1)$$

В случае тензора $\mathbf{A} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ и вектора $\mathbf{v} \in \mathbb{R}^{I_n}$ произведение по n -ой моде даёт $\mathbf{C} \in \mathbb{R}^{I_1 \times \dots \times I_{n-1} \times I_{n+1} \times \dots \times I_N}$:

$$\mathbf{C} = \mathbf{A} \times_n^1 \mathbf{v} = \mathbf{A} \times_n \mathbf{v}, \quad c_{i_1, \dots, i_{n-1}, i_{n+1}, \dots, i_N} = \sum_{i_n=1}^{I_n} a_{i_1, \dots, i_n, \dots, i_N} v_{i_n} \quad (2.2)$$

Произведения по выбранной моде для матрицы (2.1) и вектора (2.2) — это разные операции. Для простоты обе операции обозначаются (2.1). Подразумевается, что в случае матрицы используется вторая мода. В случае вектора используется только первая и единственная мода вектора.

2.1. Полилинейная динамическая система

Лемма 3 (О полилинейном расширении). Пусть $\mathcal{M}$ — компактное гладкое многообразие, $\mathbf{F} : \mathcal{M} \rightarrow \mathcal{M}$ — диффеоморфизм класса C^2 , $\tau \geq 1$ — временной шаг метода задержек и $h_i \in C^2(\mathcal{M}, \mathbb{R})$, $i = 1, \dots, r$. Положим

$$\mathbf{s}_i(\mathbf{v}) = (1, h_i(\mathbf{v}), h_i(\mathbf{F}^{-\tau}(\mathbf{v})), \dots, h_i(\mathbf{F}^{-(n_i-1)\tau}(\mathbf{v})))$$

и

$$\Phi(\mathbf{v}) = \bigotimes_{i=1}^r \mathbf{s}_i(\mathbf{v}) \in \mathbb{R}^{\prod_{i=1}^r (n_i+1)},$$

где $\otimes$ — кронекерово произведение. Пусть $\Psi(\mathbf{v})$ — конкатенация линейных задержек,

$$\Psi(\mathbf{v}) = (h_1(\mathbf{v}), \dots, h_1(\mathbf{F}^{-(n_1-1)\tau}(\mathbf{v})), \dots, h_r(\mathbf{v}), \dots, h_r(\mathbf{F}^{-(n_r-1)\tau}(\mathbf{v}))) \in \mathbb{R}^{\sum_{i=1}^r n_i}.$$

Тогда Φ является гладким вложением $\mathcal{M}$ тогда и только тогда, когда вложением является Ψ .

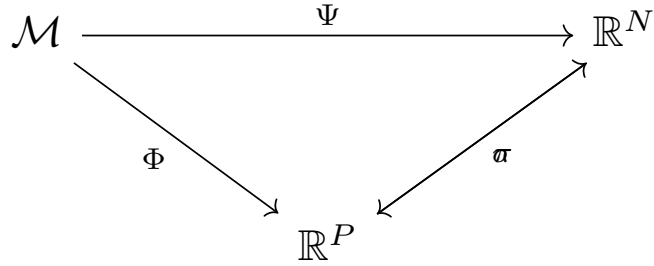
Доказательство. В кронекеровом произведении каждая координата Ψ встречается ровно один раз. Она получается, если из одного вектора $\mathbf{s}_i$ взять соответствующую задержку, а из остальных векторов — первую координату, равную единице. Пусть π — проекция на эти $\sum n_i$ координат, а σ — отображение, которое по вектору линейных задержек снова собирает векторы $\mathbf{s}_i$ и берёт их кронекерово произведение. Тогда $\Phi = \sigma \circ \Psi$ и $\pi \circ \sigma = \text{id}$ на $\mathbb{R}^{\sum n_i}$. Дифференциал удовлетворяет $d\pi \circ d\sigma = \text{id}$, поэтому $d\sigma$ инъективен и σ — гладкое вложение евклидова пространства.

Равенства $\Phi = \sigma \circ \Psi$ и $\Psi = \pi \circ \Phi$ дают инъективность одного отображения одновременно с инъективностью другого. Дифференциалы связаны соотношением $d\Phi = d\sigma \circ d\Psi$. Поскольку $d\sigma$ инъективен, $d\Phi$ инъективен в точке тогда и только тогда, когда инъективен $d\Psi$. На компактном многообразии инъективное погружение является вложением. Следовательно, Φ — вложение тогда и только тогда, когда вложение есть Ψ . □

Эквивалентность не зависит от длин окон и от периодических точек: линейные задержки восстанавливаются по полилинейным координатам гладкой проекцией π . Достаточное условие для самого Ψ даёт теорема 2 [18]. Если

$$\sum_{i=1}^r n_i \geq 2d_s + 1, \quad (2.3)$$

для каждого целого p с $1 \leq p \leq \max_i(n_i - 1)\tau$ множество точек периода p конечно. Если проекция Ψ на некоторые $2d_s + 1$ координат является вложением, то вложением является и само Ψ . Из $\Psi(\mathbf{v}) = \Psi(\mathbf{u})$ следует равенство этих координат; проекция инъективна, поэтому $\mathbf{v} = \mathbf{u}$. Если $d\Psi(\xi) = 0$, то дифференциал композиции с проекцией тоже равен нулю, поэтому $\xi = 0$. По лемме 3 вложением является и Φ . Произведения компонент для этого факта не нужны, проекция π их отбрасывает. Связь Φ и Ψ показана на рис. 2.1.



$$N = \sum_{i=1}^r n_i, \quad P = \prod_{i=1}^r (n_i + 1), \quad \pi \circ \sigma = \text{id}, \quad \Phi = \sigma \circ \Psi.$$

$$(1, x_t, x_{t-\tau}) \otimes (1, y_t, y_{t-\tau}) = (1, x, y, xy, \dots), \quad \pi(1 \cdot x \cdot 1) = x.$$

Рис. 2.1: Полилинейное расширение и отображение метода задержек. Проекция π отбрасывает произведения и возвращает задержки.

Рассматривается дискретная динамическая система задаваемая набором уравнений с полилинейными зависимостями

$$\mathbf{v}_{t+1} = \mathbf{A} \times_2 \mathbf{v}_t \times_3 \mathbf{v}_t \times_4 \dots \times_k \mathbf{v}_t,$$

где $\mathbf{A} \in \mathbb{R}^{n \times n \times \dots \times n}$ — полилинейное преобразование, $\mathbf{v}_t \in \mathbb{R}^n$ — переменная состояния, k — порядок полилинейной модели. Вектор переменных состояния — это значения некоторых системных величин в момент времени t . Предполагается, что состояние динамической системы полностью описывается этими величинами.

Для восстановления пространства состояний из экспериментальных данных используется аналогичное полилинейное преобразование в сочетании с методом задержек

$$\mathbf{s}_t = \mathbf{A} \times_2 \mathbf{x}_t \times_3 \mathbf{x}_t \times_4 \dots \times_k \mathbf{x}_t, \quad (2.4)$$

где $\mathbf{x}_t = [x_t, x_{t-1}, \dots, x_{t-n}]^T$ — вектор временной задержки из временного ряда $\mathbf{x}$ с n задержками, $\mathbf{s}_t \in \mathbb{R}^p$ — векторное представление в фазовом пространстве с $p \ll n$. Одномерный случай представлен в уравнении (2.4), в дальнейшем для упрощения предположим, что $k = 3$.

Предложенная полилинейная модель не содержит прямой линейной составляющей и части смешанных комбинаций. Константа, добавленная к вектору задержек, включает в модель линейные члены и их полилинейные комбинации:

$$\mathbf{s}_t = \mathbf{A} \times_2 \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \times_3 \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \times_4 \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix}.$$

где $\mathbf{A} \in \mathbb{R}^{p \times (n+1) \times (n+1) \times (n+1)}$. Схема полученного метода в нотации Пенроуза показана на рис.2.2.

Предложенная модель с тензором коэффициентов $\mathbf{A}$ фильтрует не только главные компоненты, как в случае PCA для одномерных временных рядов, но и учитывает их нелинейные зависимости.

Предложенный способ альтернативно представляется в классических обозначениях линейной алгебры. Пусть задано полилинейное отображение вида

$$f_A : \mathbf{X} \times \mathbf{X} \times \mathbf{X} \longrightarrow \mathbf{S},$$

где $\mathbf{X}$ — исходные $n + 1$ -мерные векторные пространства и $\mathbf{S}$ — p -мерные результирующее пространство. Тогда существует базис $\{\mathbf{e}_1, \dots, \mathbf{e}_{n+1}\}$ для $\mathbf{X}$ и

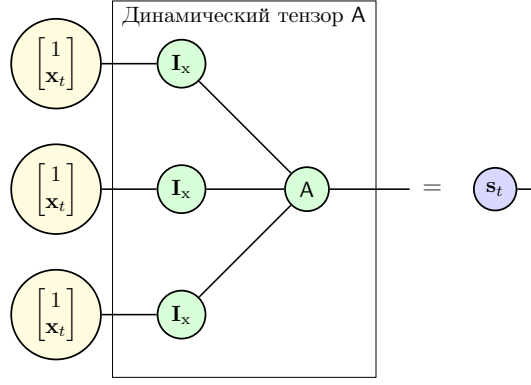


Рис. 2.2: Схема полилинейного метода снижения размерности для трехмерного случая.

базис $\{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ для $\mathbf{S}$. Таким образом, тензор $\mathbf{A}$ представляет собой набор скалярных значений, задаваемых коэффициентами функции

$$f_A(\mathbf{e}_{j_1}, \mathbf{e}_{j_2}, \mathbf{e}_{j_3}) = A_{1,j_1,j_2,j_3} \mathbf{b}_1 + \dots + A_{p,j_1,j_2,j_3} \mathbf{b}_p,$$

где $1 \leq j_k \leq n+1$. Он определяет полилинейную функцию f_A для $\mathbf{x}_t$ как

$$f_A(\mathbf{x}_t) = \sum_{j=1}^p \sum_{i_1=1}^{n+1} \sum_{i_2=1}^{n+1} \sum_{i_3=1}^{n+1} A_{j,i_1,i_2,i_3} \cdot \tilde{x}_{t,i_1} \tilde{x}_{t,i_2} \tilde{x}_{t,i_3} \cdot \mathbf{b}_j,$$

где $\tilde{\mathbf{x}}_t = (1, \mathbf{x}_t^\top)^\top = (\tilde{x}_{t,1}, \dots, \tilde{x}_{t,n+1})^\top$.

В этой записи одна из компонент вектора задержки равна константе. Константа включает в модель и сами значения временного ряда, и их полиномиальные комбинации. Такое полилинейное преобразование представляется в матричной форме:

$$\mathbf{s}_t = \mathbf{A}_{(1)} \left[1 \quad \overbrace{x_t \dots x_{t-n}}^{\text{линейная}} \quad \overbrace{x_t^2 \dots x_{t-n}^2}^{\text{билинейная}} \quad \overbrace{x_t^3 \dots x_{t-n}^3}^{\text{полилинейная}} \right]^\top$$

где $\mathbf{A}_{(1)} \in \mathbb{R}^{p \times (n+1)^3}$ — результат приведения к матрице тензора $\hat{\mathbf{A}}$ по первой моде. Таким образом предложенный метод сочетает в себе методы линейного снижения размерности и тензорный подход с полиномиальной агрегацией.

Тензорный вид выбран потому, что каноническое разложение тензора параметров $\mathbf{A}$ снижает число настраиваемых параметров. Обзор тензорных моделей

нелинейных и полилинейных систем дан в [70]; в этой главе СР-разложение применяется к тензору вложения. Разложение ранга R имеет вид $\mathbf{A} \approx \hat{\mathbf{A}} = \sum_{r=1}^R \mathbf{a}_{r,1} \otimes \mathbf{a}_{r,2} \otimes \mathbf{a}_{r,3} \otimes \mathbf{a}_{r,4}$, где $\mathbf{a}_{r,1} \in \mathbb{R}^p$ и $\mathbf{a}_{r,2}, \mathbf{a}_{r,3}, \mathbf{a}_{r,4} \in \mathbb{R}^{n+1}$. Минимальный ранг конкретного тензора в общем случае NP-труден, поэтому в расчётах R подбирается по данным. Достаточное для вложения значение R указано в теореме 5 и от длины окна не зависит. Это разложение дополнительно преобразует общую модель к следующему виду:

$$\begin{aligned}
\mathbf{s}_t &= \mathbf{A} \times_2 \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \times_3 \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \times_4 \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \\
&\quad \text{Каноническое разложение} \\
&\approx \left(\sum_{r=1}^R \mathbf{a}_{r,1} \otimes \mathbf{a}_{r,2} \otimes \mathbf{a}_{r,3} \otimes \mathbf{a}_{r,4} \right) \times_2 \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \times_3 \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \times_4 \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \\
&= \sum_{r=1}^R \mathbf{a}_{r,1} \left(\mathbf{a}_{r,2}^\top \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \right) \left(\mathbf{a}_{r,3}^\top \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \right) \left(\mathbf{a}_{r,4}^\top \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \right) \\
&= \mathbf{A}_1 \left(\mathbf{A}_2^\top \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \odot \mathbf{A}_3^\top \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \odot \mathbf{A}_4^\top \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \right)
\end{aligned}$$

где $\odot$ — поэлементное произведение Адамара. На рис. 2.3 показано графическое представление метода в нотации Пенроуза.

Предложенное разложение тензора $\hat{\mathbf{A}}$ снижает число настраиваемых параметров:

$$p(n+1)^3 \longrightarrow R(p+3(n+1)).$$

Каноническое разложение — один из трёх способов снизить число параметров полного тензора. Разложение Таккера оставляет ядро порядка три и проецирует каждую моду на промежуточную размерность R . Разложение тензорного поезда связывает моды последовательно, через ядра порядка три. Число параметров и вид модели для трёх рядов приведены в табл. 2.1. При одном ряде и единичном тензоре все три формы совпадают с методом задержек, если $n \geq 2d_s + 1$.

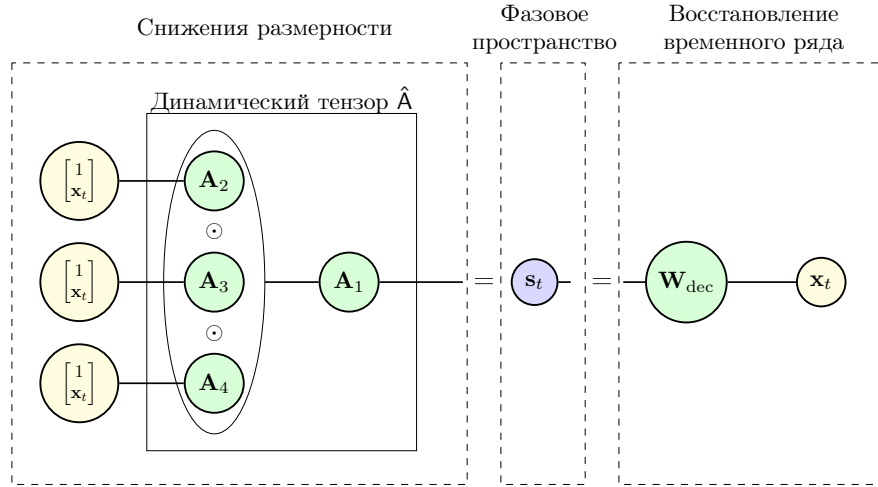
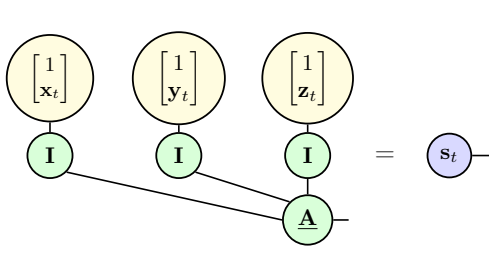


Рис. 2.3: Полилинейная динамическая система в графической нотации.

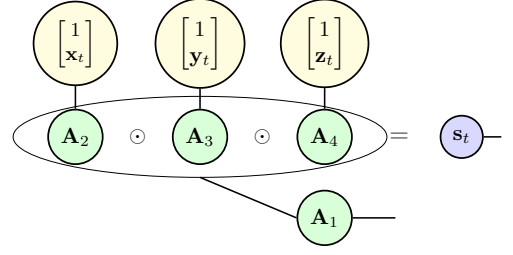
Таблица 2.1: Число параметров полилинейного отображения при разных разложениях

Разложение	Вид модели	Число параметров
Без разложения	$\mathbf{s}_t = \mathbf{A} \times_2 \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \times_3 \begin{bmatrix} 1 \\ \mathbf{y}_t \end{bmatrix} \times_4 \begin{bmatrix} 1 \\ \mathbf{z}_t \end{bmatrix}$	$p(n+1)^3$
Каноническое	$\mathbf{s}_t = \mathbf{A}_1 (\mathbf{A}_2^\top \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \odot \mathbf{A}_3^\top \begin{bmatrix} 1 \\ \mathbf{y}_t \end{bmatrix} \odot \mathbf{A}_4^\top \begin{bmatrix} 1 \\ \mathbf{z}_t \end{bmatrix})$	$R(p + 3(n+1))$
Таккера	$\mathbf{s}_t = \mathbf{A}_1 \times_2 (\mathbf{A}_2 \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix}) \times_3 (\mathbf{A}_3 \begin{bmatrix} 1 \\ \mathbf{y}_t \end{bmatrix}) \times_4 (\mathbf{A}_4 \begin{bmatrix} 1 \\ \mathbf{z}_t \end{bmatrix})$	$pR^3 + 3(n+1)R$
Тензорный поезд	$\mathbf{s}_t = \begin{bmatrix} 1 & \mathbf{x}_t^\top \end{bmatrix} \mathbf{A}_1 (\mathbf{A}_2 \times_2 \begin{bmatrix} 1 \\ \mathbf{y}_t \end{bmatrix}) (\mathbf{A}_3 \times_2 \begin{bmatrix} 1 \\ \mathbf{z}_t \end{bmatrix}) \mathbf{A}_4$	$pR + (n+1)(2R^2 + R)$

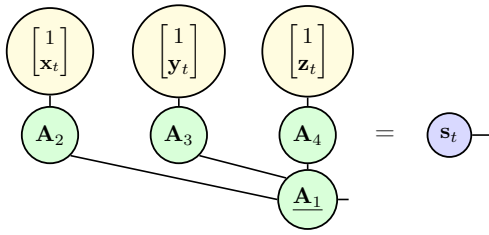
Здесь $R \leq n$ — размерность промежуточных мод. Схемы четырёх моделей в нотации Пенроуза показаны на рис. 2.4.



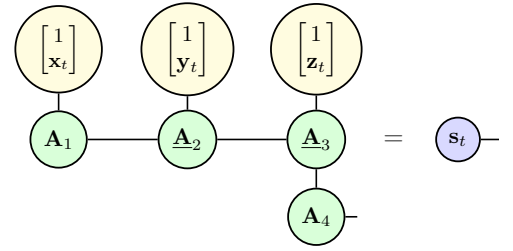
(а) без разложения



(б) каноническое разложение



(в) разложение Таккера



(г) тензорный поезд

Рис. 2.4: Схемы полилинейного отображения в нотации Пенроуза

Полилинейная модель с K модами воспроизводит полиномиальное отображение степени K . Пусть в области $U \subset \mathbb{R}^{n+1}$ истинное отображение задано полиномом степени не выше K :

$$\mathbf{s}_{t+1} = F(\mathbf{x}_t) = \sum_{|\alpha| \leq K} c_\alpha \mathbf{x}_t^\alpha, \quad \mathbf{x}_t = [1, x_t, x_{t-1}, \dots, x_{t-n}]^\top.$$

Тогда существует тензор $\mathbf{A}$ порядка $K + 1$, такой что для всех $\mathbf{x}_t \in U$

$$F(\mathbf{x}_t) = \mathbf{A} \times_2 \mathbf{x}_t \times_3 \mathbf{x}_t \times_4 \cdots \times_{K+1} \mathbf{x}_t. \quad (2.5)$$

При $K = 1$ формула (2.5) совпадает с линейной авторегрессией $\mathbf{s}_{t+1} = \mathbf{A}\mathbf{x}_t$. Такой тензор существует. Первая координата $\mathbf{x}_t$ равна единице, поэтому моном степени $d \leq K$ есть произведение d переменных координат и $K - d$ единиц. Его коэффициент записывается в одну компоненту $\mathbf{A}$ с соответствующим набором

индексов, а остальные компоненты, дающие тот же моном при перестановке индексов, полагаются равными нулю. Свёртка по модам воспроизводит полином. Представление не единственно, для дальнейшего существования.

2.2. Обобщение полилинейной модели на многомерные данные

Одномерная модель (2.4) распространяется на несколько рядов.

