Глава 10. Реконструкция уравнений: «черный ящик»

«Черный ящик» – это одновременно и наиболее трудная, и самая «соблазнительная» постановка задачи моделирования, когда априорная информация об объекте, а, следовательно, и о структуре модели отсутствует. Интрига состоит в том, что модель, способную воспроизвести наблюдаемое поведение или дать прогноз дальнейшей эволюции, нужно получить только из наблюдаемого ряда, т.е. практически «из ничего». Шансы на успех невелики, но в случае удачи «хорошая» модель становится очень ценным инструментом для характеристики объекта и понимания «механизмов» его функционирования: «почти блеф может привести к большому выигрышу». Отсутствие априорной информации вынуждает использовать универсальные структуры модельных уравнений, например, в качестве функций в правых частях используют искусственные нейронные сети, радиальные базисные функции, алгебраические многочлены и т.д. Такие модели зачастую оказываются многомерными и содержат очень много неизвестных параметров.

Учитывая, что временные ряды всех многочисленных переменных такой модели должны быть непосредственно измерены или получены по ряду наблюдаемой, этап «реконструкции» рядов недостающих переменных[1] приобретает в данной постановке исключительную важность. Его теоретическому обоснованию посвящены известные теоремы Такенса, которые упоминаются по необходимости и по традиции весьма часто (п. 10.1).

Не менее важен и труден в постановке «черного ящика» следующий этап – аппроксимация зависимости следующего значения вектора состояния от текущего (при построении модельных отображений $\mathbf{x}_{n+1} = \mathbf{f}(\mathbf{x}_n, \mathbf{c})$) или скорости изменения вектора состояния от самого вектора (при построении модельных дифференциальных уравнений $d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}, \mathbf{c})$). На практике обычно удается добиться успеха, если оказывается достаточным использование не очень большой размерности модели, грубо говоря, не более 5-6. Для построения моделей большей размерности требуются огромные объемы экспериментальных данных, а аппроксимация функций *многих* переменных намного сложнее приближения одномерных зависимостей (пп. 7.2, 9.1, 9.3). Причем трудности быстро нарастают с ростом размерности модели – это так называемое «проклятие размерности» [254], основное препятствие при моделировании множества реальных процессов.

---

[1] Говорят также о «реконструкции фазовой траектории» и «реконструкции векторов состояния».

Чтобы развеять пессимистические нотки предыдущих абзацев, отметим, что и в такой тяжелой постановке, как «черный ящик», тоже удается добиться успеха при моделировании реальных сложных объектов. Этому есть немало примеров, получен ряд красивых теоретических результатов и разработаны практические алгоритмы реконструкции, которые во многих случаях оказываются весьма эффективными для прогноза и достижения других целей моделирования.

10.1. Реконструкция фазовой траектории

В соответствии со сказанным в п. 6.1.2, при построении эмпирических моделей по временному ряду $\{\eta(t_1),\eta(t_2),...,\eta(t_N)\}$ можно использовать в качестве недостающих переменных последовательные значения наблюдаемой (вектор состояния модели $\mathbf{x}(t_i)=[\eta(t_i),\eta(t_i+\tau),...,\eta(t_i+(D-1)\tau)]$, где τ – время задержки), ее последовательные производные (вектор состояния $\mathbf{x}(t_i)=[\eta(t_i),d\eta(t_i)/dt,...,d^{D-1}\eta(t_i)/dt^{D-1}]$) и т.д. Эти подходы часто применялись с давних пор даже без специального обоснования. Так, первый из них используется с 1927 года [338] при построении широко известных авторегрессионных моделей (см. (4.12) в п. 4.4), где будущее значение наблюдаемой предсказывается по нескольким предыдущим. Это выглядит вполне разумно: если нет другой информации, кроме наблюдаемого ряда, то на что же еще опираться при прогнозе, как не на предыдущие значения наблюдаемой или какие-либо их комбинации?

В начале 1980-х годов была показана связь обоих этих эмпирических подходов с теорией динамических систем. Было доказано, что при реконструкции по скалярной временной реализации *динамической системы*[2] и метод временных задержек, и метод последовательных производных гарантируют, что в новых переменных будет получено эквивалентное[3] описание исходной динамики при достаточно большой размерности восстановленных векторов *D*. А именно, должно выполняться условие $D>2d$, где *d* – размерность множества M в фазовом пространстве исходной системы, на котором происходит моделируемое движение.[4] Эти

---

[2] При некоторых условиях гладкости, см. ниже п. 10.1.1.2.

[3] Точнее, диффеоморфное описание, п. 10.1.1.2.

[4] Множество M – компактное гладкое многообразие, а величина *d* – его топологическая размерность, см. п. 10.1.1.2. Есть обобщения теоремы на случай негладких множеств M и фрактальной размерности *d*, которые мы здесь не обсуждаем [299]. Заметим, что множество M, о котором идет речь в теоремах, *не обязательно* соответствует движению на аттракторе. Например, пусть исследуемая система имеет аттрактор – предельный цикл C, который «намотан» на тор M. Если нас интересует описание только установившегося периодического движения на цикле C, то, согласно теореме

утверждения составляют содержание знаменитых теорем Такенса [323], см. п. 10.1.1.

Подчеркнем, что теоремы относятся только к случаю, когда исходный объект – *динамическая система* (см. п. 2.2.1, с. 44). При моделировании реальных объектов использование упомянутых методов возможно и без упоминания о теоремах (как и делалось на протяжении не одного десятка лет), т.к. на практике нельзя проверить, выполняются ли условия теорем,[5] а размерность d неизвестна. Но значение теоретических результатов Такенса все же достаточно велико. Во-первых, после них стало ясно, что для достаточно широкого класса систем упомянутые методы с гарантией дают недостающие переменные, пригодные для построения динамической модели. Так что теоремы «благословляют» практическое применение методов, особенно если есть какие-либо соображения в поддержку того, что условия теорем для данной ситуации выполняются. Во-вторых, с опорой на идеи теории динамических систем впоследствии были развиты новые полезные подходы к подбору параметров алгоритма реконструкции – времени задержки τ (например, первый минимум функции взаимной информации), размерности модели D (например, метод ложных ближайших соседей) и т.д., см. п. 10.1.2.

