

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
МИНИСТЕРСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ ПО АТОМНОЙ ЭНЕРГИИ  
МИНИСТЕРСТВО ПРОМЫШЛЕННОСТИ, НАУКИ И ТЕХНОЛОГИЙ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
РОССИЙСКАЯ АССОЦИАЦИЯ НЕЙРОИНФОРМАТИКИ  
МОСКОВСКИЙ ИНЖЕНЕРНО-ФИЗИЧЕСКИЙ ИНСТИТУТ  
(ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ)

---

**НАУЧНАЯ СЕССИЯ МИФИ–2003**

**НЕЙРОИНФОРМАТИКА–2003**

**V ВСЕРОССИЙСКАЯ  
НАУЧНО-ТЕХНИЧЕСКАЯ  
КОНФЕРЕНЦИЯ**

**ЛЕКЦИИ  
ПО НЕЙРОИНФОРМАТИКЕ**

**Часть 1**

По материалам Школы-семинара  
«Современные проблемы нейроинформатики»

Москва 2003

УДК 004.032.26 (06)

ББК 32.818я5

М82

**НАУЧНАЯ СЕССИЯ МИФИ–2003. V ВСЕРОССИЙСКАЯ НАУЧНО-ТЕХНИЧЕСКАЯ КОНФЕРЕНЦИЯ «НЕЙРОИНФОРМАТИКА–2003»: ЛЕКЦИИ ПО НЕЙРОИНФОРМАТИКЕ. Часть 1.** – М.: МИФИ, 2003. – 188 с.

В книге публикуются тексты лекций, прочитанных на Школе-семинаре «Современные проблемы нейроинформатики», проходившей 29–31 января 2003 года в МИФИ в рамках V Всероссийской конференции «Нейроинформатика–2003».

Материалы лекций связаны с рядом проблем, актуальных для современного этапа развития нейроинформатики, включая ее взаимодействие с другими научно-техническими областями.

Ответственный редактор

*Ю. В. Тюменцев*, кандидат технических наук

ISBN 5–7262–0471–9

© *Московский инженерно-физический институт  
(государственный университет), 2003*

## Содержание

<b><i>Н. Г. Макаренко. Эмбедология и нейропрогноз</i></b>	<b>86</b>
Введение . . . . .	87
Отображения, функции и типичность . . . . .	89
Многообразия . . . . .	94
Погружения и вложения . . . . .	101
Трансверсальность . . . . .	105
Эмбедология и теорема Такенса . . . . .	107
Прогноз как аппроксимация: сети, но не нейронные сети . . . . .	118
Нейропрогноз: обучение отображению . . . . .	124
Пределы предсказуемости: хаотическая динамика и эффект Эдипа в мягких системах . . . . .	132
Литература . . . . .	137
Дополнения . . . . .	141

**Н. Г. МАКАРЕНКО**

Институт математики, Алма-Ата, Казахстан

**E-mail: makarenko@math.kz**

**ЭМБЕДОЛОГИЯ И НЕЙРОПРОГНОЗ**

**Аннотация**

Лекция представляет собой попытку понятно рассказать о методах алгоритмического моделирования — *Эмбедологии* или реконструкции модели из наблюдаемых скалярных временных рядов. Топологическое вложение ряда в евклидово пространство подходящей размерности приводит к схеме нелинейного локального или глобального прогноза. Последний сводится к задаче наилучшей аппроксимации нелинейной функции от многих переменных. Оптимальным аппроксиматором, во многих случаях, является нейронная сеть.

**N. MAKARENKO**

Institute of Mathematics, Kazakhstan, Alma-Ata

**E-mail: makarenko@math.kz**

**EMBEDOLOGY AND FORECASTING BY NEURAL NETWORKS**

**Abstract**

This lecture represents an attempt of comprehensible introduction into algorithmic modelling methods — Embedology reconstruction of a model from observed scalar time series. A topological time series embedding in Euclidian space of appropriate dimension in a scheme of nonlinear local or global forecast results. The last is reduced to a task of best approximation of nonlinear multivariable function. In such cases a neural network is often an optimal approximator.

## Введение

Подходите к вашим задачам с правильного конца и начинайте с ответов. Тогда в один прекрасный день вы, возможно, найдете правильный вопрос.

---

*Р. ван Гулик  
«Отшельник в журавлиных  
перьях»*

*Непостижимая эффективность математики в естественных науках*<sup>1 2</sup> обязана главным образом моделям, заданным дифференциальными уравнениями. Аналитика, однако, избыточна в смысле определений: она содержит случаи, которым не соответствует никакая реальность и не дает нам средств, с помощью которых мы могли бы различить *действительное* и *возможное*. Для того чтобы сделать это, приходится обращаться к опыту и те, кто возмущается такими экскурсами за пределы «фундаментальной науки» — явные лицемеры! Почему же тогда не попытаться построить или точнее реконструировать модель прямо из наблюдений? В истории науки уже были попытки получить модель непосредственно из данных. Наиболее удачной из них считается гелиоцентрическая модель эллиптического движения планет. Три знаменитых закона были получены *Иоганном Кеплером* из Вюртенберга (1571–1630 гг.) в результате чудовищных по объему ручных вычислений, проделанных над наблюдениями пражского астронома *Тихо Браге* (1546–1601 гг.) Модель Кеплера — не только замечательный пример получения *явных знаний из таблиц данных*. Она имела замечательные предсказательные возможности, реализованные Кеплером в так называемых *Рудольфовых таблицах* — наперед рассчитанных эфемеридах нескольких планет.

На практике наблюдения чаще всего представлены скалярными временными рядами. Можно предположить, что отсчеты ряда являются *нелинейной проекцией*<sup>[1]</sup> движения фазовой точки некоторой динамической

---

<sup>1</sup>Если ссылка на примечание представляет собой число, заключенное в круглые скобки, например, (3), то она обозначает номер примечания, помещенного в конце данного текущего раздела. Ссылка в квадратных скобках, например, [3], обозначает номер дополнения в конце лекции. Ссылка в виде числа, не заключенного в скобки — обычное подстраничное примечание. — Прим. ред.

<sup>2</sup>См. статью *Е. Вигнера*, с тем же названием в сборнике [59].

системы, продуцирующей ряд — нечто вроде теней на стене пещеры, если следовать известной аллегории Платона<sup>3</sup> [2].

В таком случае:

- *Что можно сказать о неизвестной системе на основе созерцания динамики ее тени?*
- *Можно ли по проекции восстановить образ системы и, если да, то в каком смысле?*

На первый взгляд кажется, что эти вопросы бессмысленны. Действительно, располагая, например, решением трехмерной системы уравнений лишь для одной, скажем  $x(t)$ , координаты, мы не имеем никакой информации о функциях  $y(t)$  и  $z(t)$ . Однако, в 1980 году Паккард с коллегами [17] предположил, что для *топологической идентификации*<sup>[3]</sup> такой системы достаточно измерить одновременно три *произвольных* переменных, связанных с движением системы. Для этого они должны быть независимыми в некотором (*операционном*) смысле. Такими «координатами» могут быть, например, значение единственной переменной  $x$  и ее первых двух производных. Тройка  $(x, \dot{x}, \ddot{x})$  заменяет оригинал  $(x(t), y(t), z(t))$  так, что построенный в новых «координатах» фазовый портрет системы оказывается похожим на оригинал с *точностью до гримасы клоуна* — непрерывных, но нелинейных деформаций. Эта эвристическая идея была формализована в 1981 году в знаменитой статье Ф. Такенса [20] в форме теоремы о *типичном вложении* временного ряда в  $R^n$ . Так возник новый способ построения модели из наблюдаемого сигнала (или реализации), который тут же инициировал совершенно новую область численных методов топологической динамики — *эмбедологию*<sup>4</sup> [4]. С ее помощью наконец-то удалось подойти к проблеме моделирования динамики с «правильного конца»: модель началась с «ответа» — наблюдений, а затем ставился «правильный» вопрос: что их продуцирует?

*Эмбедология* позволила сразу же получить нелинейный многомерный вариант традиционного авторегрессионного прогноза временных рядов. Поиск предиктора свелся к проблеме аппроксимации нелинейной функции

<sup>3</sup> Платон в одном из своих знаменитых диалогов рассказал притчу об узниках, находящихся в пещере и прикованных там так, что они могут наблюдать только тени от проходящих людей и проносимых предметов. В связи с этим задается вопрос — могут ли узники понять, что происходит снаружи, наблюдая за этими тенями? И Платон отвечает на этот вопрос утвердительно.

<sup>4</sup> Название данной области [20] происходит от английского слова *embedology*, производенного, в свою очередь, от слова *embedding* (вложение).

нескольких переменных. Для глобального варианта прогноза таким предиктором является нейронная сеть по своему *modus operandi*<sup>5</sup>. Кулинарная книга *Эмбедологии* предлагает рецепты острых «ресторанных» блюд, содержащих компоненты из дифференциальной топологии, фрактальной геометрии и эргодической теории гладких динамических систем. Они выглядят экзотикой с точки зрения практика-экспериментатора, привыкшего к меню традиционной «линейной» столовой, а *Пособия для Чайников* по приготовлению «блюд» эмбедологии нередко приводят в уныние. Поэтому, настоящая *Лекция* предлагает вместо рецептов — принципы, по которым готовятся основные деликатесы, включая нейропрогноз.

Автором намеренно был выбран стиль, являющийся компромиссом между лекцией и конспектом из различных источников. Лекционный стиль позволяет не слишком заботиться о строгости изложения, прибегая в особенно туманных местах к уловке св. Августина: «*Что мне, если кто не понимает!*» Форма конспекта, с другой стороны, избавляет от необходимости достичь определенной цели на заданном числе страниц. Издержки при этом сводятся к тому, что какие-либо авторские претензии на систематическую и безусловную завершенность будут по меньшей мере самообманом.

Автор искренне благодарен своим юным коллегам *Оле Круглун, Кате Данилкиной, Ерболу Куандыкову и Светлане Ким*, помощь и терпение которых позволило закончить рукопись.

### Отображения, функции и типичность

В отблесках пространств и  
отображений очертания  
мысленных структур, слов и, быть  
может, даже чисел становятся  
нечеткими, приобретая, что  
несомненно и крайне существенно,  
цвет морской волны.

---

*А. Лиепиньш*  
«Топологические пространства и  
их отображения»

---

<sup>5</sup>Способ действия (лат.). — (Прим. ред.)

Мы будем использовать следующие стандартные обозначения:

- Запись  $x \in X$  означает, что  $x$  принадлежит  $X$ .
- Запись  $\{x | d(a, x) < r\}$  читается: множество таких  $x$ , для которых расстояние от точки  $a$  меньше  $r$ .
- Запись  $A \subset B$  ( $A \subseteq B$ ) означает, что  $A$  содержится в  $B$  (содержится или совпадает с  $B$ ).
- *Объединением* множеств  $A$  и  $B$  называют множество  $C = A \cup B$ , состоящее из элементов  $c \in C$ , каждый из которых принадлежит  $A$  или (и)  $B$ .
- *Пересечением*  $A$  и  $B$  называют множество  $C = A \cap B$ , каждый элемент  $c$  которого входит и в  $A$ , и в  $B$ .

Соответствие между различными множествами удобно описывать функциями или отображениями:  $f(*)$  или  $f : X \rightarrow Y$ . Подмножество  $\{f(x)\} \in Y$  — это область значений функции, а  $\{x\} \in X$  — ее область определения.

Метрическое пространство  $(\mathbf{M}, d)$  — это множество  $\mathbf{M}$  вместе с вещественнозначной функцией  $d$ , которая удовлетворяет аксиомам метрики. Для  $\mathbf{M} \equiv R^n$ , где  $R^n$  —  $n$ -мерное евклидово пространство (обозначается как  $n = \dim M$ ), часто используют  $L_2$ -метрику

$$d(x, y) = \left[ \sum_i (x_i - y_i)^2 \right]^{1/2}.$$

Пусть  $(\mathbf{M}, d)$  — метрическое пространство и  $r > 0$  — действительное число. Тогда *открытым шаром* в точке  $a \in \mathbf{M}$  называется подмножество  $B(a, r) \subset \mathbf{M} : B(a, r) = \{x \in \mathbf{M} | d(a, x) < r\}$ . Шар называется *замкнутым*, если  $d(a, x) \leq r$ .

Отображение  $f : A \rightarrow B$  называют *сюръекцией* или *отображением на*, если образ всего  $A$  совпадает с  $B$ , т. е. каждый элемент из  $B$  является образом по крайней мере одного элемента из  $A$ . Если  $A = B$ , то элемент  $x$ , удовлетворяющий условию  $x = f(x)$  ( $f : A \rightarrow A$ ), называют *неподвижной точкой* отображения  $f$ . Отображение  $\varphi : A \rightarrow B$  называют *инъективным* или *взаимно однозначным*, если  $\forall y \in B$  существует не более одного элемента  $x \in A$  такого, что  $y = \varphi(x)$ . Тогда  $\varphi(x) = \varphi(y)$  влечет  $x = y$ . Если  $B = R^n$ , то  $f$  называют *функцией*: впрочем, термины «функция» и «отображение» обычно считают синонимами. Обратную к  $f(x)$  функцию будем обозначать  $f^{-1}(x)$ . Отображение, являющееся инъективным и сюръективным, называется *биекцией*. Теперь сформулируем условие непрерывности:

отображение  $f : X \rightarrow Y$  непрерывно в  $x \in X$ , если для любой окрестности  $V$ , открытой в  $Y$  и содержащей точку  $f(x)$ , найдется такая открытая в  $X$  окрестность  $U(x)$ , что  $f(U) \subset V$ . Для метрических пространств это определение совпадает с классическим, записанным на  $(\varepsilon - \delta)$ -языке. Отображение  $\varphi$  открытого множества  $U$  из  $R^n$  в  $R^m$  называется *гладким*, если оно имеет непрерывные частные производные всех порядков: говорят, что  $\varphi$  имеет класс  $C^r$ , если это справедливо для производных до порядка  $r$  включительно. В более общем случае, когда область определения  $f$  не открыта, отображение  $\varphi : X \rightarrow R^n$  называют гладким, если для каждой точки  $x \in X$  существует открытая окрестность  $U \subset R^n$  и гладкое отображение  $F : U \rightarrow R^n$  такое, что  $F = f$  на  $U \cap X$  (рис. 1).

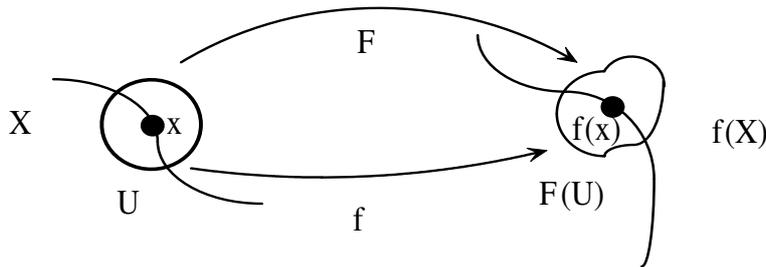


Рис. 1. Определение гладкого отображения

Известно важное свойство гладких функций: если дифференциал такой функции в точке  $x$  имеет некоторое свойство, то сама функция будет иметь такое же свойство по меньшей мере в окрестности точки  $x$ . Например, если производная  $f'(x_0)$  функции  $f$  положительна,  $f'(x_0) > 0$ , то в малой окрестности точки  $x_0$  функция  $f$  возрастает<sup>(1)</sup>. Это свойство называют обычно *принципом общего положения* или *принципом (догмой) линеаризации*, поскольку дифференциал функции является линейной частью отображения<sup>(2)</sup>. Принципу общего положения удовлетворяет не каждая функция, поэтому важно сделать математически точными утверждения о выполнимости некоторого свойства для «почти каждой» функции<sup>(3)</sup>. Вообще говоря, любые свойства математических объектов можно классифицировать как *обычные (usual)* и *необычные (unusual)*. Стандартный прием выразить распространенность некоторого свойства, выполняющегося на подмножестве  $A \subseteq B$ , сводятся к тому, чтобы показать, что обычные свойства соответствуют подмножеству с большей мерой. Однако, пространство функций

является бесконечномерным и поэтому здесь невозможно прибегнуть к замене выражения *почти каждая* на утверждение *с вероятностью единица*. Другой способ сделать это основан на топологии.

Рассмотрим два множества  $A$  и  $B$  таких, что  $A \subset B$ . Говорят, что  $A$  *плотно* в  $B$ , если:

- каждая точка  $B$  либо принадлежит  $A$ , либо является точкой накопления  $A$ , (т. е. каждая точка  $x$ , принадлежащая или не принадлежащая  $A$ , содержит в своей окрестности по меньшей мере одну точку  $y \neq x \mid y \in B$ ), или
- каждое открытое подмножество точек  $B$  содержит точку  $A$ .

**Пример 1.** Множество нецелых чисел плотно и открыто в  $R$ , а множество рациональных чисел плотно, но не открыто на вещественной оси.

Кажется разумным связать «обычные» свойства с множеством  $A$ , которое плотно в  $B$ . Однако, такой путь не приводит к результату, совместимому с интуицией: в некоторых случаях дополнение  $A$  в  $B$  тоже плотное подмножество в  $B$ .

**Пример 2.** Подмножество рациональных чисел плотно в  $R$ , но таким же является и его дополнение — подмножество иррациональных чисел. Однако, интуитивно ясно, что иррациональное число на  $R$  — более обычный объект. Чтобы адекватно отслеживать такие ситуации, было введено понятие *остаточного множества* (*residual set*)<sup>(4)</sup>. Подмножество  $A$  *остаточное* в  $B$ , если оно является пересечением конечного или счетного числа открытых и плотных подмножеств в  $B$ . Любое счетное пересечение остаточных множеств плотно, и любое «респектабельное»<sup>6</sup> пространство содержит остаточное множество. Такие пространства называют *пространствами Бэра*. Например, любое полное метрическое пространство является бэрдовским. В пространстве Бэра дополнение остаточного множества не может быть остаточным: в противном случае пустое множество пришлось бы считать плотным.

**Пример 3.** Пусть  $Q = \{q_1, q_2, \dots, q_n, \dots\}$  — подмножество рациональных чисел в  $R$ , а  $U_n = R - \{q_n\}$  открытое плотное подмножество. Тогда иррациональные числа — это счетное пересечение всех  $U_n$  и, следовательно, остаточное подмножество в  $R$ ; с другой стороны очевидно, что  $Q$  не является таковым.

<sup>6</sup>Т. е. «хорошее» пространство, в котором есть все для работы — например, полное, метрическое. Или банахово, с нормой и метрикой.