Чтобы упростить теорию и прояснить ее связь с вычислительным экспериментом, пусть многомерный временной ряд содержит три типа измерений, поступающих с трехосного акселерометра. Пусть $\mathbf{x}, \mathbf{y}, \mathbf{z}$ — временные ряды ускорения вдоль пространственных осей x, y, z . Сигнал с каждой оси в отдельности восстанавливает фазовое пространство динамической системы согласно теореме Такенса, используя вложение с временной задержкой, как в уравнении (1.4). На рис. 2.5 показан такой пример: два наблюдения одной траектории дают схожие восстановленные пространства. Расширение на многомерный случай выполня-

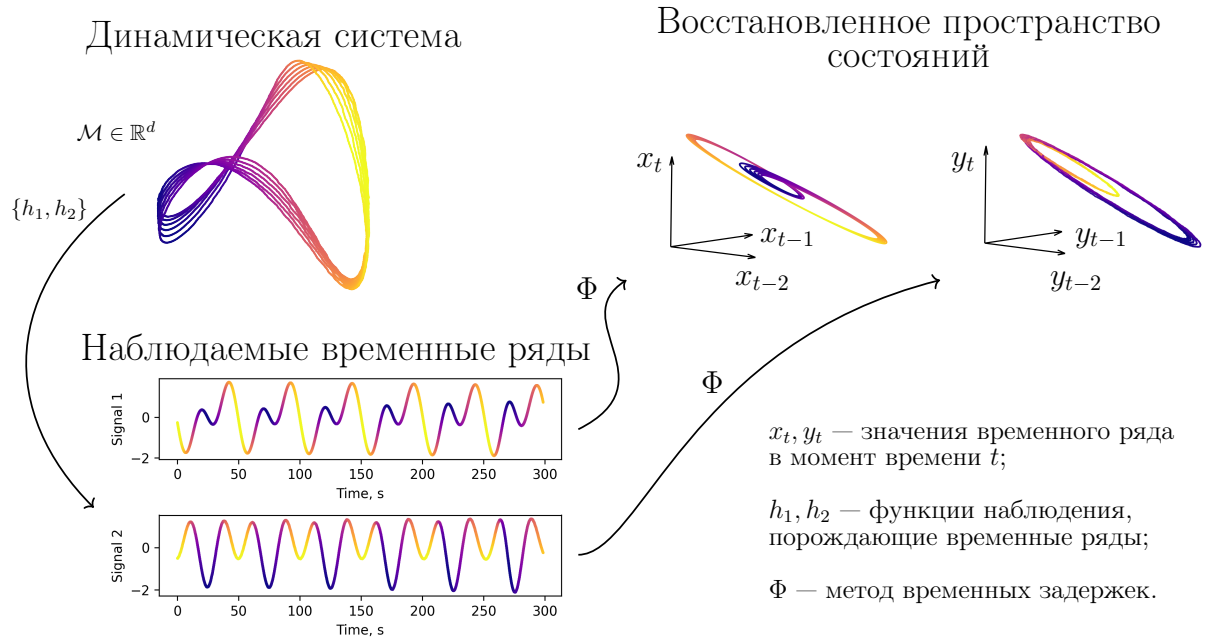


Рис. 2.5: Восстановление пространства состояний по двум наблюдениям акселерометра. Цветом отмечены близкие участки исходной траектории

ется исходя из предположения о существовании полилинейных отображений

между измерениями

$$\mathbf{X} = \mathbf{I}\mathbf{X}, \quad \mathbf{X} = \mathbf{W}_y \mathbf{Y}, \quad \mathbf{X} = \mathbf{W}_z \mathbf{Z},$$

где $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ — матрицы траекторий в исходном фазовом пространстве, $\mathbf{W}_y, \mathbf{W}_z$ — матрицы преобразований, $\mathbf{I}$ — единичная матрица. Полилинейная модель модифицируется следующим образом:

$$\mathbf{s}_t = \mathbf{A} \times_2 (\mathbf{I}\mathbf{x}_t) \times_3 (\mathbf{W}_y \mathbf{y}_t) \times_4 (\mathbf{W}_z \mathbf{z}_t) = \hat{\mathbf{A}} \times_1 \mathbf{x}_t \times_2 \mathbf{y}_t \times_3 \mathbf{z}_t,$$

где $\hat{\mathbf{A}} = \mathbf{A} \times_1 \mathbf{I} \times_2 \mathbf{W}_y \times_3 \mathbf{W}_z$ — тензор весовых коэффициентов, учитывающий линейные связи, а $\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t$ — векторы переменных состояния с каждой оси в момент времени t . В краткой форме уравнение преобразуется к виду:

$$\mathbf{s}_t = \hat{\mathbf{A}} \times_2 \mathbf{x}_t \times_3 \mathbf{y}_t \times_4 \mathbf{z}_t. \quad (2.6)$$

Константа при векторах состояния сохраняет вклад каждого ряда отдельно от их комбинаций:

$$\mathbf{s}_t = \hat{\mathbf{A}} \times_2 \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \times_3 \begin{bmatrix} 1 \\ \mathbf{y}_t \end{bmatrix} \times_4 \begin{bmatrix} 1 \\ \mathbf{z}_t \end{bmatrix}.$$

где $\hat{\mathbf{A}} \in \mathbb{R}^{p \times (n+1) \times (n+1) \times (n+1)}$. На рисунке 2.6 представлено графическое изображение метода в нотации Пенроуза.

Теорема 4 (О включении линейной векторной авторегрессии). Пусть имеется m наблюдаемых временных рядов, удовлетворяющих линейной векторной авторегрессии порядка p_{ar} :

$$\mathbf{x}_{t+1} = \sum_{k=1}^{p_{\text{ar}}} \mathbf{A}_k \mathbf{x}_{t-k+1} + \boldsymbol{\varepsilon}_t, \quad \mathbf{x}_t = (x_{1,t}, \dots, x_{m,t})^\top \in \mathbb{R}^m, \quad \mathbf{A}_k \in \mathbb{R}^{m \times m},$$

где $\boldsymbol{\varepsilon}_t$ — центрированный шум с конечной дисперсией. Упорядочим регрессор по лагам, а внутри лага — по компонентам ряда, и дополним его константой:

$$\mathbf{z}_t = (1, \mathbf{x}_t^\top, \mathbf{x}_{t-1}^\top, \dots, \mathbf{x}_{t-p_{\text{ar}}+1}^\top)^\top \in \mathbb{R}^q, \quad q = mp_{\text{ar}} + 1.$$

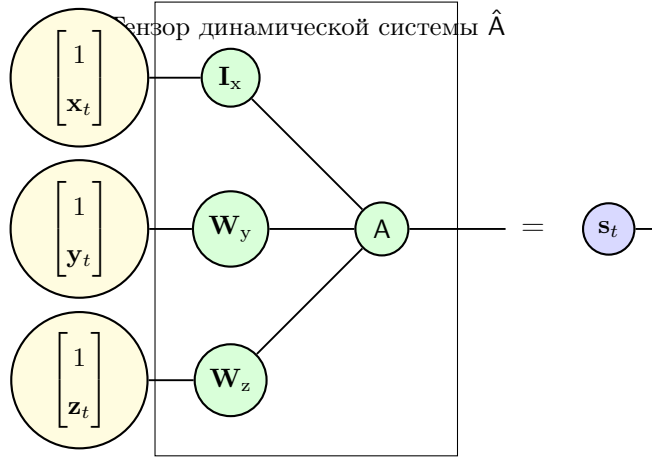


Рис. 2.6: Графическое представление тензорной динамической системы. Граф показывает, как результирующий тензор неявно включает в себя отображения временных рядов друг на друга. Математическая форма показана в уравнении (2.6).

Тогда матрица

$$\mathbf{W} = [\mathbf{0}_{m \times 1}, \mathbf{A}_1, \dots, \mathbf{A}_{p_{\text{ar}}}] \in \mathbb{R}^{m \times q}$$

удовлетворяет $\mathbf{W}\mathbf{z}_t = \sum_{k=1}^{p_{\text{ar}}} \mathbf{A}_k \mathbf{x}_{t-k+1}$. Рассмотрим полилинейную модель порядка $K \geq 1$

$$\mathbf{s}_t = \mathbf{B} \times_2 \mathbf{z}_t \times_3 \mathbf{z}_t \times_4 \cdots \times_{K+1} \mathbf{z}_t$$

с тензором $\mathbf{B}$, у которого первая мода имеет размер p , а каждая из следующих K мод — размер q . Существует такой тензор, что детерминированная часть авторегрессии воспроизводится точно. При $p = m$

$$\mathbf{s}_t = \mathbf{W}\mathbf{z}_t = \mathbf{x}_{t+1} - \boldsymbol{\varepsilon}_t.$$

При $p > m$ то же равенство выполнено для первых m компонент $\mathbf{s}_t$, а остальные компоненты равны нулю.

Доказательство. По определению произведения по модам

$$\mathbf{s}_t = \mathbf{B}_{(1)} \mathbf{z}_t^{\otimes K},$$

где $\mathbf{B}_{(1)} \in \mathbb{R}^{p \times q^K}$ — развёртка по первой моде, а $\mathbf{z}_t^{\otimes K}$ — кронекерова степень. Первая координата $\mathbf{z}_t$ равна единице. Кронекерова степень содержит K копий

каждой компоненты $\mathbf{z}_t$: переменную даёт один из K сомножителей, а остальные сомножители дают единицу. Помимо этих копий есть константа и произведения двух и более компонент. Ненулевыми оставляем столбцы только одной линейной копии: при $p = m$ они совпадают со столбцами $\mathbf{W}$, при $p > m$ — со столбцами матрицы $\begin{bmatrix} \mathbf{W} \\ \mathbf{0}_{(p-m) \times q} \end{bmatrix}$. Остальные столбцы, в том числе столбцы прочих линейных копий, равны нулю, поэтому множитель K не возникает. При $p = m$ получается

$$\mathbf{s}_t = \mathbf{W}\mathbf{z}_t = \sum_{k=1}^{p_{\text{ar}}} \mathbf{A}_k \mathbf{x}_{t-k+1} = \mathbf{x}_{t+1} - \boldsymbol{\varepsilon}_t.$$

При $p > m$ то же равенство выполнено для первых m компонент, а остальные равны нулю. Класс полилинейных моделей содержит линейные векторные авторегрессии: обнуление весов при нелинейных мономах оставляет детерминированную часть $\text{VAR}(p_{\text{ar}})$. $\square$

Тензор $\hat{\mathbf{A}}$ отбирает не только главные компоненты, как в случае PCA для одномерных временных рядов, но и фильтрует их в соответствии с полилинейными зависимостями с другими временными рядами. Дополнительные ряды дают дополнительные информативные компоненты.

То же отображение записывается в классических обозначениях линейной алгебры. Пусть

$$f_A : \mathbf{X} \times \mathbf{Y} \times \mathbf{Z} \longrightarrow \mathbf{S}$$

где $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ и $\mathbf{S}$ — конечномерные векторные пространства с размерностью вложения $n + 1$ (с добавленной константой) и размерностью результирующего пространства p . Существуют базисы $\{\mathbf{e}_{k,1}, \dots, \mathbf{e}_{k,n+1}\}$ для каждого из $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ и базис $\{\mathbf{b}_{i,1}, \dots, \mathbf{b}_{i,p}\}$ для $\mathbf{S}$. Таким образом, тензор $\mathbf{A}$ представляет собой набор скалярных величин:

$$f_A(\mathbf{e}_{x_{j_1}}, \mathbf{e}_{y_{j_2}}, \mathbf{e}_{z_{j_3}}) = A_{1,j_1,j_2,j_3} \mathbf{b}_1 + \dots + A_{p,j_1,j_2,j_3} \mathbf{b}_p,$$

где $1 \leq j_k \leq n+1$, $1 \leq k \leq 3$, а $\tilde{\mathbf{x}}_t$, $\tilde{\mathbf{y}}_t$ и $\tilde{\mathbf{z}}_t$ — векторы задержек, дополненные единицей в первой координате.

Это определяет полилинейную функцию f_A для $\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t$ как

$$f_A(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t) = \sum_{j=1}^p \sum_{i_1=1}^{n+1} \sum_{i_2=1}^{n+1} \sum_{i_3=1}^{n+1} A_{j,i_1,i_2,i_3} \cdot \tilde{x}_{t,i_1} \tilde{y}_{t,i_2} \tilde{z}_{t,i_3} \cdot \mathbf{b}_j.$$

Одна из компонент вектора задержек равна константе. Константа включает в модель как сами значения рядов, так и их комбинации. Развёртка полилинейного преобразования в матричной форме имеет вид полиномиальной регрессии

$$\mathbf{s}_t = \mathbf{A}_{(1)} \begin{bmatrix} 1 & \overbrace{x_t \dots z_{t-n}}^{\text{линейная часть}} & \overbrace{x_t y_t \dots y_{t-n} z_{t-n}}^{\text{билинейная часть}} & \overbrace{x_t y_t z_t \dots x_{t-n} y_{t-n} z_{t-n}}^{\text{полилинейная часть}} \end{bmatrix}^T \quad (2.7)$$

где $\mathbf{A}_{(1)} \in \mathbb{R}^{p \times (n+1)^3}$ — результат *развертки* тензора $\hat{\mathbf{A}}$ по первой моде.

Полиномиальное представление объединяет методы линейного снижения размерности и тензорные методы с нелинейной агрегацией.

Параметры тензора оптимизируются каноническим разложением $\hat{\mathbf{A}}$. Это разложение представляет тензор в виде суммы R внешних произведений $\hat{\mathbf{A}} = \sum_{r=1}^R \mathbf{a}_{r,1} \otimes \mathbf{a}_{r,2} \otimes \mathbf{a}_{r,3} \otimes \mathbf{a}_{r,4}$, где R минимально, $\mathbf{a}_{r,1} \in \mathbb{R}^p$, $\mathbf{a}_{r,2}, \mathbf{a}_{r,3}, \mathbf{a}_{r,4} \in \mathbb{R}^{n+1}$ — факторные векторы, $1 \leq r \leq R$. Это снижает число параметров модели до линейного относительно числа мод исходного тензора.

Вычисление минимального ранга заданного тензора — NP-трудная задача. Поэтому при оптимизации R выбирается по данным, что приводит к аппроксимации $\hat{\mathbf{A}} \approx \sum_{r=1}^R \mathbf{a}_{r,1} \otimes \mathbf{a}_{r,2} \otimes \mathbf{a}_{r,3} \otimes \mathbf{a}_{r,4}$. Достаточный ранг, при котором такая аппроксимация сохраняет вложение, от длины окна не зависит: его даёт теорема 5. Это разложение в дальнейшем изменяет общую модель к следующему

виду

$$\begin{aligned}
\mathbf{s}_t &= \hat{\mathbf{A}} \times_2 \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \times_3 \begin{bmatrix} 1 \\ \mathbf{y}_t \end{bmatrix} \times_4 \begin{bmatrix} 1 \\ \mathbf{z}_t \end{bmatrix} \\
&\quad \text{Каноническое разложение} \\
&\approx \left(\sum_{r=1}^R \mathbf{a}_{r,1} \otimes \mathbf{a}_{r,2} \otimes \mathbf{a}_{r,3} \otimes \mathbf{a}_{r,4} \right) \times_2 \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \times_3 \begin{bmatrix} 1 \\ \mathbf{y}_t \end{bmatrix} \times_4 \begin{bmatrix} 1 \\ \mathbf{z}_t \end{bmatrix} \\
&= \sum_{r=1}^R \mathbf{a}_{r,1} \left(\mathbf{a}_{r,2}^\top \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \right) \left(\mathbf{a}_{r,3}^\top \begin{bmatrix} 1 \\ \mathbf{y}_t \end{bmatrix} \right) \left(\mathbf{a}_{r,4}^\top \begin{bmatrix} 1 \\ \mathbf{z}_t \end{bmatrix} \right) \\
&= \mathbf{A}_1 \left(\mathbf{A}_2^\top \begin{bmatrix} 1 \\ \mathbf{x}_t \end{bmatrix} \odot \mathbf{A}_3^\top \begin{bmatrix} 1 \\ \mathbf{y}_t \end{bmatrix} \odot \mathbf{A}_4^\top \begin{bmatrix} 1 \\ \mathbf{z}_t \end{bmatrix} \right)
\end{aligned} \tag{2.8}$$

где $\odot$ — это произведение Адамара (поэлементное).

(2.8) — частный случай $r = 3$ отображения, которое ниже называется канонической проекцией.

Теорема 5 (О многомерном полилинейном вложении). Пусть $\Psi : \mathcal{M} \rightarrow \mathbb{R}^{n_1+1} \times \dots \times \mathbb{R}^{n_r+1}$ — гладкое вложение компактного многообразия размерности d_s ,

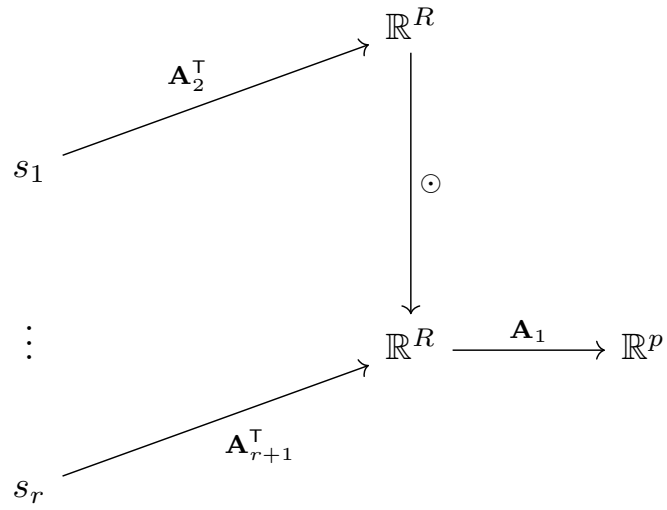
$$\Psi(\mathbf{v}) = [s_1(\mathbf{v}), \dots, s_r(\mathbf{v})], \quad s_i(\mathbf{v}) = \begin{bmatrix} 1 \\ \mathbf{x}_i(\mathbf{v}) \end{bmatrix}.$$

Каноническая проекция ранга R задаётся фактор-матрицами $\mathbf{A}_1 \in \mathbb{R}^{p \times R}$ и $\mathbf{A}_{i+1} \in \mathbb{R}^{(n_i+1) \times R}$.

$$\mathbf{g}(\mathbf{v}) = \mathbf{A}_2^\top s_1(\mathbf{v}) \odot \dots \odot \mathbf{A}_{r+1}^\top s_r(\mathbf{v}), \quad \mathbf{s}(\mathbf{v}) = \mathbf{A}_1 \mathbf{g}(\mathbf{v}).$$

Если $R \geq 2d_s + 1$, то множество матриц $\mathbf{A}_2, \dots, \mathbf{A}_{r+1}$, для которых $\mathbf{g}$ является гладким вложением $\mathcal{M}$ в $\mathbb{R}^R$, непусто и открыто. Если дополнительно $p \geq 2d_s + 1$, то множество наборов $(\mathbf{A}_1, \dots, \mathbf{A}_{r+1})$, для которых $\mathbf{s}$ является гладким вложением $\mathcal{M}$ в $\mathbb{R}^p$, непусто и открыто. Для погружения необходимы $R \geq d_s$ и $p \geq d_s$. Обе границы не зависят от длин окон $n_1, \dots, n_r$.

Композиция канонической проекции показана на рис. 2.7.



$$s_i = \begin{bmatrix} 1 \\ \mathbf{x}_i \end{bmatrix}, \quad g_j = \prod_{i=1}^r (\mathbf{A}_{i+1}^\top s_i)_j, \quad \mathbf{s} = \mathbf{A}_1 \mathbf{g}.$$

$$R \geq d_s, \quad p \geq d_s \quad (\text{погружение}), \quad R \geq 2d_s + 1, \quad p \geq 2d_s + 1 \quad (\text{вложение}).$$

Рис. 2.7: Каноническая проекция. Линейные формы окон перемножаются, ранг равен числу координат произведения.

Доказательство. Отображения $\mathbf{g}$ и $\mathbf{s}$ уже определены на $\mathcal{M}$: первое принимает значения в $\mathbb{R}^R$, второе — в $\mathbb{R}^p$. Дифференциал гладкого отображения в $\mathbb{R}^R$ имеет ранг не выше R . Погружение требует инъективный дифференциал, то есть ранг d_s , поэтому $R \geq d_s$. То же сравнение для $\mathbf{s}$ даёт $p \geq d_s$. Длины окон в эти неравенства не входят: они увеличивают пространство, из которого берётся проекция, и не увеличивают размерность пространства, в которое она попадает.

Достаточная граница берётся из теоремы о проекциях. По условию Ψ — вложение, так что $N = \Psi(\mathcal{M})$ — компактное гладкое подмногообразие размерности d_s . Первая координата каждого окна на N тождественно равна единице, и N диффеоморфно своей проекции N' на остальные координаты. Теорема 2 работы [17] утверждает, что при целевой размерности, большей удвоенной размерности динамической системы, почти всякое гладкое отображение пространства взаимно однозначно и является погружением. В доказательстве этой теоремы пробным пространством служат линейные отображения. Для гладкого многообразия размерность равна d_s , поэтому при $R \geq 2d_s + 1$ почти всякое линейное отображение пространства переменных координат в $\mathbb{R}^R$ задаёт вложение N' .

Каноническая проекция линейной не является: её j -я координата есть произведение линейных проекций форм,

$$g_j(\mathbf{v}) = \prod_{i=1}^r (c_{ij} + \ell_{ij}(\mathbf{v})).$$

Здесь c_{ij} — вклад константы i -ой компоненты, а ℓ_{ij} — линейное преобразование. Выберем $c_{ij} \neq 0$ и заменим ℓ_{ij} на $\varepsilon \ell_{ij}$. Произведение раскладывается в многочлен по ε :

$$g_j^\varepsilon(\mathbf{v}) = C_j + \varepsilon L_j(\mathbf{v}) + \varepsilon^2 \rho_j(\mathbf{v}, \varepsilon), \quad C_j = \prod_{i=1}^r c_{ij}, \quad L_j = \sum_{k=1}^r \ell_{kj} \prod_{i \neq k} c_{ij}.$$

Множители $\prod_{i \neq k} c_{ij}$ отличны от нуля, а ℓ_{kj} независимы. Значит, линейное отображение $L = (L_1, \dots, L_R)$ можно сделать любым линейным отображением переменных координат. По теореме 2 его можно выбрать так, чтобы ограничение на N' было вложением.

Остаток ρ_j — многочлен от ε и координат $\Psi(\mathbf{v})$, поэтому отображение $\varepsilon \mapsto (g^\varepsilon - C)/\varepsilon$ непрерывно в топологии C^1 и при $\varepsilon = 0$ совпадает с L . Вложения компактного многообразия образуют открытое множество в C^1 . Следовательно, при всех достаточно малых $\varepsilon \neq 0$ отображение $(g^\varepsilon - C)/\varepsilon$ есть вложение. Умножение на такое ε и сдвиг на C — диффеоморфизмы $\mathbb{R}^R$, поэтому вложением является и g^ε . Соответствующие фактор-матрицы существуют, и множество подходящих матриц $\mathbf{A}_2, \dots, \mathbf{A}_{r+1}$ непусто.

Отображение из пространства этих матриц в $C^1(\mathcal{M}, \mathbb{R}^R)$ непрерывно: на $\Psi(\mathcal{M})$ координаты $\mathbf{g}$ и их первые производные полиномиально зависят от элементов матриц.

Если $\mathbf{g}$ уже является вложением в $\mathbb{R}^R$ и $p \geq 2d_s + 1$, непустота множества матриц $\mathbf{A}_1$, для которых $\mathbf{s} = \mathbf{A}_1\mathbf{g}$ — вложение, следует из теоремы 2. Открытость этого множества следует из непрерывности отображения $\mathbf{A}_1 \mapsto \mathbf{A}_1\mathbf{g}$ в топологии C^1 . Произведение непустых открытых множеств непусто и открыто. $\square$

Замечание 1. Открытость в теореме 5 означает, что вложение сохраняется при малом изменении фактор-матриц. Функции наблюдения при этом фиксированы: пространство задержек Ψ уже предполагается вложением, что при условии (2.3) даёт теорема 2 работы [18]. Метод главных компонент выбирает одну матрицу, а не случайную. Теорема обосновывает класс канонических моделей, а не конкретный численный алгоритм, и не утверждает, что подходящие матрицы имеют полную меру. Для системы Лоренца размерность пространства состояний равна 3, поэтому достаточное условие теоремы даёт $R \geq 7$ и $p \geq 7$. Сам аттрактор имеет ёмкостную размерность около 2, и теорема 2.3 работы [17], применённая непосредственно к аттрактору, допускает меньший ранг. В табл. 2.3 ошибка прогноза контрольного ряда при тензорном вложении перестаёт убывать после размерности 4–5. Оба числа далеки от длины окна.