10.1.1. Теоремы Такенса

Поясним сначала нестрого содержание теорем на простом примере (п. 10.1.1.1), а затем приведем их математическую формулировку и обсудим некоторые детали на более строгом языке (п. 10.1.1.2). В данной главе мы будем обозначать вектор состояния исходной системы $\mathbf{y}$ в отличие от восстановленных векторов – $\mathbf{x}$. Обозначение d относится *всегда* к размерности множества M в фазовом пространстве исходной системы (это не обязательно размерность всего пространства, т.е. векторов $\mathbf{y}$), а D – это размерность восстановленных векторов $\mathbf{x}$ и, следовательно, модели.

10.1.1.1. Пояснительный пример. Пусть объект представляет собой трехмерную динамическую систему с непрерывным временем. Вектор состояния $\mathbf{y} = (y_1, y_2, y_3)$. Пусть движение происходит на предельном цикле (рис.10.1,а), т.е. на множестве M размерности $d = 1$.

Такенса, для реконструкции достаточно использовать $D > 2$ модельных переменных. Если же нужно описать движение на всем торе M, включая и переходные процессы на нем, то нужно $D > 4$ переменных. Обычно на практике имеется одна реализация, которая соответствует установившемуся процессу. Поэтому и говорят часто о реконструкции аттрактора.

[5] Насколько вообще можно говорить о выполнении условий математической теоремы применительно к реальному объекту.

Если бы наблюдались все три переменных y_1, y_2, y_3 (все компоненты вектора **y**), то можно было бы сразу приступать к аппроксимации зависимости следующего состояния $\mathbf{y}(t+\tau)$ от текущего $\mathbf{y}(t)$. Эта зависимость однозначна, т.к. **y** – вектор состояния. Последнее означает, что каждый раз, когда наблюдается некоторое значение вектора $\mathbf{y} = \mathbf{y}^*$, через фиксированный интервал времени каждый раз будет наблюдаться одно и то же событие (одно и то же будущее). Если же наблюдаются не все компоненты вектора состояния, то ситуация может быть сложнее. Поставим вопрос: сколько и каких переменных достаточно использовать для эквивалентного описания исходной динамики, а какие варианты не подходят для этого?

Поскольку множество M, на котором происходит рассматриваемое движение, одномерно ($d = 1$), то должна существовать такая *скалярная* динамическая переменная, с помощью которой можно описать это движение. Например, замкнутую кривую M (рис.10.1,а) можно отобразить на окружность (рис.10.1,б). Важно, что при этом можно *взаимно однозначно* связать векторы $\mathbf{y}(t)$ на цикле M с углом поворота точки на окружности $\phi(t)$.[6] Благодаря взаимной однозначности, переменная ϕ полностью задает состояние системы: значение фазы ϕ^* всегда соответствует одному и тому же значению вектора $\mathbf{y}^*$ в тот же момент времени. Располагая наблюдаемой ϕ, можно строить одномерную ($D = 1$) *динамическую модель* с $x_1 = \phi$.

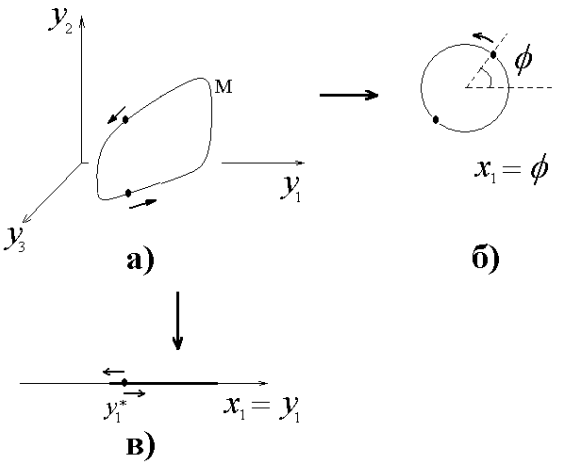

Рис.10.1. Одномерное представление динамики на предельном цикле (а) с помощью отображения на окружность (б) и проекции на координатную ось (в). Размерность фазового пространства исходной системы равна 3, размерность многообразия M – замкнутой кривой $d = 1$, размерность «восстановленных» векторов в обоих случаях $D = 1$. Два разных исходных состояния (черные кружки на рисунке а) соответствуют двум разным точкам на окружности (рисунок б) и одной и той же точке на оси (рисунок в). Отображение цикла на окружность взаимно однозначно, а на отрезок – нет

Но далеко не каждая переменная подходит для одномерного динамического описания. Так, если наблюдается лишь одна из компонент вектора **y**, например y_1, то замкнутая кривая отображается на отрезок (это

[6] Эта переменная – «свернутая» фаза колебаний, рассматривавшаяся в п. 6.4.3.

просто проекция на ось y_1). Это отображение не взаимно однозначно: каждой точке отрезка с координатой $y_1^*(t)$ соответствуют, по меньшей мере, два вектора состояния $\mathbf{y}(t)$, отличающиеся направлением дальнейшего движения – в сторону увеличения или уменьшения y_1, т.е. влево или вправо на оси $y_1(t)$, см. рис.10.1,а,в. Таким образом, задание y_1 не определяет однозначно состояния системы. Если наблюдается некоторое значение $y_1 = y_1^*$, то через фиксированный интервал времени может реализоваться один из двух (или более) вариантов будущего. Поэтому одномерное описание наблюдаемого движения при использовании переменной $x_1 = y_1$ невозможно.

Итак, при использовании размерности модели *D*, равной размерности наблюдаемого движения *d*, построение динамической модели может оказаться как возможным (если «повезло»), так и невозможным. Оба варианта типичны.[7] Это утверждение мы проиллюстрировали на простом примере, но оно справедливо и в общем случае.