Современная метрическая точка зрения на типичность заключается в идее *превалентности*: свойство превалентно <sup>(5)</sup> для отображений из некоторого гладкого конечнопараметрического семейства, если множество значений параметра, соответствующих отображениям, для которых свойство не выполняется, имеет лебегову меру нуль.

Ради простоты мы будем понимать выражение «с точностью до предположения о типичности» в следующем не очень строгом смысле. Свойство типично для функции  $f$ , если оно выполняется в каждой точке соответствующего пространства и открытой окрестности этой точки. Кроме того, если оно не выполняется, например, в точке  $p$ , всегда можно найти достаточно близкую к  $p$  точку, в которой это свойство выполняется. Аналогом является трактовка: свойство *типично*, если оно справедливо для  $f$ , либо, если это не так, становится справедливым для  $f + \delta f$ , где порядок возмущения  $\delta f$  заранее оговаривается.

### Примечания

1. Более точно, если дифференциал функции  $f$  в точке  $x_0$  строго возрастает, то и сама функция  $f$  строго возрастает в окрестности точки  $x_0$ .
2. Тривиальный пример информативности линеаризации: заменим в уравнении  $f(x, y) = 0$  функцию  $f$  на  $df$ . Уравнение  $f'_x dx + f'_y dy = df(x_0, y_0) = 0$  — линейное и условие  $f'_y(x_0, y_0) = 0$  эквивалентно существованию его решения. Но именно тогда  $y = \varphi(x)$  существует и является решением исходного уравнения.
3. Примером могут служить утверждения: почти каждая интегрируемая функция  $f : [0, 1] \rightarrow R$  удовлетворяет условию

$$\int_0^1 f(x) dx \neq 0$$

или почти каждая непрерывная функция  $f : [0, 1] \rightarrow R$  нигде не дифференцируема.

4. Иногда его называют *густым* множеством.
5. Точнее, пусть  $V$  — полное метрическое пространство. Запись  $S + b$  означает трансляцию множества  $S \subset B$  на вектор  $b \in B$ . Говорят,

что мера  $\mu$  *трансверсальна* борелеву множеству  $S \subset B$ , если (i) существует компактное множество  $A \subset B$ , для которого  $0 < \mu(A) < \infty$  и (ii)  $\mu(S + b) = 0, \forall b \in B$ . Борелево множество  $S \subset B$  называется *недостаточным* (*shy*), если существует мера, трансверсальная  $S$ . Дополнение к множеству *shy* называют *превалентным*. Все превалентные множества плотны.

**Путеводитель по литературе.** Краткую сводку определений, касающихся функций и отображений, можно найти в карманном справочнике [1]. Необходимые детали содержатся в учебниках [2, 3] и популярном курсе [4]. Доступное введение в идеи типичности можно найти в книге [13] и приложении к монографии [12]. Строгое определение превалентности приведено в [14].

## Многообразия

Облик их близок к облику  
просветленных стихий, но  
форма не струистая, как у тех, и  
лишенная способности телесного  
взаимопроникновения.

---

Д. Андреев  
«Роза Мира»

Рассмотрим группу преобразований плоскости  $R^2$

$$x' = ax + by + c, \quad y' = px + qy + s, \quad \|A\| = \begin{pmatrix} a & b \\ p & q \end{pmatrix},$$

заданную матрицей  $\|A\|$ . Если  $\|A\|$  — ортогональна, то обычную геометрию можно определить как совокупность тех свойств, которые сохраняются (инвариантны) относительно действия этой группы. Например, заведомо сохраняются расстояния между точками и углы между прямыми. *Аффинная геометрия* получается, если ортогональность матрицы  $\|A\|$  заменить на более скромное требование  $\det A \neq 0$ . Следующим обобщением будет группа, заданная в виде:

$$x' = f(x, y), \quad y' = g(x, y),$$

где  $f$  и  $g$  — нелинейные, но  $C^n$ -гладкие функции. Если же такими свойствами обладают обратные преобразования  $(f^{-1}, g^{-1})^7$ , мы получим *геометрию гладких многообразий*. Это совокупность свойств, инвариантных относительно растяжений и деформаций, т. е. *гомеоморфизмов* <sup>(1)</sup>. В этой геометрии куб, шар и конус неразличимы. Здесь тело человека можно превращать в тела жирафа, зайца или верблюда — все они здесь «топологические синонимы».

Рассмотрим сферу  $S^2$  и бесконечное число ее копий, не различимых с точностью до гомеоморфизмов. Весь этот класс эквивалентных объектов называют *многообразием*. Привычная сфера лишь типичный представитель этого класса, его конкретная реализация в евклидовом пространстве  $R^3$ . Визуализация сферы, к которой мы привыкли, это *поверхность уровня* для функции  $f(x, y, z) = x^2 + y^2 + z^2$ , т. е.  $f^{-1}$ .

Важно подчеркнуть, что форма вовсе не обязательный атрибут многообразия, а просто его геометрическая визуализация <sup>(2)</sup>. Примером многообразия, *негомеоморфного*  $S^2$ , является тор  $T^2$ : действительно, не существует непрерывных деформаций, переводящих тор в сферу. Гомеоморфная копия  $T^2$  может выглядеть совсем не похожей на «бублик». В качестве примера возьмем единичный квадрат

$$X = \{(x, y) \mid 0 \leq x, y \leq 1\}.$$

Эта запись означает, что  $X$  — множество пар точек  $x$  и  $y$  таких, что любая из них имеет координаты, заключенные в отрезке  $[0, 1]$ . Теперь «склеим» правую сторону  $x$  с левой (это запишем, как  $(0, y) \sim (1, y)$ ).

Мы не делаем этого в действительности, а считаем, что указанные пары сторон совпадают, т. е. они «отождествлены». После этого отождествим верхнюю и нижнюю окружности полученного цилиндра. То, что получилось в итоге — это и есть копия тора <sup>(3)</sup>. Сферу также можно получить «склежкой» круга или квадрата, у последнего надо отождествить пары смежных сторон — вместо клея можно представить себе «замки-молнии». Нетривиальная склейка «по диагонали»:  $(0, y) \sim (1, y)$  и  $(x, 0) \sim (1 - x, 1)$  приводит к экзотическому преобразованию — «бутылке Клейна». Чтобы выполнить второе отождествление, нам необходимо попасть в  $R^4$ !

Гомеоморфизмы позволяют «выпрямлять» куски многообразия на области в  $R^n$ : небольшой участок сферы хорошо моделируется плоской областью. Эта идея давно использована в картографии: любой известный

<sup>7</sup>Их существование гарантирует нам группа.

способ построения карт — это проекция кусков сферы  $S^2$  на плоские области  $R^2$ . Почти очевидно, что невозможно получить непрерывное и взаимно однозначное отображение всей сферы на один связный кусок  $R^2$ . Поэтому сферу разбивают на отдельные куски, которые частично перекрываются, и для каждого из них строится карта. Перекрытия позволяют согласовать отдельные карты географического атласа при их «склейке» в сферу.

Согласование — это установление взаимно однозначных соотношений между разными картами и, тем самым, разными областями  $S^2$ . Даже для окружности недостаточно одной карты. Действительно, можно попытаться использовать уравнение  $x^2 + y^2 = 1$ , чтобы картографировать  $S^1$  одним связным куском  $R^1$ . Однако, небольшое размышление показывает, что такая процедура возможна лишь для верхней и нижней полуплоскости отдельно. Можно использовать полярные координаты, но и в этом случае одного «куска» оси  $\varphi$  недостаточно: когда точка окружности делает один оборот,  $\varphi$  скачком меняется на  $2\pi$ . При этом нарушается непрерывность гомеоморфизма. Идею гомеоморфизма часто используют для определения многообразия: многообразием называют топологическое пространство локально гомеоморфное  $R^m$ . Так, например, двойной конус на рис. 2 (слева) не является многообразием. Действительно, окрестность общей вершины двух конусов не похожа на фрагмент плоскости, она двухлистная. Напротив, сфера (рис. 2, справа) — локально похожа на фрагмент евклидовой плоскости.

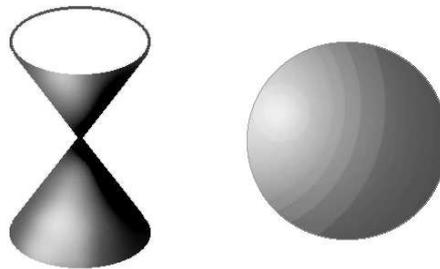


Рис. 2. Примеры топологических пространств: слева (два конуса с общей вершиной) — не многообразие; справа (сфера) — многообразие

В этом смысле гомеоморфизмы являются клеем для изготовления многообразий, которые можно считать теперь объектами, склеенными из от-

дельных кусков  $R^n$ . Изготовление многообразий таким способом сохраняет близость точек: близкие точки многообразия отображаются в близкие точки плоской карты. На карте можно выбрать подходящую систему координат и тогда понятие близости обретает смысл расстояния. Само многообразие не наделено координатами, поэтому понятие близости здесь носит несколько призрачный оттенок. Ситуация становится более определенной, если ввести понятие *окрестности*.

*Окрестность точки  $x$*  многообразия можно определить, указав выделенный класс подмножеств (их называют *открытыми*), все точки которых «достаточно близки к любой точке». Класс замкнут относительно операций объединения и пересечения. А именно, пересечение *конечного* числа открытых множеств и объединение *любого* их числа также принадлежит указанному классу. Именно он определяет *топологию* многообразия. Если в нем всего два объекта: все множество и пустое множество, то топология *тривиальна*. Объявив открытыми все точки множества, мы получим *дискретную топологию*. Важно, *что* объявить открытым!

Пусть многообразие — обычное  $R^n$  с евклидовой метрикой. Открытые окрестности, *индуцирующие метрическую топологию*, — это открытые шары:  $\sum x_i^2 < \rho$ . Не все множества позволяют ввести топологию, например, два конуса с общей вершиной на рис. 2. Окрестность общей точки не допускает «выпрямление» на плоскость. Аналог этой ситуации — множество  $xy = 0$  в  $R^2$ . Попытка ввести карту, т. е. построить гомеоморфизм на  $R^1$ , ведет к нелепости: наше множество — координатные оси; и если мы отобразим открытое на оси  $x$  множество, например,  $a < x < b$  в  $R^1$ , считая, что точка  $(0, 0)$  принадлежит  $(a, b)$ , то в  $R^1$  не останется места для точек, содержащих  $(0, 0)$  по оси  $y$ . С другой стороны, нельзя объявить открытыми множества

$$(a < x < b; y = 0) \quad \text{и} \quad (c < y < d; x = 0),$$

так как их пересечение — точка  $(0, 0)$  должна быть открытой. Но ее образ в  $R^1$  — отдельная точка, т. е. замкнутое пространство! Такие множества не являются многообразиями, поскольку для них не выполняется *аксиома отделимости Хаусдорфа*, справедливая для многообразий: любая пара точек может быть разделена непересекающимися окрестностями.

Пусть  $M$  — многообразие,  $U$  — подмножество в  $M$ , а  $\Phi$  — биекция  $U$  в  $R^n$ . Тогда  $n$ -мерной картой «с» на  $M$  называют пару  $(U, \Phi)$ . Здесь  $U$  — *область карты*, а  $\Phi$  — *координатное отображение*. Иными словами, мы

просто приписываем каждой точке  $p \in U$   $n$  чисел:

$$(\Phi_1(p), \Phi_2(p), \dots, \Phi_n(p)) = (x_1, x_2, \dots, x_n),$$

которые называются координатами  $p$  в карте  $c = (U, \Phi)$ . Отображение  $\Phi^{-1}$  называют *параметризацией*  $U$ .

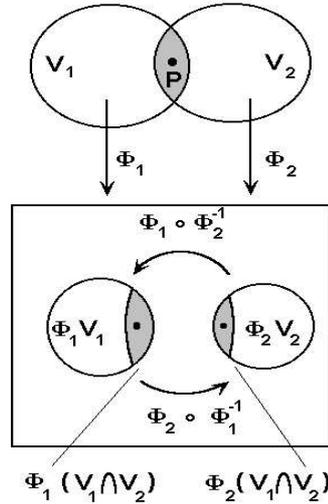


Рис. 3. Согласование карт на многообразии

Пусть точку  $p$  содержат две пересекающиеся окрестности  $V_1$  и  $V_2$  (см. рис. 3). Имеем две карты:  $c = (V_1, \Phi_1)$  и  $c' = (V_2, \Phi_2)$ . Эти карты  $C^r$ -согласованы, если:  $\Phi_1(V_1 \cap V_2)$  и  $\Phi_2(V_1 \cap V_2)$  открыты в  $R^n$  и функции

$$\Phi_1 \circ \Phi_2^{-1} : \Phi_2(V_1 \cap V_2) \rightarrow \Phi_1(V_1 \cap V_2)$$

$$\Phi_2 \circ \Phi_1^{-1} : \Phi_1(V_1 \cap V_2) \rightarrow \Phi_2(V_1 \cap V_2)$$

являются  $C^r$ -гладкими функциями, т. е.  $C^r$ -диффеоморфизмами. Полный набор согласованных карт для всего многообразия называют  $C^r$ -атласом, а само многообразие  $C^r$ -дифференцируемым.

Отображение  $\varphi : X \rightarrow Y$  двух гладких многообразий или двух произвольных множеств в  $R^n$  называют *диффеоморфизмом*, если  $\varphi$  и  $\varphi^{-1}$  —

гладкие гомеоморфизмы (см. рис. 4). Окружность в  $R^2$  диффеоморфна эллипсу, но не треугольнику, которому она лишь гомеоморфна. Диффеоморфизмы определяют класс эквивалентных диффеоморфных пространств.

Введем на многообразии некоторые понятия. *Кривая на  $M$* , проходящая через точку  $p$  — это отображение  $\mathfrak{S} \equiv [0, 1] \rightarrow M$ , сопоставляющее каждой  $\lambda \in \mathfrak{S}$  точку  $c(\lambda) \in M$ , так что  $c(0) = p$ . Ясно, что  $c$  — это параметризация: две кривые различны, даже если их образы совпадают, но одинаковым точкам  $p$  в  $M$  соответствуют разные  $\lambda$ .

*Функция на  $M$*  — это отображение  $M \rightarrow R^1$ , сопоставляющее каждой точке  $p \in M$  вещественное число. Поскольку точка  $p$  всегда содержится в некоторой окрестности  $V$ , карта окрестности  $(V, \Phi)$  превращает  $f : M \rightarrow R^1$  в функцию, заданную на  $R^n : f \circ \Phi^{-1} : R^n \rightarrow R^1$ . Если она дифференцируема на  $R^n$ , то говорят, что она дифференцируема на  $M$ .

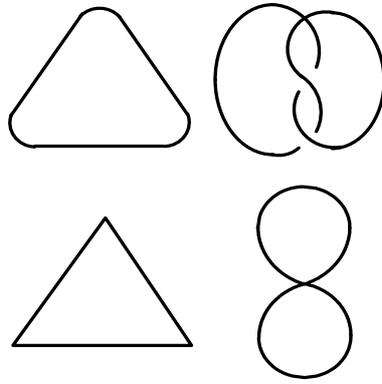


Рис. 4. Две фигуры сверху диффеоморфны; нижняя пара — нет

Определим касательный вектор к  $p \in M$ . Хотелось бы обойтись без координат, т. е. дать внутреннее определение. Идея стрелки, «приклеенной» к  $p \in M$ , дает подходящее представление, недостатком которого является необходимость иметь некоторое объемлющее пространство, в которое выходит конец стрелки.

Пусть через точку  $p \in M$  проходит кривая  $\mathfrak{S} \rightarrow M : \lambda \rightarrow c(\lambda)$  и  $c(0) = p$ . Будем изучать смещение точки  $p$  с помощью произвольной скалярной функции  $f$ : мы просто используем в качестве аргумента этой функции точки, принадлежащие кривой. Интуитивно ясно, что в инфинитези-

мальной окрестности  $p$  изменение функции  $f(c(\lambda)) - f(c(0))$  не зависит от самой функции — это просто сдвиг  $p$  и больше ничего! Но сдвиг — это оператор, который выражает общее свойство всех кривых, проходящих через  $p$ , а именно, то что все они имеют одно и то же изменение для произвольных функций, моделирующих смещение  $p$ . Этот оператор  $d/d\lambda$  называют *касательным вектором к  $p$  на  $M$* . Он определяет класс эквивалентности кривых (т. е. то общее, что они имеют), проходящих через точку  $p$ .

### Примечания

1. Точнее, *гомеоморфизмом* называется отображение  $f$ , для которого  $f$  и  $f^{-1}$  непрерывны. Пожалуй, наиболее оригинальное применение этого свойства — непрерывности гомеоморфизмов — нашел французский художник *Оноре Домье* (1808–1879 гг.). В серии карикатур он небольшими деформациями преобразовал физиономию *Луи Филиппа* в грушу! Королевский суд не смог указать, на каком именно рисунке кончается искусство и начинается оскорбление Его Величества! Рассмотрим два интервала: замкнутый —  $[0, 1]$  и открытый —  $(0, 1)$ . Если выбросить точку 0 из первого, получится полуоткрытый интервал  $(0, 1]$ , который интуитивно состоит из одного куска. Но если выбросить любую точку из  $(0, 1)$  — мы получим два куска. Поэтому эти два интервала не гомеоморфны.
2. *Имея дело с формами, мы приобретаем здоровое неуважение к их авторитету . . .* — очень точная метафора топологии, принадлежащая *М. Шубу*<sup>8</sup>.
3. Тор гомеоморфен следующему подпространству в  $R^3$ :

$$(x, y, z) \in R^3 \mid (\sqrt{x^2 + y^2} - 2)^2 + z^2 = 1.$$

**Путеводитель по литературе.** С основными понятиями топологии можно познакомиться в популярной книге [4], которую хорошо дополняет брошюра [5]. Строгие топологические определения можно найти в учебниках [3, 6–8, 10–11].

<sup>8</sup>*Майкл Шуб (Michael Shub)* — это тополог, работающий в области динамических систем. См. его страничку URL: <http://www.research.ibm.com/people/s/shub/>  
Цитата взята из его письма *Ральфу Фрубергу (Ralf Fruberg)*, написанного в 1969 году; приводится по книге *М. Хирша* [6] (гл. 9, с. 244).

**Погружения и вложения**

... и по виду их и по устройению их  
казалось, будто колесо находилось  
в колесе...