Это представление обладает двумя важными свойствами. Во-первых, веса оцениваются с использованием метода обратного распространения ошибки, а

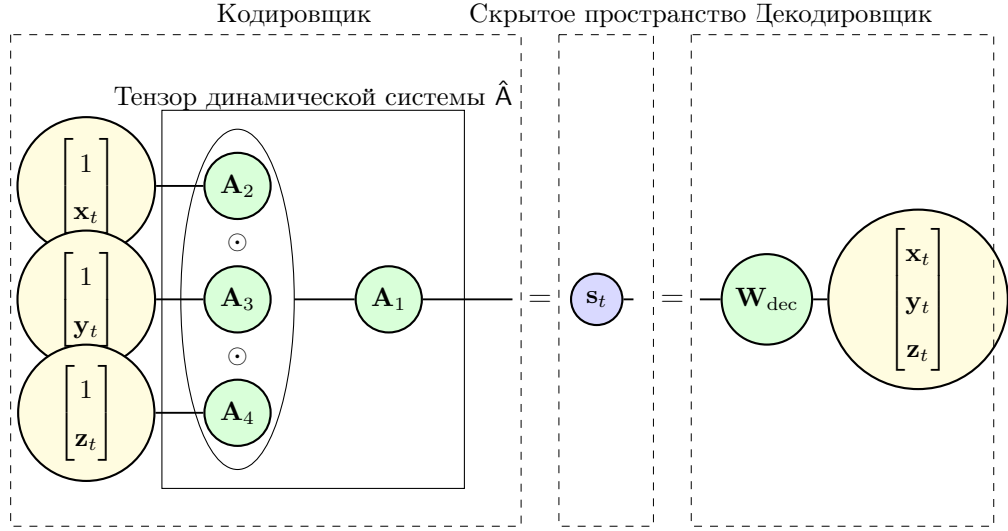


Рис. 2.8: Схема обучения автоэнкодера. Левая часть (кодировщик) представляет собой полилинейное отображение в подпространство состояний, правая часть (декодировщик) — линейное отображение обратно в исходное фазовое пространство. Математическая форма показана в уравнении (2.8).

во-вторых, оно уменьшает количество настраиваемых параметров с $O(p * (n + 1)^3)$ до $O(R * (p + 3n))$. Рисунок 2.8 поясняет предложенный метод.

2.3. Анализ полилинейных методов снижения размерности

Основная задача эксперимента — оценить качество реконструкции фазового пространства и сравнить его с одномерными и многомерными подходами.

Качество восстановления пространства состояний оценивается не только по точности восстановления исходных рядов, но и по точности линейной проекции на *контрольный временной ряд*. *Контрольный временной ряд* принадлежит той же системе, обладает полилинейными связями с использованными рядами и не входит в модель восстановления пространства состояний. Примером такой группы временных рядов являются проекции на оси при движении в трехмерном пространстве. Дополнительным условием для таких проекций является несводимость к двумерному случаю, например, движение по плоскости. Такой подход в эксперименте используется для того, чтобы отделить эффект

от увеличения числа параметров модели от действительно качественного восстановления пространства состояний.

Анализируются динамические системы с тремя доступными временными рядами. Два из них используются при восстановлении пространства состояний, а третий используется для анализа его качества через линейное отображение.

Прогноз будущего состояния для каждого временного ряда выполняется отдельно с помощью d_m многомерных линейных моделей.

$$\mathbf{s}_{t+1} = \mathbf{G}(\mathbf{s}_t) = \mathbf{G}\mathbf{s}_t,$$

Параметры отображения и прогноза находятся из двух функций ошибки. Ошибка восстановления значений временных рядов

$$\mathcal{L}_{\text{rec}} = \frac{1}{T} \sum_{t=1}^T (\|\mathbf{x}_t - \hat{\mathbf{x}}_t\|_2^2 + \|\mathbf{y}_t - \hat{\mathbf{y}}_t\|_2^2 + \|\mathbf{z}_t - \hat{\mathbf{z}}_t\|_2^2)$$

минимизируется по фактор-матрицам и ядрам разложения. Ошибка прогноза в подпространстве состояний

$$\mathcal{L}_{\text{pred}} = \frac{1}{T-1} \sum_{t=1}^{T-1} \|\mathbf{s}_{t+1} - \mathbf{W}_{\text{ДС}}\mathbf{s}_t\|_2^2$$

минимизируется по матрице $\mathbf{W}_{\text{ДС}}$. Для оценки точности восстановления пространства состояний и долгосрочного прогнозирования временных рядов используются среднеквадратичная ошибка (rMSE) и коэффициент детерминации (R^2). Модель прогнозирует последовательно значение до 25 точки в будущее.

В этом примере для анализа восстановленных пространств используется аттрактор Лоренца. Исследуемые переменные определяются системой дифференциальных уравнений

$$\begin{cases} \frac{dx}{dt} = \sigma(y - x), \\ \frac{dy}{dt} = x(r - z) - y, \\ \frac{dz}{dt} = xy - bz \end{cases} \quad (2.9)$$

со следующими параметрами $\sigma = 10$, $r = 28$, $b = 8/3$. Временные ряды x , y используются для реконструкции фазового пространства. Временной ряд z используется как **контрольный**. Для моделирования использовался шаг по времени $dt = 0.0075$ и количество точек в траектории 10 000. Половина точек, т.е. 5 000, использовалась для обучения моделей снижения размерности и прогнозирования, остальные — для оценки моделей. В таблицах U-GPVE, M-GPVE и T-GPVE обозначают одномерное, многомерное и тензорное глобальные вложения по главным компонентам. Метрики приведены в табл. 2.2 и табл. 2.3.

Таблица 2.2: Прогноз $x_{t+\Delta t}$ системы Лоренца; ряд x_t входит в реконструкцию

Размерность	U-GPVE		M-GPVE		T-GPVE	
	R^2	rMSE	R^2	rMSE	R^2	rMSE
2	0,944	0,0290	0,944	0,0288	0,945	0,0288
3	0,945	0,0284	0,945	0,0286	0,947	0,0284
4	0,945	0,0284	0,945	0,0284	0,945	0,0282
5	0,945	0,0284	0,945	0,0282	0,945	0,0281
6	0,945	0,0284	0,945	0,0282	0,945	0,0282
7	0,945	0,0284	0,945	0,0282	0,945	0,0281

При отсутствии шума все методы показывают одинаковые результаты на тестовой выборке. Этот пример показывает, что результирующее фазовое пространство достаточно хорошо восстанавливается для задачи прогнозирования временного ряда при $p = 3$.

Таблица 2.3: Прогноз контрольного ряда $z_{t+\Delta t}$ системы Лоренца

Размерность	U-GPVE		M-GPVE		T-GPVE	
	R^2	rMSE	R^2	rMSE	R^2	rMSE
2	0,020	0,110	0,0278	0,131	0,2951	0,109
3	0,025	0,115	0,0210	0,131	0,2944	0,109
4	0,026	0,121	0,0345	0,127	0,6233	0,079
5	0,027	0,125	0,4122	0,099	0,6277	0,079
6	0,028	0,128	0,4128	0,099	0,6206	0,080
7	0,031	0,131	0,4133	0,099	0,6444	0,077

Для ряда z_t полилинейное отображение даёт больший R^2 при меньшей размерности, чем многомерное вложение: прирост определяется компонентой $x_t y_t$, уравнение (2.7). Поворот траектории вокруг одной из осей воспроизводит табл. 2.2.

На рис. 2.9 приведены фазовые портреты исходной системы, полилинейной модели и автоэнкодера Купмана [36]. Горизонт сравнения — одно время Ляпунова $T_\lambda = 1/\lambda_{\max}$: за этот интервал расхождение двух близких траекторий возрастает в e раз. Метрики на этом горизонте собраны в табл. 2.4. На коротком интервале все три модели точны. Полилинейная модель содержит около 3000 параметров против примерно 9000 у автоэнкодера Купмана и сохраняет форму распределения точек. При оптимизации выделяются квадратичные и полилинейные члены x_{t-1}^2 , $x_{t-1}x_{t-2}$, $x_{t-1}y_{t-1}$ и линейные члены, что согласуется с системой (2.9).

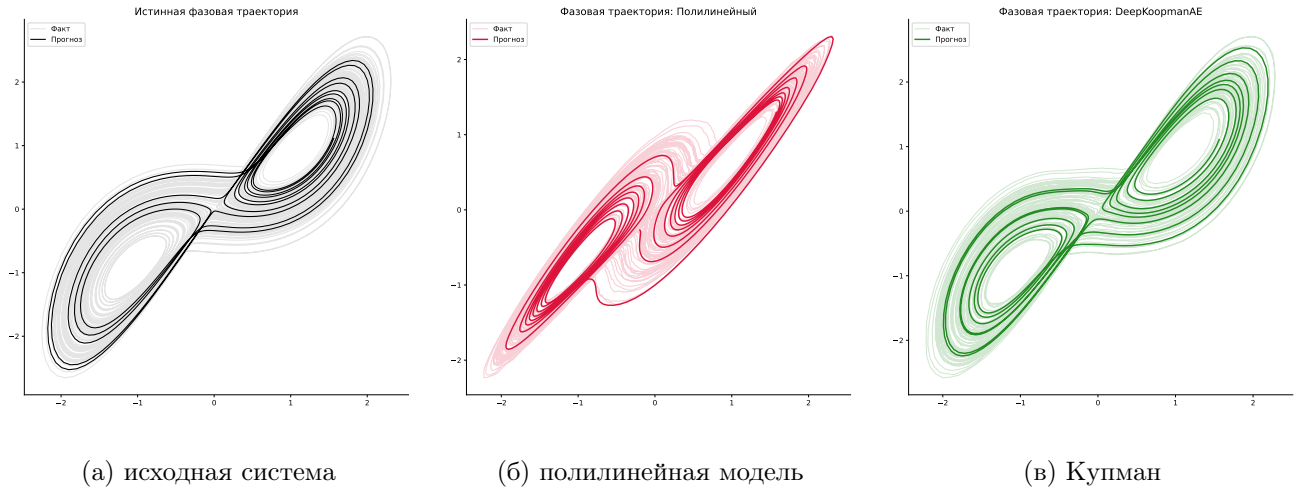


Рис. 2.9: Фазовые портреты аттрактора Лоренца: исходная система, полилинейная модель и автоэнкодер Купмана

Три вложения сравниваются при одной длине окна, $n = 50$ в (1.4), и одинаковом числе точек. В выборку входят 10 участников, ходьба и спуск по лестнице без смены типа активности; длительность записи не меньше 10 минут. Схема эксперимента показана на рис. 2.10: качество пространства оценивается и по рядам, вошедшим в реконструкцию, и по контрольному ряду.

Таблица 2.4: Реконструкция аттрактора Лоренца на горизонте одного времени Ляпунова

Модель	MSE	MAE	KL	Число параметров
Линейная	0,0027	0,0422	0,0220	1500
Тензорная	0,0009	0,0232	0,0119	3000
Автоэнкодер Купмана	0,0002	0,0131	0,0054	9000



Рис. 2.10: Схема эксперимента с акселерометром: ряды x_t и y_t входят в реконструкцию, ряд z_t остаётся контрольным

При ходьбе полилинейная модель превосходит комбинацию линейных моделей, табл. 2.5 и табл. 2.6. При спуске по лестнице различие меньше, табл. 2.9 и табл. 2.10.

Таблица 2.5: Прогноз $x_{t+\Delta t}$ и $y_{t+\Delta t}$ при ходьбе; ось x входит в реконструкцию

Размерность	U-GPVE		M-GPVE		T-GPVE	
	R^2	rMSE	R^2	rMSE	R^2	rMSE
3	0,003	0,414	0,054	0,401	0,133	0,384
4	0,046	0,404	0,099	0,39	0,206	0,367
5	0,125	0,386	0,188	0,37	0,269	0,352
6	0,17	0,376	0,244	0,356	0,337	0,334
7	0,196	0,37	0,293	0,343	0,394	0,318
8	0,224	0,363	0,327	0,334	0,447	0,302
9	0,255	0,358	0,356	0,327	0,475	0,294
10	0,277	0,353	0,368	0,324	0,51	0,284

Таблица 2.6: Прогноз контрольного ряда $z_{t+\Delta t}$ при ходьбе

Размерность	U-GPVE		M-GPVE		T-GPVE	
	R^2	rMSE	R^2	rMSE	R^2	rMSE
3	0,064	0,561	0,222	0,513	0,226	0,512
4	0,112	0,545	0,253	0,503	0,269	0,498
5	0,162	0,528	0,31	0,484	0,301	0,487
6	0,188	0,52	0,35	0,469	0,35	0,47
7	0,205	0,514	0,371	0,461	0,392	0,455
8	0,223	0,507	0,384	0,456	0,421	0,443
9	0,27	0,494	0,39	0,453	0,432	0,439
10	0,296	0,486	0,395	0,451	0,44	0,436

В случае ходьбы одинаковые значения метрики достигаются при разных размерностях фазового пространства. Для одномерного вложения $R^2 = 0,277$ достигается при размерности 10, для многомерного — при размерности 7, для тензорного — $R^2 = 0,269$ уже при размерности 5; при размерности 10 тензорное вложение даёт $R^2 = 0,51$. На контрольном ряде прирост R^2 составляет около 0,045 при всех размерностях.

На размерностях 3, 6 и 9 те же три вложения сопоставлены с автоэнкодером Купмана [36], табл. 2.7 и табл. 2.8. Значения R^2 и rMSE для U-GPVE, M-GPVE и T-GPVE совпадают с табл. 2.5 и табл. 2.6. Новое в этом сравнении — автоэнкодер и число параметров. У одномерного, многомерного и тензорного вложений число параметров одного порядка, 10^2 . У автоэнкодера оно на порядок больше, 10^3 – 10^4 . На рядах, вошедших в реконструкцию, тензорное вложение даёт больший R^2 при меньшем числе параметров. На контрольном ряде автоэнкодер немного точнее.

Таблица 2.7: Ходьба, ряды использованы при реконструкции

Размерность	U-GPVE		M-GPVE		T-GPVE		Купман		Параметры
	R^2	rMSE	R^2	rMSE	R^2	rMSE	R^2	rMSE	
3	0,003	0,414	0,054	0,401	0,133	0,384	0,091	0,398	$162/170/312/5 \cdot 10^3$
6	0,170	0,376	0,244	0,356	0,337	0,334	0,278	0,349	$342/360/318/8 \cdot 10^3$
9	0,255	0,358	0,356	0,327	0,475	0,294	0,402	0,312	$540/560/324/12 \cdot 10^3$

Таблица 2.8: Ходьба, контрольный ряд

Размерность	U-GPVE		M-GPVE		T-GPVE		Купман		Параметры
	R^2	rMSE	R^2	rMSE	R^2	rMSE	R^2	rMSE	
3	0,064	0,561	0,222	0,513	0,226	0,512	0,251	0,498	$150/300/350/3 \cdot 10^3$
6	0,188	0,520	0,350	0,469	0,350	0,470	0,387	0,451	$300/600/650/5 \cdot 10^3$
9	0,270	0,494	0,390	0,453	0,432	0,439	0,461	0,421	$450/900/950/7 \cdot 10^3$

Аналогичный эксперимент для движения вниз по лестнице также показывает устойчивый рост метрик для всех подходящих участников. Разность R^2 между одномерным, многомерным и предложенным подходами составляет около 0,03.

Таблица 2.9: Прогноз $x_{t+\Delta t}$ и $y_{t+\Delta t}$ при спуске по лестнице

Размерность	U-GPVE		M-GPVE		T-GPVE	
	R^2	rMSE	R^2	rMSE	R^2	rMSE
3	0,021	0,338	0,053	0,331	0,053	0,332
5	0,069	0,33	0,107	0,321	0,124	0,319
7	0,116	0,322	0,16	0,311	0,167	0,31
9	0,155	0,315	0,198	0,304	0,217	0,301
11	0,198	0,307	0,242	0,296	0,256	0,293
13	0,236	0,3	0,271	0,289	0,299	0,285
15	0,263	0,294	0,301	0,283	0,327	0,279

Таблица 2.10: Прогноз контрольного ряда $z_{t+\Delta t}$ при спуске по лестнице

Размерность	U-GPVE		M-GPVE		T-GPVE	
	R^2	rMSE	R^2	rMSE	R^2	rMSE
3	0,021	0,519	0,067	0,508	0,072	0,507
5	0,04	0,514	0,1	0,498	0,111	0,495
7	0,049	0,512	0,121	0,492	0,126	0,491
9	0,056	0,51	0,127	0,49	0,146	0,485
11	0,069	0,507	0,139	0,487	0,156	0,482
13	0,085	0,503	0,155	0,482	0,177	0,475
15	0,087	0,502	0,17	0,477	0,189	0,471

Глава 3

Полилинейные методы объединения пространств состояний

В данной главе ставится задача объединения пространств состояний нескольких временных рядов. Сравниваются вложение глобальных компонент, объединение исходных векторов задержек, объединение скрытых представлений и полилинейное отображение. Каноническое разложение снижает число параметров тензора связей. Приводится вычислительный эксперимент на рядах акселерометра при ходьбе и спуске по лестнице.

3.1. Задача объединения пространств состояний

Построение пространства состояний по нескольким рядам опирается на теорему Такенса и теорему Уитни о вложении [14, 16, 17]. Наблюдаемый ряд задаётся функцией h , а вектор задержек — отображением $\Phi_{h,f}$. Альтернатива задержкам — метод производных [15]: состояние восстанавливается по значениям $x_t, \dot{x}_t, \ddot{x}_t, \dots$

Параметры вложения выбираются по данным. Задержка τ оценивается по первому минимуму взаимной информации [19]. Размерность n оценивается по доле ложных ближайших соседей [20, 21]. Практическая реализация этих критериев собрана в пакете TISEAN [71]. Для многомерного ряда главные компоненты траекторной матрицы отбираются методом главных компонент и сингулярным спектральным анализом [3, 22, 24].

Объединение нескольких рядов выполняется двумя способами. Первый строит общее вложение, в котором компоненты всех рядов входят в один вектор [25]. Второй строит вложение каждого ряда отдельно и затем согласует полученные подпространства. Линейное согласование не удерживает произведения компонент разных рядов. Полилинейное отображение и каноническое разложение тензора [49] сохраняют такие произведения при меньшем числе

параметров, чем полный тензор. Тензорные авторегрессионные модели решают близкую задачу прогноза многомерного ряда, но не строят общее фазовое пространство.

3.2. Нейросетевые модели построения пространства состояний

При нелинейной динамической системе используются нейросетевые методы. Первый способ объединяет исходные векторы задержек $[\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t]^\top$. Второй объединяет уже полученные представления $[\mathbf{h}_{x,t}, \mathbf{h}_{y,t}, \mathbf{h}_{z,t}]^\top$. Во втором случае у каждого ряда своя архитектура. Пример классической архитектуры показан на рис. 3.1, пример с индивидуальными подпространствами — на рис. 3.2.

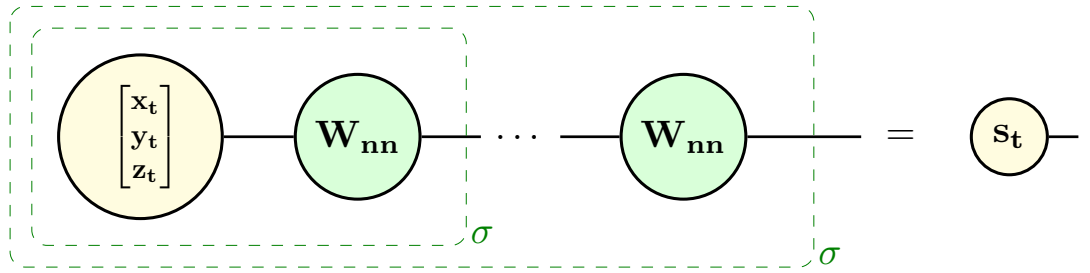


Рис. 3.1: Схема полносвязной нейронной сети с объединением исходных векторов задержек

Подобная задача формально представима, как минимизация функции ошибки восстановления исходных временных рядов. Например, минимизируется среднеквадратичная ошибка между восстановленным из фазового подпространства и исходным временным рядом

$$\hat{\theta}, \hat{\phi} = \arg \min_{\theta, \phi} \frac{1}{n} \sum_{i=1}^n \mathcal{L}(\mathbf{X}, g_{\phi}(f_{\theta}(\mathbf{X})))$$

где θ — параметры сети снижения размерности $f_{\theta} : \mathbb{R}^n \rightarrow \mathbb{R}^p$, ϕ — параметры сети восстановления ряда $g_{\phi} : \mathbb{R}^p \rightarrow \mathbb{R}^n$, $\mathcal{L}$ — функция потерь, например среднеквадратичная ошибка.

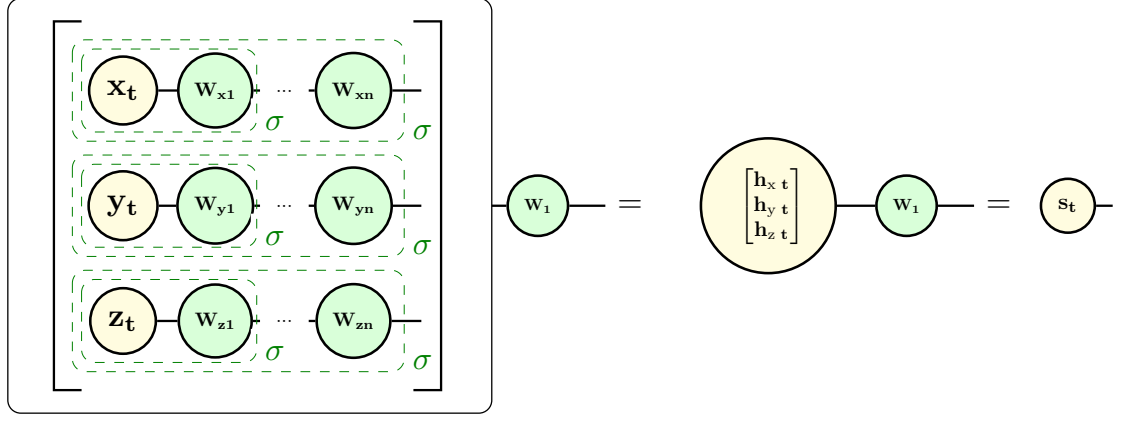


Рис. 3.2: Схема полносвязной нейронной сети с объединением преобразованных фазовых пространств каждого временного ряда, где σ — функция активации

Чтобы определить полилинейное отображение для реконструкции динамических систем, необходимо ввести тензорное произведение вектора по выбранной моде. Произведение тензора $\mathbf{A} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ и вектора $\mathbf{v} \in \mathbb{R}^{I_n}$ по n -й моде даёт тензор $\mathbf{C} \in \mathbb{R}^{I_1 \times \dots \times I_{n-1} \times I_{n+1} \times \dots \times I_N}$:

$$\mathbf{C} = \mathbf{A} \times_n^1 \mathbf{v} = \mathbf{A} \times_n \mathbf{v}, \quad c_{i_1, \dots, i_{n-1}, i_{n+1}, \dots, i_N} = \sum_{i_n=1}^{I_n} a_{i_1, \dots, i_n, \dots, i_N} v_{i_n}$$

Для объединения пространства состояний из разных временных рядов используется полилинейное преобразование вида

$$\mathbf{s}_t = \mathbf{W} \times_2 \mathbf{h}_{x,t} \times_3 \mathbf{h}_{y,t} \times_4 \mathbf{h}_{z,t}, \quad (3.1)$$

где $\mathbf{h}_{x,t}, \mathbf{h}_{y,t}, \mathbf{h}_{z,t} \in \mathbb{R}^p$ — промежуточные нейросетевые представления, $\mathbf{s}_t \in \mathbb{R}^p$ — вектор в фазовом пространстве, $p \ll n$. Уравнение (3.1) задаёт полилинейное объединение представлений разных рядов.