Что меняется если использовать двухмерное описание? Здесь оказываются возможными и типичными те же две ситуации. Они проиллюстрированы на рис.10.2. Рассмотрим в качестве наблюдаемых две компоненты исходного вектора состояния – y_1 и y_2, т.е. модельный вектор $\mathbf{x} = (y_1, y_2)$. Это соответствует проекции замкнутой кривой на плоскость (y_1, y_2). В проекции можно наблюдать замкнутую кривую без самопересечений (рис.10.2,а) или с самопересечениями (рис.10.2,б), в зависимости от формы замкнутой кривой и ее ориентации в пространстве. В первом случае связь исходной замкнутой кривой и ее проекции взаимно однозначна, т.е. двухмерные векторы **x** полностью определяют состояние и могут использоваться для построения

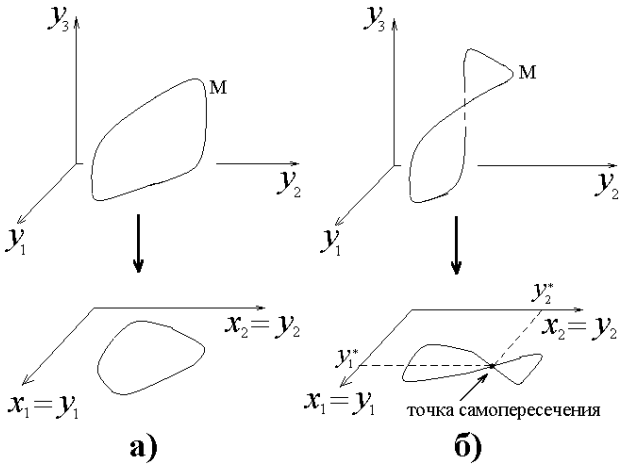

Рис.10.2. Два случая при проецировании одномерного многообразия из трехмерного пространства в двухмерное: а) взаимно однозначное отображение, б) не взаимно однозначное (в проекции появилась точка самопересечения кривой)

[7] Типичность здесь и ниже понимается в смысле «грубости», «случая общего положения», т.е. ситуация не меняется при «малом шевелении» исходной системы и наблюдаемой. Более строгую формулировку см. в п. 10.1.1.2.

динамической модели. Во втором случае это не так. Точка самопересечения y_1^*, y_2^* на плоскости (y_1, y_2), рис.10.2,б, соответствует двум различным состояниям исходной системы (т.е. связь **x** и **y** не взаимно однозначна). Поэтому, когда наблюдаются значения y_1^*, y_2^*, то будущее по ним однозначно предсказать нельзя. Значит, с помощью такого вектора **x** не удастся осуществить динамическое описание процесса.[8]

Аналогичная ситуация имеет место, если в качестве двух переменных модели используются не y_1 и y_2, а две переменных, полученных любым из упомянутых выше методов. Пусть наблюдаемая $\eta = h(\mathbf{y})$, где h – произвольная гладкая (измерительная) функция, а компоненты вектора **x** получены методом временных задержек $\mathbf{x}(t) = (\eta(t), \eta(t+\tau))$. При этом, в зависимости от h и τ, может иметь место ситуация, аналогичная рис.10.2,а, – нет самопересечений на плоскости (x_1, x_2), но столь же типична и ситуация рис.10.2,б, когда динамическое моделирование невозможно. Таким образом, даже число переменных большее, чем размерность исходного движения, $D = 2 > d = 1$, *не гарантирует* возможности динамического описания.

Наконец, рассмотрим трехмерное представление. Например, когда для наблюдаемой $\eta = h(\mathbf{y})$ векторы восстанавливаются в виде $\mathbf{x}(t) = (\eta(t), \eta(t+\tau), \eta(t+2\tau))$. Образ исходной замкнутой кривой в трехмерном пространстве (x_1, x_2, x_3) – также замкнутая кривая. И в типичном случае она тоже не имеет самопересечений. Это означает, что имеет место взаимно однозначное соответствие между векторами **x** и **y**. Исходное движение на предельном цикле может быть эквивалентно описано с помощью векторов **x**. Вообще говоря, самопересечение кривой в пространстве x_1, x_2, x_3 может иметь место, но это ситуация вырожденная, нетипичная. Интуитивно не трудно согласиться с тем, что самопересечение кривой в трехмерном пространстве, грубо говоря, крайне маловероятно.

Итак, в рассмотренном примере эквивалентное описание динамики достигается *с гарантией* только при восстановлении в пространстве размерности $D > 2d$.[9] Это и есть основное утверждение теорем Такенса. Подчеркнем, что это – *достаточное* условие. В отдельных случаях может оказаться возможным эквивалентное описание даже при $D = d$ (при удачной наблюдаемой). При реконструкции уравнений на практике теорема Такенса служит лишь для дополнительной психологической

---

[8] Это возможно только локально – в удалении от точки самопересечения.

[9] Эквивалентное описание движения на предельном цикле гарантируется при $D = 3$, даже если этот цикл «живет» в бесконечномерном фазовом пространстве.

уверенности, ведь согласно ей существует конечная размерность модели *D*, при которой вероятен успех динамического описания! Реально же приходится пробовать различные значения *D*, начиная с малых, и стараться получить «хорошую» модель как можно меньшей размерности, чтобы избежать трудностей, связанных с упомянутым выше «проклятием размерности».

10.1.1.2. Математические детали. Сформулируем теоремы более строго. Для этого потребуется ввести ряд терминов и обозначений. Пусть объект представляет собой динамическую систему:

$$\begin{aligned} &\mathbf{y}(t_0 + t) = \Phi_t(\mathbf{y}(t_0)), \\ &\boldsymbol{\eta}(t) = \mathbf{h}(\mathbf{y}(t)), \end{aligned} \tag{10.1}$$

где $\mathbf{y}$ – вектор состояния, Φ_t – оператор эволюции, $\mathbf{h}$ – измерительная функция.[10] Вектор наблюдаемых конечномерен $\boldsymbol{\eta} \in R^m$, причем мы будем говорить только о наиболее распространенном (и наиболее сложном для моделирования) случае скалярного временного ряда $\eta(t_i)$, $m = 1$.