*Книга Пророка Иезекииля  
«Гл. 1, ст. 16»*

Пусть  $f : X \rightarrow Y$  гладкое отображение двух многообразий и  $\dim X < \dim Y$ . Этому отображению можно поставить в соответствие отображение касательных пространств  $df : T_x(X) \rightarrow T_{f(x)}(Y)$ , «приклеенных» к точкам  $x$  и  $f(x)$ , соответственно <sup>(1)</sup>. Его называют *дифференциалом* отображения  $f$  и обозначают  $df$ . Понятно, что в нашем примере  $df$  не может быть сюръекцией, потому что оно не в состоянии «накрыть» лишние координаты в  $Y$ . Самое большее, на что мы можем претендовать, это на свойство инъекции  $df$ . В этом случае,  $f$  называют *иммерсией* или *погружением*  $X$  в  $Y$ . Например, окружность  $S^1 : r(t) = (\sin t, \cos t)$  можно погрузить в  $R^2$  так, что ее образом будет «восьмерка»:  $r(t) = (\sin t, 2 \sin 2t)$ . При этом взаимно-однозначное соответствие точек образа и прообраза, конечно, отсутствует (рис. 5), но каждому вектору скорости  $\dot{r}(t)$  на  $S^1$  соответствует единственный касательный вектор «восьмерки».

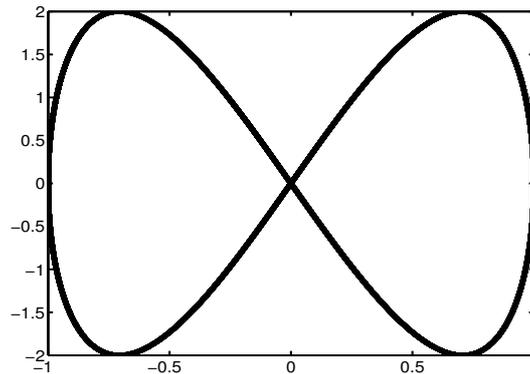


Рис. 5. Погружение окружности в плоскость

Можно изменить параметризацию, используя замену координат

$$t \rightarrow 2 \operatorname{arctg} u : \tilde{r}(u) = (\sin(2 \operatorname{arctg} u), 2 \sin(4 \operatorname{arctg} u))$$

и тогда кривая пройдет через начало координат только один раз, однако, взаимнооднозначное соответствие между кривой и окружностью отсутствует — у кривой появились свободные концы.

Каноническая иммерсия получается, если евклидово пространство  $R^k$  погрузить в  $R^m$  ( $k < m$ ). При этом точка  $x \in R^k$  линейно отображается в точку  $f(x) \in R^m$ :

$$(x_1, x_2, \dots, x_k) \rightarrow (x_1, x_2, \dots, x_k, \underbrace{0, \dots, 0}_{m-k}).$$

В этом случае образ иммерсии — подпространство в  $R^m$ . К сожалению, для многообразий, которые являются только локально евклидовыми, это не так. Например, в случае «восьмерки» окрестность точки самопересечения не является открытой в  $R^2$ . Если использовать параметризацию:

$$\hat{r}(v) = (-\sin(2 \operatorname{arctg} v), 2 \sin(4 \operatorname{arctg} v)),$$

образы  $\tilde{r}(u)$  и  $\hat{r}(v)$  совпадают, однако, композиция  $r \circ \hat{r}^{-1} : R \rightarrow R$  не является непрерывным отображением.

Отображение

$$\varphi(t) = ((1 + e^{-t}) \cos t, (1 + e^{-t}) \sin t)$$

определяет иммерсию вещественной прямой  $R$  в плоскость  $R^2$ : образ  $\varphi(e)$  накручивается на единичную окружность  $x^2 + y^2 = 1$  (см. рис. 6).

В общем случае гладкое отображение  $F$  компактного гладкого многообразия  $A$  называют *иммерсией*, если дифференциал отображения  $dF(x)$  является взаимно-однозначным в каждой точке  $A$  <sup>(2)</sup>. При этом безразлично, имеет ли само  $F$  такое свойство.

Дифференцируемое *вложение*  $A$  — это гладкий диффеоморфизм из  $A$  на собственный образ  $F(A)$ , т. е. гладкое взаимно-однозначное отображение, обратное для которого  $F^{-1}(A)$  также гладко. Если  $A$  — компактное многообразие <sup>(3)</sup>, то отображение  $F$  является вложением тогда и только тогда, когда  $F$  — взаимно-однозначная иммерсия <sup>(4)</sup>. Для *топологического вложения* диффеоморфизм заменяется на гомеоморфизм.

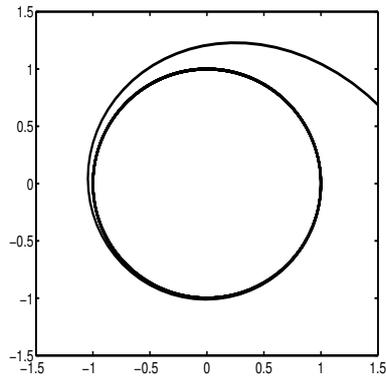


Рис. 6. Иммерсия прямой в плоскость

Знаменитая *теорема Уитни* показывает, что всякое компактное многообразие можно считать подмногообразием евклидова пространства достаточно большой размерности. Теорема Уитни гласит:

*Пусть  $M$  — гладкое компактное  $k$ -мерное многообразие. Тогда существует вложение  $F : M \rightarrow R^n$  для подходящего выбора размерности  $n$ , а именно  $n = 2k + 1$ .*

**Важное замечание.** Вложение может быть реализовано *почти любой* функцией, удовлетворяющей упомянутым условиям. Это означает, что если  $A$  — гладкое многообразие размерности  $k$ , то множество отображений в  $R^{2k+1}$ , которые являются вложениями  $A$ , образуют остаточное (густое) множество в классе всех непрерывных отображений <sup>(5)</sup>.

Условие теоремы Уитни для размерности  $R^n$  можно трактовать следующим образом <sup>(6)</sup>: для взаимной однозначности образа и прообраза отображения следует задать  $k$  координат; для такого же соответствия между касательными векторами необходимо еще  $k$  чисел. Наконец, единица гарантирует дополнительную степень свободы для того, чтобы избежать ситуацию с самопересечением <sup>(7)</sup>. Геометрические истоки условия Уитни восходят к *догме трансверсальности*, о которой рассказывается в следующем разделе. Существует важное обобщение теоремы Уитни на случай, когда  $A$  не является многообразием. Эта ситуация встречается, например, для *фрактальных аттракторов*, для которых не существует разумного определения топологической размерности. Вместо нее используют так называемую *box-*

*размерность*, которая обычно не является целым числом. В этом случае [20] для компактного подмножества  $A \in R^k$  с *box*-размерностью  $d$  почти каждое гладкое отображение  $F : R^k \rightarrow R^n$ , где  $n$  — целое число, большее чем  $2d$ , является взаимно-однозначной иммерсией каждого компактного гладкого многообразия  $C$ , содержащего  $A$ .

### Примечания

1. Объединение касательных пространств, построенных в каждой точке  $x \in X$ , т.е.  $T_X = \bigcup T_x$  называют *касательным расслоением*  $X$ . Оно не является векторным пространством, так как невозможно ввести операцию сложения векторов из двух разных касательных пространств.
2. Дифференциал  $dF(x)$  задается обычно матрицей Якоби. Поскольку  $dF(x)$  — линейное отображение, иммерсия эквивалентна условию, что якобиан имеет полный ранг в касательном пространстве.
3. Множество  $A$  называется *компактным*, если оно ограничено и замкнуто. Ограниченность означает, что  $A$  содержится в шаре заданного радиуса, а замкнутость — что оно содержит все свои предельные точки.
4. Представим себе бесконечную нить, накрученную на тор: это инъективная иммерсия  $R \rightarrow T^2$ . Однако бесконечно удаленным точкам нити нельзя поставить в соответствие точки тора, следовательно инъективная иммерсия не является вложением без условия компактности.
5. Иными словами, если данное гладкое отображение есть вложение, его малое возмущение также является вложением. С другой стороны, любое гладкое отображение независимо от того, вложение это или нет, «лежит» произвольно близко от вложения.
6. Другая «физическая» интерпретация сводится к структуре фазового пространства: для идентификации динамической системы необходимо  $k$  координат,  $k$  импульсов и время.
7. Тор  $T^2$  требует, согласно теореме Уитни,  $R^5$  для вложения. Этот избыток размерностей гарантирует, что не только привычный «бублик»,

но и любые его диффеоморфные копии будут вложениями! Кроме того, условие  $2k+1$  гарантирует, что любое гладкое отображение можно аппроксимировать вложением.

**Путеводитель по литературе.** Строгое изложение иммерсий и вложений содержится в [6–8]. Лучшее введение в эту область можно найти в книгах [10, 11] и сборнике [15], обзоре [16] и статье [9]. Наконец, фракталам и фрактальным аттракторам посвящена монография [23] и лекция [24].

## Трансверсальность

Помни, что в Искраженном Мире все правила ложны, в том числе и правило, перечисляющее исключения, в том числе и наше определение, подтверждающее правило.

---

*Р. Шекли*  
«Обмен разумов»

В этом разделе речь пойдет о геометрической ипостаси типичности: о взаимном расположении геометрических тел в пространстве в рамках концепции «общего положения».

Две линии в  $R^3$  в общем случае не пересекаются: в самом деле, достаточно слегка «пошевелить» две пересекающиеся прямые, чтобы достичь этого общего случая. Напротив, в  $R^2$  такое «малое шевеление» не меняет ситуацию, так что пара прямых на плоскости, *как правило*, пересекается. Пара плоскостей в  $R^3$  либо параллельна, либо пересекается вдоль линии — второй случай, очевидно, устойчив относительно «малых шевелений». Линия и плоскость имеют, *как правило*, общую точку — линия, параллельная плоскости не устойчива. Эти наблюдения легко обобщаются на  $R^n$ : пусть  $R^k$  и  $R^l$  — два подпространства в  $R^n$ . В общем случае они не пересекаются, если  $(k + l) \leq n$ ; в противном случае они пересекаются так, что размерность их «общей части» равна  $k + l - n$ . На рис. 7 приведены примеры трансверсальных пересечений (вверху) и нетрансверсальных (внизу).

**Важное замечание.** Может случиться, что пересечение вообще отсутствует. Такая ситуация также считается трансверсальной<sup>(1)</sup>. Поскольку ло-

кальной версией многообразий являются касательные пространства, приведенные рассуждения легко обобщаются для взаимных конфигураций двух подмногообразий  $K$  и  $L$  в  $M$ . Говорят, что  $K$  и  $L$  пересекаются в  $M$  *транскверсально*, если для каждой  $p \in K \cap L$

$$T_p K + T_p L = T_p M$$

и это записывают как  $K \pitchfork L$ . Иначе говоря, при транскверсальном пересечении сумма касательных пространств слагаемых равна полному касательному пространству многообразия <sup>(2)</sup>.

Адаптируем это определение для отображений.

Пусть  $Z \subset M$ ,  $f : N \rightarrow M$  — гладкое отображение и  $p \in N$  — точка. Говорят, что  $f$  — *транскверсально*  $Z$  в  $p$  ( $f \pitchfork Z$ ), если либо  $f(p) \notin Z$ , либо  $T_p f(T_p N) + T_{f(p)} Z = T_{f(p)} M$ . Таким образом, условие транскверсальности необходимо для описания общей типичной ситуации. В локальном варианте его можно редуцировать в *принцип общего положения*.

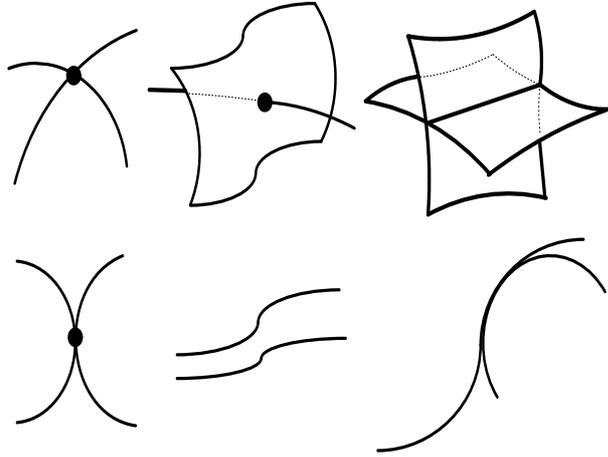


Рис. 7. Примеры взаимного расположения геометрических тел в  $R^3$  (вверху) и  $R^2$  (внизу): транскверсальными являются 2-й и 3-й пример в  $R^3$

Рассмотрим теперь связь между транскверсальностью и условием теоремы Уитни. Пусть по-прежнему  $R^k$  и  $R^l$  — два подпространства в  $R^n$ , которые пересекаются транскверсально. Тогда верхняя часть рис. 7 учит нас, что

в случае трансверсальности размерности  $k, l, n$  удовлетворяют неравенству  $(k + l - n) \geq 0$ . Следовательно, при отсутствии трансверсального пересечения должно выполняться неравенство  $k + l - n < 0$ . Случай  $k = l = d$  соответствует самопересечению или его отсутствию для одного геометрического тела, например, кривой. Тогда условие  $n > 2d$  гарантирует нам, что самопересечение либо отсутствует, либо не является трансверсальным. В последнем случае оно легко ликвидируется малым шевелением. Самый простой способ удовлетворить этому неравенству – положить  $n = 2d + 1$ . Это и есть условие теоремы Уитни.

### Примечания

1. Пусть  $L \equiv R$ , а  $K \equiv R^2$ ,  $M \equiv R^3$ . Тогда в точке их трансверсального пересечения мы имеем тройку касательных векторов: два из них принадлежат плоскости, а третий – линии. Эта тройка – локальная модель  $R^3$ . Таким образом, трансверсальное пересечение моделирует (дает скелет) пространства, в котором происходит пересечение.
2. Здесь используется известный в математике и в жизни «принцип развесистой клюквы»: *развесистая клюква не обязательно является развесистой и не обязательно является клюквой.*

**Путеводитель по литературе.** Строгое изложение идей трансверсальности можно найти в учебниках [6–8, 10, 11]. Понятное для физиков введение содержится в [9, 16].

### Эмбедология и теорема Такенса

... Ничто так не ускользает от изображения в слове и в то же время ничто так настоятельно не требует передачи на суд людей, как некоторые вещи, существование которых недоказуемо, да и маловероятно.

---

*Альберт Второй*  
*«Трактат о кристаллических духах», книга 1, раздел 28. Цит. по: Герман Гессе. «Игра в бисер».*

Напомним, что динамической системой в фазовом пространстве  $M$  называется однопараметрическая группа диффеоморфизмов  $f^t(x) : M \rightarrow M$ . Параметр группы  $t$  — это время, которое может принимать непрерывные или дискретные значения. В качестве  $M$  выступают обычные евклидовы пространства  $R^n$  или гладкие многообразия. Диффеоморфизмы  $f^t$  могут быть заданы в виде дискретных или непрерывных преобразований; групповое свойство означает:

$$f^{t_2}(f^{t_1}(x)) = f^{t_1+t_2}(x).$$

При этом  $f^0(x) = x$  — единица группы; если  $x_1 = f^t(x_2)$ , то  $x_2 = f^{-t}(x_1)$ .

Если динамическая система задана дифференциальным уравнением  $\dot{x} = F(x)$ , то в первом приближении

$$\begin{aligned} x(t+dt) &= x + dx = x + Fdt = f^{dt}(x) = \\ &= f^0(x) + \frac{df^0(x)}{dt}dt + \dots = x + \frac{df^0(x)}{dt}dt. \end{aligned}$$

Функцию  $F(x) = \left. \frac{df^t(x)}{dt} \right|_{t=0}$  называют *векторным полем*, которое задает касательную к траектории в каждой точке фазового пространства.

*Траекторией или орбитой* динамической системы называют последовательность точек

$$\dots, f^{-3}(x), f^{-2}(x), f^{-1}(x), x, f(x), f^2(x), \dots$$

*Неподвижной точкой*  $x_0$  называется траектория, которая для всех  $t$  удовлетворяет условию:  $f^t(x_0) = x_0$ .

*Периодической* называется траектория, не являющаяся неподвижной точкой, и такая, что для некоторого  $P$  и произвольного  $t$  выполнено равенство  $f^{t+P}(x) = f^t(x)$ .

Точка  $p$  называется  $\omega$ -*предельной точкой* для  $x$ , если  $f^{t_i}(x) \rightarrow p$  при  $t_i \rightarrow \infty$ .

Пусть  $A \in M$  и  $f^t(A)$  — множество образов всех точек из  $A$ . Множество  $A$  *инвариантно* относительно  $f$ , если  $f^t(A) = A$  для всех  $t$ .

Замкнутое инвариантное множество  $A \subset M$  называется *притягивающим*, если для него существует окрестность  $U$  такая, что для всех  $x \in U$ ,  $f^t(x) \rightarrow A$  при  $t \rightarrow \infty$ . Инвариантное притягивающее множество, которое содержит все  $\omega$ -предельные точки называют *аттрактором*.

**Важное замечание.** Если аттрактор достаточно велик и содержит только конечное число неподвижных точек и периодических орбит, то поведение фазовых кривых на аттракторе и рядом с ним является хаотическим<sup>(1)</sup>. Хаотический аттрактор ведет себя как гибкий, хотя не вполне управляемый, информационный процессор, обрабатывающий информацию о начальных данных. Фазовые траектории разбегаются в одних (неустойчивых) направлениях и сжимаются в других. Если сжатие преобладает (система диссипативна), то аттрактор копирует сам себя: сечение фазового потока приобретает самоподобную (странную) структуру канторова множества с дробной размерностью<sup>(2)</sup>. Информация порождается не только каскадом бифуркаций, приводящих к нарушению симметрии, но и последовательными итерациями, приводящими к все более тонкому разрешению геометрической структуры. Такая «аппаратурная реализация» может исполнять очень сложный функциональный репертуар, меняя поведение от относительно простого, квазипериодического, до стохастического. Наиболее важным свойством хаотического аттрактора является *существенная зависимость от начальных условий*. Говорят, что  $f^t$  обладает упомянутым свойством, если для точки  $x \in M$  и некоторого числа  $\delta > 0$  существует время  $T > 0$  и такая точка  $y$  из окрестности  $x$ , что расстояние между образами  $d(f^T(x), f^T(y)) > \delta$ . Иными словами, погрешности начальных данных экспоненциально растут в фазовом потоке, так что начиная с некоторого момента времени, будущее состояние системы становится непредсказуемым.