Модель не содержит отдельно линейных членов и части смешанных комбинаций. Константа при каждом представлении включает эти члены:

$$\mathbf{s}_t = \mathbf{W} \times_2 \begin{bmatrix} 1 \\ \mathbf{h}_{x,t} \end{bmatrix} \times_3 \begin{bmatrix} 1 \\ \mathbf{h}_{y,t} \end{bmatrix} \times_4 \begin{bmatrix} 1 \\ \mathbf{h}_{z,t} \end{bmatrix},$$

где $\mathbf{W} \in \mathbb{R}^{p \times (n+1) \times (n+1) \times (n+1)}$. На рис. 3.3 показано графическое представление предлагаемого метода в графической нотации Пенроуза.

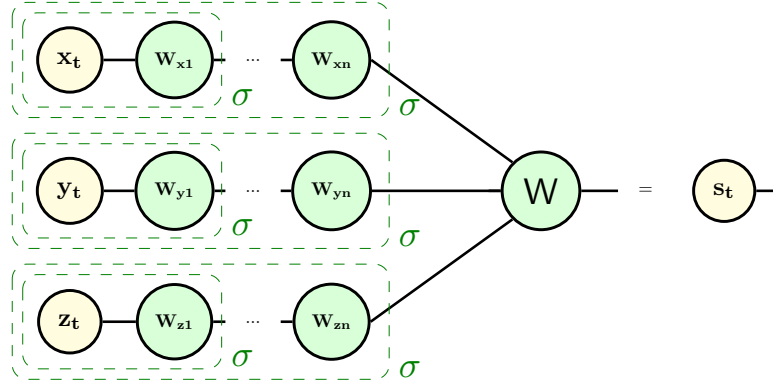


Рис. 3.3: Схема нейронной сети с полилинейным объединением фазовых пространств каждого временного ряда

Тензор W отбирает главные компоненты, как PCA для одномерных рядов, и полилинейно объединяет нейросетевые представления нескольких рядов.

Каноническое разложение используется для оптимизации W . Это разложение представляет тензор суммой R внешних произведений: $W = \sum_{r=1}^R \mathbf{w}_{r,1} \otimes \mathbf{w}_{r,2} \otimes \mathbf{w}_{r,3} \otimes \mathbf{w}_{r,4}$, где R — ранг тензора, $\mathbf{w}_{r,1} \in \mathbb{R}^p$, $\mathbf{w}_{r,2}, \mathbf{w}_{r,3}, \mathbf{w}_{r,4} \in \mathbb{R}^{n+1}$ — векторы-факторы $1 \leq r \leq R$.

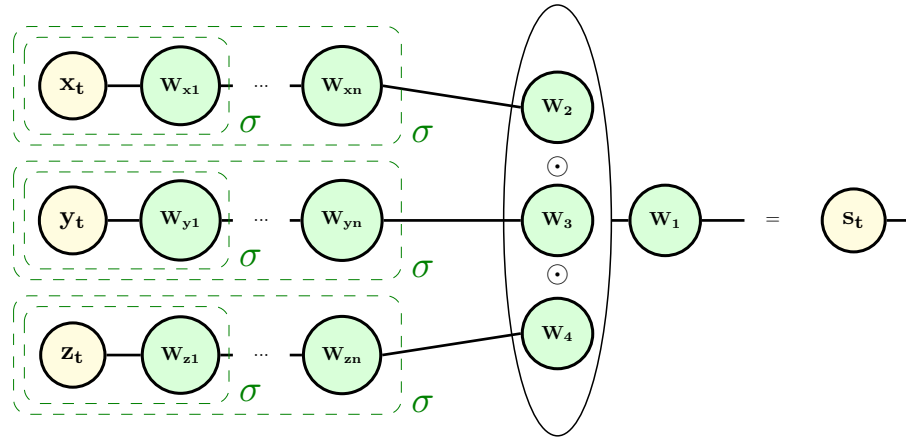


Рис. 3.4: Схема нейронной сети с полилинейным объединением фазовых пространств каждого временного ряда с каноническим разложением

В общем случае вычисление ранга R — NP-трудная задача. Поэтому ранг R подбирается для каждой задачи, и тензор аппроксимируется суммой $W \approx$

$\sum_{r=1}^R \mathbf{w}_{r,1} \otimes \mathbf{w}_{r,2} \otimes \mathbf{w}_{r,3} \otimes \mathbf{w}_{r,4}$. Это разложение дополнительно преобразует общую модель к следующему виду:

$$\begin{aligned}
\mathbf{s}_t &= \mathbf{W} \times_2 \begin{bmatrix} 1 \\ \mathbf{h}_{x,t} \end{bmatrix} \times_3 \begin{bmatrix} 1 \\ \mathbf{h}_{y,t} \end{bmatrix} \times_4 \begin{bmatrix} 1 \\ \mathbf{h}_{z,t} \end{bmatrix} \\
&\quad \text{Каноническое разложение} \\
&\approx \left(\sum_{r=1}^R \mathbf{w}_{r,1} \otimes \mathbf{w}_{r,2} \otimes \mathbf{w}_{r,3} \otimes \mathbf{w}_{r,4} \right) \times_2 \begin{bmatrix} 1 \\ \mathbf{h}_{x,t} \end{bmatrix} \times_3 \begin{bmatrix} 1 \\ \mathbf{h}_{y,t} \end{bmatrix} \times_4 \begin{bmatrix} 1 \\ \mathbf{h}_{z,t} \end{bmatrix} \\
&= \sum_{r=1}^R \mathbf{w}_{r,1} \left(\mathbf{w}_{r,2}^\top \begin{bmatrix} 1 \\ \mathbf{h}_{x,t} \end{bmatrix} \right) \left(\mathbf{w}_{r,3}^\top \begin{bmatrix} 1 \\ \mathbf{h}_{y,t} \end{bmatrix} \right) \left(\mathbf{w}_{r,4}^\top \begin{bmatrix} 1 \\ \mathbf{h}_{z,t} \end{bmatrix} \right) \\
&= \mathbf{W}_1 \left(\mathbf{W}_2^\top \begin{bmatrix} 1 \\ \mathbf{h}_{x,t} \end{bmatrix} \odot \mathbf{W}_3^\top \begin{bmatrix} 1 \\ \mathbf{h}_{y,t} \end{bmatrix} \odot \mathbf{W}_4^\top \begin{bmatrix} 1 \\ \mathbf{h}_{z,t} \end{bmatrix} \right)
\end{aligned}$$

где $\odot$ — поэлементное произведение Адамара, $\mathbf{W}_1, \dots, \mathbf{W}_4$ — фактор-матрицы мод. Изменение модели с таким разложением показано на рис. 3.4.

3.3. Анализ методов объединения пространств состояний

Сравниваются четыре способа объединения рядов: вложение глобальных компонент, объединение исходных векторов задержек, объединение скрытых представлений и полилинейное отображение. Вложение глобальных компонент — базовая линейная проекция траекторной матрицы. Число параметров сетей лежит в диапазоне 500–3000 и зависит только от длины окна задержек. Точность реконструкции и прогноза оценивается по rMSE и R^2 .

Используется набор акселерометра и гироскопа из [72]: 24 участника, частота 50 Гц, шесть типов активности. В эксперимент включены ходьба, спуск и подъём по лестнице; устройство находилось в кармане.

Сравнивается точность прогноза четырёх подходов при одинаковом объёме данных, сопоставимом числе параметров и фиксированной длине окна $n = 10$. Были использованы данные участников с непрерывным движением длительно-стью не менее 10 минут. Из данных были выбраны сегменты временных рядов

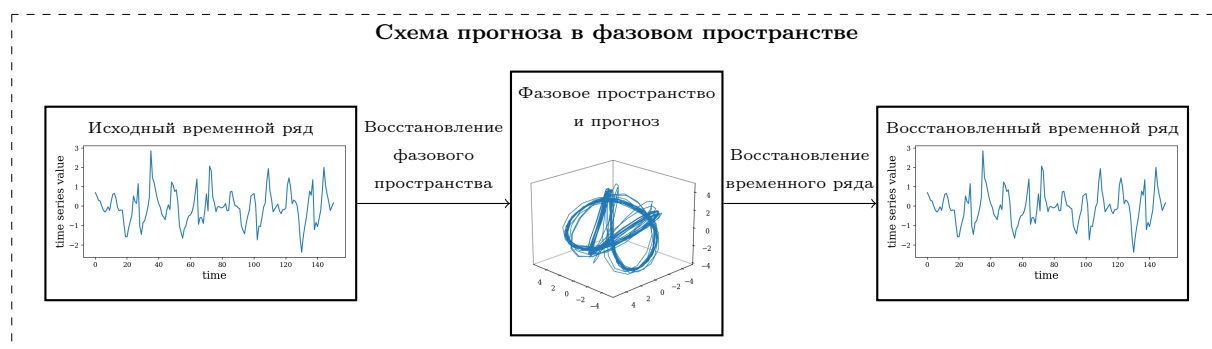


Рис. 3.5: Схема вычислительного эксперимента с оценкой качества на восстановленном временном ряду.

Таблица 3.1: Восстановление ряда при ходьбе

Размерность	Глобальное		Исходное		Скрытое		Полилинейное	
	R^2	rMSE	R^2	rMSE	R^2	rMSE	R^2	rMSE
3	0,160	0,308	0,154	0,310	0,125	0,314	0,198	0,299
4	0,205	0,299	0,203	0,301	0,170	0,306	0,254	0,289
5	0,260	0,280	0,250	0,291	0,207	0,298	0,319	0,276
6	0,280	0,275	0,298	0,282	0,247	0,291	0,361	0,267
7	0,350	0,240	0,346	0,273	0,283	0,284	0,411	0,255

Таблица 3.2: Восстановление ряда при спуске по лестнице

Размерность	Глобальное		Исходное		Скрытое		Полилинейное	
	R^2	rMSE	R^2	rMSE	R^2	rMSE	R^2	rMSE
3	0,062	0,296	0,061	0,297	0,053	0,298	0,072	0,296
4	0,101	0,292	0,098	0,291	0,086	0,293	0,102	0,291
5	0,143	0,271	0,141	0,284	0,122	0,287	0,144	0,284
6	0,180	0,271	0,178	0,278	0,155	0,281	0,183	0,278
7	0,210	0,271	0,214	0,271	0,189	0,275	0,218	0,271

без смены типа активности в пределах одного окна.

Точность восстановления при разных размерностях фазового пространства приведена в табл. 3.1 и 3.2. Точность прогнозирования напрямую зависит от размерности фазового подпространства. С увеличением размерности значения R^2 увеличиваются, а $rMSE$ уменьшаются для всех трех подходов. Это справедливо как для задачи восстановления временного ряда, соответствующего ходьбе, так и для спуска по лестнице.

Различие сильнее выражено для ходьбы. Для размерности 7 значение R^2 полилинейного подхода равно 0,411 и превышает 0,350 у вложения глобальных компонент, 0,346 у объединения исходного пространства и 0,283 у объединения скрытого пространства. Для спуска по лестнице методы дают близкие и менее точные результаты. Ранг канонического разложения в этих экспериментах фиксирован. Автоматический подбор ранга по данным и тензорная регуляризация остаются отдельной задачей.

Глава 4

Разреженное восстановление системы обыкновенных дифференциальных уравнений

В данной главе ставится задача восстановления аналитического вида дифференциальных уравнений по наблюдаемым рядам. Вводится тензорное расширение метода SINDy: взаимодействия функций-кандидатов задаются полилинейным отображением, слагаемые отбираются разреживанием фактор-матриц. Приводится сравнение с методом SINDy, нейросетевыми моделями и оператором Купмана. Вычислительный эксперимент выполнен на аттракторе Лоренца и на нормальной форме бифуркации Хопфа.

4.1. Задача восстановления системы дифференциальных уравнений

Динамическая система задаётся системой обыкновенных дифференциальных уравнений [73]. Задача восстановления явного вида уравнений возникает, когда истинные зависимости неизвестны или известны частично.

Методы восстановления уравнений [74–76] требуют априорного задания структуры системы. Функции и их комбинации выбираются по этому заданию, а параметры оцениваются по данным [45, 77]. На практике структура заранее неизвестна, и метод должен определять и структуру, и параметры [43, 78].

Нейросетевые модели дают высокую точность прогноза [38–41, 45, 79, 80], но не дают явного вида уравнений. Сравнение подходов приведено в табл. 4.1.

Явный вид уравнений восстанавливает метод разреженной идентификации нелинейных динамических систем (SINDy) [42]. Задача ставится как разреженная линейная регрессия в пространстве функций-кандидатов [27]. По наблюдаемым рядам строится матрица признаков, и регуляризация отбирает малое число слагаемых. Метод применялся к системе Лоренца, осциллятору Ван дер Поля, уравнению Курамото–Сивашинского, а также к задачам гидродинамики

и биологии [44, 77, 81, 82]. Обзор разреженных моделей и их использования в управлении дан в [83]. Устойчивость отбора при шуме повышают частотный анализ и разложения [84] и байесовские подходы [85]; оба подхода по-прежнему перебирают библиотеку кандидатов. Сетевая структура наблюдений учитывается отдельным штрафом в регрессии [86]. При высокой размерности и сложных нелинейных взаимодействиях прямой перебор библиотеки становится вычислительно затратным. Разреживание обнуляет коэффициенты по одному и не возвращает уже отброшенные функции, поэтому в модели остаются слагаемые, слабо улучшающие аппроксимацию, а слагаемые, соответствующие закону системы, отбрасываются.

Тензорное отображение [49] заменяет перебор комбинаций параметризации взаимодействий. Каноническое разложение [87] снижает число параметров. На этом построен метод T-SINDy: разреженная модель задаётся низкоранговой структурой и разреживанием фактор-матриц, а не пороговым отсечением отдельных коэффициентов.

При неполных наблюдениях аттрактор восстанавливается по теореме Такенса [16–18] и методу задержек [22]. В отличие от автоэнкодера, T-SINDy записывает уравнения в исходных переменных.

Рассматривается динамическая система

$$\frac{d}{dt}\mathbf{v}(t) = \mathbf{f}_{\text{ds}}(\mathbf{v}(t)),$$

где $\mathbf{f}_{\text{ds}} : \mathbb{R}^{d_s} \rightarrow \mathbb{R}^{d_s}$ — гладкая функция, $\mathbf{v}(t) \in \mathbb{R}^{d_s}$ — точка пространства состояний.

Истинные переменные $\mathbf{v}(t)$ недоступны. Наблюдаемый ряд $\mathbf{x}(t) \in \mathbb{R}^{d_o}$ связан с $\mathbf{v}(t)$ неизвестной функцией наблюдения $\mathbf{h} : \mathbb{R}^{d_s} \rightarrow \mathbb{R}^{d_o}$, причём $d_o < d_s$. Восстановление аналитического вида по наблюдениям есть поиск функции $\mathbf{f}_{\text{obs}}$:

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{f}_{\text{obs}}(\mathbf{x}(t)), \quad \mathbf{x}(t) = \mathbf{h}(\mathbf{v}(t)). \quad (4.1)$$

Наблюдения дискретны: $\mathbf{x}_1, \dots, \mathbf{x}_T$ в моменты $t_1 < \dots < t_T$ с шагом Δt .

Таблица 4.1: Сравнение подходов к восстановлению уравнений динамической системы

Метод	Преимущества	Недостатки
SINDy [42]	Уравнения в исходных переменных; разреженный отбор; малый объём данных	Заданная библиотека; чувствительность к шуму; нужны истинные переменные; число параметров растёт экспоненциально по размеру библиотеки
Нейросетевые (Neural ODE [38], PINN)	Аппроксимация сложных систем; устойчивость к шуму	Нет уравнений в исходных переменных; подбор архитектуры; большой объём данных
DMD / eDMD / Koopman	Линейные уравнения в поднятом пространстве; малый объём данных	Уравнения в скрытом пространстве; при нелинейном расширении те же ограничения, что у SINDy
T-SINDy	Уравнения в исходных переменных; линейный рост числа параметров по размеру библиотеки	Заданная библиотека; умеренный объём данных; зависимость от ранга разложения

Конечно-разностная схема связывает $\mathbf{f}_{\text{obs}}$ с рядом:

$$\frac{\mathbf{x}_{t+1} - \mathbf{x}_t}{\Delta t} \approx \mathbf{f}_{\text{obs}}(\mathbf{x}_t).$$

Зависимость $\mathbf{f}_{\text{obs}}$ аппроксимируется параметрической функцией $\hat{\mathbf{f}}(\cdot, \Xi)$. Задача (4.1) записывается как

$$\min_{\Xi} \frac{1}{T-1} \sum_{t=1}^{T-1} \left\| \mathbf{x}_{t+1} - \hat{\mathbf{f}}(\mathbf{x}_t, \Xi) \right\|_2^2 + \mathcal{R}(\Xi), \quad (4.2)$$

где $\mathcal{R}(\Xi)$ — регуляризатор, например L_1 -норма или групповая норма. Регуляризатор задаёт и число ненулевых компонент итоговой матрицы.

Метод SINDy восстанавливает аналитический вид системы, если наблюдаемые ряды совпадают с истинными переменными: $\mathbf{v}(t) \equiv \mathbf{x}(t) = \mathbf{x}_t = [x_{1,t}, \dots, x_{d_o,t}]^\top$.

Для разреженной регрессии строятся матрица данных $\mathbf{X}_t \in \mathbb{R}^{(T-1) \times d_o}$ и мат-

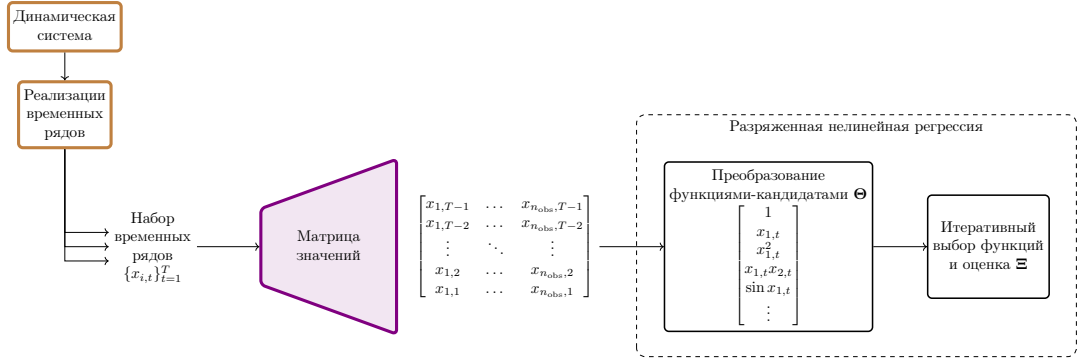


Рис. 4.1: Схема метода SINDy

рица прогнозируемых значений $\mathbf{X}_{t+1} \in \mathbb{R}^{(T-1) \times d_o}$:

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{x}_{T-1}^\top \\ \mathbf{x}_{T-2}^\top \\ \vdots \\ \mathbf{x}_1^\top \end{bmatrix} = \begin{bmatrix} x_{1,T-1} & \dots & x_{d_o,T-1} \\ x_{1,T-2} & \dots & x_{d_o,T-2} \\ \vdots & \ddots & \vdots \\ x_{1,1} & \dots & x_{d_o,1} \end{bmatrix}, \quad \mathbf{X}_{t+1} = \begin{bmatrix} \mathbf{x}_T^\top \\ \mathbf{x}_{T-1}^\top \\ \vdots \\ \mathbf{x}_2^\top \end{bmatrix} = \begin{bmatrix} x_{1,T} & \dots & x_{d_o,T} \\ x_{1,T-1} & \dots & x_{d_o,T-1} \\ \vdots & \ddots & \vdots \\ x_{1,2} & \dots & x_{d_o,2} \end{bmatrix}. \quad (4.3)$$

Библиотека функций-кандидатов $\{\theta_1, \dots, \theta_{n_\theta}\}$ состоит из полиномов, тригонометрических функций и их комбинаций. Она преобразует матрицу данных:

$$\Theta(\mathbf{X}_t) = [\theta_1(\mathbf{X}_t), \dots, \theta_{n_\theta}(\mathbf{X}_t)] \in \mathbb{R}^{(T-1) \times (n_\theta d_o)},$$

где столбец $\theta_j(\mathbf{X}_t)$ получен применением θ_j к строкам $\mathbf{X}_t$. Решение SINDy есть линейная комбинация кандидатов:

$$\mathbf{X}_{t+1} = \Theta(\mathbf{X}_t) \Xi, \quad (4.4)$$

где $\Xi \in \mathbb{R}^{(n_\theta d_o) \times d_o}$ — матрица коэффициентов.

Малое число ненулевых компонент Ξ находится из задачи (4.2) последовательным пороговым методом наименьших квадратов: на каждом шаге решается задача наименьших квадратов, затем коэффициенты ниже порога обнуляются. Схема метода показана на рис. 4.1.

4.2. Методы восстановления вида системы дифференциальный уравнений

Метод SINDy использует истинные переменные системы. Если доступны только ряды из (4.1), пространство состояний строится методом задержек. По теореме Такенса [16] при размерности вложения $n \in \mathbb{N}$ и задержке $\tau \in \mathbb{N}$ вектор

$$\mathbf{z}_t = [x_t, x_{t-\tau}, \dots, x_{t-(n-1)\tau}]^\top \in \mathbb{R}^n$$

восстанавливает пространство состояний.