Многообразие. Пусть движение системы происходит на многообразии M конечной размерности *d*.[11] Многообразие – это обобщение понятия гладкой поверхности в евклидовом пространстве (подробности см. в [58, 112, 299] и [115, с. 244-248]). Грубо, *d*-мерное многообразие M – это поверхность, которую локально в окрестности каждой точки можно параметризовать с помощью *d* евклидовых координат. Другими словами, любую точку $p \in \mathrm{M}$ и ее локальную окрестность $U(p)$ можно взаимно однозначно и взаимно непрерывно отобразить на *d*-мерный фрагмент (например, шар) пространства R^d. Полученный образ $\Psi : U \to \Psi(U)$ называют *картой* окрестности, а само непрерывное отображение Ψ – гомеоморфизмом. Примерами двухмерных многообразий в трехмерном

---

[10] Если объект – отображение $\mathbf{y}(t_{n+1}) = \mathbf{F}(\mathbf{y}(t_n))$, то оператор эволюции $\Phi_{\Delta t}(\mathbf{y}(t_0))$ есть просто функция **F**. Если объект – система ОДУ $d\mathbf{y}/dt = \mathbf{F}(\mathbf{y}(t))$, то функция $\Phi_t(\mathbf{y}(t_0))$ – результат интегрирования ОДУ на интервале шириной *t*. Если исходная система – уравнение в частных производных $\partial\mathbf{y}/\partial t = \mathbf{F}(\mathbf{y}, \partial\mathbf{y}/\partial\mathbf{r}, \partial^2\mathbf{y}/\partial\mathbf{r}^2, ...)$, где $\mathbf{r}$ – пространственная координата, то $\mathbf{y}$ есть вектор *бесконечномерного* пространства функций $\mathbf{y}(\mathbf{r})$, а Φ_t – оператор, действующий в этом пространстве.

[11] Это может иметь место даже для бесконечномерной системы. Для описания многомерных систем вводится полезное понятие *инерциального многообразия* [115]. Не строго, это многообразие минимальной размерности *d*, каждая точка которого полностью определяет состояние исходной системы. Величины, определяющие положение точек на многообразии (*d* штук), называют *параметрами порядка*.

евклидовом пространстве являются сфера, тор, бутылка с ручкой и т.п. (рис.10.3,а), но не (двойной) конус (рис.10.3,б).

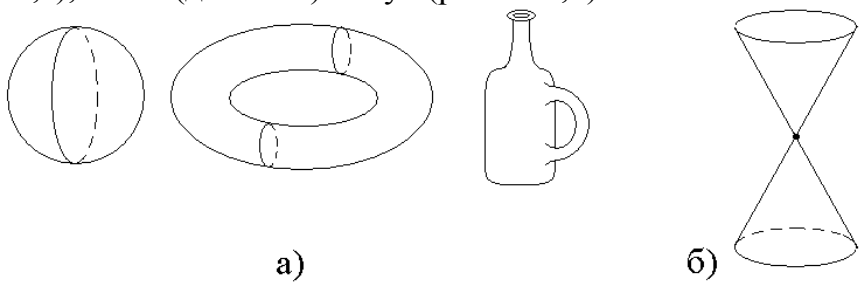

Рис.10.3. Примеры множеств, которые являются (а) и не являются (б) многообразиями

Если Ψ – n раз дифференцируемое отображение с таким же обратным, то говорят, что M принадлежит классу C^n. В случае $n \geq 1$ отображение Ψ называют *диффеоморфизмом*. Если многообразие M отображается с помощью диффеоморфизма Ψ на многообразие $\mathrm{S} \in R^D$, $D \geq d$, то M и S называют *диффеоморфными* друг другу. При этом говорят, что S – *вложение* многообразия M в евклидово пространство R^D. Ниже потребуется еще, чтобы многообразие M было ограничено и замкнуто. Первое условие означает, что его можно заключить в шар конечного радиуса, а второе – то, что все предельные точки многообразия M принадлежат ему же. Такое многообразие M в конечномерном пространстве называют *компактным*.

Постановка вопроса и обозначения. Каждой фазовой траектории системы (10.1) $\mathbf{y}(t)$, $0 < t < \infty$, на многообразии M соответствует временная реализация наблюдаемой величины η: $\eta(t) = h(\mathbf{y}(t))$, $0 < t < \infty$. Вектор $\mathbf{y}(t_0)$ однозначно определяет будущее поведение системы (10.1), в частности, всю реализацию $\eta(t)$, $t \geq t_0$. Можно ли по фрагменту временной реализации $\eta(t)$ в окрестности момента t_0 однозначно определить положение системы на многообразии M в момент t_0 и, следовательно, всю будущую эволюцию? Другими словами, можно ли по значениям $\eta(t)$ на ограниченном интервале времени «восстановить» состояние системы? Это ключевой вопрос, и теоремы Такенса дают положительный ответ на него при выполнении ряда условий.

Сначала введем необходимые обозначения для строгого рассмотрению метода задержек. Поставим в соответствие вектору $\mathbf{y}(t)$ D-мерный вектор $\mathbf{x}(t) = [\eta(t), \eta(t+\tau), ..., \eta(t+(D-1)\tau)]$. Зависимость $\mathbf{x}$ от $\mathbf{y}$ (имеются в виду значения векторов в один и тот же момент времени t) описывается *однозначным* отображением $\Psi : \mathrm{M} \to R^D$, которое выражается через оператор эволюции $\Phi_t : \mathrm{M} \to \mathrm{M}$ и функцию $h : \mathrm{M} \to R$ следующим образом:

$$\mathbf{x}(t) = \Psi(\mathbf{y}(t)) \equiv \begin{bmatrix} \Psi_1(\mathbf{y}(t)) \\ \Psi_2(\mathbf{y}(t)) \\ \dots \\ \Psi_D(\mathbf{y}(t)) \end{bmatrix} = \begin{bmatrix} h(\mathbf{y}(t)) \\ h(\Phi_\tau(\mathbf{y}(t))) \\ \dots \\ h(\Phi_{(D-1)\tau}(\mathbf{y}(t))) \end{bmatrix}. \qquad (10.2)$$

Гладкость Ψ (непрерывность, дифференцируемость, существование и непрерывность производных высокого порядка) определяется гладкостью Φ_τ и h. Образом многообразия M под действием Ψ является некоторое множество $S \subset R^D$.