Предположим теперь, что мы верим в существование динамической системы  $f^t$ , объясняющей наши наблюдения, но не имеем ее уравнений в явном виде. То, что нам доступно — это множество наблюдаемых функций  $h(t) \equiv h(x(t)) : M \rightarrow R$ , которые можно представить себе как нелинейные проекции фазовой траектории на вещественную ось. Обычно такими функциями являются экспериментальные временные ряды<sup>(3)</sup>. Мы будем предполагать, что они обладают непрерывностью<sup>(4)</sup> и может быть даже гладкостью, например,  $h(t)$  может принадлежать классу  $C^r(M, R)$ . Во всяком случае, когда фазовая точка смещается под действием  $f : x \rightarrow f(x)$ , это движение должно индуцировать смещение в отсчетах временного ряда:  $h(t) = h(x(t)) \rightarrow h(f(x(t)))$ . Можно ли в этом случае восстановить динамику  $f^t$  в каком либо разумном смысле? Иными словами, какую информацию о неизвестной системе можно извлечь из наблюдений ее «тени», которую отбрасывают траектории? Или, в более общем смысле, можно ли восстановить модель непосредственно из экспериментальных данных? Впервые ответ на этот вопрос был дан в 1980 году в пионерской работе

группы Паккарда и др. [17]: «Эвристическая идея метода реконструкции состоит в том, что для описания состояния трехмерной системы в любой момент времени достаточно измерения любых трех независимых величин, где термин “независимый” пока не определен формально, но будет определен операционно». В качестве такой тройки в этой же работе было предложено использовать сам ряд  $h(t)$  и две его производные. Полученные новые координаты  $(h, \dot{h}, \ddot{h})$  позволяли восстановить фазовый портрет трехмерной модельной системы, заданной в явном виде. В качестве ряда было взято решение системы трех дифференциальных уравнений для одной из координат<sup>(5)</sup>. Несмотря на привлекательность самой идеи — реконструкция фазовой геометрии системы из временного ряда только одной координаты<sup>(6)</sup>, математическая корректность самой процедуры вызвала сомнения. Действительно, согласно теореме Уитни, трехмерное многообразие невозможно даже погрузить в  $R^3$  так, чтобы образ был компактным подмножеством с точностью до предположения о типичности<sup>(7)</sup>. Ситуация изменилась в 1981 году, когда Такенс [20] формализовал эвристику Паккарда в форме своей известной теоремы. Я приведу ее в строгой форме, а затем дам неформальные разъяснения.

**Теорема Такенса.** Пусть  $D^r$  — множество  $C^r$ -диффеоморфизмов  $\{f^t\}$ ,  $f^t : M \rightarrow M$ , компактного  $d$ -мерного многообразия  $M$  и пусть  $C^r(M, R)$  — множество наблюдаемых функций  $h(t), h : M \rightarrow R$ . Определим для  $m \geq 2d + 1$  отображение запаздывающих координат  $F_{f,h} : M \rightarrow R^m$ :

$$F_{f,h}(\vec{x}) = (h(\vec{x}), h(f(\vec{x})), \dots, h(f^{m-1}(\vec{x}))).$$

Тогда множество  $(f^t, h)$ , для которого  $F_{f,h}$  является вложением, открыто и всюду плотно в  $D^r(M) \times C^r(M, R)$  для  $r \geq 1$ .

Первая часть теоремы содержит наши символы веры<sup>(8)</sup>, сводящиеся фактически к тому, что достаточно гладкие (т. е. класса  $C^r(M, R)$ ) наблюдения — временной ряд  $h(t)$  — продуцируются  $C^r(M)$ -гладкой динамической системой, которая имеет компактный  $d$ -мерный аттрактор.

Вторая часть теоремы учит нас, как, используя временной ряд  $h(t)$ , построить копию аттрактора в евклидовом пространстве  $R^m$  подходящей размерности  $m = 2d + 1$  (см. рис. 8). Для этого следует сперва взять  $m$  отсчетов временного ряда, начиная с произвольного номера  $n$ :

$$h_n, h_{n+1}, \dots, h_{n+m-1}.$$

Они рассматриваются как компоненты  $m$ -мерного вектора и образуют точку образа  $z_n \in R^m$ :

$$\begin{aligned} z_n &= (h(f^n(\vec{x}_0)), h(f^{n+1}(\vec{x}_0)), \dots, h(f^{n+m-1}(\vec{x}_0))) = \\ &= (h(\vec{x}_n), h(f(\vec{x}_n)), \dots, h(f^{m-1}(\vec{x}_n))) = F(\vec{x}_n). \end{aligned}$$

Следующая точка получается сдвигом  $m$ -мерного набора на один отсчет, а итеративное продолжение этой процедуры даст упорядоченную последовательность точек — копию истинной орбиты! Действительно, поскольку  $F_{f,h}$  — гладкое и обратимое преобразование, можно определить отображение  $\sigma = F \circ f \circ F^{-1}$  <sup>(9)</sup>. Так как точка  $z_n$  является образом  $F$ , применение  $\sigma$  в  $R^m$  дает:

$$\begin{aligned} \sigma(z_n) &= F \circ f \circ F^{-1}(z_n) = F \circ f \circ F^{-1}(F(\vec{x}_n)) = \\ &= F \circ f(\vec{x}_n) = F(\vec{x}_{n+1}) = z_{n+1}. \end{aligned}$$

Заметим, что теорема справедлива для диффеоморфизмов  $f^t$ , имеющих лишь конечное число неподвижных точек и периодических орбит с периодом меньше  $m$ . В противном случае, для такой периодической точки нарушается даже условие иммерсии (погружения)  $F_{f,h}$ .

Теорема гарантирует, что копия будет *вложением*, т. е. сохранит свойство оригинала с точностью до диффеоморфизмов — непрерывных и дифференцируемых преобразований. Вспомним, что для (топологического) вложения можно использовать любую непрерывную функцию; сдвиг — самая простая функция из возможных. На рис. 9 приведен пример вложения в  $R^3$  временного ряда чисел Вольфа (рис. 10).

**Важное замечание.** Именно это свойство позволяет оценивать многие необходимые динамические инварианты (размерность, ляпуновские показатели, энтропию) для неизвестной в явном виде системы по диффеоморфной копии ее аттрактора. Реконструкция находится в привычном  $R^m$  и вполне доступна для численных манипуляций! Более того, отображение  $\sigma = F \circ f \circ F^{-1}$  <sup>(9)</sup> :  $z_{n+1} = \sigma(z_n)$  лежит в основе современной схемы нелинейного многомерного прогноза. Конечно,  $z_{n+1}$  является вектором, однако уравнение можно переписать для проекции на первую координату:

$$h_{n+1} = \Phi(h_n, h_{n-1}, \dots, h_{n-m+1}).$$

Таким образом, прогноз сводится к поиску наилучшей аппроксимации нелинейной, но непрерывной функции  $\sigma$  от  $m$  переменных. Отображение

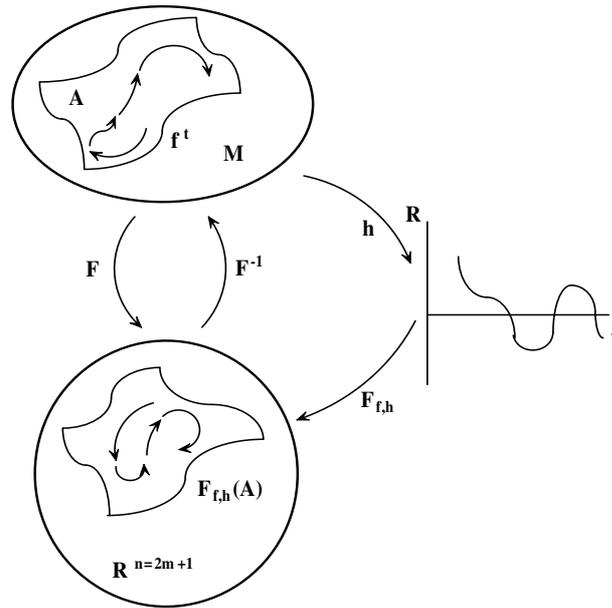


Рис. 8. Реконструкция копии аттрактора  $A$  из временного ряда  $h(t)$  согласно тереме Такенса

$F$ , которое реализует вложение, называют *отображением запаздывающих координат*, а пространство  $R^m$  называют *пространством вложения*. Хотя справедливость теоремы не зависит от величины сдвига (*лага*) между отсчетами — в приведенной выше формулировке он был выбран равным единице — на практике требуется выполнение условия «независимости» новых запаздывающих координат. В противном случае реконструкция «коллапсирует» к диагональной гиперплоскости в пространстве вложения<sup>(10)</sup>. Несмотря на простоту их получения, запаздывающие координаты Такенса в аналитическом смысле хуже варианта *Паккарда* с производными: они не позволяют получить простую форму уравнений движения<sup>(11)</sup>.

Для практического применения теоремы Такенса нужна, прежде всего, оценка *бок-размерности* аттрактора  $d$ . Для ее получения используют простую разновидность метода «проб и ошибок». Поскольку мы не знаем истинной размерности аттрактора, для оценки  $d$  воспользуемся его копией.

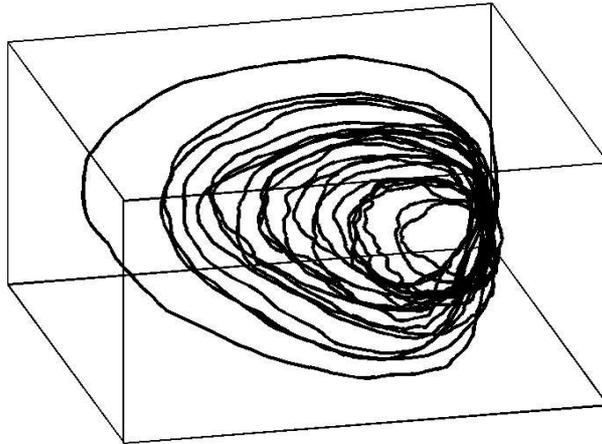


Рис. 9. Вложение временного ряда чисел Вольфа в  $R^3$ . Показан фазовый портрет, построенный в запаздывающих координатах; здесь  $x, x + \text{лаг}, x + 3\text{лаг}$  играют роль  $x, y, z$

А именно, начнем строить реконструкции аттрактора в  $R^d$ , для  $d = 2, 3, \dots$ . Для каждой такой пробной копии будем вычислять некоторую меру, зависящую от размерности. Чаще всего для этих целей используют усредненное число пар  $\varepsilon$ -близких точек копии, подсчитанное в любой подходящей метрике. Мера меняется в зависимости от пробной размерности, но эта зависимость стабилизируется, как только мы достигаем «правильного» значения размерности. Используя затем масштабное поведение размерности от разрешения  $\varepsilon$ , легко оценить значение показателя этой зависимости, так называемую *корреляционную размерность*  $\nu$ . Она и будет нижней оценкой размерности  $d$  аттрактора<sup>(12)</sup>. Практическое применение теоремы Такенса, т. е. реконструкция копии аттрактора из временных рядов с помощью топологических вложений, получило широкое распространение: весь комплекс необходимых для этого процедур, вместе с искусством получения численных оценок называют *эмбедологией*. Этот термин ввел *Тим Зауер и др.* в статье [16], синонимом является термин «*алгоритмическое моделирование*», предложенное в [22].

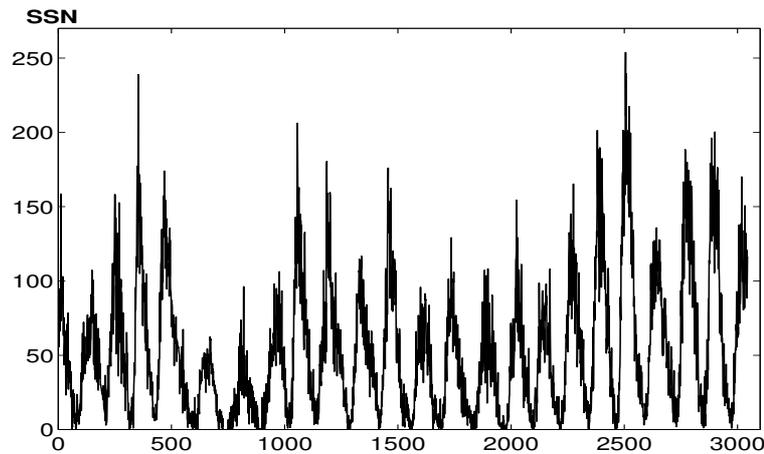


Рис. 10. Временной ряд чисел Вольфа: по вертикали — среднемесячные числа Вольфа (SSN — Sun Spot Numbers), по горизонтали — порядковые номера

Следует помнить, что корректность реконструкции методами эмбедологии существенно зависит от *Кredo идеального экспериментатора* [19] (см. с. 115). Самый болезненный для совести «эмбедолога» вопрос — это проблема существования аттрактора. Для того чтобы избежать в этом месте мучительных раздумий, обычно прибегают к нехитрой уловке, полагая, что *... все что происходит, если это действительно происходит, должно находиться на аттракторе. В противном случае, мы ничего бы и не видели!*

#### Примечания

1. Этот принцип не только не доказан, но даже и не формализован.
2. С физической точки зрения размерность множества — это число параметров, различающих точки этого множества. Для множества экзотической структуры это понятие может быть формализовано как размерность типичной плоскости, на которую множество можно взаимнооднозначно ортогонально спроектировать.

3. Часто используют прилагательное «наблюдаемая» как существительное: *наблюдаемая*  $h(t)$ . Более точно, ее называют *детерминированно порожденной* [18], если существует гладкая динамическая система  $\dot{x} = F(x)$  в  $M = R^n$ , начальная точка  $x_0 \in M$  и ограниченное решение  $t \rightarrow x_0(t)$  вместе с функцией  $h(t) : R^n \rightarrow R$  так, что  $h(t_i) = h(x_0(i))$ .
4. Обычно достаточно условия липшиц-непрерывности, т. е.  $|h(t + \varepsilon) - h(t)| \leq C\varepsilon^H$ ,  $H \in [0, 1]$ .
5. Прием *Паккарда* сводится к следующему изменению координат [25]: для динамической системы

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}),$$

$\mathbf{x} = (x_1, x_2, x_3)$ , определим вектор реконструкции  $\mathbf{y}$  следующим образом:

$$\begin{aligned} y_1 &= x_1, \\ y_2 &= \dot{x}_1 = \dot{y}_1 = F_1, \\ y_3 &= \dot{y}_2 = \ddot{x}_1 = \dot{F}_1 = \frac{\partial F_1}{\partial x_i} \frac{dx_i}{dt} = \mathbf{F}\nabla_1. \end{aligned}$$

В этом случае уравнения движения:

$$\begin{aligned} \dot{y}_1 &= y_2, \\ \dot{y}_2 &= y_3, \\ \dot{y}_3 &= G(y_1, y_2, y_3) = (\mathbf{F}\nabla)^2 F_1 \end{aligned}$$

упрощаются, так как вместо трех отдельных функций от трех переменных мы имеем дело лишь с одной функцией  $G$ .

6. Статья [17] так и называлась: «*Геометрия из временных рядов*».
7. Условие  $n \geq 2d + 1$  можно редуцировать к  $n = d$  только в случае *линейных*  $f^t$  и  $h(t)$  [21]. Но это исключительный случай!
8. Условия теоремы Такенса резюмирует *Кредо идеального экспериментатора* [19], которое предполагает, что в фазовом пространстве  $M$ :

- существует гладкая динамическая система  $f^t : M \rightarrow M$  с конечномерным аттрактором  $A$ ;
  - аттрактор имеет единственную инвариантную меру  $\mu$ : она описывает частоту посещаемости различных частей аттрактора, т. е. время, которое проводит фазовая траектория в каждом фрагменте аттрактора;
  - мера эргодична (частота посещаемости пропорциональна объему фрагмента) и инвариантна под действием  $f^t$ , т. е.  $\mu(B) = \mu(f^t(B))$  для любого  $B \subseteq A$ ;
  - начальная точка фазовой траектории — типичная точка в смысле меры  $\mu$ .
9. Это просто запись последовательных процедур, которые следует читать слева направо: сперва мы отображаем точку из  $R^m$  в фазовое пространство  $M$  с помощью  $F^{-1}$ , затем сдвигаем полученный образ в  $M$  с помощью отображения  $f^t$  и наконец, отображаем полученную фазовую точку с помощью  $F$  назад в  $R^m$ .
10. Это легко понять для вложения временного ряда в  $R^2$ : если  $h_n \simeq h_{n+1}$  — пара координат  $(h_n, h_{n+1})$  принадлежит диагонали!
11. В общем случае, запаздывающие координаты получают сдвигом на некоторый лаг  $\tau$ . Следовательно, для трехмерной системы имеем:

$$\begin{aligned}y_1(t) &= x_1(t), \\y_2(t) &= x_1(t - \tau), \\y_3(t) &= x_1(t - 2\tau).\end{aligned}$$

Тогда уравнения движения имеют вид:  $\dot{y}_i = H_i(\mathbf{y})$ , для  $i = 1, 2, 3$ , но функции  $H_i$ , в отличие от дифференциальных координат Паккарда нельзя явно выразить через первоначальные функции  $F_i(\mathbf{x})$  [25].