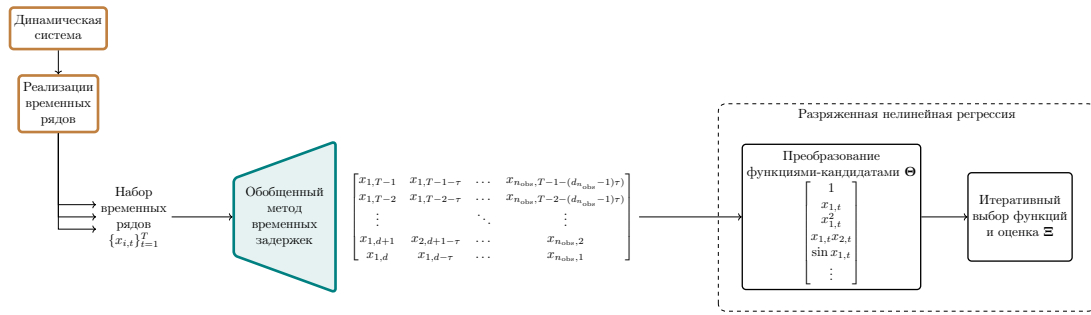


Рис. 4.2: Схема метода SINDy для наблюдаемых временных рядов

Параметры задержек выбираются двумя способами. Первый не зависит от итоговой задачи: n оценивается по доле ложных ближайших соседей, τ — по первому минимуму взаимной информации или по первому нулю автокорреляции. Второй включает n и τ в оптимизируемые переменные. Критерии, например число параметров и точность прогноза, противоречат друг другу, поэтому ищется множество Парето-оптимальных решений.

При $d_o > 1$ строится объединённый вектор задержек

$$\mathbf{z}_t = [x_{1,t}, x_{1,t-\tau}, \dots, x_{1,t-(n_1-1)\tau}, \dots, x_{d_o,t-(n_{d_o}-1)\tau}]^\top \in \mathbb{R}^{n_{\text{tot}}}, \quad (4.5)$$

где $n_{\text{tot}} = \sum_{i=1}^{d_o} n_i$.

Матрица данных метода задержек $\mathbf{Z}_t \in \mathbb{R}^{(T-\tau n-1) \times n_{\text{tot}}}$ при равных длинах

ОКОН имеет вид

$$\mathbf{Z}_t = \begin{bmatrix} \mathbf{z}_{T-1}^\top \\ \vdots \\ \mathbf{z}_{\tau n}^\top \end{bmatrix} = \begin{bmatrix} x_{1,T-1} & x_{1,T-1-\tau} & \dots & x_{d_o,T-1-(n_{d_o}-1)\tau} \\ \vdots & & \ddots & \vdots \\ x_{1,\tau n} & x_{1,\tau n-\tau} & \dots & x_{d_o,1} \end{bmatrix}.$$

По ней решается задача (4.2). Матрица прогноза при этом короче на длину окна: $\mathbf{X}_{t+1} \in \mathbb{R}^{(T-\tau n-1) \times d_o}$, последние τn строк отбрасываются. Схема показана на рис. 4.2.

Метод задержек распространяет SINDy на ряды, не совпадающие с истинными переменными. Размерность n_{tot} при этом часто избыточна из-за выбора n и τ . Избыточность создаёт две трудности. Компоненты $\mathbf{z}_t$ сильно коррелированы, матрица $\Theta(\mathbf{Z}_t)$ плохо обусловлена, и оценки Ξ неустойчивы. Число параметров растёт экспоненциально с размерностью входа и с числом функций в Θ .

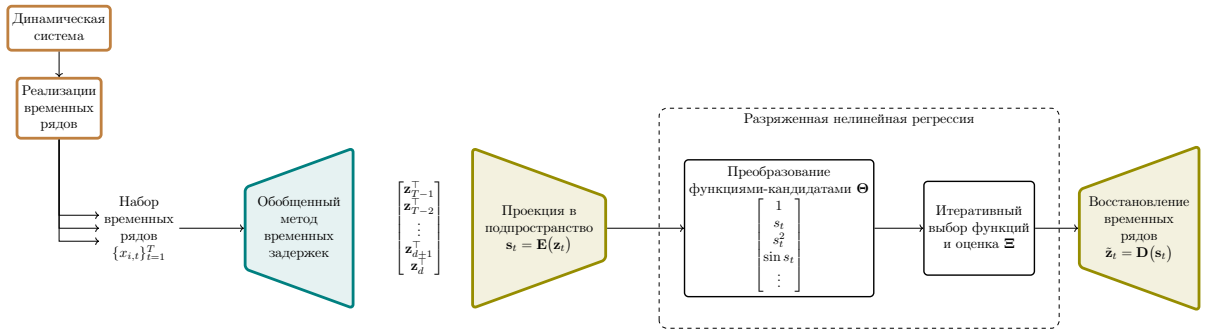


Рис. 4.3: Схема автоэнкодера для наблюдаемых временных рядов

Отображение из $\mathbb{R}^{n_{\text{tot}}}$ в подпространство $\mathbb{R}^{d_{\text{lat}}}$, $d_{\text{lat}} \ll n_{\text{tot}}$, задаётся автоэнкодером. По ряду $\mathbf{x}_t \in \mathbb{R}^{d_o}$ построены векторы $\mathbf{z}_t \in \mathbb{R}^{n_{\text{tot}}}$ согласно (4.5). Автоэнкодер имеет вид

$$\mathbf{s}_t = E(\mathbf{z}_t; \mathbf{W}_{\text{enc}}) \in \mathbb{R}^{d_{\text{lat}}}, \quad \tilde{\mathbf{z}}_t = D(\mathbf{s}_t; \mathbf{W}_{\text{dec}}) \in \mathbb{R}^{n_{\text{tot}}},$$

где $\mathbf{W}_{\text{enc}}$ и $\mathbf{W}_{\text{dec}}$ — параметры кодировщика и декодера.

В подпространстве аналитический вид восстанавливается методом SINDy.

Библиотека Θ строится по $\mathbf{s}_t$, и решается задача (4.4):

$$\mathbf{S}_{t+1} = \Theta(\mathbf{S}_t) \Xi,$$

где $\mathbf{S}_t$ и $\mathbf{S}_{t+1}$ составлены из векторов $\mathbf{s}_t$ так же, как матрицы в (4.3). Схема показана на рис. 4.3. Уравнения зависят от архитектуры автоэнкодера и не записаны в исходных переменных.

4.3. Полилинейный метод выбора функций кандидатов

В этой главе перебор библиотеки заменяется тензорным отображением, построенным по пространству состояний каждого ряда. Метод сохраняет запись уравнений в исходных переменных. Число параметров растёт линейно за счёт тензорного разложения.

Для k -й компоненты $x_{k,t}$ строится вектор задержек длины n_k с шагом τ_k по (4.5). Каждая матрица задержек дополняется столбцом единиц и преобразуется функциями-кандидатами. Получается набор из K матриц признаков $\mathbf{Y}_{k,t} \in \mathbb{R}^{(T-\tau n_k-1) \times (n_k+1)}$, $k = 1, \dots, K$. Число матриц складывается из d_o матриц задержек и $n_\theta d_o$ матриц, преобразованных кандидатами:

$$K = d_o + n_\theta d_o.$$

При $k = 1, \dots, d_o$

$$\mathbf{Y}_{k,t} = \begin{bmatrix} 1 & x_{k,T-1} & x_{k,T-1-\tau} & \dots & x_{k,T-1-(n_{d_o}-1)\tau} \\ 1 & x_{k,T-2} & x_{k,T-2-\tau} & \dots & x_{k,T-2-(n_{d_o}-1)\tau} \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & x_{k,n} & x_{k,n-\tau} & \dots & x_{k,1} \end{bmatrix},$$

а при $k = d_o + 1, \dots, K$

$$\mathbf{Y}_{k,t} = \begin{bmatrix} 1 & \theta_k(x_{k,T-1}) & \theta_k(x_{k,T-1-\tau}) & \dots & \theta_k(x_{k,T-1-(n_{d_o}-1)\tau}) \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & \theta_k(x_{k,n}) & \theta_k(x_{k,n-\tau}) & \dots & \theta_k(x_{k,1}) \end{bmatrix}.$$

Индекс k в θ_k нумерует матрицу $\mathbf{Y}_{k,t}$, а не элемент библиотеки. Порядковый номер в библиотеке равен $k - d_o$.

Полилинейное отображение как полная регрессия. В SINDy коэффициенты матрицы Ξ добавляются и обнуляются по порогу. В предлагаемом обобщении та же регрессия записывается тензором $\mathbf{W}$:

$$x_{i,t+1} = \sum_{j_1=1}^{n_1+1} \cdots \sum_{j_K=1}^{n_K+1} \mathbf{W}_{j_1, \dots, j_K, i} \prod_{k=1}^K y_{k, j_k, t}, \quad \mathbf{W} \in \mathbb{R}^{(n_1+1) \times \cdots \times (n_K+1) \times d_o}, \quad (4.6)$$

где $y_{k, j_k, t}$ — j_k -й элемент строки признаков k -й модальности в момент t . Уравнение (4.6) есть линейная регрессия по всем комбинациям значений матриц кандидатов.

Пример. Пусть ряд имеет длину $T = 8$ и две компоненты $x_{1,t}$, $x_{2,t}$, библиотека состоит из одного преобразования $\{\sin(\cdot)\}$, а $n_1 = n_2 = 2$, $\tau_1 = \tau_2 = 1$. Матрицы признаков

$$\mathbf{Y}_{1,t} = \begin{bmatrix} 1 & x_{1,7} & x_{1,6} \\ 1 & x_{1,6} & x_{1,5} \\ 1 & x_{1,5} & x_{1,4} \\ 1 & x_{1,4} & x_{1,3} \\ 1 & x_{1,3} & x_{1,2} \\ 1 & x_{1,2} & x_{1,1} \end{bmatrix}, \quad \mathbf{Y}_{2,t} = \begin{bmatrix} 1 & x_{2,7} & x_{2,6} \\ 1 & x_{2,6} & x_{2,5} \\ 1 & x_{2,5} & x_{2,4} \\ 1 & x_{2,4} & x_{2,3} \\ 1 & x_{2,3} & x_{2,2} \\ 1 & x_{2,2} & x_{2,1} \end{bmatrix},$$

$$\mathbf{Y}_{3,t} = \begin{bmatrix} 1 & \sin(x_{1,7}) & \sin(x_{1,6}) \\ 1 & \sin(x_{1,6}) & \sin(x_{1,5}) \\ 1 & \sin(x_{1,5}) & \sin(x_{1,4}) \\ 1 & \sin(x_{1,4}) & \sin(x_{1,3}) \\ 1 & \sin(x_{1,3}) & \sin(x_{1,2}) \\ 1 & \sin(x_{1,2}) & \sin(x_{1,1}) \end{bmatrix}, \quad \mathbf{Y}_{4,t} = \begin{bmatrix} 1 & \sin(x_{2,7}) & \sin(x_{2,6}) \\ 1 & \sin(x_{2,6}) & \sin(x_{2,5}) \\ 1 & \sin(x_{2,5}) & \sin(x_{2,4}) \\ 1 & \sin(x_{2,4}) & \sin(x_{2,3}) \\ 1 & \sin(x_{2,3}) & \sin(x_{2,2}) \\ 1 & \sin(x_{2,2}) & \sin(x_{2,1}) \end{bmatrix},$$

матрица прогноза

$$\mathbf{X}_{t+1} = \begin{bmatrix} x_{1,8} & x_{2,8} \\ x_{1,7} & x_{2,7} \\ x_{1,6} & x_{2,6} \\ x_{1,5} & x_{2,5} \\ x_{1,4} & x_{2,4} \\ x_{1,3} & x_{2,3} \end{bmatrix}.$$

Внешнее произведение четырёх векторов длины 3 даёт признак длины 81. Модель имеет вид

$$\mathbf{X}_{t+1} = \begin{bmatrix} 1 & \mathbf{L}_t & \mathbf{B}_t & \cdots & \mathbf{T}_t \end{bmatrix} \mathbf{W}_{(K+1)},$$

где $\mathbf{L}_t$ — линейная часть с компонентами $[1, x_{1,t}, x_{1,t-1}, x_{2,t}, x_{2,t-1}, \sin(x_{1,t}), \sin(x_{1,t-1}), \sin(x_{2,t}), \sin(x_{2,t-1})]$, $\mathbf{B}_t$ — билинейная часть, парные произведения этих компонент, $\mathbf{T}_t$ — произведения более высокого порядка, например $x_{1,t}x_{2,t} \sin(x_{1,t}) \sin(x_{2,t})$, а $\mathbf{W}_{(K+1)} \in \mathbb{R}^{d_o \times (n+1)^K}$ — развёртка тензора по моде прогноза.

Каноническое разложение для снижения числа параметров. Полный тензор содержит $d_o \prod_{k=1}^K (n_k + 1)$ элементов. Каноническое разложение ранга $R \ll \min\{n_1, \dots, n_K\}$ записывает его суммой внешних произведений:

$$\mathbf{W} = \sum_{r=1}^R \mathbf{w}_{1,r} \otimes \mathbf{w}_{2,r} \otimes \cdots \otimes \mathbf{w}_{K,r} \otimes \mathbf{w}_{\text{out},r}, \quad (4.7)$$

где $\mathbf{w}_{k,r} \in \mathbb{R}^{n_k+1}$, $\mathbf{w}_{\text{out},r} \in \mathbb{R}^{d_o}$.

Подстановка (4.7) в (4.6) даёт регрессию

$$\mathbf{X}_{t+1} = \left(\bigodot_{k=1}^K (\mathbf{Y}_{k,t} \mathbf{W}_k) \right) \mathbf{W}_{\text{out}},$$

где $\mathbf{W}_k \in \mathbb{R}^{(n_k+1) \times R}$, $\mathbf{W}_{\text{out}} \in \mathbb{R}^{R \times d_o}$, $\odot$ — произведение Адамара. Число параметров равно $R(\sum_{k=1}^K (n_k + 1) + d_o)$ и линейно по K и R . Обнуление отдельных коэффициентов и обнуление компоненты фактора сопоставлены на рис. 4.4. Схема метода показана на рис. 4.5.

(а) отдельные коэффициенты

ξ_{11}	0	ξ_{13}
0	ξ_{22}	ξ_{23}
ξ_{31}	ξ_{32}	0

нулю равны три элемента

(б) нулевая компонента фактора

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \begin{pmatrix} b_1 & 0 & b_3 \end{pmatrix}$$

$a_1 b_1$	0	$a_1 b_3$
$a_2 b_1$	0	$a_2 b_3$
$a_3 b_1$	0	$a_3 b_3$

нулю равен целый столбец

Рис. 4.4: Отбор слагаемых. В (а) обнуляются отдельные коэффициенты, в (б) нуль в факторе удаляет целый столбец.

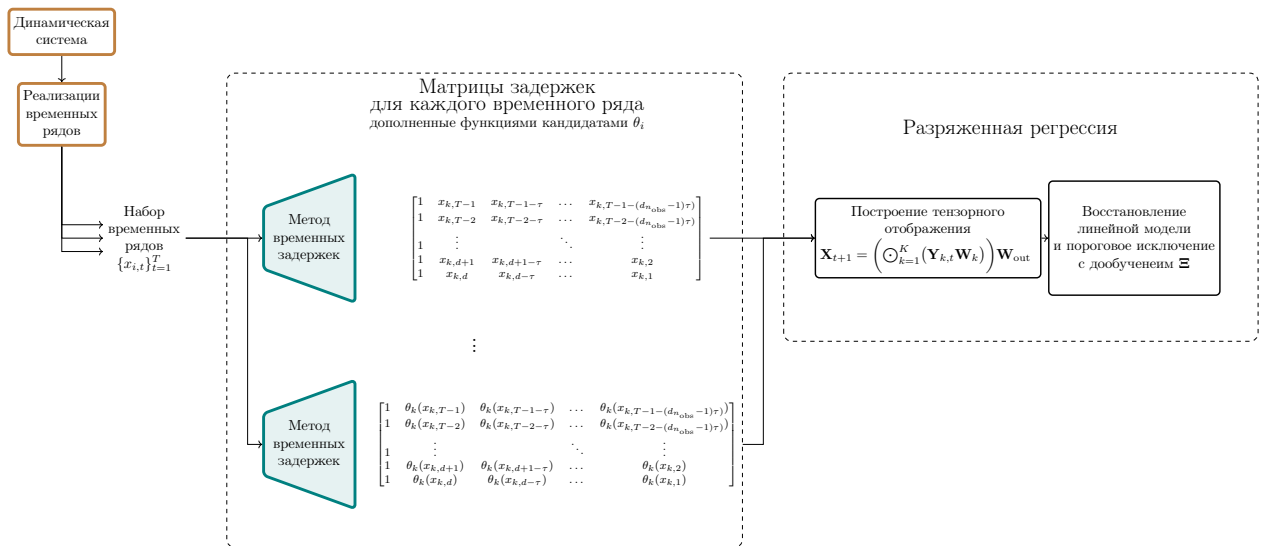


Рис. 4.5: Схема тензорного метода SINDy для наблюдаемых временных рядов

После нескольких итераций элементы фактор-матриц с $|\mathbf{W}_k| < \varepsilon$ обнуляются, и модель дообучается при фиксированных нулях минимизацией (4.2). Дополнительное отсечение $|\mathbf{W}_{(K+1)}| < \delta$ оставляет малое число слагаемых. Порог ε действует на факторы канонического разложения, порог δ — на развёртку тензора, восстановленную из уже разреженных факторов.

Метод сглаживает шум за счёт низкого ранга, заменяет экспоненциальный перебор линейным числом параметров и сохраняет уравнения в исходных переменных. Условие Крускала гарантирует единственность самого разложения тензора и не описывает, какие мономы исчезают при пороге по факторам.

4.4. Анализ восстановленных системы дифференциальных уравнений

Эксперимент проверяет восстановление уравнений по неполным данным на системах, заданных обыкновенными дифференциальными уравнениями. Автоэнкодер в сравнение не входит: у него нет уравнений в исходных переменных.

Рассматривается аттрактор Лоренца

$$\begin{cases} \dot{x} = \sigma(y - x), \\ \dot{y} = x(\rho - z) - y, \\ \dot{z} = xy - \beta z, \end{cases} \quad \sigma = 10, \rho = 28, \beta = \frac{8}{3}.$$

Ряд имеет длину $T = 20\,000$ и шаг $\Delta t = 0,01$. В обучение входят $x(t)$ и $y(t)$, компонента $z(t)$ не наблюдается. К сигналам добавляется независимый гауссовский шум с уровнем от 1% до 10% стандартного отклонения соответствующей компоненты. Для каждого уровня шума перебираются длина окна $n \in \{3, 6, 9, 12, 15\}$ и начало обучающей выборки $t_0 \in \{2000, 2500, \dots, 5000\}$. Сдвиг $t_0 = 2000$ отсекает переходный процесс. Строятся четыре матрицы признаков: задержки x_t и y_t и те же задержки после преобразования $\sin$. Разреженность задаётся L_1 -

регуляризатором факторных матриц:

$$\mathcal{L} = \frac{1}{T} \|\mathbf{X}_{t+1} - \hat{\mathbf{X}}_{t+1}\|_F^2 + \lambda \sum_{k=1}^4 \|\mathbf{W}_k\|_1,$$

где $\lambda = 10^{-4}$. После оптимизации элементы факторных матриц с модулем меньше $\varepsilon = 0,01$ обнуляются, и модель дообучается без изменения нулей.

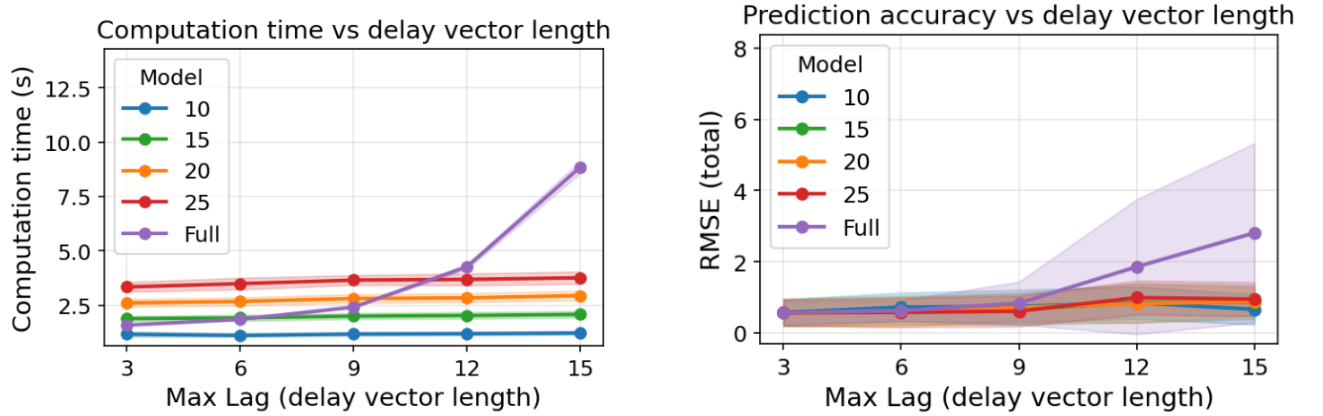


Рис. 4.6: Время обучения и среднеквадратичная ошибка прогноза в зависимости от длины окна

На рис. 4.6 время обучения классического SINDy быстро растёт с n . У предлагаемого метода рост медленнее и зависит от ранга канонического разложения. Ошибка всех подходов минимальна при $n = 6-9$ и далее растёт из-за избыточной размерности.

Наиболее значимые слагаемые дают уравнения

$$\begin{aligned} x(t+1) &= 0,01 x_t + 0,98 y_t, \\ y(t+1) &= 0,9 y_t + 0,9 x_t + 0,25 \sin(y_t). \end{aligned}$$

Метод выделяет доминирующие линейные члены и нелинейную поправку $\sin(y_t)$, через которую учитывается скрытая компонента $z(t)$.

Вторая система — нормальная форма бифуркации Хопфа:

$$\begin{cases} \dot{x} = \mu x + \omega y - Ax(x^2 + y^2), \\ \dot{y} = -\omega x + \mu y - Ay(x^2 + y^2), \end{cases} \quad \mu = 0,1, \omega = 1, A = 1.$$

Наблюдается только $x(t)$. Влияние $y(t)$ восстанавливается через задержки и нелинейные преобразования x . Длина ряда $T = 20\,000$, шаг $\Delta t = 0,01$. Шум — аддитивный гауссовский, уровни 0%, 1%, ..., 9% стандартного отклонения x . Для каждого уровня перебираются $n \in \{3, 6, 9, 12, 15\}$ и $t_0 \in \{2000, 2500, \dots, 5000\}$.