Поставленный выше вопрос формулируется теперь так: является ли Ψ диффеоморфизмом? Если да, то S – это вложение M и любому вектору **x** из S соответствует только один вектор **y** из M.[12] В таком случае $\mathbf{x}(t)$ может служить вектором состояния для описания динамики на M и уравнения (10.1) можно переписать в терминах новых переменных:

$$\mathbf{x}(t_0 + t) = \varphi_t(\mathbf{x}(t_0)), \qquad (10.3)$$

где новый оператор эволюции $\varphi_t(\mathbf{x}) = \Psi(\Phi_t(\Psi^{-1}(\mathbf{x})))$. Причем в силу диффеоморфизма локальные свойства динамики (устойчивость и типы неподвижных точек и т.д.) сохраняются. Любой фазовой траектории $\mathbf{y}(t)$ на M взаимно однозначно соответствует траектория $\mathbf{x}(t)$ на S. Если система (10.1) имеет аттрактор в M, то система (10.3) имеет аттрактор в S; такие характеристики, как фрактальная размерность и ляпуновские показатели для этих аттракторов одинаковы. Другими словами, система (10.3) на многообразии M и система (10.1) на многообразии S – это два «почти одинаковых» представления одной и той же динамической системы.

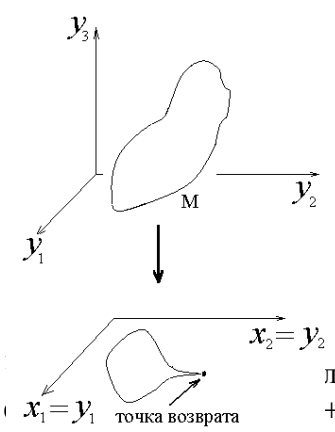

Рис.10.4. При проецировании одномерного многообразия M может иметь место

Ясно, что отображение Ψ (10.2) является диффеоморфизмом не всегда. Так, на рис.10.2,б представлен пример, когда гладкое отображение Ψ (проецирование) одномерного многообразия M имеет неоднозначное обратное отображение Ψ^{-1} (рис.10.2,б). Возможна и другая нежелательная ситуация, представленная на рис.10.4: Ψ^{-1}

ли в разные моменты времени t встречается один и тот же $+\tau),\dots,\eta(t+(d-1)\tau)]$, то за ним всегда следует одно и то же бы обоснование прогностическому *методу аналогов*, . Лоренцем. Метод основан на поиске фрагментов временного ые «похожи» на текущий (с точностью до конечной величины), и использовании какой-либо комбинации следовавших за ствe прогноза. В более современной трактовке – это локальные модели, см. п.10.2.1.4.

однозначно, но *не дифференцируемо*. Последнее имеет место в *точке возврата* на множестве S (в проекции), в окрестности которой двухмерный вектор (y_1, y_2) не может использоваться для описания динамики с помощью ОДУ, т.к. точка возврата оказывается неподвижной точкой, и кривая S не может быть предельным циклом. Дифференциальные свойства M и S оказываются различными из-за недифференцируемости отображения Ψ^{-1}.

Формулировка теоремы. Возвращаясь к системе (10.1) и отображению (10.2), можно сказать, что для разных Φ, M, h, d, D и τ можно встретиться с любой из упомянутых ситуаций. Множества самопересечений и множества возврата на $S = \Psi(M)$ могут быть весьма обширными, что крайне нежелательно. Но можно встретить и «хорошую» ситуацию вложения (рис.10.2,а). Приведенный ниже результат, впервые был строго получен голландским математиком Флорисом Такенсом [323], а затем обобщен в работе [299]. Он говорит о том, при каких условиях обеспечивается вложение исходного компактного d-мерного многообразия M в пространство R^D с помощью отображения (10.2).[13]

<u>Теорема 1.</u> Пусть M – компактное d-мерное многообразие класса C^2. Для *почти любой*[14] пары Φ_t, h (где Φ_t и h дважды непрерывно дифференцируемы на M) отображение Ψ: $M \to R^D$, определяемое (10.2), есть диффеоморфизм[15] для *почти* любого[16] $\tau > 0$ и $D > 2d$.

Обсуждение. Итак, если взять достаточно большую размерность векторов временных задержек **x** (10.2), то получим (за исключением вырожденных случаев) вложение многообразия M и сможем использовать **x** в качестве векторов состояния динамической модели. Возможна следующая наглядная интерпретация того, что размерность D должна быть больше именно величины $2d$ [115]. Чтобы установить неоднозначность

---

[13] Теорема Такенса тесно связана с теоремой Уитни, которая касается произвольных отображений, из курсов дифференциальной геометрии. Первая отличается тем, что относится к специальному случаю отображений (10.2), определяющихся оператором эволюции динамической системы [299].

[14] Термин «почти любая пара» понимается у Такенса в смысле типичности. Например, если для какого-либо Φ_t отображение (10.2) не дает вложения, то найдется такая сколь угодно малая вариация $\Phi_t + \delta\Phi_t$, что вложение будет достигнуто. Более строго, типичные свойства выполняются на пересечении открытых и всюду плотных множеств. Метрическим аналогом типичности является превалентность (prevalence) [299].

[15] Следовательно, отображение (10.2) дает вложение многообразия M. Пространство R^D, в котором содержится образ $S = \Psi(M)$ называют *пространством вложения*.

[16] Например, в случае наличия предельного цикла внутри M величина τ не должна быть равна периоду колебаний на этом цикле. Подробнее см. [299].

отображения Ψ^{-1}, надо на многообразии M найти такие два вектора $\mathbf{y}_1$ и $\mathbf{y}_2$, для которых $\Psi(\mathbf{y}_1) = \Psi(\mathbf{y}_2)$. Последнее равенство представляет собой систему D уравнений с $2d$ переменными (по d компонент каждого вектора $\mathbf{y}_1$ и $\mathbf{y}_2$, задающих положение на M). Грубо, эта система не имеет решений *в типичном случае*, если число уравнений больше числа неизвестных, т.е. $D > 2d$. Это и есть утверждение теоремы Такенса.

Еще раз подчеркнем, что условие $D > 2d$, наложенное на размерность вектора **x**, – достаточное, но не необходимое. Т.е. при соблюдении условия диффеоморфизм гарантирован с точностью до типичности, но если «повезет», то можно получить «хорошую реконструкцию» и при меньшей размерности, см. рис.10.1,а,б, где вложение одномерного многообразия достигнуто при $D = 1$ и является случаем общего положения.

Какими могут быть те нетипичные случаи, когда теоремы не справедливы? Упомянем два примера [115].

1) Измерительная функция – константа: $h(\mathbf{y}) = a$. Это гладкая функция, но она отображает всю динамику в одну точку. Малым шевелением измерительной функции это устраняется – добавкой к a «малой» функции от **y** (конечно, не просто константы).