12. Пусть  $\mu$  — борелева вероятностная мера с ограниченным носителем в  $R^n$  с метрикой  $|\cdot|$ . Определим корреляционную функцию  $C_d(\varepsilon)$ , как взвешенную с помощью  $\mu$  долю пар векторов  $(x, y) \in R^d \subseteq R^n$ , таких что  $|x - y| \leq \varepsilon$ . Тогда, *корреляционная размерность*  $\nu$  меры  $\mu$  определяется выражением:

$$\nu(\mu) = \sup_s \left\{ \int \int |x - y|^{-s} d\mu(x) d\mu(y) = \int_0^\infty \varepsilon^{-s} dC_d(\varepsilon) < \infty \right\}.$$

Масштабное поведение размерности следует из следующего построения. Разобьем компактную область в  $R^d$ , содержащую  $N$  точек, на  $N(\varepsilon)$  непустых кубов, и пусть  $\mu_i$  — число точек в  $i$ -ом кубе. Тогда с точностью до  $O(1)$ :

$$C_d(\varepsilon) \approx N^{-2} \sum_i \mu_i^2 = N(\varepsilon) N^{-2} \langle \mu^2 \rangle.$$

Здесь  $\langle * \rangle$  — среднее по всем непустым боксам. Используя неравенство Шварца, получаем:

$$\begin{aligned} N(\varepsilon) N^{-2} \langle \mu^2 \rangle &\approx N^{-2} N(\varepsilon) \langle \mu \rangle^2 = \\ &= N^{-2} N^{-1}(\varepsilon) \left[ \sum \mu_i \right]^2 = 1/N(\varepsilon) = \varepsilon^d. \end{aligned}$$

Таким образом, в пределах малых  $\varepsilon$ :

$$C_d(\varepsilon) \propto \varepsilon^\nu; \quad \nu \equiv d = \lim \log C_d(\varepsilon) / \log \varepsilon.$$

**Путеводитель по литературе.** В основе эмбедологии лежат статьи [17–18]; историческая предтеча алгоритмического моделирования изложена в обзорной статье [22]. Понятное и строгое изложение теоремы Такенса дано в работе [21]. В [19] обсуждаются *Кредо идеального экспериментатора* (см. с. 115), впервые введенное в [18]. Фракталы и их связь с динамическими системами изложены в [23] и лекции [24]. Современное изложение теории гладких эргодических динамических систем можно найти в [25–33]. Разнообразные определения аттракторов приводятся в статье *Милнора* [34]. В обзорах [35–36] изложена техника алгоритмического моделирования, главные первоисточники содержатся в сборнике репринтов оригинальных работ [15]. Существует довольно много компьютерных программ для алгоритмического моделирования, наиболее полезными из них, по моему мнению, являются пакеты *Tisean* (Linux, Dos) [37] и *TSTOOL* (Windows MatLab) [38].

**Прогноз как аппроксимация: сети, но не нейронные сети**

Если бы это было так, это бы еще ничего, а если бы ничего, оно бы так и было, но так как это не так, так оно и не этак! Такова логика вещей!

---

*Л. Кэрролл*  
«Алиса в Зазеркалье»

Напомним прежде всего схему обычного авторегрессионного <sup>(1)</sup> линейного прогноза временных рядов. Обозначим наблюдаемый временной ряд (время дискретно) как  $x(n)$  ( $n = 1, 2, \dots, N$ ), а его предсказание — как  $\hat{x}(n)$ . Предположим, что наша последняя известная точка ряда  $x(n-1)$ . Тогда линейное авторегрессионное предсказание на один шаг вперед определяется выражением:

$$\hat{x}(n) = a_1 x(n-1) + \dots + a_p x(n-p),$$

где  $a_i$  ( $i = 1, 2, \dots, p$ ) — весовые коэффициенты, а  $p$  — порядок авторегрессии <sup>(2)</sup>. Веса находятся из условия, чтобы ошибки предсказания:

$$w(n) = x(n) - \hat{x}(n) = x(n) - a_1 x(n-1) - \dots - a_p x(n-p)$$

для  $n = 1, \dots, N-p$  были случайными некоррелированными величинами. В общем случае  $AR$ -процесс  $p$ -го порядка можно записать в виде:

$$\sum_{j=0}^p a_j x(n-j) = w(n),$$

где  $a_0 = 1, a_1, a_2, \dots, a_p < 0$ . Введем оператор сдвига на один отсчет:  $B$  ( $Bx(n) = x(n-1)$ ). Тогда уравнение для ошибки предсказания принимает вид:

$$(1 - a_1 B - \dots - a_p B^p)x(n) = Q_p(B)x(n) = w(n)$$

и его решением будет выражение, в котором заключена суть  $AR$ -модели:  $x(n) = Q_p^{-1}(B)w(n)$ . Мы получили просто *Черный Ящик*, который преобразует белый шум в отсчет временного ряда!

Для дальнейшего важно, что  $AR$ -прогноз основан на *линейной* комбинации прошлых значений. Наиболее очевидное обобщение заключается в

том, чтобы заменить линейную комбинацию на нелинейную функцию. Но именно этот вариант и предлагает эмбедология! Действительно, согласно теореме Такенса в пространстве вложения  $R^m$  существует универсальная динамическая модель:

$$x_{i+k} = \Phi(\mathbf{z}_i),$$

где  $k$  — искомая координата вектора  $\mathbf{z}_{i+1}$ . Таким образом, задача прогноза в эмбедологии приобретает следующую форму: для данного временного ряда и полученной для него реконструкции в  $R^m$ , известны  $(N - m + 1)$  значений<sup>9</sup> векторов  $\mathbf{z}_i = x_i, x_{i+1}, x_{i+2}, \dots, x_{i+(m-1)}$ .

Следовательно, для каждого индекса  $i = 1, 2, \dots, N - m + 1$  мы имеем значения функции:  $x_{i+m} = \Phi(\mathbf{z}_i)$ . Так, например, если  $m = 3$ ,  $N = 6$ , мы получим три «примера»:

$$x_4 = \Phi(x_3, x_2, x_1),$$

$$x_5 = \Phi(x_4, x_3, x_2),$$

$$x_6 = \Phi(x_5, x_4, x_3).$$

Следовательно, для прогноза необходимо найти оценки для  $x_{6+i}$ , начиная с  $i = 1$ , т. е.  $x_7 = \Phi(x_6, x_5, x_4)$ .

К сожалению, мы имеем лишь самую общую информацию о функции  $\Phi(\mathbf{z}_i)$ : эта функция непрерывна и, возможно, дифференцируема. Хуже того, она обычно определена не на всем пространстве векторов  $\mathbf{z}_i$ , а только на поверхности некоторой неизвестной размерности, меньше чем  $m$ . В этих условиях, задача аппроксимации  $\Phi$  решается только на уровне технической строгости. Удивительно, что в ряде случаев результаты прогноза тем не менее вполне удовлетворительны.

Принято делить нелинейные методы прогноза на *локальные* и *глобальные*. В *локальных* методах, функция  $\Phi$  аппроксимируется в локальной окрестности каждой фазовой точки. Поскольку сшивать эти отдельные приближения не требуется, локальный метод называют иногда *непараметрической регрессией*. Идея состоит в следующем. Для каждой точки  $\mathbf{z}_i^0$  находят  $k$  ее ближайших соседей. После этого, поведение функции  $\Phi$  аппроксимируется полиномом степени  $n$  в окрестности точки  $\mathbf{z}_i^0$ . Коэффициенты полинома находят методом наименьших квадратов из условия:

<sup>9</sup>Здесь, ради простоты, мы полагаем лаг  $\tau = 1$ ; если это не так, число таких векторов  $(N - (m + 1)\tau)$ .

$\sum_{i=1}^k |P_n(\mathbf{z}_i^0) - \Phi|^2 \rightarrow \min^{(3)}$ . В *глобальных* методах динамика аппроксимируется сразу во всем  $\mathbf{z}$ -пространстве. При использовании полиномов, минимизируется функционал *полной* ошибки:

$$\sum_{i=1}^N |P_n(\mathbf{z}_i^0) - \Phi|^2 \rightarrow \min .$$

На практике чаще используют глобальные методы с *локальными* свойствами — радиальные базисные функции и нейросетевые технологии. Рассмотрим теперь общие принципы аппроксимации непрерывной многомерной функции  $f(\mathbf{x})$  с помощью некоторой аппроксимирующей функции  $F(\mathbf{w}, \mathbf{x})$  с фиксированным числом параметров  $\mathbf{w}$ ;  $\mathbf{x}, \mathbf{w}$  — вещественные векторы вида  $\mathbf{x} = x_1, x_2, \dots, x_n$  и  $\mathbf{w} = w_1, w_2, \dots, w_m$ . При выборе  $F$  возникает задача нахождения параметров  $\mathbf{w}$ , которые обеспечивают наилучшую возможную аппроксимацию функции  $f$  на основе конечного множества известных «примеров»<sup>(4)</sup>. Качество аппроксимации можно измерить с помощью некоторой функции  $\rho$ , измеряющей расстояние  $\rho[f(\mathbf{x}), F(\mathbf{w}, \mathbf{x})]$  между аппроксимацией  $F(\mathbf{w}, \mathbf{x})$  и  $f(\mathbf{x})$ . Тогда, если  $f(\mathbf{x})$  — непрерывная функция, определенная на множестве  $X$ ,  $F(\mathbf{w}, \mathbf{x})$  — аппроксимирующая функция, непрерывно зависящая от  $\mathbf{w} \in P$  и  $X$ , задача аппроксимации заключается в поиске параметров  $\mathbf{w}^*$ , таких что

$$\rho[F(\mathbf{w}^*, \mathbf{x}), f(\mathbf{x})] < \rho[F(\mathbf{w}, \mathbf{x}), f(\mathbf{x})]$$

для всех  $\mathbf{w}$  из  $P$ .

Как правило, функции  $F$  выбираются из некоторого класса  $A = \{F_i\}$ , который является подмножеством некоторого общего множества функций  $\mathfrak{R} \supset A$ , содержащего и функцию  $f$ . Поэтому, полезно определить расстояние между функцией  $f$  и подмножеством  $A$ , как  $\rho(f, A) = \inf_{F_i \in A} \|f - F_i\|$ . Если точная нижняя граница  $\inf$  достигается для некоторой  $F_* \in A$ , то именно эта функция и будет *наилучшей аппроксимацией* для  $f$ <sup>(5)</sup>. Теперь задача аппроксимации формулируется так: для  $f \in \mathfrak{R}$  и  $A \in \mathfrak{R}$  найти наилучшую аппроксимацию  $f$  функциями из  $A$ . Для решения задачи необходимо конечно, чтобы функции  $F_i$ , т. е. подмножество  $A$  было *всюду плотным* в  $\mathfrak{R}$ . В этом случае, какова бы ни была  $f \in \mathfrak{R}$ , в ее окрестности всегда найдутся подходящие «строительные блоки», из которых мы и построим ее лучшую аппроксимацию<sup>(6)</sup>. К счастью, существуют теоремы, которые помогают нам найти такие  $A$ . Например, из курса анализа (*теорема Вейерштрасса*) мы знаем, что любую непрерывную функцию, заданную в

ограниченном интервале, можно сколь угодно точно аппроксимировать полиномами. Следовательно, множество полиномов (т. е. подмножество  $A$ ) всюду плотно в пространстве  $\mathfrak{R} \equiv C[X]$  непрерывных функций. Самым известным обобщением этого утверждения является *теорема Стоуна*.

**Теорема Стоуна.** Пусть  $X$  компактное метрическое пространство,  $C[X]$  — множество непрерывных функций, определенных на  $X$  и  $A$  подалгебра<sup>(7)</sup> в  $C[X]$  со следующими свойствами:

- функция  $f(x) = 1$  принадлежит  $A$ ;
- для любых двух точек  $x \neq y$  из  $X$  существует функция  $f \in A$ , такая что  $f(x) \neq f(y)$ .

Тогда  $A$  плотно в  $C[X]$ .

Теорема Стоуна значительно расширяет класс доступных строительных блоков: действительно, достаточно взять компактное подмножество функций, различающих точки и образующих подалгебру относительно «школьных» алгебраических операций и мы получим возможность аппроксимировать практически любую непрерывную функцию! Такими функциями могут быть, например, тригонометрические многочлены, радиальные базисные функции и др.

Почти любая схема аппроксимации может быть «отображена» на некоторый тип «сети» или графа, который можно условно считать «нейронной сетью». Такие сети являются просто графическим обозначением широкого класса алгоритмов. Чтобы продемонстрировать, как задача аппроксимации выглядит в «сетевой» формулировке, рассмотрим несколько вариантов для возможного выбора  $F(\mathbf{w}, \mathbf{x})$ :

- линейный случай:  $F(\mathbf{w}, \mathbf{x}) = \mathbf{w} \cdot \mathbf{x}$ , где  $\mathbf{w}$  — матрица размера  $m \times n$ ,  $\mathbf{x}$  —  $n$ -мерный вектор; этот вариант соответствует сети без скрытых слоев;
- классическая схема аппроксимации с базисными функциями  $\varphi_i(\mathbf{x})$ , т. е.  $F(\mathbf{w}, \mathbf{x}) = \mathbf{w} \cdot \varphi_i(\mathbf{x})$ . Этот вариант соответствует сети с одним скрытым слоем; разложение по ортогональным полиномам или радиальным базисным функциям;
- схема со вложенными сигмоидами, которую мы рассмотрим в следующем разделе.

## Примечания

1. *Регрессия* — это взгляд назад, на прошлое какого-то временного ряда, *авторегрессия* — это собственное прошлое. Идеи авторегрессионной (*AR*)-модели восходят к пионерской работе английского статистика Дж. Юла (1927 г.), посвященной исследованию вариаций солнечных пятен [39–40]. Предтечей *AR*-моделей была модель скользящего среднего (*Moving average* — *MA*), предложенная Е. Слущким, в которой случайный процесс  $X_n$  моделировался как линейная комбинация белого шума  $R_n$ :  $X_n = \sum_i C_i R_{n-i}$ . *AR*-модель выражает будущее значение через линейные комбинации своего прошлого:  $X_n = R_n + \sum_i B_i R_{n-i}$ , где  $B_i$  — постоянные коэффициенты. Комбинация *AR*- и *MA*-моделей называется *ARMA*-моделью [40].
2. Для нахождения порядка авторегрессии  $p$  используют различные подходы. Чаще всего используются критерии Акаике [40]. Согласно первому из них — *Окончательной ошибке предсказания (ООП)*,  $p$  выбирается так, чтобы средняя дисперсия ошибки на каждом шаге предсказания была минимальна. Второй — *Информационный критерий Акаике (ИКА)* оценивает порядок модели посредством минимизации некоторой информационной функции. Если *AR*-процесс имеет гауссовские статистики, то  $ИКА[p] = N \ln(\rho_p) + 2p$ , где  $\rho_p$  — оценка дисперсии белого шума, которая используется в качестве ошибки линейного предсказания. Следует упомянуть еще критерий *Риссанена — ДМО*, который основан на минимизации *длины минимального описания модели*:  $ДМО[p] = N \ln(\rho_p) + p \ln N$ .
3. Поскольку число соседей  $k$  должно быть больше числа коэффициентов полинома (это число в  $R^m$  растет как  $m^n$ ), обычно используют полиномы степени  $n \leq 2$ . Дело еще и в том, что при чрезмерном увеличении числа ближайших соседей в их число попадают фазовые точки с далеких соседних траекторий реконструкции; динамический смысл локальной окрестности поэтому теряется.
4. Вообще говоря, следует различать три разные проблемы [45]:
  - задача выбора аппроксимации (или задача представления): какие классы функций  $f(\mathbf{x})$  могут быть эффективно аппроксимированы и какими аппроксимирующими функциями  $F(\mathbf{w}, \mathbf{x})$ ?
  - проблема выбора алгоритма поиска оптимальных значений параметров  $\mathbf{w}$  для данного  $F$ ;

- проблема эффективной реализации алгоритма в параллельном, по-возможности аналоговом аппаратном средстве.
5. Множество  $A$  называют *множеством существования*, если для каждой  $f \in \mathfrak{R}$  существует *по меньшей мере* одна наилучшая аппроксимация. Наконец, множество  $A$  называют *чебышёвским*, если существует *только одна* такая аппроксимация [46].
  6. Для полного комфорта крайне желательно, чтобы множество  $A$  было *чебышёвским*, тогда лучшая аппроксимация будет единственной. Однако, на практике приходится ограничить себя более скромным требованием,  $A$  — просто *множество существования*. Соответствующие теоремы, говорят нам, что для этого  $A$  должно быть компактным [46].
  7. *Алгебра* — это множество элементов  $E$  вместе со скалярным полем  $\mathcal{F}$ , замкнутое относительно операций сложения (+) и умножения ( $\times$ ) между своими элементами, а также умножения ( $\cdot$ ) на скаляр из  $\mathcal{F}$ . Множество  $E$  вместе с  $\mathcal{F}$ , (+) и ( $\cdot$ ) образует линейное пространство. Кроме того, для любых  $f, g, h$  из  $E$  и  $a \in \mathcal{F}$ ,

$$\begin{aligned} f \times g \in E; \quad f + g \in E; \quad a \cdot f \in E; \\ f \times (g + h) = f \times g + f \times h; \quad f \times (g \times h) = (f \times g) \times h; \\ a \cdot (f \times g) = (a \cdot f) \times g = f \times (a \cdot g). \end{aligned}$$

*Подалгебра* — это подмножество  $S \subset E$ , которое является линейным подпространством в  $E$  и замкнуто относительно операции умножения ( $\times$ ): если  $f, g \in S$ , то  $f \times g \in S$ .

**Путеводитель по литературе.** Исходные идеи  $AR$ -моделей великолепно описаны в учебнике [39]. В монографии [40] дано подробное описание техники этих моделей. Статья [41] чрезвычайно полезна в качестве введения в нелинейный многомерный локальный и глобальный прогноз. Пионерскую работу [42] Фармера и Сидоровича хорошо дополняют детали, приведенные в учебнике [43]. Некоторые интересные вещи содержит обзорная статья [44]. Замечательный обзор [45] и препринт [46], которым я и следовал, посвящен методам аппроксимации в применении к нейронным сетям.

## Нейропрогноз: обучение отображению

Она (Программа) способна не только *предвидеть, что случится*, если **Что-То** случится, но, кроме того точно предсказывает, **Что** случится, если **То** ни капельки не случится, то есть вовсе не произойдет.

Станислав Лем  
«Экстелопедия Вестранда»

Вернемся вновь к основному соотношению для прогноза:  $x_{i+m} = \Phi(\mathbf{z}_i)$ . В правой части стоит  $\Phi(\mathbf{z}_i)$  — функция многих переменных. Нельзя ли представить эту функцию как некоторую, возможно нелинейную, комбинацию более простых функций, содержащих, например, не более одной переменной? Простой пример показывает, что такие попытки не выглядят безнадёжными. Рассмотрим функцию двух переменных  $f(x, y) = xy$ . Пусть  $g(\star) = \exp(\star)$  и  $h(\star) = \log(\star)$  — две вспомогательных функции. Тогда соотношение

$$f(x, y) = xy = \exp(\log(x) + \log(y)) = g(h(x) + h(y))$$

показывает, что функцию двух переменных можно выразить через *суперпозицию* функций только *одной* переменной. Вопрос о том, существуют ли вообще *истинные* функции многих переменных, составлял 13-ю проблему Гильберта: *можно ли произвольную непрерывную функцию  $n$  переменных получить с помощью операций сложения, умножения и суперпозиции непрерывных функций двух переменных?*<sup>(1)</sup> Ответ найденный А. Н. Колмогоровым был положительным:

**Теорема Колмогорова.** Существуют фиксированные возрастающие непрерывные функции  $h_{pq}(x)$ , определенные на  $I = [0, 1]$  такие, что любая непрерывная функция  $f$  на  $I^n = I \times I \times \dots \times I$  может быть записана в форме<sup>10</sup>

$$f(x_1, x_2, \dots, x_n) = \sum_{q=1}^{2n+1} g_q \left[ \sum_{p=1}^n h_{qp}(x_p) \right],$$

<sup>10</sup>Заметим, что здесь речь идет о *точном* представлении функции, а не о ее *аппроксимации*.