При $n = 9$, нулевом шуме и $t_0 = 2000$ восстановленное уравнение имеет вид

$$x(t+1) = -0,0001 + 0,3484 x_{t-3} + 0,6578 \sin(x_{t-3}).$$

Модель оставляет x_{t-3} и $\sin(x_{t-3})$. Прогноз при нулевом шуме показан на рис. 4.7.

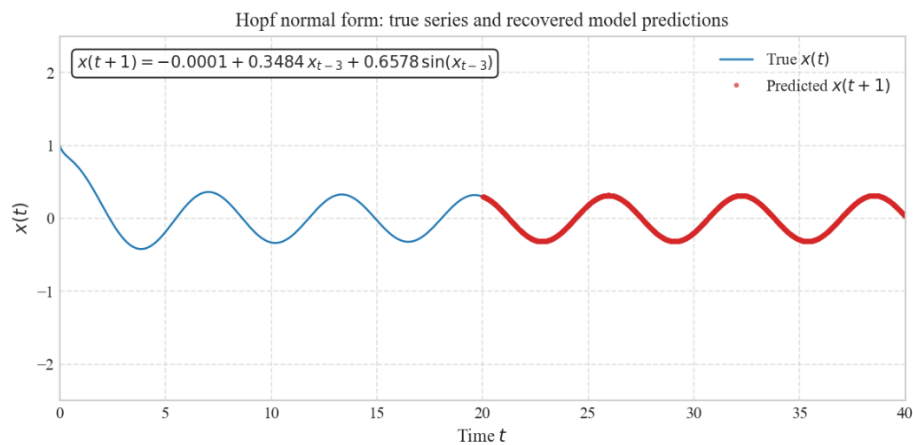


Рис. 4.7: Восстановленное уравнение и прогноз временного ряда для нормальной формы Хопфа

Таблица 4.2: Точность и время обучения для системы Хопфа, средние значения при $n = 15$

Модель	Среднее RMSE	Среднее время, с
SINDy	0,0144	5,3
T-SINDy, $R = 10$	0,0094	1,27
T-SINDy, $R = 15$	0,0098	2,18
T-SINDy, $R = 20$	0,0102	2,99
T-SINDy, $R = 25$	0,0100	3,87

В табл. 4.2 приведены средние по начальным индексам и уровням шума при фиксированной длине окна $n = 15$. Оба метода восстанавливают динамическую систему с сопоставимой точностью. Время обучения T-SINDy меньше, чем у SINDy.

Метод T-SINDy объединяет временные задержки, тензорное отображение функций-кандидатов и разреженную регрессию. Каноническое разложение заменяет экспоненциальный перебор комбинаций линейным ростом числа параметров. Уравнения остаются записанными в исходных переменных. На аттракторе Лоренца по рядам x и y восстанавливаются доминирующие линейные члены и поправка, связанная со скрытой компонентой. На нормальной форме Хопфа при одной наблюдаемой компоненте время обучения меньше, чем у SINDy, при сопоставимой ошибке прогноза.

Глава 5

Перенос пространства состояний внутри класса динамических систем

В данной главе ставится задача восстановления пространства состояний динамической системы по короткому временному ряду и вектору статических параметров. Системы одного класса описываются общим законом эволюции и различаются статическими параметрами. Вводится полилинейное отображение вектора задержек и вектора статических параметров, общее для всех систем класса. Затем переход между пространствами состояний двух систем задаётся матрицей Якоби, линейной по приращению параметров. Доказывается липшицевость параметризованной проекции. Приводятся вычислительные эксперименты на рядах акселерометра.

Рассматривается конечный набор динамических систем $\{\mathbf{E}_i\}_{i=1}^{N_E}$ одного класса. Классом называются системы, описываемые одним набором ОДУ и отличающиеся статическими параметрами, например математические маятники разной длины и массы. Каждая система $\mathbf{E}_i$ характеризуется вектором статических параметров $\mathbf{c}_i \in \mathcal{C} \subset \mathbb{R}^{d_c}$ и временным рядом $\{x_{i,t}\}_{t=1}^{T_i}$. Функция эволюции $\mathbf{E}$ неизвестна и задаёт будущее состояние системы

$$x_{i,t+1} = \mathbf{E}(x_{i,t}, \mathbf{c}_i),$$

где зависимость от $\mathbf{c}_i$ выделяет систему $\mathbf{E}_i$ внутри класса.

Так как исходное пространство состояний неизвестно, по конечному ряду строится отображение в пространство состояний, сохраняющее зависимости системы:

$$\mathbf{s}_{i,t} = \tilde{\mathbf{E}}(x_{i,1}, \dots, x_{i,T_i}, \mathbf{c}_i, \mathbf{w}_i),$$

где $\mathbf{w}_i \in \mathbb{R}^{N_w}$ — настраиваемые параметры модели, $\mathbf{s}_{i,t} \in \mathbb{R}^p$, p — размерность пространства состояний. Предполагается, что восстановленное пространство состояний диффеоморфно истинному аттрактору системы $\mathbf{E}_i$.

Предполагается, что существует гладкое отображение $\mathbf{P} : \mathbb{R}^p \rightarrow \mathbb{R}^p$, связывающее пространства состояний двух систем класса:

$$\mathbf{s}_{j,t} = \mathbf{P}(\mathbf{s}_{i,t}, \mathbf{c}_i, \mathbf{c}_j, \mathbf{w}_P) + \varepsilon_t,$$

где ε_t — шум, $\mathbf{w}_P \in \mathbb{R}^{N_w}$ — настраиваемые параметры отображения. Различия пространств состояний полностью определяются статическими параметрами $\mathbf{c}_i$ и $\mathbf{c}_j$. Таким образом, значения $\mathbf{c}_i$ и $\mathbf{c}_j$ параметризуют изменение координат пространств состояний.

Задача: для новой системы $\mathbf{E}_{N_E+1}$ с известным вектором статических параметров $\mathbf{c}_{N_E+1}$ и коротким временным рядом $\{x_{N_E+1,t}\}_{t=1}^{T_{\text{short}}}$, недостаточным для построения пространства состояний, найти отображение $\tilde{\mathbf{E}}_{N_E+1}$ такое, что восстановленное пространство состояний $\hat{\mathbf{S}}_{N_E+1}$ аппроксимирует пространство $\mathbf{S}_{N_E+1}$, которое было бы получено по полному ряду. Задача представима в виде минимизации расстояния между известными пространствами состояний и их оценками:

$$\min_{\mathbf{w}_P} \sum_{i \neq j} \|\mathbf{S}_j - \mathbf{P}(\mathbf{S}_i, \mathbf{c}_i, \mathbf{c}_j, \mathbf{w}_P)\|^2.$$

Задача сводится к восстановлению отображения $\mathbf{P}$ по набору систем. По этому отображению пространства состояний новых систем строятся по коротким рядам и статическим параметрам.

Ниже отображение $\mathbf{P}$ задаётся двумя способами. В первом проекция каждой системы получается свёрткой общего тензора с вектором статических параметров. Во втором проекция новой системы получается из проекции ближайшей по параметрам системы матрицей Якоби, линейной по приращению параметров. Оба перехода показаны на рис. 5.1.

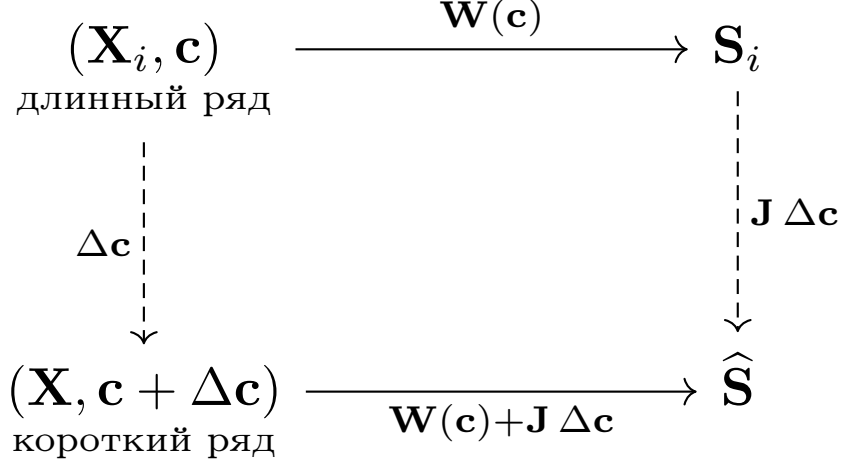


Рис. 5.1: Переход к системе с коротким рядом. Приращение статических параметров переносится матрицей Якоби.

5.1. Полилинейная параметризация семейства проекций статическими параметрами

Пространство состояний каждой системы восстанавливается методом задержек. По ряду $\{x_{i,t}\}_{t=1}^{T_i}$ строится траекторная матрица $\mathbf{X}_i$ вида (1.4) со столбцами $\mathbf{x}_{i,t} = [x_{i,t-n+1}, \dots, x_{i,t}]^\top$. Базовая параметрическая модель имеет вид

$$\mathbf{s}_{i,t} = \mathbf{A}(\mathbf{c}_i) \mathbf{x}_{i,t}, \quad (5.1)$$

где $\mathbf{A}(\mathbf{c}_i) \in \mathbb{R}^{p \times n}$ — матрица проекции, зависящая от статических параметров. При фиксированном $\mathbf{c}_i$ это проекция вида (1.6) с матрицей $\mathbf{W}_i = \mathbf{A}(\mathbf{c}_i)$, отдельной для каждой системы.

Модель (5.1) не задаёт, как матрица проекции меняется при изменении $\mathbf{c}_i$. Зависимость от параметров задаётся полилинейным отображением, связывающим пространства состояний систем класса:

$$\mathbf{s}_{i,t} = \mathbf{A} \times_2 \mathbf{x}_{i,t} \times_3 \mathbf{c}_i, \quad (5.2)$$

где $\mathbf{A} \in \mathbb{R}^{p \times n \times d_c}$ — тензор параметров модели, а $\times_2$ и $\times_3$ — произведения по модам (2.2). Тензор $\mathbf{A}$ один для всех систем класса; произведения по модам задают

взаимодействие задержек и статических параметров.

Модель (5.2) не содержит отдельно линейных по $\mathbf{x}_{i,t}$ и линейных по $\mathbf{c}_i$ членов. Константа, добавленная к вектору задержек и к вектору статических параметров, включает эти члены в модель:

$$\mathbf{s}_{i,t} = \mathbf{B} \times_2 \begin{bmatrix} 1 \\ \mathbf{x}_{i,t} \end{bmatrix} \times_3 \begin{bmatrix} 1 \\ \mathbf{c}_i \end{bmatrix} = \mathbf{B}(\mathbf{c}_i) \begin{bmatrix} 1 \\ \mathbf{x}_{i,t} \end{bmatrix}, \quad (5.3)$$

где $\mathbf{B} \in \mathbb{R}^{p \times (n+1) \times (d_c+1)}$ — тензор параметров модели, $\mathbf{B}(\mathbf{c}_i) \in \mathbb{R}^{p \times (n+1)}$ — матрица проекции, полученная свёрткой тензора с дополненным вектором статических параметров. Схема отображения в нотации Пенроуза представлена на рис. 5.2.

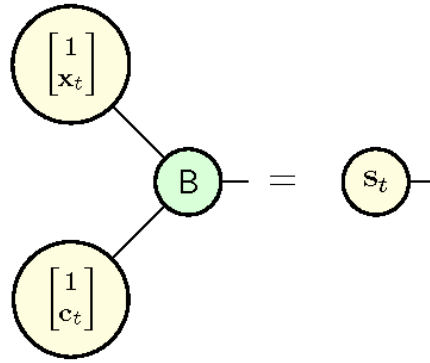


Рис. 5.2: Схема полилинейного отображения вектора задержек и вектора статических параметров. Входные данные взаимодействуют через тензор $\mathbf{B}$, формируя вектор состояния в пространстве низкой размерности.

Тензор $\mathbf{B}$ задаёт не только линейную проекцию, как в методе главных компонент, но и полилинейную зависимость задержек и статических параметров. Число операций остаётся линейным по размеру тензора. Следующая теорема показывает, что матрица проекции непрерывно зависит от статических параметров.

Теорема 6 (Липшицева непрерывность параметризованного полилинейного отображения). Пусть $\mathcal{C} \subset \mathbb{R}^{d_c}$ — множество допустимых статических характеристик. Рассмотрим семейство полилинейных отображений

$$\{ \mathbf{B}(\mathbf{c}) : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^p \}_{\mathbf{c} \in \mathcal{C}},$$

заданных как

$$\mathbf{B}(\mathbf{c}) = \mathbf{B}_0 \times_3 \begin{pmatrix} 1 \\ \mathbf{c} \end{pmatrix},$$

где $\mathbf{B}_0 \in \mathbb{R}^{p \times (n+1) \times (d_c+1)}$ — фиксированный тензор. Тогда отображение

$$\Psi : \mathcal{C} \rightarrow \mathbb{R}^{p \times (n+1)}, \quad \mathbf{c} \mapsto \mathbf{B}(\mathbf{c})$$

липшицево. Существует константа $L > 0$ такая, что для любых $\mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C}$

$$\|\mathbf{B}(\mathbf{c}_1) - \mathbf{B}(\mathbf{c}_2)\|_F \leq L \|\mathbf{c}_1 - \mathbf{c}_2\|_2,$$

где $\|\cdot\|_F$ — норма Фробениуса матрицы $p \times (n+1)$, а $\|\cdot\|_2$ — евклидова норма вектора параметров.

Доказательство. Рассмотрим разность значений отображения Ψ для двух произвольных векторов параметров $\mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C}$. Произведение по третьей моде линейно, поэтому

$$\mathbf{B}(\mathbf{c}_1) - \mathbf{B}(\mathbf{c}_2) = \mathbf{B}_0 \times_3 \begin{pmatrix} 1 \\ \mathbf{c}_1 \end{pmatrix} - \mathbf{B}_0 \times_3 \begin{pmatrix} 1 \\ \mathbf{c}_2 \end{pmatrix} = \mathbf{B}_0 \times_3 \begin{pmatrix} 0 \\ \mathbf{c}_1 - \mathbf{c}_2 \end{pmatrix}.$$

Элемент результирующей матрицы с индексами (α, β) равен $\sum_{\gamma} (\mathbf{B}_0)_{\alpha\beta\gamma} v_{\gamma}$, где $\mathbf{v} = (0, \mathbf{c}_1 - \mathbf{c}_2)^{\top}$. Тогда

$$\left(\sum_{\gamma} (\mathbf{B}_0)_{\alpha\beta\gamma} v_{\gamma} \right)^2 \leq \left(\sum_{\gamma} (\mathbf{B}_0)_{\alpha\beta\gamma}^2 \right) \|\mathbf{v}\|_2^2.$$

Суммирование по α и β даёт

$$\|\mathbf{B}(\mathbf{c}_1) - \mathbf{B}(\mathbf{c}_2)\|_F^2 \leq \|\mathbf{B}_0\|_F^2 \cdot \|\mathbf{v}\|_2^2,$$

и после извлечения квадратного корня

$$\|\mathbf{B}(\mathbf{c}_1) - \mathbf{B}(\mathbf{c}_2)\|_F \leq \|\mathbf{B}_0\|_F \cdot \|\mathbf{v}\|_2.$$

Евклидова норма вектора с нулевой первой координатой совпадает с нормой разности параметров:

$$\|\mathbf{v}\|_2 = \|\mathbf{c}_1 - \mathbf{c}_2\|_2.$$

Следовательно,

$$\|\mathbf{B}(\mathbf{c}_1) - \mathbf{B}(\mathbf{c}_2)\|_F \leq L \|\mathbf{c}_1 - \mathbf{c}_2\|_2, \quad L = \|\mathbf{B}_0\|_F.$$

Тензор $\mathbf{B}_0$ фиксирован, поэтому $L < \infty$ и отображение Ψ липшицево. $\square$

Для новой системы формируется алгоритм восстановления пространства состояний, состоящий из двух этапов — обучения и восстановления.

Этап обучения. Для каждого ряда $\{x_{i,t}\}_{t=1}^{T_i}$ обучающего набора $\{\mathbf{E}_i\}_{i=1}^{N_E}$ строится траекторная матрица $\mathbf{X}_i$ вида (1.4), и каждый её столбец дополняется константой, как в (5.3). Для каждого столбца и вектора статических параметров строится векторизованное внешнее произведение

$$\mathbf{z}_{i,t} = \text{vec} \left(\begin{bmatrix} 1 \\ \mathbf{x}_{i,t} \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{c}_i \end{bmatrix}^\top \right) \in \mathbb{R}^{(n+1)(d_c+1)}.$$

Векторы $\mathbf{z}_{i,t}$ всех систем и моментов времени образуют общую матрицу. К ней применяется метод главных компонент (1.5); матрица проекции на p главных компонент есть развёртка $\mathbf{B}_{(1)} \in \mathbb{R}^{p \times (n+1)(d_c+1)}$ тензора параметров по первой моде. Развёртка сворачивается в тензор $\mathbf{B} \in \mathbb{R}^{p \times (n+1) \times (d_c+1)}$.

Этап восстановления. Для новой системы с вектором статических параметров $\mathbf{c}_{N_E+1}$ и коротким рядом $\{x_{N_E+1,t}\}_{t=1}^{T_{\text{short}}}$ вычисляется матрица проекции

$$\mathbf{B}(\mathbf{c}_{N_E+1}) = \mathbf{B} \times_3 \begin{bmatrix} 1 \\ \mathbf{c}_{N_E+1} \end{bmatrix}.$$

По аналогии с (1.6) она задаёт пространство состояний короткого ряда:

$$\mathbf{S}_{N_E+1} = \mathbf{B}(\mathbf{c}_{N_E+1}) \tilde{\mathbf{X}}_{N_E+1}, \quad (5.4)$$

где $\tilde{\mathbf{X}}_{N_E+1}$ — траекторная матрица (1.4) нового ряда, дополненная строкой единиц. Обратное отображение в пространство ряда задаётся псевдообратной матрицей $\mathbf{B}(\mathbf{c}_{N_E+1})^\dagger$.

Процедура строит точки пространства состояний по статическим параметрам новой системы и короткому ряду. Обучение требует только одного тензора на весь класс.

5.2. Параметризация перехода между пространствами состояний матрицей Якоби

В модели (5.3) проекция новой системы получается из общего тензора. В этом разделе проекция новой системы получается из проекции уже известной системы класса.

Пусть заданы два временных ряда $\{x_{1,t}\}_{t=1}^N$ и $\{x_{2,t}\}_{t=1}^N$, порождённые системами одного класса со статическими параметрами $\mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C}$. После построения траекторных матриц (1.4) и проекций (1.6) с матрицами $\mathbf{W}_1$ и $\mathbf{W}_2$ получаются траектории $\{\mathbf{s}_{1,t}\}_{t=1}^k \subset \mathbb{R}^p$ и $\{\mathbf{s}_{2,t}\}_{t=1}^k \subset \mathbb{R}^p$. Предполагается, что существует гладкое отображение $\mathbf{P} : \mathbb{R}^p \rightarrow \mathbb{R}^p$, связывающее два пространства состояний: $\mathbf{s}_{2,t} = \mathbf{P}(\mathbf{s}_{1,t}) + \varepsilon_t$, где ε_t — шум. Линейная часть отображения $\mathbf{P}$ в окрестности точки $\mathbf{s}_{1,0}$ задаётся матрицей Якоби

$$\mathbf{s}_{2,t} = \mathbf{P}(\mathbf{s}_{1,0}) + \mathbf{J}(\mathbf{s}_{1,t} - \mathbf{s}_{1,0}) + o(\|\mathbf{s}_{1,t} - \mathbf{s}_{1,0}\|), \quad \mathbf{J} = \begin{bmatrix} \frac{\partial P_1}{\partial s_{1,1}} & \cdots & \frac{\partial P_1}{\partial s_{1,p}} \\ \vdots & \ddots & \vdots \\ \frac{\partial P_p}{\partial s_{1,1}} & \cdots & \frac{\partial P_p}{\partial s_{1,p}} \end{bmatrix} \in \mathbb{R}^{p \times p},$$

где P_k — k -я координата отображения $\mathbf{P}$, $s_{1,l}$ — l -я координата вектора $\mathbf{s}_{1,t}$.

Для аффинного отображения $\mathbf{P}$ матрица Якоби не зависит от точки, в окрестности которой вычисляются частные производные, а для центрированных траекторий остаток и сдвиг отсутствуют: $\mathbf{s}_{2,t} = \mathbf{J}\mathbf{s}_{1,t}$. Такую зависимость удобно параметризовать статическими параметрами систем.

Предполагается, что матрица Якоби $\mathbf{J}$ зависит от разности статических параметров $\mathbf{c}_i, \mathbf{c}_j \in \mathcal{C}$ через параметризованное отображение $\mathcal{J} : \mathcal{C} \rightarrow \mathbb{R}^{p \times p}$. Тогда при наличии пар параметр–матрица $\{(\mathbf{c}_i, \mathbf{J}_i)\}_{i=1}^{N_E}$ строится линейная по приращению параметров модель матрицы Якоби

$$\mathbf{J}_{i \rightarrow j} = \mathbf{I} + \mathbf{J} \times_3 \Delta \mathbf{c}_{i,j}, \quad \Delta \mathbf{c}_{i,j} = \mathbf{c}_j - \mathbf{c}_i, \quad (5.5)$$

где $\mathbf{J} \in \mathbb{R}^{p \times p \times d_c}$ — тензор параметров, $\mathbf{I} \in \mathbb{R}^{p \times p}$ — единичная матрица, $\mathbf{J} \times_3 \Delta \mathbf{c}_{i,j}$ — произведение по третьей моде (2.2). Единичная матрица выде-

ляет тождественный переход: если статические параметры описывают одну и ту же систему, $\Delta \mathbf{c}_{i,j} = 0$, преобразование пространства состояний не требуется и $\mathbf{J}_{i \rightarrow j} = \mathbf{I}$. С учётом параметризации (5.5) проекция системы $\mathbf{E}_j$ выражается через проекцию системы $\mathbf{E}_i$:

$$\mathbf{s}_{j,t} = \mathbf{W}_j \mathbf{x}_{j,t} = \mathbf{J}_{i \rightarrow j} \mathbf{W}_i \mathbf{x}_{j,t} = (\mathbf{I} + \mathbf{J} \times_3 \Delta \mathbf{c}_{i,j}) \mathbf{W}_i \mathbf{x}_{j,t}. \quad (5.6)$$

Модели (5.3) и (5.6) связаны. В первой матрица проекции $\mathbf{B}(\mathbf{c})$ аффинна по $\mathbf{c}$, и переход от $\mathbf{c}_i$ к $\mathbf{c}_j$ задаётся общим тензором. Во второй переход записан относительно проекции $\mathbf{W}_i$ опорной системы, а тензор $\mathbf{J}$ удерживает линейную по $\Delta \mathbf{c}_{i,j}$ поправку. Вторая модель не требует общего тензора для всего класса и опирается на ближайшую по параметрам систему.