2) Система, состоящая из двух подсистем с однонаправленной связью $d\mathbf{y}_1/dt = \mathbf{F}(\mathbf{y}_1, \mathbf{y}_2)$, $d\mathbf{y}_2/dt = \mathbf{G}(\mathbf{y}_2)$, при наблюдении только за ведущей подсистемой – $\eta = h(\mathbf{y}_2)$. В наблюдаемой нет информации об $\mathbf{y}_1$, поэтому вложение невозможно. Устраняется добавкой зависимости от $\mathbf{y}_1$ к η.

Подобные теоремы. В [299] доказан более общий вариант теоремы 1 – для случая *фильтрованного* вложения, когда в качестве координат вектора **x** берутся не просто последовательные значения наблюдаемой, а их линейные комбинации, которые можно рассматривать как результат действия линейного нерекурсивного фильтра

Такенс доказал подобную теорему для использования последовательных производных наблюдаемой в качестве компонент вектора состояния:

$$\mathbf{x}(t) = \begin{bmatrix} \eta(t) \\ d\eta(t)/dt \\ \dots \\ d^{D-1}\eta(t)/dt^{D-1} \end{bmatrix}, \tag{10.4}$$

где $D > 2d$. Она формулируется полностью аналогично теореме 1, но с более жесткими требованиями к гладкости функций Φ_t и h. А именно, требуется существование непрерывных производных D-го порядка каждой из них, чтобы существовали производные, используемые в (10.4).

Если производные аппроксимировать конечными разностями, то (10.4) сводится к фильтрованному вложению [229].

На практике всегда есть шумы, а к такому случаю теоремы Такенса не относятся, хотя имеются некоторые их обобщения [319, 207]. Тем не менее, теоремы представляют ценность для решения практических задач моделирования, о чем уже говорилось в начале п. 10.1 (см. с. 256).

10.1.2. Практические алгоритмы реконструкции

10.1.2.1. Метод временных задержек. Это наиболее популярный из методов реконструкции. Из скалярного ряда $\{\eta_i\}_{i=1}^{N}$ формируется ряд векторов $\{\mathbf{x}_i = (\eta_i, \eta_{i+l}, ..., \eta_{i+(D-1)l})\}_{i=1}^{N-(D-1)l}$. Величина задержки $\tau = l\Delta t$ теоретически может быть почти любой, но на практике часто нежелательны как слишком малые l, которые ведут к сильной корреляции компонент[17] вектора состояния, так и слишком большие l, при которых может слишком усложниться структура восстановленного аттрактора. Поэтому предлагалось выбирать величину τ, равной первому нулю автокорреляционной функции [229], первому минимуму функции взаимной информации [227], и пр. [269]. Используется неравномерное вложение, когда временные интервалы между компонентами вектора состояния не одинаковы, что уместно для систем с несколькими характерными временными масштабами [220, 249]. Для неоднородной динамики (например, с чередующимися интервалами почти периодического и очень сложного движений) предложено переменное вложение, при котором набор временных задержек и даже размерность вектора $\mathbf{x}$ зависят от его положения в пространстве состояний [249]. Каждый из подходов хорош для своего круга систем, нередко ведет к сложным вычислениям и не гарантирует успеха в общем случае [115].

Как выбрать размерность модели D, опираясь на анализ наблюдаемого ряда? Есть различные методы оценки: метод ложных соседей [258], метод главных компонент [199], метод Грассбергера – Прокаччиа [234], метод хорошо приспособленного базиса [108]. Два первых мы рассмотрим ниже (пп. 10.1.2.2, 10.1.2.3). Однако часто приходится просто перебирать разные пробные значения размерности D, начиная с малых и постепенно увеличивая, и строить модельные уравнения для каждого D, пока не будет получена «хорошая» модель. В таком случае подбор размерности и даже временных задержек может становиться частью единого процесса моделирования, а не отдельной замкнутой первой стадией.

---

[17] При малом интервале выборки и $l = 1$ восстановленная фазовая траектория вытягивается вдоль главной диагонали $x_1 = x_2 = ... = x_D$ (практически совпадает с ней).

10.1.2.2. Метод ложных соседей. Это метод целочисленной оценки размерности аттрактора сверху. Он основан на проверке того свойства, что фазовая траектория, восстановленная в пространстве достаточной размерности не должна иметь самопересечений. Проиллюстрируем метод на простом примере восстановления фазовой траектории по временной реализации синусоиды $\eta(t) = \sin t$, рис.10.5,а.

При D = 1 (т.е. $\mathbf{x}(t) = \eta(t)$) восстановленное множество лежит на отрезке прямой, рис.10.5,б. На нем точка с временным индексом k имеет соседей – точки l и s, соответствующие двум разным состояниям системы (разные по знаку производные $\eta(t)$). В двухмерном пространстве ($\mathbf{x}(t) = [\eta(t), \eta(t+\tau)]$) все точки «разойдутся», но точки k и l слабо, а k и s – сильно, рис.10.5,в. По этому признаку k и l называют «истинными», а k и s – «ложными» соседями.

Один из вариантов алгоритма таков. При пробной размерности D для каждого восстановленного вектора $\mathbf{x}_k$ отыскивают одного (самого близкого) соседа; увеличив D на 1, определяют, какие из соседей оказались ложными (сильно разошлись), а какие – истинными. Подсчитывают отношение числа ложных соседей к общему числу восстановленных векторов. Откладывают это число в зависимости от D, рис.10.5,г. Если при увеличении D это относительное число самопересечений уменьшается до нуля при некотором значении D^*, то последнее и есть оценка размерности пространства, в котором достигается вложение траектории моделируемого движения. На практике, начиная с некоторой «правильной» размерности D^*, число ложных соседей становится достаточно малым (но нестрого нулевым из-за шумов и т.д.) и далее не уменьшается. Такое значение D^* принимают в качестве оценки размерности модели. В случае с синусоидой – это число 2, рис.10.5,г. Подробности см., например, в [115].