где  $g_q$  — должным образом выбранные непрерывные функции одной переменной.

Эту теорему удалось улучшить. В частности, было показано, что функции  $g_q$  можно заменить на одну универсальную непрерывную функцию  $g$ , которую можно представить в виде абсолютно сходящегося ряда Фурье, а функции  $h_{pq}$  можно «расщепить» на произведения  $l_p h_q$ , где  $(l_p, p = 1, 2, \dots, n)$  — рациональные независимые числа, а  $h_q, (q = 1, 2, \dots, 2n + 1)$  — непрерывные неубывающие функции на  $I = [0, 1]$ . Тогда, представление любой непрерывной функции  $f \in C[I^n]$  можно переписать в виде:

$$f(x_1, x_2, \dots, x_n) = \sum_{q=1}^{2n+1} g \left[ \sum_{p=1}^n l_p h_q(x_p) \right].$$

Ради наглядности, запишем это представление для случая функции двух переменных:

$$f(x, y) = g(l_1 h_1(x) + l_1 h_2(x) + l_1 h_3(x) + l_1 h_4(x) + l_1 h_5(x) + l_2 h_1(y) + l_2 h_2(y) + l_2 h_3(y) + l_2 h_4(y) + l_2 h_5(y)).$$

На рис. 11 приведена сетевая форма этого примера. Она похожа на многослойную нейронную сеть, которую, однако, почти невозможно реализовать на практике. Число «нейронов» растет здесь с увеличением числа переменных, но дело не в этом. Основная проблема заключается в том, что функции  $h_q$  в формуле Колмогорова *непрерывны*, но не являются *гладкими* <sup>(2)</sup>. Поскольку длина машинного кода функции обратно пропорциональна ее гладкости, мы получим очень плохую точность реализации колмогоровского представления.

Кроме того, *хорошее* сетевое представление должно иметь фиксированные элементы и изменяемые параметры. Сеть *Колмогорова* на рис. 11 не является сетью такого типа: форма  $g$  зависит от функции  $f$ , которую необходимо представить (функции  $h_q$  не зависят от  $f$ ).

С другой стороны, давно известна не точная, а аппроксимационная схема представления функции многих переменных в форме суперпозиции функций от одной переменной:

$$F(\mathbf{w}, \mathbf{x}) = \sigma \left( \sum_n w_n \sigma \left( \sum_i v_i \sigma \left( \dots \sigma \left( \sum_j u_j x_j \right) \dots \right) \right) \right),$$

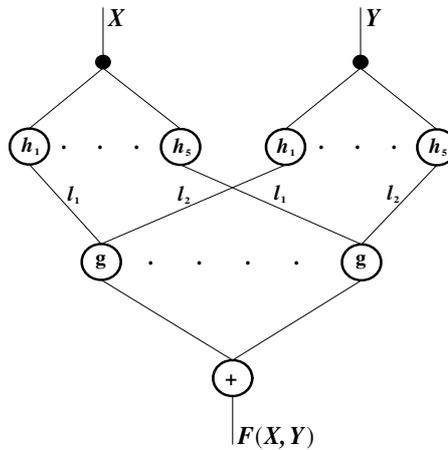


Рис. 11. Сеть, соответствующая колмогоровскому представлению функции двух переменных

которую называют обычно *схемой с обратным распространением*. Она соответствует многослойной нейронной сети с элементами, которые суммируют свои входы с весами  $w, v, u, \dots$  и затем выполняют *сигмоидальное* преобразование этой суммы<sup>(3)</sup>. Такая форма (схема вложенных нелинейных функций) необычна с точки зрения классической теории аппроксимации непрерывных функций<sup>(4)</sup>. Ее истоки лежат в том факте, что с помощью соотношения:

$$F(\mathbf{w}, \mathbf{x}) = \sigma\left(\sum_n w_n \sigma\left(\sum_j u_j x_j\right)\right),$$

где  $\sigma$  — *линейная* пороговая функция, можно выразить любую *булеву* функцию<sup>(5)</sup>. В теории нейронных сетей класс функций:

$$F(\mathbf{w}, \mathbf{x}) = \sum_{i=1}^m c_i \sigma(\mathbf{x} \cdot \mathbf{w} + \theta_i),$$

которые соответствуют одному скрытому слою, принято рассматривать как аппроксимирующие функции для функций  $f \in C[U \subset R^n]$ . Действительно, функции  $\sigma$  образуют плотное множество, но к сожалению оно не является *множеством существования*<sup>(6)</sup>, и следовательно нельзя гарантировать

единственности лучшей аппроксимации. Впрочем, это не так страшно для практических приложений.

Любопытно, что аппроксимационные свойства нейросети определяются не конкретным выбором функции активации, будь то сигмоида или что-то другое, а свойством нелинейности этой функции. А. Н. Горбань [48–50] доказал замечательную теорему, обобщающую теорему Стоуна на нейронные сети. Оказывается, если подалгебра  $E$  в теореме Стоуна замкнута относительно нелинейной операции, она остается всюду плотной в  $C[X]$ . Это эквивалентно утверждению об универсальных аппроксимационных свойствах любой нелинейности [49]: *с помощью линейных операций и каскадного соединения можно из произвольных нелинейных элементов получить требуемый результат с любой наперед заданной точностью!*

Нейросетевая схема аппроксимации функции многих переменных идеально подходит для реализации глобального <sup>(7)</sup> нелинейного прогноза. Сеть учится отображению запаздывающих координат  $\Phi(\mathbf{z}_i) \rightarrow x_{i+m}$  на задачнике, составленном из прошлых значений временного ряда. Рассмотрим, в качестве примера, форму такого задачника. Пусть  $N = 10$ ,  $m = 4$  и лаг  $\tau = 2$ . Тогда таблица для обучения выглядит следующим образом:

$x_1$	$x_3$	$x_5$	$x_7$	$x_9$
$x_2$	$x_4$	$x_6$	$x_8$	$x_{10}$
$x_3$	$x_5$	$x_7$	$x_9$	$\hat{x}_{11}$
$x_4$	$x_6$	$x_8$	$x_{10}$	$\hat{x}_{12}$

В правом крайнем столбце стоят *ответы*. Поскольку лаг равен двум, одношаговый прогноз позволяет получить сразу два предсказанных значения:  $\hat{x}_{11}$  и  $\hat{x}_{12}$ .

На рис. 12 приведен пример нейропрогноза ежемесячных значений временного ряда чисел Вольфа. Прогноз начинается с цикла № 20 и заканчивается серединой цикла № 24. Каждый цикл прогнозировался отдельно. Фактически, до середины текущего 23-го цикла делался эпигноз<sup>11</sup>. Параметры вложения  $m = 7$ ,  $\tau = 129$ . Использовалась коррекция, описанная в Примечании 7.

<sup>11</sup>Здесь речь идет о том, что в прогнозе для тестирования используют уже известные значения ряда. Т. е. вы делаете вид, что их не знаете, и предсказываете, чтобы проверить метод. Мотивировка простая: для проверки корректности прогноза нескольких 11-летних циклов вперед можно и не дожить. Вот для такого варианта прогноза используют термин *эпигноз*. Иногда приходится делать прогноз в прошлое, продолжая какие-то ряды за пределы интервала времени, когда они действительно есть. Такой прогноз называют *палегноз*.

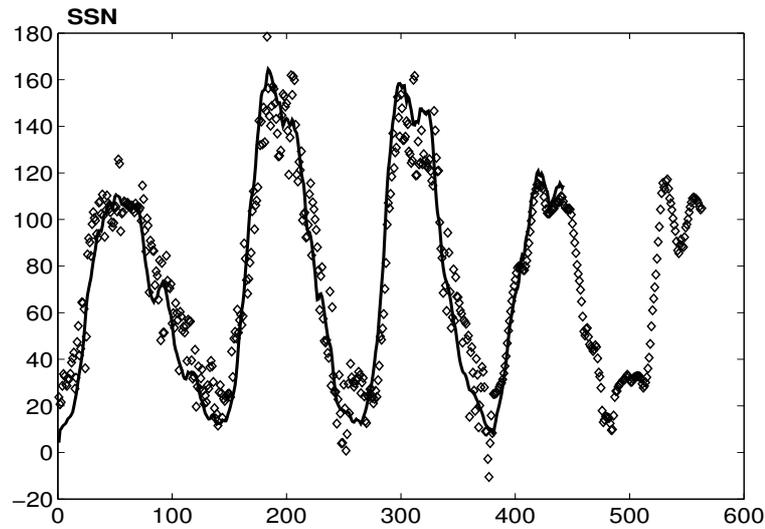


Рис. 12. Нейропрогноз временного ряда ежемесячных чисел Вольфа (SSN) с цикла № 20. Сплошная линия — реальные значения, ромбики — прогноз.

**Важное замечание.** Размерность вложения не слишком влияет на качество нейропрогноза: теорема Такенса требует лишь выполнения условия  $m > 2d$ . Правильный выбор лага  $\tau$  очень существен: большое  $\tau$  позволяет получить долгосрочный прогноз «за один шаг» — на рис. 12 прогноз был сделан сразу на 129 точек. Это значительно уменьшает ошибку, которая растет с числом итераций в «одношаговом» варианте. Обычно выбор лага связан с характерным масштабом времени динамической системы, например с ее периодом. Проблемы возникают, когда в системе два или более таких масштабов. Лаг, связанный с одним из них, не позволяет получить оптимальное вложение, отслеживающее динамику на неучтенных масштабах. Выход был предложен А. Меес и др. [53] в форме так называемого *неоднородного вложения*. Идея заключается в замене запаздывающих векторов однородного вложения:

$$\mathbf{z}(t) = (x(t), x(t - \tau), \dots, x(t - (m - 1)\tau)) \in R^m$$

на векторы:

$$\hat{\mathbf{z}}(t) = (x(t), x(t - l_1), x(t - l_2), \dots, x(t - l_{m-1})),$$

где лаги  $(l_1, \dots, l_{m-1})$  — различные числа. Любая схема прогноза основана на минимизации функционала ошибки предсказания. Его нельзя использовать дважды: для нахождения модели (аппроксимирующей функции) и параметров вложения. Для получения второго, дополнительного функционала авторы [53] предложили использовать *принцип минимальной длины описания*, формализующий средневековое правило *Бритвы Оккама* <sup>(8)</sup>. Идея заключается в следующем. Предположим, что вам надо передать кому-либо экспериментальный временной ряд, измеренный с некоторой точностью. Вы можете послать исходные данные и в этом случае длина описания данных будет равна количеству бит, необходимых для бинарного кодирования данных. Однако можно поступить иначе: построить динамическую модель для данных, позволяющую прогнозировать ряд. Если вы договорились о классе модели, то вам достаточно передать ее параметры, начальное значение и допустимую ошибку предсказания. Тогда, полная длина описания будет состоять из описания самой модели (или данных, кодированных с помощью модели), начального значения и ошибки. Чем сложнее модель, тем длиннее ее описание, но тем короче оставшаяся часть сообщения. Необходимый нам функционал содержит как раз эти два слагаемых:

$$\begin{aligned} (\text{Длина описания}) &\approx (\text{число данных}) \times \\ &\times \log(\text{ошибка предсказания}) + \\ &+(\text{штраф за число и точность параметров модели}). \end{aligned}$$

Когда число параметров модели увеличивается, ошибка предсказания уменьшается: согласно *МДО-принципу*, оптимальная модель доставляет минимум этому функционалу в заданном классе моделей. Именно этот функционал и позволяет вычислить набор лагов для неоднородного вложения <sup>(9)</sup>. Наши эксперименты показали, что такое вложение в большинстве случаев существенно улучшает нейропрогноз.

#### Примечания

1. Если отказаться от условия непрерывности, получится тривиальный результат, связанный в сущности с многомерным обобщением известного примера *Кантора*, который показал, как построить отображение единичного квадрата в единичный отрезок: эти два множества *равномощны*.

2. Гладкость функции измеряется ее *гельдеровским показателем*: говорят, что функция  $f(x)$  имеет локальный *показатель Гельдера*  $\alpha \in [0, 1]$  в точке  $x$ , если для достаточно малого числа  $h > 0$  существует такая константа  $C$ , что  $|f(x+h) - f(x)| \leq Ch^\alpha$ . Значению  $\alpha = 1$  соответствует существование одной производной. Пусть функция  $f(x) \in C^r$  задана в единичном  $n$ -мерном кубе. Если все ее частные производные порядка  $r$  имеют показатель Гельдера  $\alpha$ , то *сложность* функции можно измерить величиной [45]  $\chi^{-1} = n/(r + \alpha)$ . Таким образом, сложность функции прямо пропорциональна числу переменных и обратно пропорциональна суммарной гладкости функции и ее производных. Известно, что *не все* функции с показателем  $\chi_0$  могут быть представлены суперпозицией функций с характеристикой  $\chi > \chi_0$ . В частности, существуют  $C^r$ -функции  $n$  переменных, которые нельзя представить с помощью суперпозиции  $C^r$ -функций от *менее* чем  $n$  переменных. Таким образом, теорема Колмогорова — просто патологический случай!
3. Функция активации  $\sigma$  называется *сигмоидальной*, так как форма тех функций, которые выбираются на практике, обычно напоминают по форме деформированную букву  $S$ . Примером может служить широко распространенная *логистическая функция*  $y = (1 + \exp(-x))^{-1}$ . Она имеет два преимущества. Во-первых, ее производная  $dy/dx$  очень просто выражается через  $y$ , что существенно упрощает численную минимизацию ошибки при коррекции весов. Во-вторых, она естественным образом применима в задачах *байесовской классификации* [57]. Формально говоря, сигмоидальными называют функции, удовлетворяющие двум условиям:  $\lim_{t \rightarrow \infty} \sigma(t) = 1$  и  $\lim_{t \rightarrow -\infty} \sigma(t) = 0$ .
4. Недавно появилась работа [56], в которой предлагается нейронная сеть, согласованная с традициями теории аппроксимации.
5. Иными словами, отображение  $S : \{0, 1\}^N \rightarrow \{0, 1\}$  может быть записано как *дизъюнкция конъюнкций*, что в переводе на язык пороговых элементов и есть приведенное в тексте выражение.
6. Существенное множество должно быть замкнутым. Этому условию удовлетворяют радиальные базисные функции, но не сигмоидные [46].

7. При этом полагают, что функция  $\Phi$  универсальна, т. е. справедлива для всей истории временного ряда. Таким образом, предполагается стационарность ряда. Однако ряды, с которыми приходится иметь дело, обычно нарушают это условие. В этом случае полагают, что отображение  $\Phi(\mathbf{z}_i, \mathbf{c})$  зависит от вектора скрытых или бифуркационных параметров  $\mathbf{c}$ : их изменение и управляет перестройкой динамики. Для того чтобы учесть влияние  $\mathbf{c}$ , разбивают ряд на фрагменты с приблизительно одинаковым поведением, полагая, что на каждом таком участке значение параметра фиксировано. После эпингоза каждого фрагмента полученную ошибку используют для настройки корректора, исправляющего расхождение. Его используют затем для коррекции истинного долгосрочного прогноза [52].
8. Уильям Оккам (*William Occam*, 1285–1349) английский монах-францисканец, философ, теолог-логик, основатель одной из форм номинализма. Логику, еще мальчиком изучал во францисканском монастыре, теологию — в Оксфордском университете. Оккам настаивал на строго рациональных оценках, различии между необходимым и случайным, очевидностью и степенью вероятности. Принцип, известный сейчас под наименованием бритвы Оккама, является вариантом средневекового правила «не следует допускать множественность без необходимости». Оккам использовал его для исключения многих сущностей, которые использовались для объяснения реальности схоластами. В оригинале принцип звучит так: “*Pluralitas non est ponenda sine necessitate*”, т. е. «Не следует умножать сущности без необходимости».
9. Кроме того, МДО-принцип можно использовать еще для нахождения параметров радиальных базисных функций и оптимального числа нейронов сети [56].

**Путеводитель по литературе.** С теоремой Колмогорова и ее историей можно познакомиться по работе [47]. Эта теорема и нейросетевое обобщение теоремы Стоуна приведены в статьях А. Н. Горбаня [48–50]. Методы нелинейного прогноза обсуждаются в книгах [30, 43] и в сборнике «Пределы предсказуемости» (см. ссылку [58]). Много полезных замечаний об ошибках предсказания можно найти в замечательном обзоре [51]. Коррекция прогноза была предложена в статье [52], а неоднородное вложение —

в работе [53]. Принцип минимальной длины описания обсуждается в лекциях С.А.Шумского [54] и Риссанена [55]. Его применение к оптимизации нейронной сети предложено в [56]. Байесовы аргументы в пользу логистической функции обсуждаются в препринте [57].

### Пределы предсказуемости: хаотическая динамика и эффект Эдипа в мягких системах

Что было, то и будет; и что  
делалось, то и будет делаться, и  
нет ничего нового под солнцем.

*Книга Екклесиаста  
«Гл. 1 ст. 9»*

Мы начнем с простого, но поучительного примера прогноза хаотической системы. Фазовым пространством здесь является единичный отрезок  $M = [0, 1]$ , а динамика определяется дискретным кусочно-линейным отображением  $x(n+1) = 2x(n) \pmod{1}$ . Таким образом, каждая последующая координата фазовой точки получается удвоением предыдущего значения. Если полученное число больше 1, то эта единица отбрасывается — это и означает символ  $\pmod{1}$  <sup>(1)</sup>. Легко найти общее решение  $x(n) = 2^n x(0)$ . Воспользуемся тем, что каждое число из интервала  $[0, 1]$  можно записать в форме двоичной дроби:  $x(0) = 0, a_1 a_2 a_3 \dots a_n \dots$ , где  $a_i$  — принимают значения 0 либо 1 <sup>(2)</sup>. Умножение на 2 в двоичной арифметике эквивалентно переносу запятой вправо на одну позицию и отбрасыванию единицы  $\pmod{1}$ . Поэтому такое отображение часто называют *сдвигом Бернулли*. Его график приведен на рис. 13.