Если для системы $\mathbf{E}_j$ матрица $\mathbf{W}_j$ неизвестна, а оценить её нельзя из-за короткого ряда, то проекция получается преобразованием известной матрицы $\mathbf{W}_i$ по вектору $\Delta \mathbf{c}_{i,j}$. Задача совместного обучения проекций и параметризованной модели матрицы Якоби формулируется как задача оптимизации

$$\min_{\mathbf{W}, \mathbf{J}} \sum_{i=1}^{N_E} \sum_{j=1}^{N_E} \|\mathbf{X}_i - \mathbf{W}_i^\dagger \mathbf{J}_{j \rightarrow i} \mathbf{J}_{i \rightarrow j} \mathbf{W}_i \mathbf{X}_i\|^2,$$

где $\mathbf{W}_i^\dagger$ — псевдообратная матрица для $\mathbf{W}_i$. Критерий требует, чтобы переход в пространство системы $\mathbf{E}_j$ и обратно восстанавливал траекторную матрицу $\mathbf{X}_i$.

Для новой системы формируется алгоритм восстановления пространства состояний, состоящий из двух этапов — обучения и восстановления.

Этап обучения. Для каждой системы $\mathbf{E}_i$ из набора $\{\mathbf{E}_i\}_{i=1}^{N_E}$ по ряду $\{x_{i,t}\}_{t=1}^{T_i}$ строится траекторная матрица $\mathbf{X}_i$ и решается задача построения проекции (1.6) с матрицей $\mathbf{W}_i$ в подпространство размерности p . В подпространствах состояний $\{\mathbf{S}_i\}_{i=1}^{N_E}$ для всех пар (i, j) вычисляются матрицы Якоби, минимизирующие расхождение преобразованных пространств: $\mathbf{J}_{i \rightarrow j} = \arg \min_{\mathbf{J}} \|\mathbf{S}_j - \mathbf{J} \mathbf{S}_i\|^2$. Параметры тензорной модели $\mathbf{J}$ находятся из задачи

$$\min_{\mathbf{J}} \sum_{i=1}^{N_E} \sum_{j=1}^{N_E} \|\mathbf{S}_j - (\mathbf{I} + \mathbf{J} \times_3 \Delta \mathbf{c}_{i,j}) \mathbf{S}_i\|^2.$$

Этап восстановления. Для системы $\mathbf{E}_{N_E+1}$ с вектором статических параметров $\mathbf{c}_{N_E+1}$ и коротким рядом $\{x_{N_E+1,t}\}_{t=1}^{T_{\text{short}}}$ выбирается ближайшая по параметрам система $\mathbf{E}_i$. По (5.5) оценивается матрица $\mathbf{J}_{i \rightarrow N_E+1}$. Вместе с матрицей $\mathbf{W}_i$ она задаёт пространство состояний короткого ряда:

$$\mathbf{S}_{N_E+1} = \mathbf{J}_{i \rightarrow N_E+1} \mathbf{W}_i \mathbf{X}_{N_E+1}, \quad (5.7)$$

где $\mathbf{X}_{N_E+1}$ — траекторная матрица (1.4) нового ряда.

Процедура строит точки пространства состояний новой системы по статическим параметрам, короткому ряду и проекции ближайшей системы.

5.3. Анализ полилинейной параметризации семейства проекций

Вычислительный эксперимент оценивает качество пространства состояний, восстановленного по короткому ряду новой системы. Использованы записи акселерометра одной модели мобильного телефона [72]: 24 испытуемых, частота 50 Гц, телефон в правом кармане на уровне пояса, ходьба по прямым маршрутам с одной частотой шага в течение 5 минут. Вектор статических параметров содержит рост, вес, пол и возраст. Описание измерений, маршрутов и типов активности приведено в [72]. Ряды делятся на две группы: $N_E = 10$ испытуемых используются для обучения, остальные $N_{\text{Test}} = 14$ — для оценки точности. Это разделение общее для обоих экспериментов главы.

Сравниваются четыре модели. Полилинейная модель (5.3) оценивает единый тензор $\mathbf{B}$ по рядам обучающей выборки; прямое и обратное отображения тестового ряда задаются обученным тензором и вектором статических параметров по (5.4). Модель ближайшей системы строит отдельную проекцию (1.6) для каждого обучающего ряда и отображает тестовый ряд проекцией ближайшей по статическим параметрам системы. Единая линейная модель выравнивает ряды по началу ходьбы, добавляет к траекторной матрице вектор статических параметров как экзогенные переменные и строит одно общее подпространство для всех рядов. Модель на полных данных строит проекцию по полному ряду

тестового испытуемого и служит верхней оценкой качества. Для модели ближайшей системы и полилинейной модели $\tau = 1$, $n = 35$; для единой линейной модели $\tau = 1$, $n = 55$.

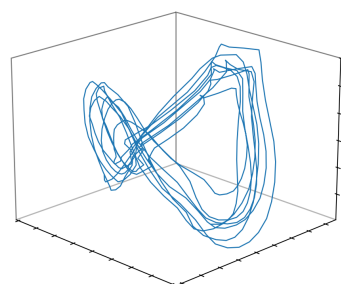
Качество измеряется коэффициентом детерминации R^2 восстановления ряда обратным отображением из пространства состояний. Сравнение приведено для размерностей пространства состояний $p = 4, \dots, 9$ в табл. 5.1.

Таблица 5.1: Коэффициент детерминации R^2 восстановления временных рядов при полилинейной параметризации проекций

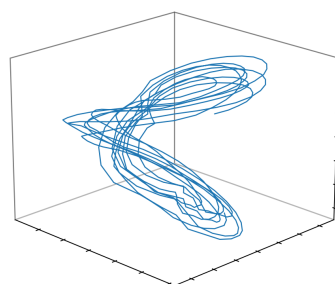
Размерность p	Полные данные	Ближайшая система	Единая линейная модель	Полилинейное отображение
4	0,43	0,37	0,16	0,39
5	0,49	0,42	0,26	0,42
6	0,53	0,45	0,36	0,47
7	0,58	0,48	0,39	0,51
8	0,61	0,51	0,42	0,54
9	0,64	0,54	0,46	0,57

По табл. 5.1 полилинейное отображение превосходит единую линейную модель и модель ближайшей системы при всех p и остаётся близким к модели, обученной на полных данных. Полилинейная модель также имеет меньше параметров, чем набор проекций для поиска ближайшей системы. При $p = 5$, $n = 35$, $\tau = 1$ набор проекций всех обучающих систем содержит 1750 параметров, полилинейная модель — 700, при большем R^2 .

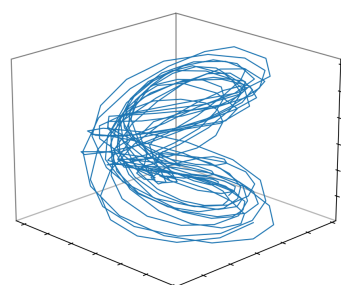
На рис. 5.3 показаны восстановленные фазовые траектории одного тестового ряда. Траектория полилинейного отображения, рис. 5.3,б, ближе к траектории модели на полных данных, рис. 5.3,а. Модель ближайшей системы, рис. 5.3,в, и единая линейная модель, рис. 5.3,г, искажают геометрию траектории. Полилинейное отображение учитывает связь задержек и статических параметров и восстанавливает пространство состояний новых систем по короткому ряду.



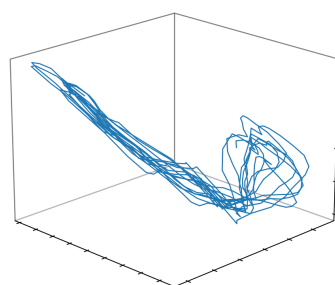
(a)



(б)



(в)



(г)

Рис. 5.3: Фазовые траектории тестового ряда при полилинейной параметризации проекций: (а) модель на полных данных, (б) полилинейное отображение, (в) модель ближайшей системы, (г) единая линейная модель

5.4. Анализ перехода между пространствами состояний матрицей Якоби

Второй эксперимент оценивает переход между пространствами состояний матрицей Якоби на том же наборе рядов акселерометра и при том же разделении: $N_E = 10$ рядов используются для построения модели матрицы Якоби, $N_{\text{Test}} = 14$ — для оценки точности. Параметры построения пространства состояний: вектор задержек с шагом $\tau = 1$ и длиной $n = 35$, размерность подпространства $p = 6$.

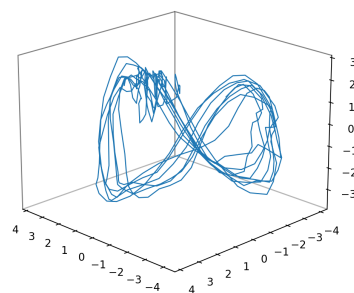
Сравниваются те же альтернативы, что и в предыдущем разделе: модель ближайшей по статическим параметрам системы, единая линейная модель с экзогенными переменными, обученная на всех рядах, и модель на полных данных. Предложенный переход строит пространство состояний тестового ряда по (5.7): проекция ближайшей системы преобразуется матрицей $\mathbf{J}_{i \rightarrow N_E+1}$, оценённой по приращению статических параметров. Качество измеряется коэффициентом детерминации R^2 восстановления ряда обратным отображением. Сравнение приведено для размерностей $p = 3, \dots, 7$ в табл. 5.2.

Таблица 5.2: Коэффициент детерминации R^2 восстановления временных рядов при переходе матрицей Якоби

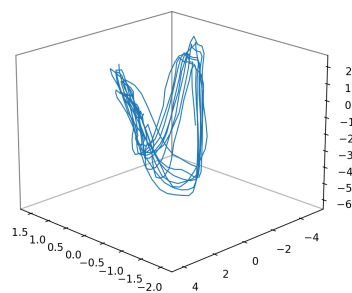
Размерность p	Полные данные	Ближайшая система	Единая линейная модель	Матрица Якоби
3	0,392	0,307	0,281	0,338
4	0,466	0,318	0,354	0,407
5	0,518	0,351	0,428	0,450
6	0,565	0,379	0,467	0,488
7	0,608	0,409	0,505	0,521

Коэффициент детерминации перехода матрицей Якоби на тестовой выборке выше, чем у модели ближайшей системы и единой линейной модели, при всех размерностях $p = 3, \dots, 7$ и ниже, чем у модели на полных данных.

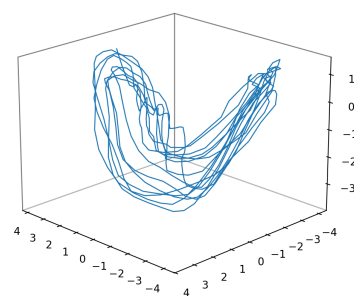
Для иллюстрации изменения пространств состояний одного тестового ряда



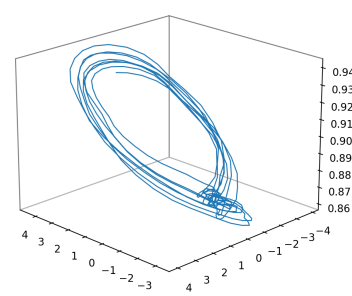
(а)



(б)



(в)



(г)

Рис. 5.4: Фазовые траектории тестового ряда при переходе матрицей Якоби: (а) модель на полных данных, (б) переход матрицей Якоби, (в) модель ближайшей системы, (г) единая линейная модель

на рис. 5.4 показаны фазовые траектории в пространстве размерности $p = 3$. Траектория, построенная на полных данных, рис. 5.4,а, сравнивается с траекторией перехода матрицей Якоби, рис. 5.4,б, и с траекторией ближайшей по статическим параметрам системы, рис. 5.4,в. Переход матрицей Якоби даёт траекторию, повёрнутую на меньший угол относительно траектории по полным данным, чем модель ближайшей системы. Единая линейная модель теряет геометрические свойства и порождает самопересечение, рис. 5.4,г; вне этой области траектория более гладкая.

Переход матрицей Якоби точнее сохраняет геометрию исходной динамической системы: модель ближайшей системы поворачивает траекторию, единая линейная модель искажает топологию и вызывает самопересечения.

В главе предложены два способа восстановления пространства состояний системы класса по короткому ряду и статическим параметрам. Полилинейное отображение задаёт проекцию всех систем класса одним тензором, свёрнутым с вектором статических параметров; матрица проекции липшицева по параметрам. Переход матрицей Якоби преобразует проекцию ближайшей системы линейной по приращению параметров поправкой. На рядах акселерометра новых испытуемых оба способа дают более высокий R^2 , чем модель ближайшей системы и единая линейная модель, и приближаются к модели, построенной по полным рядам.

Заключение

Основные результаты диссертационной работы заключаются в следующем.

В главе 1 поставлена задача восстановления пространства состояний по наблюдаемым временным рядам. Введены динамическая система, вектор задержек и теорема Такенса. Рассмотрены линейные методы — метод задержек и метод главных компонент — и нелинейные методы: оператор Купмана, нейросетевые модели и разреженная идентификация уравнений. Выделены ограничения этих методов: выбор параметров вложения, избыточность многомерного ряда и отсутствие явных уравнений.

В главе 2 введено полилинейное отображение векторов задержек. Лемма о полилинейном расширении показывает, что оно задаёт гладкое вложение исходного многообразия тогда и только тогда, когда вложением является отображение многомерного метода задержек. Теорема о многомерном полилинейном вложении устанавливает, что каноническое разложение сохраняет это вложение на непустом открытом множестве фактор-матриц, если ранг и размерность восстановленного пространства не меньше $2d_s + 1$; для погружения необходимы ранг и размерность не меньше d_s . Доказано, что класс полилинейных моделей векторов задержек, дополненных константой, содержит линейные векторные авторегрессии. Для снижения числа параметров введены каноническое разложение, разложение Таккера и тензорный поезд. На аттракторе Лоренца и на рядах акселерометра тензорное вложение точнее прогноза контрольной компоненты, чем одномерное и многомерное вложения по главным компонентам.

В главе 3 введено полилинейное объединение нейросетевых представлений нескольких рядов. Каноническое разложение задаёт тензор связей. Полилинейное объединение сохраняет произведения компонент разных рядов, которые линейное объединение не удерживает. На рядах ходьбы этот способ даёт более высокий R^2 , чем вложение глобальных компонент, объединение исходных векторов задержек и объединение скрытых представлений.

В главе 4 введён метод T-SINDy. Временные задержки задают координаты, тензорное отображение параметризует взаимодействия функций-кандидатов, разреживание фактор-матриц отбирает слагаемые. Порог действует на фактор-матрицы канонического разложения, а не на отдельные коэффициенты. Уравнения записываются в исходных переменных, число параметров растёт линейно по длине вложения. На аттракторе Лоренца и на нормальной форме бифуркации Хопфа уравнения восстанавливаются по неполному наблюдению. При длине окна $n = 15$ на системе Хопфа время обучения меньше, чем у метода SINDy, при сопоставимой ошибке прогноза.

В главе 5 пространство состояний системы класса восстановлено по короткому ряду и статическим параметрам. Семейство проекций параметризовано полилинейным отображением вектора задержек и вектора статических параметров; переход между пространствами состояний двух систем задан матрицей Якоби, линейной по приращению параметров. Доказана липшицевость параметризованной проекции по статическим параметрам. На рядах акселерометра новых испытуемых коэффициент детерминации выше, чем у модели ближайшей системы и у единой линейной модели, и приближается к модели, построенной по полным рядам.

Список основных обозначений

$\mathbb{R}, \mathbb{N}, \mathbb{Z}$ — множества действительных, натуральных и целых чисел

$\mathcal{T}$ — интервал времени фазовой траектории

$\mathcal{M}$ — гладкое многообразие размерности d_s , фазовое пространство истинной системы

d_s — размерность многообразия состояний

$\mathbf{v}(t) \in \mathbb{R}^{d_s}$ — состояние динамической системы в момент t

$\mathbf{f} : \mathbb{R}^{d_s} \rightarrow \mathbb{R}^{d_s}$ — векторное поле, правая часть системы ОДУ

$\mathbf{c} \in \mathbb{R}^{d_c}$, $\mathbf{c}_i$ — вектор статических параметров системы; d_c — его размерность

$\mathcal{C} \subset \mathbb{R}^{d_c}$ — множество допустимых статических параметров

$\mathbf{h} : \mathcal{M} \rightarrow \mathbb{R}^{d_o}$ — функция наблюдения; при $d_o = 1$ обозначается h

d_o — число наблюдаемых рядов ($d_o = 1$ в одномерном случае, $d_o \geq 2$ в многомерном)

$\mathbf{F}, \mathbf{F}_{\Delta t}$ — дискретный оператор эволюции, $\mathbf{v}_{k+1} = \mathbf{F}_{\Delta t}(\mathbf{v}_k)$

Δt — шаг дискретизации

τ — лаг вектора задержек

n — размерность вложения, длина вектора задержек; n_i — длина вложения i -го ряда

N — длина временного ряда

T — число отсчётов в сумме критерия

T_i — правый конец интервала наблюдения i -й системы

$\mathbf{x}_k \in \mathbb{R}^n$ — вектор задержек $[x_{k-(n-1)\tau}, \dots, x_k]^\top$

$\mathbf{X} \in \mathbb{R}^{n \times m}$ — траекторная матрица; $m = N - n + 1$ — число векторов задержек

$\Phi_{h,F,\tau}$ — отображение Такенса, функция лаговых координат

$\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}$ — множители сингулярного разложения траекторной матрицы

$\mathbf{W} \in \mathbb{R}^{p \times n}$ — матрица линейной проекции (РСА)

p — размерность редуцированного фазового пространства, $p \ll n$

p_{ar} — порядок векторной авторегрессии

$\mathbf{s}_t \in \mathbb{R}^p$ — точка редуцированного фазового пространства

$\mathbf{S}$ — матрица, столбцы которой суть $\mathbf{s}_t$

$\mathbf{A}$, $\mathbf{B}$, $\mathbf{W}$, $\mathbf{J}$ — тензоры параметров полилинейных моделей; в T-SINDy $\mathbf{W}$ — тензор коэффициентов полилинейной регрессии

$\mathbf{A}_{(1)}$, $\mathbf{B}(\mathbf{c})$ — развёртка тензора в матрицу и матрица отображения после свёртки с $\mathbf{c}$

$\times_n$ — произведение по n -й моде

$\otimes$ — внешнее произведение

$\odot$ — поэлементное произведение Адамара

R — ранг CP-разложения

$\mathbf{J} \in \mathbb{R}^{p \times p}$ — матрица Якоби перехода между фазовыми подпространствами

$\mathbf{P}$ — гладкое отображение между фазовыми пространствами ансамбля систем

$\mathbf{E}_i$ — неизвестная функция эволюции i -й системы ансамбля; N_E — число систем в обучении

$\mathbf{w} \in \mathbb{R}^{N_w}$ — вектор настраиваемых параметров модели

$\Theta(\mathbf{x})$ — библиотека функций-кандидатов метода SINDy

θ_j , $j = 1, \dots, n_\theta$ — функция-кандидат; n_θ — число функций в библиотеке

Ξ — матрица коэффициентов разреженной модели SINDy

n_{tot} — суммарная длина объединённого вектора задержек

n_k — длина вектора задержек k -й модальности в T-SINDy; размер этой моды тензора равен $n_k + 1$

d_{lat} — размерность латентного пространства автоэнкодера

$\mathbf{z}(t)$, $\mathbf{z}_t$ — объединённый вектор задержек, при необходимости дополненный константой

$\mathbf{h}_t$ — скрытое состояние рекуррентной сети; $\mathbf{h}_{x,t}$ — нейросетевое представление ряда x

$\mathcal{K}$ — оператор Купмана; $\mathbf{K}$ — его конечномерная матрица

$\|\cdot\|$ — евклидова норма; $\|\cdot\|_F$ — норма Фробениуса

Список иллюстраций

1.1	Упрощённая схема получения фазового пространства в одномерном случае. Показаны три подхода: одномерный метод временных задержек, прямой выбор n и Глобальное вложение по главным значениям.	17
1.2	Упрощённая схема получения фазового пространства в многомерном случае. Показаны четыре подхода: метод временных задержек, прямой выбор n , главные глобальные компоненты или реконструкция с использованием одного выбранного временного ряда. Примечание: на этой схеме матрица траектории не транспонирована, в отличие от рисунка 1.1.	18
1.3	Пример восстановленного пространства состояний из различных вложений: исходные координаты (слева), метод задержек с временным рядом x_t (в центре), метод задержек с временным рядом y_t (справа). Чтобы подчеркнуть линейную зависимость между результирующими многообразиями, все рисунки показаны в одинаковых масштабах по осям.	20
2.1	Полилинейное расширение и отображение метода задержек. Проекция π отбрасывает произведения и возвращает задержки. . . .	26
2.2	Схема полилинейного метода снижения размерности для трехмерного случая.	28
2.3	Полилинейная динамическая система в графической нотации. . .	30
2.4	Схемы полилинейного отображения в нотации Пенроуза	31
2.5	Восстановление пространства состояний по двум наблюдениям акселерометра. Цветом отмечены близкие участки исходной траектории	32

2.6	Графическое представление тензорной динамической системы. Граф показывает, как результирующий тензор неявно включает в себя отображения временных рядов друг на друга. Математическая форма показана в уравнении (2.6).	34
2.7	Каноническая проекция. Линейные формы окон перемножаются, ранг равен числу координат произведения.	38
2.8	Схема обучения автоэнкодера. Левая часть (кодировщик) представляет собой полилинейное отображение в подпространство состояний, правая часть (декодировщик) — линейное отображение обратно в исходное фазовое пространство. Математическая форма показана в уравнении (2.8).	41
2.9	Фазовые портреты аттрактора Лоренца: исходная система, полилинейная модель и автоэнкодер Купмана	44
2.10	Схема эксперимента с акселерометром: ряды x_t и y_t входят в реконструкцию, ряд z_t остаётся контрольным	45
3.1	Схема полносвязной нейронной сети с объединением исходных векторов задержек	50
3.2	Схема полносвязной нейронной сети с объединением преобразованных фазовых пространств каждого временного ряда, где σ — функция активации	51
3.3	Схема нейронной сети с полилинейным объединением фазовых пространств каждого временного ряда	52
3.4	Схема нейронной сети с полилинейным объединением фазовых пространств каждого временного ряда с каноническим разложением	52
3.5	Схема вычислительного эксперимента с оценкой качества на восстановленном временном ряду.	54
4.1	Схема метода SINDy	59
4.2	Схема метода SINDy для наблюдаемых временных рядов	60