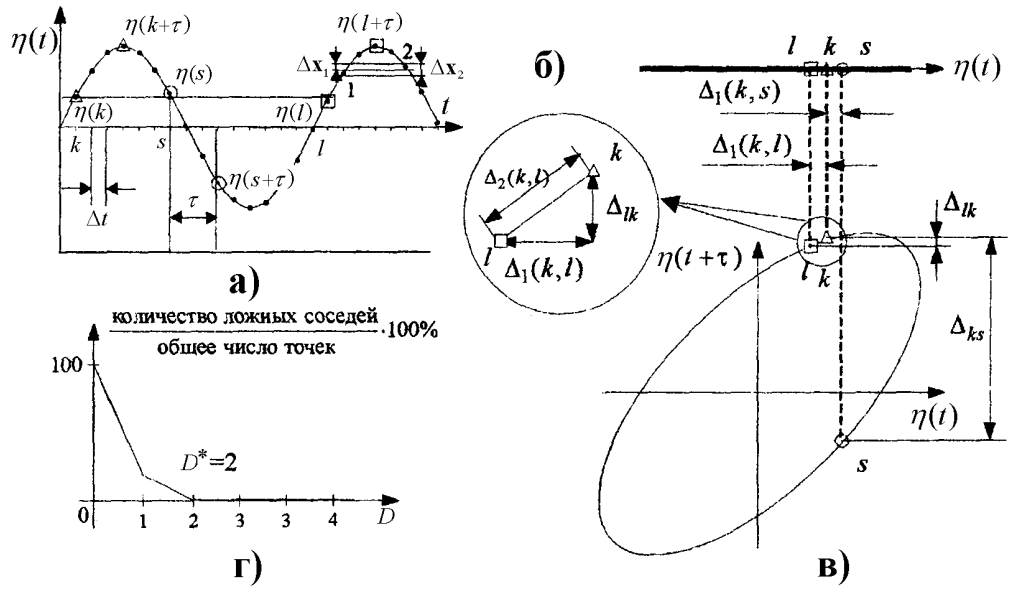

Рис.10.5. Иллюстрация метода ложных соседей. а) временная реализация $\eta(t)$, значками помечены точки $\eta(k)$, $\eta(s)$, $\eta(l)$, соответствующие моментам времени k, s, l и близким значениям η, а также точки, сдвинутые относительно них на $\tau = 3\Delta t$. б) траектория, восстановленная в одномерном пространстве, в) траектория, восстановленная в двухмерном пространстве, г) график зависимости отношения числа ложных соседей к общему числу точек в восстановленной траектории от пробной размерности восстановленных векторов D

10.1.2.3. Метод главных компонент. Он может использоваться как для оценки размерности, так и для восстановления векторов состояния. Метод используется в разных областях и имеет много названий, в приложении к реконструкции был предложен в [199]. Идея состоит в повороте осей координат в многомерном пространстве и выборе малого числа направлений, вдоль которых, в основном, развивается движение.

Для простоты обозначений положим, что среднее значение наблюдаемой η равно нулю. Векторы $\mathbf{w}(t_i) = (\eta_i, \eta_{i+1}, ..., \eta_{i+w-1})$ строятся в пространстве достаточно большой размерности k, причем время задержки полагается равным интервалу выборки $l = 1$. Компоненты этих векторов в случае малого интервала выборки сильно коррелированы. На рис.10.6 для иллюстрации представлен гармонический сигнал и восстановление фазовой траектории в трехмерном пространстве (k = 3).

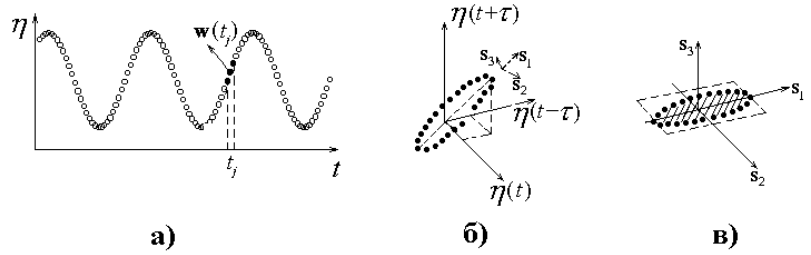

Рис.10.6. Гармонические колебания при отсутствии шума: а) восстановление векторов **w** методом задержек по скалярному ряду (размерность $k = 3$); б) восстановленная траектория – плоский эллипс, вытянутый вдоль главной диагонали пространства R^k; в) восстановленная траектория в новой системе координат (после преобразования поворота), составляющая векторов вдоль направления $\mathbf{s}_3$ равна нулю

В этом пространстве производится преобразование поворота, причем направления новых осей координат (это $\{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3\}$ на рис.10.6,б,в) выбираются в соответствии с теми направлениями, вдоль которых наиболее интенсивно развивается движение. Количественно эти характерные направления и протяженность вдоль них можно определить, анализируя матрицу ковариаций **Θ** компонент вектора **w**. Это – квадратная матрица порядка k: $\Theta_{i,j} = \sum_{n=0}^{N-k} \eta_{i+n}\eta_{j+n}$, $i, j = 1,...,k$. Она симметрична, вещественна, положительно определена. Следовательно, ее собственные векторы образуют полный ортонормированный базис пространства R^k, а ее собственные значения являются неотрицательными величинами. Обозначим собственные значения в порядке невозрастания $\sigma_1^2, \sigma_2^2, ..., \sigma_k^2$, а соответствующие собственные векторы – $\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_k$. Переход к базису $\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_k$ можно выполнить путем преобразования $\mathbf{x}'(t_i) = \mathbf{S}^{\mathrm{T}} \cdot \mathbf{w}(t_i)$, где **S** – матрица со столбцами $\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_k$. Это преобразование известно в теории информации как преобразование Карунена – Лоэва. Нетрудно показать, что при этом переходе матрица ковариаций компонент векторов $\mathbf{x}'$ примет диагональный вид:

$$\mathbf{\Theta}' = \mathbf{S}^{\mathrm{T}}\mathbf{\Theta}\mathbf{S} = \begin{bmatrix} \sigma_1^2 & 0 & ... & 0 \\ 0 & \sigma_2^2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \sigma_k^2 \end{bmatrix},$$

т.е. компоненты векторов $\mathbf{x}'$ некоррелированы, что является признаком «хорошей» реконструкции. Диагональные элементы матрицы ковариаций в новом базисе σ_i^2 – это средние квадраты проекций траектории $\mathbf{w}(t_i)$ на координатные оси $\{\mathbf{s}_i\}$. Они определяют протяженность траектории вдоль соответствующего направления. Ранг матрицы **Θ** равен числу ненулевых собственных значений (для ситуации, представленной на рис.10.6,б, ненулевые только σ_1^2 и σ_2^2) и равен размерности подпространства, в котором происходит движение.