Рассмотрим некоторое начальное значение, например 0.10111001, заданное с точностью до восьми знаков. Первая цифра после запятой (1) означает, что начальная точка находится в правой половине единичного интервала ( $0,5 < x(0) < 1$ ); вторая цифра (0) означает, что начальная точка находится в интервале  $0,5 < x(0) < 0,75$ ; и т. д. Следовательно, каждая итерация (а это вариант интегрирования для уравнения!) уточняет «адрес» начальной точки. Выпишем первые три шага итерации для некоторого на-

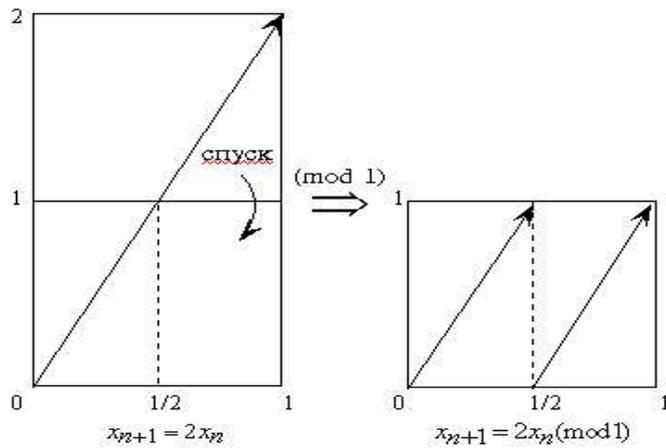


Рис. 13. Сдвиг Бернулли

чального значения:

$$\begin{aligned} x(0) &= 0.10111001, \\ x(1) &= 0.0111001, \\ x(2) &= 0.111001, \\ x(3) &= 0.11001. \end{aligned}$$

Мы видим, что каждый последующий шаг динамики не что иное, как *уточнение начального условия!* Предположим теперь, что точность начальных данных ограничена — в нашем примере восемью разрядами. Иными словами, мы не знаем, какой символ стоит в девятом разряде: 0 или 1. Тогда через восемь последовательных итераций эта девятая цифра окажется сразу после запятой, и мы не будем знать, справа или слева от середины интервала окажется фазовая точка. Самое удивительное в том, что наша простая модель не содержит каких либо случайных членов — она полностью детерминирована! Итак, чему же учит сдвиг Бернулли?

- Любой прогноз сводится к уточнению начальных данных, точность которых всегда ограничена!
- Некоторые детерминированные системы имеют ограниченный горизонт предсказуемости<sup>(3)</sup>. Такие системы называют *хаотическими*.

Рис. 13 подсказывает причину хаотичности. Две  $\varepsilon$ -близкие начальные точки на горизонтальной оси уже после первой итерации окажутся разделенными расстоянием  $2\varepsilon$  (так как коэффициент наклона прямых на рис. 13 равен двум), после второй —  $2^2\varepsilon$  и т. д. Через конечное число итераций расстояние между ними достигнет размеров фазового пространства ( $= 1$ ). Число необходимых для этого итераций и определяют границу прогноза. Для измерения скорости «разбегания» близких траекторий в общем случае используют так называемые *показатели Ляпунова*. Пусть отображение задано в виде:  $x(n+1) = f(x(n))$ ,  $M = [0, 1]$ . Тогда две близкие точки  $x(0)$  и  $(x(0) + \varepsilon)$  после  $N$  итераций окажутся на расстоянии:

$$\varepsilon e^{N\lambda(x_0)} \approx |f^N(x_0 + \varepsilon) - f^N(x_0)|$$

друг от друга. В пределе  $\varepsilon \rightarrow 0$ ,  $N \rightarrow \infty$  получим:

$$\begin{aligned} \lambda(x_0) &= \lim_{N \rightarrow \infty} \lim_{\varepsilon \rightarrow 0} \frac{1}{N} \ln \left| \frac{f^N(x_0 + \varepsilon) - f^N(x_0)}{\varepsilon} \right| = \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \ln \left| \frac{df^N(x_0)}{dx_0} \right|, \end{aligned}$$

где  $\lambda(x_0)$  — коэффициент растяжения (или сжатия) расстояния между двумя точками за одну итерацию. Его называют *локальным ляпуновским показателем* <sup>(4)</sup>.

Используя это понятие, нетрудно оценить среднее время предсказуемости  $T_m$  нашего отображения. Очевидно, наша способность предсказывать кончается, когда расстояние между близкими точками достигнет размеров фазового пространства  $M$  — этот размер у нас равен 1, т. е.  $\varepsilon \times \exp(\lambda T_m) \sim 1$ . Отсюда следует, что  $T_m \propto (1/\lambda) \ln(\varepsilon^{-1})$  <sup>(5)</sup>. Эту величину часто называют *горизонтом предсказуемости*. Поскольку точность начальных данных всегда конечна, поведение систем с хаотической динамикой всегда *ограниченно предсказуемо*. Следует заметить, что не всегда следует полагаться на оценки ляпуновских показателей. Для реконструкций, построенных из временных рядов со значительной шумовой компонентой, невозможно получить даже оценку самого простого максимального положительного показателя. В этом случае более полезны оценки горизонта предсказуемости, основанные на эвристических соображениях.

В заключение упомянем еще об одном интересном классе систем, для которых *будущее предвидимо, но не предсказуемо!* Это так называемые *мягкие системы*, существенным элементом которых является человеческий

фактор. Такие системы приходится рассматривать, например, в социологии, демографии и экономике. Представим себе такую ситуацию: в некотором условном государстве футуролог, астролог или эмбедолог предсказывает государственный переворот, инициированный военными. Правительство, получив эту информацию, тут же использует ее. Например, подвергает аресту всех заговорщиков. Сценарий предсказанного Будущего тут же меняется! Или, в другой ситуации, узнав о предсказанном вам благоприятном событии, вы «самоосуществили» его, предприняв ряд шагов в нужном направлении, вещь, которую Вы никогда бы не сделали, не зная предсказания! <sup>(6)</sup>. Обобщение подобных примеров резюмировал *И. В. Бестужев-Лада* в форме следующего постулата [58]:

**Постулат 1.** *О принципиальной невозможности безусловного предсказания будущих состояний объектов, поддающихся видоизменению действиями на основе решений, принимаемых с учетом подобного прогноза. Изменения сценария Будущей Реальности действиями, учитывающими предсказание этой Реальности в мягких системах, называют эффектом Эдина <sup>(7)</sup>. Казалось бы, что предсказание поведения таких систем заведомо обречено на провал. Однако неумение или нежелание управлять социальными или экономическими явлениями, которые в принципе поддаются управлению, позволяет считать их квазистихийными или квазиестественными, а следовательно и применять к ним методы прогноза хаотических природных систем. В этом состоит суть Постулата 2 И. В. Бестужева-Лады [58]. Оба постулата кажутся весьма правдоподобными, но, к сожалению, чисто эвристическими. Поэтому следует быть очень осмотрительным, применяя методы эмбедологии и прогноза к мягким системам — полученный результат вполне может оказаться квазистихийным. В прогностике особенно полезен известный совет Козьмы Пруtkова: «Имея в виду какое-либо предприятие, помысли сперва, точно ли оно тебе удастся».*

#### Примечания

1. Арифметика вычетов по модулю единица используется для того, чтобы избежать потери фазовых точек при эволюции системы. Так, например, точка с координатой  $1, 3$  не покидает  $M$ , так как  $1, 3 = 0, 3$  по модулю единица.
2. Для того чтобы обеспечить единственность разложения, условимся выбирать разложение кончающееся нулями для дробей, имеющих два двоичных представления.

3. Этот эффект связан не с видом уравнения, а с типом нелинейности в них.
4. Для сдвига Бернулли  $\lambda(x) = \ln 2$ . Для многомерных систем каждой пространственной координате соответствует свой показатель. Знак  $\lambda(x) > 0$  означает разбегание траекторий,  $\lambda(x) < 0$  — сжатие. Аттрактору — неподвижной точке в  $R^3$ , соответствуют, например, три отрицательных показателя  $(-, -, -)$ , предельному циклу  $(0, -, -)$ .
5. Величина  $\varepsilon$ , которая задает точность начального состояния, влияет на горизонт предсказуемости лишь логарифмически. Для многомерной системы, с размерностью аттрактора  $d$  и максимальным *положительным* ляпуновским показателем  $\lambda_m^+$ , существует более точная формула для оценки горизонта [29,39]:

$$T_m = \frac{\Delta I}{d\lambda_m^+},$$

где  $\Delta I$  — неопределенность начальных данных. Если мы имеем дело с реконструкцией аттрактора из временного ряда, содержащего  $N$  равновероятных отсчетов, то  $\Delta I = \log_2 N$ .

6. Вообще говоря, трудно представить себе наше существование, если бы, например, каждый человек точно знал все этапы своей будущей реальности. Теоретик легального марксизма, русский экономист и философ *Петр Бернгардович Струве* (1870–1944) считал, например, прогноз в социальной области вещь весьма безнравственной: *«Я считаю чудовищной идею научной этики — разве мыслимо такое, чтобы то, что есть, могло обосновывать то, что должно быть»*.
7. *Эдип*, в греческой мифологии — сын фиванского царя *Лая* и его жены *Иокасты*. Лаю была предсказана *Аполлоном* смерть от собственного сына, поэтому новорожденного Эдипа с проколотыми булавкой сухожилиями — имя *Эдип* — означает *с опухшими ногами* — отнесли в горы. Его усыновил бездетный царь Коринфа — *Полиб*. Юношей Эдип узнал от дельфийского оракула, что ему предстоит убить своего отца и жениться на матери. Считая настоящим отцом Полиба, Эдип, пытаясь изменить предсказание, ушел из Коринфа. Путешествуя, он встретил и убил в ссоре мужчину, который и был его родным отцом.

Не зная этого, Эдип пришел в Фивы и избавил город от чудовища Сфинкс. Благодарные жители сделали его царем и женили на вдовой Иокасте. Конец истории о человеке, пытавшемся изменить предсказанную судьбу очень грустный. Узнав через некоторое время всю правду, Иокаста повесилась на своем поясе, Эдип — выколол себе глаза пряжкой этого пояса и был изгнан из Фив.

**Путеводитель по литературе.** *Сдвиг Бернулли* и другие дискретные преобразования, обладающие свойством детерминированного хаоса, обсуждаются в [23, 32–33]. Свойства ляпуновских показателей описаны в [26–30, 33]. На web-страничке [53] можно найти много полезных статей, связанных с пределами предсказуемости. Наконец, сборник [58] содержит хорошие обзоры на русском языке по нелинейному прогнозу.

### Литература

1. Фор Р., Кофман А., Дени-Папен М. Современная математика. – М.: Мир, 1966. – 271 с.
2. Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа. – М.: Наука, 1989. – 496 с.
3. Келли Дж. Л. Общая топология. – М.: Наука, 1968. – 432 с.
4. Стинрод Н., Чинн У. Первые понятия топологии. – М.: Мир, 1967. – 224 с.
5. Прасолов В. В. Наглядная топология. – М.: МЦНМО, 1995. – 111 с.
6. Хириш М. Дифференциальная топология. – М.: Мир, 1979. – 280 с.
7. Милнор Дж., Уоллес А. Дифференциальная топология. – Платон, 1997. – 277 с.
8. Мищенко А. С., Фоменко А. Т. Дифференциальная геометрия и топология. – М.: МГУ, 1980. – 439 с.
9. Макаренко Н. Г. Многообразия, вложения, погружения и трансверсальность // В сб. *Проблемы Солнечной активности*. – Ленинград: ФТИ, 1991. – с. 13–27.
10. *Chillingworth D.* Differential topology with a view to applications. – Pitman Press, 1976. – 291 pp.
11. *Guillemin V., Pollak A.* Differential topology. – Prentice-Hall, Inc, Englewood Cliffs, New Jersey, 1974. – 219 pp.
12. *Рихтмайер Р.* Принципы современной математической физики. – М.: Мир, 1984 – т.2 – 486 с.

13. *Ильяшенко Ю. С., Лу В.* Нелокальные бифуркации. – М.: МЦНМО-ЧеРо, 1999. – 415 с.
14. *Hunt B. R., Sauer T., Yorke J. A.* Prevalence: A translation-invariant “almost every” on infinite-dimensional spaces // *Bull. Amer. Math. Soc., (New Series)*. – 1992. – v. 27 – pp. 217–238.
15. *Ott E., Sauer T., Yorke J. A.* Coping with chaos: Analysis of chaotic data and the exploitation of chaotic systems. – John Wiley and Sons, 1994. – 432 pp.
16. *Sauer T., Yorke J. A., Casdagli M.* Embedology // *J. Statist. Phys.*– 1991. – v. 65. – pp. 579–616.
17. *Packard N. H., Crutchfield J. P., Farmer J. D., Shaw R. S.* Geometry from a time series // *Phys. Rev. Lett.* – 1980. – v. 45. – pp. 712–716.
18. *Takens F.* Nonlinear dynamics and turbulence // Ed. *Barenblatt G. J., Jooss G., Joseph D. D.* – N. Y. Pitman. – 1983. – pp. 314–333.
19. *Афраймович В. С., Рейман А. М.* Размерности и энтропии в многомерных системах // *Нелинейные волны. Динамика и эволюция*. – М.: Наука, 1989. – с. 238–262.
20. *Takens F.* Detecting strange attractors in turbulence // *Lecture Notes in Math.* – 1981. – v. 898. – pp. 366–381.
21. *Noakes L.* The Takens embedding Theorem // *Inter. J. Bifurcation and Chaos.* – 1991. – v. 1. – pp. 867–872.
22. *Rapp P. E., Schah T. I., Mees A. I.* Models of knowing and the investigation of dynamical systems // *Physica D.* – 1999. – v. 132. – pp. 133–149.
23. *Кроновер П. М.* Фракталы и хаос в динамических системах. – М.: ПОСТМАРКЕТ, 2000. – 350 с.
24. *Макаренко Н. Г.* Фракталы, аттракторы, нейронные сети и все такое // *Лекции по нейроинформатике* – М.: МИФИ, 2002. – ч. 2. – с. 121–169.
25. *Gilmore R.* Topological analysis of chaotic dynamical systems // *Rev. Mod. Phys.* – 1998. – v. 70 – pp. 1456–1529.
26. *Ruelle D.* Chaotic evolution and strange attractors. The statistical analysis of time series for deterministic nonlinear systems. – Cambridge Univ. Press., 1989. – 96 p.
27. *Abarbanel H. D. I., Rabinovich M. I., Sushchik M. M.* Introduction to nonlinear dynamics for physicists. – World Scientific. – 1993. – v. 53 – 158 p.
28. *Eckmann J. P., Ruelle D.* Ergodic theory of chaos and strange attractors // *Rev. Mod. Phys.* – 1985. – v. 57. – pp. 617–656.
29. *Schaw R.* Strange attractors, chaotic behavior, and information flow // *Z. Naturforsch.* – 1981. – v. 36a. – pp. 80–112.

30. *Малинецкий Г. Г., Потанов А. Б.* Современные проблемы нелинейной динамики. – М.: Эдиториал УРСС, 2000. – 335 с.
31. *Малинецкий Г. Г.* Хаос. Структуры. Вычислительный эксперимент: Введение в нелинейную динамику. – М.: Эдиториал УРСС, 2002. – 254 с.
32. *Данилов Ю. А.* Лекции по нелинейной динамике – М.: ПОСТМАРКЕТ, 2001. – 184 с.
33. *Шустер Г.* Детерминированный хаос – М.: Мир, 1988. – 240 с.
34. *Milnor J.* On the concept of attractor // *Commun. Math. Phys.*– 1985. – v. 99. – pp. 177–195.
35. *Schreiber T.* Interdisciplinary application of nonlinear time series methods // *Phys. Rep.* – 1999. – v. 308. – No. 2  
URL: <http://xyz.lanl.gov/chao-dyn/980700>
36. *Parlitz U.* Nonlinear time series analysis  
URL: <http://WWWDP1.Physik.Uni-Goettingen.DE/~ulli/pub95.html>
37. *TISEAN* Nonlinear time series analysis  
URL: <http://WWW.mpipks-dresden.mpg.de/tisean/>
38. *TSTOOL* Software package for nonlinear time series analysis  
URL: <http://WWW.physik3.gwdg.de/tstool/>
39. *Тутубалин В. Н.* Теория вероятности и случайные процессы. – М.: МГУ, 1992. – с. 342–353.
40. *Марпл-мл. С. Л.* Цифровой спектральный анализ и его приложения. – М.: Мир, 1990. – 584 с.
41. *Serio C.* Towards long-term prediction // *Il Nuovo Cimento.* – 1992. – v. 107B. – pp. 681–701.
42. *Farmer J. D., Sidorovich J. J.* Predicting chaotic time series // *Phys. Rev. Lett.* – 1987. – v. 59. – pp. 845–848.
43. *Малинецкий Г. Г.* Синергетика, предсказуемость и детерминированный хаос: Пределы предсказуемости. – М.: ЦентрКом, 1997. – 247 с.
44. *Малинецкий Г. Г., Курдюмов С. П.* Нелинейная динамика и проблемы прогноза // *Вестник Российской Академии наук.* – 2001. – т. 71, № 3. – с. 210–232.
45. *Poggio T., Girosi F.* A Theory of networks for approximation and learning // *MIT AI Lab. Technical Report.* – 1989. – Memo No 1140. – Paper No 31.  
URL: <http://citeseer.nj.nec.com/poggio89theory.html>
46. *Girosi F., Poggio T.* Networks and the best approximation property // *MIT AI Lab. Technical Report.* – 1989. – Memo No 1164. – Paper No 45.  
URL: <http://www.ai.mit.edu/people/poggio>

47. Арнольд В. И. О представлении функций нескольких переменных суперпозицией функций меньшего числа переменных // *Математическое просвещение*. – 1958. – вып. 3. – с. 41–61.
48. Горбань А. Н. Возможности нейронных сетей // В сб. *Нейроинформатика*. – Новосибирск: Наука, 1998. – с. 18–45.
49. Горбань А. Н. Обобщенная аппроксимационная теорема и вычислительные возможности нейронной сети // *Сибирский журнал вычислит. математики*. – 1998. – т. 1. – с. 11–24.
50. Gorban A. N. Approximation of continuous functions of several variables by an arbitrary nonlinear continuous function of one variable, linear functions, and their superpositions // *Appl. Math. Lett.* – 1998. – v. 11. – pp. 45–49.
51. Smith L. A. The Maintenance of uncertainty  
URL: <http://www.maths.ox.ac.uk/~lenny/papers.html>
52. Judd K., Small M. Towards long-term prediction // *Physica D*. – 2000. – v. 136. – pp. 31–44.
53. Judd K., Small M., Mees A. Achieving good nonlinear models: Keep it simple, vary the embedding, and get the dynamics right // In: *Nonlinear Dynamics and Statistics*, edited by A. I. Mees. – 2001. – Birkhäuser, Boston. – pp. 65–80.
54. Шумский С. А. Байесова регуляризация обучения // *Лекции по нейроинформатике*. – М.: МИФИ, 2002. – ч. 2. – с. 30–93.
55. Rissanen J. Lectures on statistical modelling theory  
URL: <http://www.cs.tut.fi/~rissanen>
56. Small M., Tse C. K. Minimum description length neural networks for time series prediction  
URL: <http://www.eie.polyu.edu.hk/en-small/pubs.html>
57. Jordan M. I. Why the logistic function? A tutorial discussion on probabilities and neural networks  
URL: <ftp://psyche.mit.edu/pub/jordan/uai.ps>
58. Бестужев-Лада И. В. Будущее предвидимо, но не предсказуемо: Эффект Эдипа в социальном прогнозировании // В сб.: *Пределы предсказуемости*. – М.: ЦентрКом, 1997. – 247 с.
59. Вигнер Е. Этюды о симметрии. – М.: Мир, 1971. – 320 с.