4.3	Схема автоэнкодера для наблюдаемых временных рядов	61
4.4	Отбор слагаемых. В (а) обнуляются отдельные коэффициенты, в (б) нуль в факторе удаляет целый столбец.	65
4.5	Схема тензорного метода SINDy для наблюдаемых временных рядов	65
4.6	Время обучения и среднеквадратичная ошибка прогноза в зависимости от длины окна	67
4.7	Восстановленное уравнение и прогноз временного ряда для нормальной формы Хопфа	68
5.1	Переход к системе с коротким рядом. Приращение статических параметров переносится матрицей Якоби.	72
5.2	Схема полилинейного отображения вектора задержек и вектора статических параметров. Входные данные взаимодействуют через тензор $\mathbf{V}$, формируя вектор состояния в пространстве низкой размерности.	73
5.3	Фазовые траектории тестового ряда при полилинейной параметризации проекций: (а) модель на полных данных, (б) полилинейное отображение, (в) модель ближайшей системы, (г) единая линейная модель	80
5.4	Фазовые траектории тестового ряда при переходе матрицей Якоби: (а) модель на полных данных, (б) переход матрицей Якоби, (в) модель ближайшей системы, (г) единая линейная модель . . .	82

Список таблиц

2.1	Число параметров полилинейного отображения при разных раз- ложениях	30
2.2	Прогноз $x_{t+\Delta t}$ системы Лоренца; ряд x_t входит в реконструкцию .	43
2.3	Прогноз контрольного ряда $z_{t+\Delta t}$ системы Лоренца	43
2.4	Реконструкция аттрактора Лоренца на горизонте одного времени Ляпунова	45
2.5	Прогноз $x_{t+\Delta t}$ и $y_{t+\Delta t}$ при ходьбе; ось x входит в реконструкцию .	46
2.6	Прогноз контрольного ряда $z_{t+\Delta t}$ при ходьбе	46
2.7	Ходьба, ряды использованы при реконструкции	47
2.8	Ходьба, контрольный ряд	47
2.9	Прогноз $x_{t+\Delta t}$ и $y_{t+\Delta t}$ при спуске по лестнице	48
2.10	Прогноз контрольного ряда $z_{t+\Delta t}$ при спуске по лестнице	48
3.1	Восстановление ряда при ходьбе	54
3.2	Восстановление ряда при спуске по лестнице	54
4.1	Сравнение подходов к восстановлению уравнений динамической системы	58
4.2	Точность и время обучения для системы Хопфа, средние значения при $n = 15$	68
5.1	Коэффициент детерминации R^2 восстановления временных рядов при полилинейной параметризации проекций	79
5.2	Коэффициент детерминации R^2 восстановления временных рядов при переходе матрицей Якоби	81

Список литературы

1. *Pearson Karl*. On Lines and Planes of Closest Fit to Systems of Points in Space // *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. — 1901. — Vol. 2, no. 11. — Pp. 559–572.
2. *Hotelling Harold*. Analysis of a Complex of Statistical Variables into Principal Components // *Journal of Educational Psychology*. — 1933. — Vol. 24, no. 6. — Pp. 417–441.
3. *Jolliffe Ian T., Cadima Jorge*. Principal Component Analysis: A Review and Recent Developments // *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. — 2016. — Vol. 374, no. 2065. — P. 20150202.
4. *Benner Peter, Gugercin Serkan, Willcox Karen*. A Survey of Projection-Based Model Reduction Methods for Parametric Dynamical Systems // *SIAM Review*. — 2015. — Vol. 57, no. 4. — Pp. 483–531.
5. *Eckart Carl, Young Gale*. The Approximation of One Matrix by Another of Lower Rank // *Psychometrika*. — 1936. — Vol. 1, no. 3. — Pp. 211–218.
6. *Sirovich Lawrence*. Turbulence and the Dynamics of Coherent Structures. Part I: Coherent Structures // *Quarterly of Applied Mathematics*. — 1987. — Vol. 45, no. 3. — Pp. 561–571.
7. *Berkooz Gal, Holmes Philip, Lumley John L*. The Proper Orthogonal Decomposition in the Analysis of Turbulent Flows // *Annual Review of Fluid Mechanics*. — 1993. — Vol. 25. — Pp. 539–575.
8. Turbulence, Coherent Structures, Dynamical Systems and Symmetry / Philip Holmes, John L. Lumley, Gahl Berkooz, Clarence W. Rowley. — 2 edition. — Cambridge: Cambridge University Press, 2012.
9. *Moore Bruce C*. Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction // *IEEE Transactions on Automatic Control*. — 1981. — Vol. 26, no. 1. — Pp. 17–32.

10. *Juang Jer-Nan, Pappa Richard S.* An Eigensystem Realization Algorithm for Modal Parameter Identification and Model Reduction // *Journal of Guidance, Control, and Dynamics*. — 1985. — Vol. 8, no. 5. — Pp. 620–627.
11. *Antoulas Athanasios C., Beattie Christopher A., Gugercin Serkan.* Interpolatory Methods for Model Reduction. — Philadelphia: Society for Industrial and Applied Mathematics, 2020.
12. *Bellman Richard.* Adaptive Control Processes: A Guided Tour. — Princeton: Princeton University Press, 1961.
13. *Peherstorfer Benjamin, Willcox Karen.* Data-Driven Operator Inference for Nonintrusive Projection-Based Model Reduction // *Computer Methods in Applied Mechanics and Engineering*. — 2016. — Vol. 306. — Pp. 196–215.
14. *Whitney Hassler.* Differentiable Manifolds // *Annals of Mathematics*. — 1936. — Vol. 37, no. 3. — Pp. 645–680.
15. Geometry from a Time Series / Norman H. Packard, James P. Crutchfield, J. Doyne Farmer, Robert S. Shaw // *Physical Review Letters*. — 1980. — Vol. 45, no. 9. — Pp. 712–716.
16. *Takens Floris.* Detecting Strange Attractors in Turbulence // *Dynamical Systems and Turbulence, Warwick 1980* / Ed. by David A. Rand, Lai-Sang Young. — Springer, 1981. — Vol. 898 of *Lecture Notes in Mathematics*. — Pp. 366–381.
17. *Sauer Timothy D., Yorke James A., Casdagli Martin.* Embedology // *Journal of Statistical Physics*. — 1991. — Vol. 65, no. 3–4. — Pp. 579–616.
18. *Deyle Ethan R., Sugihara George.* Generalized Theorems for Nonlinear State Space Reconstruction // *PLoS ONE*. — 2011. — Vol. 6, no. 3. — P. e18295.
19. *Fraser Andrew M., Swinney Harry L.* Independent Coordinates for Strange Attractors from Mutual Information // *Physical Review A*. — 1986. — Vol. 33, no. 2. — Pp. 1134–1140.

20. *Kennel Matthew B., Brown Reggie, Abarbanel Henry D. I.* Determining Embedding Dimension for Phase-Space Reconstruction Using a Geometrical Construction // *Physical Review A*. — 1992. — Vol. 45, no. 6. — Pp. 3403–3411.
21. *Cao Liangyue.* Practical Method for Determining the Minimum Embedding Dimension of a Scalar Time Series // *Physica D: Nonlinear Phenomena*. — 1997. — Vol. 110, no. 1–2. — Pp. 43–50.
22. *Abarbanel Henry D. I.* Analysis of Observed Chaotic Data. — New York: Springer, 1996.
23. Stabilizing Embedology: Geometry-Preserving Delay-Coordinate Maps / Armin Eftekhari, Han Lun Yap, Michael B. Wakin, Christopher J. Rozell // *Physical Review E*. — 2018. — Vol. 97. — P. 022222.
24. *Broomhead David S., King Gregory P.* Extracting Qualitative Dynamics from Experimental Data // *Physica D: Nonlinear Phenomena*. — 1986. — Vol. 20, no. 2–3. — Pp. 217–236.
25. *Vlachos Ioannis, Kugiumtzis Dimitris.* Nonuniform State-Space Reconstruction and Coupling Detection // *Physical Review E*. — 2010. — Vol. 82. — P. 016207.
26. *Sugihara George, May Robert M.* Nonlinear Forecasting as a Way of Distinguishing Chaos from Measurement Error in Time Series // *Nature*. — 1990. — Vol. 344. — Pp. 734–741.
27. *Tibshirani Robert.* Regression Shrinkage and Selection via the Lasso // *Journal of the Royal Statistical Society: Series B*. — 1996. — Vol. 58, no. 1. — Pp. 267–288.
28. *Koopman Bernard O.* Hamiltonian Systems and Transformation in Hilbert Space // *Proceedings of the National Academy of Sciences*. — 1931. — Vol. 17, no. 5. — Pp. 315–318.
29. *Mezić Igor.* Spectral Properties of Dynamical Systems, Model Reduction and Decompositions // *Nonlinear Dynamics*. — 2005. — Vol. 41. — Pp. 309–325.

30. *Mezić Igor*. Analysis of Fluid Flows via Spectral Properties of the Koopman Operator // *Annual Review of Fluid Mechanics*. — 2013. — Vol. 45. — Pp. 357–378.
31. *Schmid Peter J*. Dynamic Mode Decomposition of Numerical and Experimental Data // *Journal of Fluid Mechanics*. — 2010. — Vol. 656. — Pp. 5–28.
32. On Dynamic Mode Decomposition: Theory and Applications / Jonathan H. Tu, Clarence W. Rowley, Dirk M. Luchtenburg et al. // *Journal of Computational Dynamics*. — 2014. — Vol. 1, no. 2. — Pp. 391–421.
33. *Schmid Peter J*. Dynamic Mode Decomposition and Its Variants // *Annual Review of Fluid Mechanics*. — 2022. — Vol. 54, no. 1. — Pp. 225–254.
34. *Williams Matthew O., Kevrekidis Ioannis G., Rowley Clarence W*. A Data-Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition // *Journal of Nonlinear Science*. — 2015. — Vol. 25, no. 6. — Pp. 1307–1346.
35. *Kaiser Eurika, Kutz J. Nathan, Brunton Steven L*. Sparse Identification of Nonlinear Dynamics for Model Predictive Control in the Low-Data Limit // *Proceedings of the Royal Society A*. — 2018. — Vol. 474. — P. 20180335.
36. *Lusch Bethany, Kutz J. Nathan, Brunton Steven L*. Deep Learning for Universal Linear Embeddings of Nonlinear Dynamics // *Nature Communications*. — 2018. — Vol. 9. — P. 4950.
37. *Wehmeyer Christoph, Noé Frank*. Time-Lagged Autoencoders: Deep Learning of Slow Collective Variables for Molecular Kinetics // *The Journal of Chemical Physics*. — 2018. — Vol. 148, no. 24. — P. 241703.
38. Neural Ordinary Differential Equations / Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, David Duvenaud // *Advances in Neural Information Processing Systems*. — Vol. 31. — 2018. — Pp. 6572–6583.
39. *Raissi Maziar, Perdikaris Paris, Karniadakis George E*. Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse

- Problems Involving Nonlinear Partial Differential Equations // *Journal of Computational Physics*. — 2019. — Vol. 378. — Pp. 686–707.
40. *Greydanus Samuel, Dzamba Misko, Yosinski Jason*. Hamiltonian Neural Networks // *Advances in Neural Information Processing Systems*. — Vol. 32. — 2019.
 41. *Rubanova Yulia, Chen Ricky T. Q., Duvenaud David*. Latent Ordinary Differential Equations for Irregularly-Sampled Time Series // *Advances in Neural Information Processing Systems*. — Vol. 32. — 2019.
 42. *Brunton Steven L., Proctor Joshua L., Kutz J. Nathan*. Discovering Governing Equations from Data by Sparse Identification of Nonlinear Dynamical Systems // *Proceedings of the National Academy of Sciences*. — 2016. — Vol. 113, no. 15. — Pp. 3932–3937.
 43. Data-Driven Discovery of Partial Differential Equations / Samuel H. Rudy, Steven L. Brunton, Joshua L. Proctor, J. Nathan Kutz // *Science Advances*. — 2017. — Vol. 3, no. 4. — P. e1602614.
 44. Model Selection for Dynamical Systems via Sparse Regression and Information Criteria / Niall M. Mangan, J. Nathan Kutz, Steven L. Brunton, Joshua L. Proctor // *Proceedings of the Royal Society A*. — 2017. — Vol. 473. — P. 20170009.
 45. Data-Driven Discovery of Coordinates and Governing Equations / Kathleen Champion, Bethany Lusch, J. Nathan Kutz, Steven L. Brunton // *Proceedings of the National Academy of Sciences*. — 2019. — Vol. 116, no. 45. — Pp. 22445–22451.
 46. *Carroll J. Douglas, Chang Jih-Jie*. Analysis of Individual Differences in Multidimensional Scaling via an N-Way Generalization of Eckart–Young Decomposition // *Psychometrika*. — 1970. — Vol. 35, no. 3. — Pp. 283–319.
 47. *Tucker Ledyard R.* Some Mathematical Notes on Three-Mode Factor Analysis // *Psychometrika*. — 1966. — Vol. 31, no. 3. — Pp. 279–311.

48. *Kruskal Joseph B.* Three-Way Arrays: Rank and Uniqueness of Trilinear Decompositions, with Application to Arithmetic Complexity and Statistics // *Linear Algebra and its Applications*. — 1977. — Vol. 18, no. 2. — Pp. 95–138.
49. *Kolda Tamara G., Bader Brett W.* Tensor Decompositions and Applications // *SIAM Review*. — 2009. — Vol. 51, no. 3. — Pp. 455–500.
50. *Oseledets Ivan V.* Tensor-Train Decomposition // *SIAM Journal on Scientific Computing*. — 2011. — Vol. 33, no. 5. — Pp. 2295–2317.
51. *Tikhonov D. M., Strijov V. V.* Dynamic system reconstruction from multivariate time series via multilinear map and time delay embedding // International Conference on Computational Optimization (ICOMP). — Innopolis: 2024. <https://openreview.net/forum?id=xYkLT5f600>.
52. *Тихонов Д. М.* Восстановление зашумленных значений многомерных временных рядов с помощью тензорного метода задержек и тензорного разложения // 66-я Всероссийская научная конференция Московского физико-технического института: тезисы докладов. Прикладная математика и информатика. — Москва: Физматкнига, 2024. — С. 128–130.
53. *Тихонов Д. М., Стрижов В. В.* Объединение пространств состояний динамической системы через полилинейное отображение // *Труды Института системного программирования РАН*. — 2026. — Т. 38, № 6, ч. 1. — С. 63–72. — DOI: 10.15514/ISPRAS-2026-38(6)-4. https://www.ispras.ru/proceedings/isp_38_2026_6-1/isp_38_2026_6-1_63/.
54. *Тихонов Д. М., Стрижов В. В.* Тензорное разреженное восстановление дифференциальных уравнений нелинейных динамических систем // *Моделирование и анализ информационных систем*. — 2026. — Т. 33, № 3. — С. 344–357. — DOI: 10.18255/1818-1015-2026-3-344-357. <https://www.mais-journal.ru/jour/article/view/2113>.
55. *Тихонов Д. М.* Полилинейное отображение для мультимоделирования динамических систем // *Труды МФТИ*. — 2026.

- Т. 18, № 2 (70). — С. 88–97. <https://sciup.org/polilinejnoe-otobrazhenie-dlja-multimodelirovaniya-dinamicheskikh-sist>
56. *Тихонов Д. М.* Классификация квазипериодических временных сигналов с помощью сферических гармоник // Материалы Международного молодежного научного форума «ЛОМОНОСОВ-2022». — Москва: 2022. https://conf.msu.ru/archive/Lomonosov_2022/data/25480/uid684519_report.pdf.
 57. *Hirsch Morris W., Smale Stephen, Devaney Robert L.* Differential Equations, Dynamical Systems, and an Introduction to Chaos. — 3 edition. — Amsterdam: Academic Press, 2013.
 58. *Strogatz Steven H.* Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering. — 2 edition. — Boulder: Westview Press, 2015.
 59. *Khalil Hassan K.* Nonlinear Systems. — 3 edition. — Upper Saddle River: Prentice Hall, 2002.
 60. *Kuznetsov Yuri A.* Elements of Applied Bifurcation Theory. — 3 edition. — New York: Springer, 2004. — Vol. 112 of *Applied Mathematical Sciences*.
 61. *Lorenz Edward N.* Deterministic Nonperiodic Flow // *Journal of the Atmospheric Sciences*. — 1963. — Vol. 20, no. 2. — Pp. 130–141.
 62. *Grassberger Peter, Procaccia Itamar.* Characterization of Strange Attractors // *Physical Review Letters*. — 1983. — Vol. 50, no. 5. — Pp. 346–349.
 63. Determining Lyapunov Exponents from a Time Series / Alan Wolf, Jack B. Swift, Harry L. Swinney, John A. Vastano // *Physica D: Nonlinear Phenomena*. — 1985. — Vol. 16, no. 3. — Pp. 285–317.
 64. *Casdagli Martin.* Nonlinear Prediction of Chaotic Time Series // *Physica D: Nonlinear Phenomena*. — 1989. — Vol. 35, no. 3. — Pp. 335–356.
 65. *Golub Gene H., Loan Charles F. Van.* Matrix Computations. — 4 edition. — Baltimore: Johns Hopkins University Press, 2013.

66. *Hénon Michel*. A Two-Dimensional Mapping with a Strange Attractor // *Communications in Mathematical Physics*. — 1976. — Vol. 50, no. 1. — Pp. 69–77.
67. Deep Koopman Learning of Nonlinear Time-Varying Systems / Wenjian Hao, Bowen Huang, Wei Pan et al. // *Automatica*. — 2024. — Vol. 159. — P. 111372.
68. Deep Koopman Operators for Causal Discovery / Juan Nathaniel, Carla Roesch, Jatan Buch et al. // *Communications Physics*. — 2025. — Vol. 8. — P. 513.
69. *Lathauwer Lieven De, Moor Bart De, Vandewalle Joos*. A Multilinear Singular Value Decomposition // *SIAM Journal on Matrix Analysis and Applications*. — 2000. — Vol. 21, no. 4. — Pp. 1253–1278.
70. *Favier Gérard, Kibangou Alain*. Tensor-Based Approaches for Nonlinear and Multilinear Systems Modeling and Identification // *Algorithms*. — 2023. — Vol. 16, no. 9. — P. 443.
71. *Kantz Holger, Schreiber Thomas*. Nonlinear Time Series Analysis. — 2 edition. — Cambridge: Cambridge University Press, 2004.
72. Mobile Sensor Data Anonymization / Mohammad Malekzadeh, Richard G. Clegg, Andrea Cavallaro, Hamed Haddadi // Proceedings of the International Conference on Internet of Things Design and Implementation. — 2019. — Pp. 49–58.
73. *Guckenheimer John, Holmes Philip*. Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields. — New York: Springer, 1983. — Vol. 42 of *Applied Mathematical Sciences*.
74. *Ljung Lennart*. System Identification: Theory for the User. — 2 edition. — Upper Saddle River: Prentice Hall, 1999.
75. *Billings Stephen A*. Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains. — Chichester: Wiley, 2013.

76. *Schoukens Johan, Ljung Lennart*. Nonlinear System Identification: A User-Oriented Road Map // *IEEE Control Systems Magazine*. — 2019. — Vol. 39, no. 6. — Pp. 28–99.
77. *Kaheman Kadierdan, Kutz J. Nathan, Brunton Steven L*. SINDy-PI: A Robust Algorithm for Parallel Implicit Sparse Identification of Nonlinear Dynamics // *Proceedings of the Royal Society A*. — 2020. — Vol. 476, no. 2242. — P. 20200279.
78. Exact Recovery Guarantees for Parameterized Nonlinear System Identification Problem under Sparse Disturbances or Semi-Oblivious Attacks / Haixiang Zhang, Baturalp Yalcin, Javad Lavaei, Eduardo Sontag // *Transactions on Machine Learning Research*. — 2025.
79. Discovering Governing Equations from Partial Measurements with Deep Delay Autoencoders / Joseph Bakarji, Kathleen Champion, J. Nathan Kutz, Steven L. Brunton // *Proceedings of the Royal Society A*. — 2023. — Vol. 479. — P. 20230422.
80. *Hochreiter Sepp, Schmidhuber Jürgen*. Long Short-Term Memory // *Neural Computation*. — 1997. — Vol. 9, no. 8. — Pp. 1735–1780.
81. PySINDy: A Python Package for the Sparse Identification of Nonlinear Dynamical Systems from Data / Brian M. de Silva, Kathleen Champion, Markus Quade et al. // *Journal of Open Source Software*. — 2020. — Vol. 5, no. 49. — P. 2104.
82. PySINDy: A Comprehensive Python Package for Robust Sparse System Identification / Alan A. Kaptanoglu, Brian M. de Silva, Urban Fasel et al. // *Journal of Open Source Software*. — 2022. — Vol. 7, no. 69. — P. 3994.
83. Machine Learning for Sparse Nonlinear Modeling and Control / Steven L. Brunton, Nicholas Zolman, J. Nathan Kutz, Urban Fasel // *Annual Review of Control, Robotics, and Autonomous Systems*. — 2025. — Vol. 8. — Pp. 127–152.

84. Ensemble-SINDy: Robust Sparse Model Discovery in the Low-Data, High-Noise Limit, with Active Learning and Control / Urban Fasel, J. Nathan Kutz, Bingni W. Brunton, Steven L. Brunton // *Proceedings of the Royal Society A*. — 2022. — Vol. 478, no. 2260. — P. 20210904.
85. *Fung Lloyd, Fasel Urban, Juniper Matthew*. Rapid Bayesian Identification of Sparse Nonlinear Dynamics from Scarce and Noisy Data // *Proceedings of the Royal Society A*. — 2025. — Vol. 481, no. 2307. — P. 20240200.
86. *Basiri Mohammad Amin, Khanmohammadi Sina*. SINDyG: Sparse Identification of Nonlinear Dynamical Systems from Graph-Structured Data, with Applications to Stuart–Landau Oscillator Networks // *Journal of Complex Networks*. — 2025. — Vol. 13, no. 5. — P. cnaf029.
87. Modelling Matrix Time Series via a Tensor CP-Decomposition / Jinyuan Chang, Jing He, Lin Yang, Qiwei Yao // *Journal of the Royal Statistical Society: Series B*. — 2023. — Vol. 85, no. 1. — Pp. 127–148.