**Дополнение 1**

**Р:** Мне не очень понятен термин «нелинейная проекция». Я знаю про центральную проекцию, параллельную проекцию, ортогональную проекцию, но все они порождаются проектирующими лучами-прямыми, а откуда же берется нелинейность? Это что — некая «искажающая» проекция? Если да, то как и почему искажающая?

**А:** Да, вы правы! Это *искажающая* проекция. Только в обобщенном смысле. Например, ортогональная проекция  $R^3 \rightarrow R^2$  делается просто: из тройки координат  $x_1, x_2, x_3$  удаляем, скажем  $x_3$  — вот вам и проекция на плоскость. Временной ряд считают *нелинейной проекцией* фазовой траектории динамической системы  $f^k(x) : M \rightarrow M$ , потому что последовательность скалярных отсчетов  $h_n, n = 1, 2, \dots$  порождается наблюдаемой функцией, которая и играет роль «проектора»  $h : M \rightarrow R$  как:

$$h_1 = h(x), h_2 = h(f(x)), h_3 = h(f^2(x)), \dots, h_m = h(f^{m-1}(x)) \dots$$

В общем случае  $h$  — нелинейная функция. Конечно она «искажает»! Впрочем, в простейшем варианте это могут быть искажения, вызванные возмущениями. Ну например, качается физический маятник, а мы регистрируем луч лазера отраженный от его поверхности. Поставим на пути отраженного луча аквариум, пусть в нем плавает карась и «баламутит» воду. Принимаемый сигнал искажается. Совсем другой пример — нас интересует теория внутреннего строения звезд. В простейшем варианте светимость звезды зависит от массы и радиуса. Именно эти величины входят, например, в уравнения. Но мы наблюдаем всего лишь блеск (светимость) как функцию времени. Конечно она зависит от упомянутых величин, но с помощью нелинейной функции или даже суперпозиции таких функций. Вот это-то и есть нелинейная проекция! Важно чтобы функция  $h$  была достаточно гладкой. Поскольку она зависит от времени сложным образом  $h(x(t))$ , обычно требуют *лишниц-непрерывности* относительно  $x$  [19] и (или) существования двух производных по  $t$ .

<sup>12</sup>Между редактором (Р) лекций Школы-семинара и автором (А) данной лекции несколько раз возникла дискуссия, касавшаяся принципиальных моментов, связанных с темой лекции. Отдельные фрагменты этой дискуссии, позволяющие ярче высветить упомянутые моменты, приводятся здесь в качестве дополнений к лекции.

**Дополнение 2**

**Р:** Мне тоже нравится приведенная *платоновская аллегория* — она хорошо иллюстрирует процесс восприятия окружающего мира через наблюдение за ним. Но ведь если прочитать соответствующий фрагмент «Диалогов» целиком<sup>13</sup>, то дальше-то там выводы, как Вы помните, не столь благостные, как можно заключить из примечания 3 на странице 88. Когда узника выводят, наконец, «на свет божий» и он видит то, что раньше наблюдал только через тени, то ему становится ясно, насколько же беден был весь его предыдущий «теневой» опыт!

**А:** Ну конечно, именно это и имелось в виду (см. выше Дополнение 1 о нелинейной проекции): бедность теневого опыта — это восстановление образа по тени «с точностью до». В случае эмбедологии — с точностью до топологии. Круг узник распознает как овал или эллипс, но он не должен увидеть восьмерку!

**Р:** А если имеет дело нелинейная проекция, как чаще всего и будет? Пусть, например, перед пещерой стоит тот же самый аквариум с карасем и свет в пещеру проходит только через него, через этот аквариум. Будут ли здесь основания считать, что мы сможем «правильно» реконструировать действительность по ее теням? Возможны ли такие нелинейные преобразования, при которых нельзя будет говорить даже и о реконструкции с точностью до топологии? Другими словами, *где пределы применимости методов эмбедологии?*

**А:** Эти пределы конечно, есть. Именно их и определяет *Кредо идеального экспериментатора* (см. с. 115). Ну например, если тень дает не динамическая система с аттрактором. Или если вы не выполнили условие вложения  $2d + 1$ . Оно, правда, часто избыточно.

Например, тень от листа бумаги — двумерная. Но если лист повернуть ребром, то тень одномерная. Однако ситуаций, когда лист оказывается ребром к экрану, меньше, чем всех остальных его положений.

Бублик обычно дает тень с дыркой. Его можно повернуть так, что получится прямоугольник — тоже ребром. Но опять же — таких ситуаций мало.

Теперь про то, когда не получается с точностью до топологии. Нелинейная проекция (наши наблюдения) может не быть непрерывной, точнее *липшиц-непрерывной* проекцией. Ну, например, это функция типа  $1/\log x$ . Если доопределить ее в нуле условием  $\log 0 = 0$ , она будет непрерывной, но

<sup>13</sup>См., например, с. 296–299 в книге: *Платон. Собрание сочинений* в 4-х томах. Том 3. — М.: Мысль, 1994. — (Серия «Философское наследие»).

условие Липшица нарушится. Или это фрактальная функция типа триады Коха. Тогда ничего не получится.

Хороший пример нелинейной проекции, когда все получается. В комнате по сложной траектории летает муха. Типичная муха. Вы записываете хорошим микрофоном звук ее жужжания. Вот это ваш скалярный ряд или «детерминированно порожденная наблюдаемая», как сказал бы Такенс. Если использовать эмбедологию, *можно восстановить* ее пространственную траекторию — с точностью до топологии, вложив фонограмму полета в  $R^m$ .

**Р:** «... восстановить пространственную траекторию “с точностью до топологии”». Что это значит? Платоновский узник вместо круга может увидеть эллипс, т. е. «несколько не то», что есть «на самом деле». На примере с мухой — чем же придется заплатить за «услуги» эмбедологии?

**А:** Предположим, что муха описывает в комнате пространственную спираль. Если пробная размерность вложения выбрана неверно, скажем  $R^2$ , мы получим наложение нескольких окружностей. Можно проверить, все ли в порядке, т. е. имеем ли мы дело с вложением. Для этого достаточно выбрать на полученной копии три точки и проверить их взаимное расположение для реконструкции в  $R^3$ . Если расстояния между центральной точкой и ее соседями скачком увеличиваются (ложные соседи), то, следовательно, размерность была выбрана неверно. Аналогичные эксперименты могли бы делать платоновские узники, проверяя справедливость своих реконструкций. Совсем другое дело «Кредо». Это символ веры, т. е. *догма*. Его почти бессмысленно проверять. Конечно, хорошо было бы дополнить Лекцию некоторым списком *Предубеждений* против символов веры. Они более полезны для практики, чем *Ожидания*. Это, впрочем, относится к любой новой концепции. Мне давно хотелось это сделать: в конце концов всегда легко найти то, что ищешь... Но очень трудно увидеть место, где споткнешься.

### Дополнение 3

**Р:** Хотя бы пару слов надо сказать, что такое *топологическая идентификация*. «Просто» идентификация — вещь достаточно известная, но про топологическую идентификацию такого не скажешь.

**А:** Имеется в виду, что аттрактор оригинальной системы и аттрактор копии будут гомеоморфны — топологически неразличимы. Иными словами, копия и оригинал могут быть отображены друг на друга непрерывными обратимыми преобразованиями. Термин выделен в тексте лекции курсивом,

потому что цитируется из оригинальной статьи *Паккарда* и др. [16]. Кстати, в теореме Такенса, аттрактор и его копия *диффеоморфны*, т. е. могут быть отображены друг в друга обратимыми *дифференцируемыми* преобразованиями.

#### Дополнение 4

**Р:** Один из первых вопросов, возникающий при начальном знакомстве с *эмбедологией* — о месте данного направления среди уже существующих. Чем, кроме используемого аппарата, рассматриваемый класс задач отличается от традиционной задачи *идентификации*, понимаемой достаточно широко — как получение математической модели объекта по данным наблюдений за ним? Не будет ли введение нового термина тем самым «умножением сущностей» против которого выступал *У. Оккам* (см. с. 131)? Можно, конечно, и некую подобласть знаний снабдить именем собственным, но тогда неплохо бы указать и область, в которую эта подобласть входит.

**А:** Нет, не будет. Обычно, под идентификацией<sup>14</sup> понимают формирование модели путем анализа данных «вход–выход». Таким образом, идентификация предполагает модель типа «черного ящика». Эмбедология же предполагает построение модели (из тени — образ) вложением тени (временного ряда) в евклидово пространство. Еще одно отличие — полученная реконструкция всегда *многомерна*.

**Р:** Так куда же все-таки отнести эмбедологию? Я не пытаюсь ее «принизить», мне хотелось бы просто понять ее место среди других научных областей. Ваша аргументация на тему «эмбедология — это не идентификация» меня пока не убедила. Формулируя свой вопрос, я хотел сослаться на того же самого *Л. Льюнга*, да потом как-то забыл в спешке. А ведь он пишет буквально следующее: «Решение задачи построения математических моделей динамических систем по данным наблюдений за их поведением составляет предмет теории идентификации. . .» А в предисловии высказывается еще интереснее: «Идентификация систем представляет собой ту часть теории автоматического регулирования и анализа временных рядов. . .» и т. д. Чуть ли не дословно то же самое говорится в соответствующих главах известного справочника под ред. *А. А. Красовского* по теории автоматического управления<sup>15</sup>. И обратите внимание, нигде никаких упо-

<sup>14</sup>См., например: *Льюнг Л.* Идентификация систем: Теория для пользователя: Пер. с англ. под ред. *Я. З. Цыпкина*. — М.: Наука, 1991. — 432 с.

<sup>15</sup>Справочник по теории автоматического управления / Под ред. *А. А. Красовского*. — М.: Наука, 1987. — 712 с.

минаний о парах «вход-выход», вообще об управлениях, о «черном ящике» и т. п. Совсем уж «черный ящик» будет, если мы станем искать модель в виде нейросети, которая представляет собой, как известно, композицию «неизвестно чего», т. е. набор элементов и связей между ними, не имеющих «предметной» интерпретации.

Но ведь типичная ситуация-то как раз не такая. Возьмем в качестве примера такую область как механика полета. Типичная задача идентификации в ней выглядит так. Есть модель движения летательного аппарата в виде системы дифференциальных уравнений, полученных из законов механики (и никакого «черного ящика»!) В этой системе есть неизвестные параметры (функции, на самом деле) — коэффициенты аэродинамических сил и моментов, надо найти их значения. В летном эксперименте записывается поведение самолета (значения его фазовых переменных по времени) в ответ на некоторое начальное возмущение (обычно это ступенчатое отклонение соответствующего органа управления), и из этих данных «извлекаются» значения коэффициентов.

Или возьмем полет неуправляемой ракеты, по регистрации которого опять же надо восстановить модель движения — есть только выходы (а на роль входов могут претендовать разве что начальные условия движения), внешние воздействия (например, турбулентность атмосферы) мы ни контролировать, ни измерять не можем.

А довод о том, что получаемая в эмбедологии реконструкция всегда многомерна мне непонятен — точно так же можно при желании проинтерпретировать и «классику».

**А:** Отчасти я с вами согласен. Но не во всем. Во всех ваших примерах из наблюдений строится не истинная (в некотором обобщенном смысле) модель, а некоторая правдоподобная логическая схема. От таких моделей требуется лишь «функциональное» сходство с оригиналом. В эмбедологии требуется сходство решений модели и копии с точностью до диффеоморфизмов. Она ведь претендует на восстановление геометрии фазового пространства! Разумеется, при большом желании можно попытаться поместить Слона в чайник и подобрать для новой области подходящие рамки из теории идентификации. Но, быть может, более уместно найти для Слона более достойноеместилище? Можно поступить следующим образом. В математике есть такое старое понятие, как «число Коши», введенное, кажется, *А. Лебегом*, которое относится к способу определения некоторого понятия. Его (понятие) можно просто декларировать. Но можно, согласно

О. Коши, поступить иначе. А именно, привести рецепт получения данной величины, а затем рецепт принять за определение. Скажем определением классической окрошки может служить способ ее приготовления: «Возьмите огурец, вареный картофель, яйцо, колбасу и т. д. Мелко покрошите. Залейте квасом. То, что получится и называется окрошкой». Можно воспользоваться аналогичным приемом и попробовать перечислить, что же позволяет *делать* эмбедология.

1. Из временного ряда (тени) восстановить (с точностью до топологии) и символов веры («кредо») фазовый портрет (интегральные траектории) неизвестной системы, порождающей ряд.
2. Оценить (по копии) размерность аттрактора, т. е. число неизвестных уравнений первого порядка.
3. Оценить (по копии) стохастичность неизвестной системы (ляпуновские показатели и энтропию). Ряды, которые генерируются хаотическими системами, проходят *все* тесты на случайность. И только методы эмбедологии позволяют обнаружить там детерминизм.
4. Восстановить (и это делается, правда не всегда) сами уравнения.
5. Реализовать нелинейную схему прогноза.
6. Обнаружить и оценить взаимодействие двух систем и даже определить его направленность. И это в том случае, когда коэффициент линейной корреляции близок к нулю!

Не мало, не правда ли, если в руках только скалярный временной ряд? Правда еще и Символы веры — Кредо Идеального экспериментатора (см. с. 115). В самом примитивном варианте, эмбедологией можно назвать набор алгоритмов, позволяющий из «тени» системы построить ее диффеоморфную модель.

**Р:** Я же ведь не пытаюсь доказать, что эмбедология — слабая область с не нужными никому результатами. Не пытаюсь я и утверждать, что эмбедология есть ветвь *традиционной* идентификации. Но ведь если появляется некая новая область — это же не «кошка, гуляющая сама по себе», она имеет не только различия с другими научными областями, но и сходства! И мне-то представляется, что такое сходство у эмбедологии есть с идентификацией, если ее трактовать в духе того определения из Льюнга, что я цитировал. Ведь заметьте, там речи не идет о линейности-нелинейности, применяемых методах и прочем — это уже вещи вторичные, хотя и безусловно важные. И, на мой взгляд, *сходство* эмбедологии и теории идентификации — в решаемой задаче, а *различие* (коренное!) — в используемых подходах и методах.

Тем эмбедология, на мой взгляд, и интересна — взяли старую почтенную задачу (а ей лет уж столько, сколько и науке) и подошли к ней так, как никто пока не подходил, получив за счет этого качественный рывок.

**А:** Я уже почти согласен с тем, что между идентификацией и эмбедологией очень много общего. И из этой общности, наверное, еще не все извлечено. Мне кажется вопрос, куда отнести эмбедологию, это больше вопрос *семантики*. Истинный смысл любого определения следует искать не в его формальном определении, а употреблении. Ведь именно так Платон в «Диалогах» пытался найти истинное значение некоторых слов.

Существовала, кстати, точка зрения, известная в средние века как *номинализм* (ее активно защищал, в частности, У. Оккам, а в новейшее время — английский математик Г. Х. Харди (1877–1947)), согласно которой общие имена (т. е. имена, связанные не с конкретными предметами, а с их классами) являются чисто словесными, произвольными знаками, а не относятся к объективным сущностям. Великолепный пример номинализма содержится в прелестном диалоге *Алисы и Шалтай-Болтая* из известной книги Льюиса Кэрролла<sup>16</sup>:

— Когда я беру слово, оно означает то, что я хочу, не больше и не меньше — сказал Шалтай-Болтай презрительно.

— Вопрос в том, подчинится ли оно вам, — сказала Алиса.

— Вопрос в том, кто из нас здесь хозяин, — сказал Шалтай-Болтай. — Вот в чем вопрос!

**Р:** «... между идентификацией и эмбедологией очень много общего...» На этом обсуждение «взаимоотношений» идентификации и эмбедологии можно пока и завершить.

Действительно, между ними много общего в решаемых задачах, но не в применяемых методах!

В то же время, считать эмбедологию «всего лишь» разделом теории идентификации было бы, по-видимому, несправедливо.

В самом деле, обратимся опять к упоминавшейся книге Л. Льюнга (с. 15): «Формирование моделей на основе результатов наблюдений и исследование их свойств — вот, по существу, основное содержание науки. Модели (“гипотезы”, “законы природы”, “парадигмы” и т. п.) могут быть

---

<sup>16</sup> См. Кэрролл Л. Приключения Алисы в стране чудес. Сквозь зеркало и что там увидела Алиса, или Алиса в Зазеркалье: пер. с англ. Н. М. Демуровой, С. Я. Маршака, Д. Г. Орловской и О. А. Седаковой. — СПб.: Кристалл, 1999. — 431 с. — (Серия «Библиотека мировой литературы»).

более или менее формализованными, но все обладают той главной особенностью, что связывают наблюдения в некую общую картину». А далее идет цитата, уже приводившаяся выше.

Если это высказывание понимать буквально, то получим такую широкую трактовку теории идентификации, что практически все науки надо будет считать ее разделами! А это уже явный «перебор», из него видно, что расширяя «область влияния» той или иной научной области, надо и меру знать. . .

. . . **А:** Вообще-то еще академик *В. А. Фок* говорил, что основная задача науки «упорядочить наши впечатления».

**Николай Григорьевич МАКАРЕНКО**, в. н. с., к. ф.-м. н., руководитель группы в Лаборатории компьютерного моделирования (Институт математики, Алма-Ата, Казахстан). Область научных интересов: фрактальная геометрия, вычислительная топология, алгоритмическое моделирование, детерминированный хаос, нейронные сети, физика Солнца. Имеет более 50 научных публикаций.