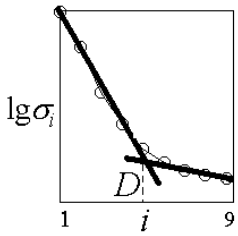

Рис.10.7. Возможный качественный характер зависимости значений собственных чисел матрицы ковариаций от их номера при k = 9. «Точка излома» D на графике служит оценкой размерности наблюдаемого

При наличии измерительного шума все σ_i^2 отличны от нуля, так как в направлениях, которые не «осваиваются» детерминированной составляющей траектории, представлена шумовая компонента. В этом случае размерность движения можно оценить как число D существенных собственных значений, иллюстрацию для произвольного случая см. на рис.10.7. Проекции вектора $\mathbf{w}(t_i)$ на соответствующие направления (т.е. первые D компонент вектора $\mathbf{x}'$) называют его *главными компонентами*. Остальные собственные значения составляют так называемый «шумовой пьедестал», а соответствующие компоненты можно отбросить. Таким образом, получим D-мерные векторы $\mathbf{x}(t_i)$ с компонентами $x_k(t_i) = \mathbf{s}_k \cdot \mathbf{w}(t_i), k = 1,...,D$.

Если на графике (рис.10.7) нет характерного излома, то следует увеличивать пробную размерность k, пока он не появится. Оценка размерности D более надежна, если при увеличении k излом наблюдается при одном и том же значении D.

Метод главных компонент – частный случай фильтрованного вложения. Он очень полезен в случае наличия измерительных шумов, поскольку позволяет их в значительной степени отфильтровать: реализация $x_1(t)$ – более плавная, чем $\eta(t)$. Примеры и обсуждение представлены также в [37].

10.1.2.4. Последовательное дифференцирование и другие методы восстановления фазовой траектории. Использование векторов (10.4) имеет привлекательные черты, связанные с их «физичностью». Многие процессы описываются модельным ОДУ (9.4) высокого порядка (п. 9.1), т.е. с последовательными производными единственной переменной. Некоторые системы ОДУ (например, система Ресслера, см. п. 10.2.2) могут быть приведены к такому виду *аналитически*. Но существенным недостатком метода является высокая чувствительность к наличию измерительных шумов, т.к. производные должны рассчитываться численно (пп. 7.4.2, 9.1).

Вообще говоря, методов восстановления векторов состояния очень много. Как уже сказано, имея только скалярный временной ряд, можно использовать метод последовательных производных или метод временных задержек. Причем для каждого из них имеются некоторые подстраиваемые параметры, значения которые можно подбирать (например, время задержки и схема численного дифференцирования). Кроме этого, есть

методы взвешенного суммирования [299, 201] и интегрирования [9, 247], целесообразные для сильно неоднородных сигналов. Часто используют главные компоненты, эмпирические моды, сопряженный сигнал и фазу (п. 6.4.3). Можно использовать и комбинации всех методов, например, часть переменных получать методом задержек, часть – интегрированием, часть – дифференцированием [201].

Если же наблюдаемая – векторная, то по каждой ее компоненте можно восстанавливать переменные с помощью любой комбинации описанных методов и число вариантов значительно возрастает [204, 209].

10.1.2.5. Выбор динамических переменных. Какой из вариантов выбора переменных предпочесть? Этот вопрос очень важен и давно привлекает к себе внимание [268, 264, 311]. Пробовать поочередно все возможные варианты и для каждого аппроксимировать зависимость вида $d\mathbf{x}/dt = \mathbf{f}(\mathbf{x},\mathbf{c})$ или $\mathbf{x}_{n+1} = \mathbf{f}(\mathbf{x}_n,\mathbf{c})$ не реально, поскольку решение задачи аппроксимации занимает часто очень значительное время и требует специальных подходов и усилий. Желательно было бы заранее выбрать относительно небольшое число наиболее подходящих вариантов и только для них проводить аппроксимацию. Для осуществления такого выбора предложен ряд процедур, основанных на предварительном анализе экспериментально полученных зависимостей, подлежащих аппроксимации [297, 314]. Эти процедуры опираются на то очевидное обстоятельство, что для построения модели нужен такой набор переменных, который обеспечивает однозначность и непрерывность зависимостей наблюдаемых (или рассчитанных по наблюдаемым) значений величин, стоящих в «левых частях» модельных уравнений, от наблюдаемых (или рассчитанных по наблюдаемым) значений динамических переменных.

Обозначим «левую часть» модельных уравнений $\mathbf{z}$. Для ОДУ вида $d\mathbf{x}/dt = \mathbf{f}(\mathbf{x},\mathbf{c})$ это $\mathbf{z}(t) = d\mathbf{x}(t)/dt$; для отображений вида $\mathbf{x}(t_{n+1}) = \mathbf{f}(\mathbf{x}(t_n),\mathbf{c})$ – это $\mathbf{z}(t_n) = \mathbf{x}(t_{n+1})$. Восстановив векторы $\mathbf{x}$ по наблюдаемой η, следует получить по ряду $\{\mathbf{x}(t_i)\}$ ряд «левых частей» $\{\mathbf{z}(t_i)\}$. Для ОДУ это делают численным дифференцированием ряда $\{\mathbf{x}(t_i)\}$, а для отображений – сдвигом $\{\mathbf{x}(t_i)\}$ на один шаг по времени. Далее нужно проверить, соответствуют ли близким значениям $\mathbf{x}$ близкие (одновременные) значения $\mathbf{z}$. Остановимся на одной из таких процедур, которая состоит в следующем [314].

Область V, внутри которой содержится множество векторов $\{\mathbf{x}(t_i)\}$, разбивается на одинаковые «гиперкубические» ячейки со стороной δ (рис.10.8,а). Из них выбираются все ячейки $s_1,...,s_M$, содержащие больше одного вектора $\mathbf{x}(t_i)$ каждая, т.е. имеющие хотя бы две точки в нижней клетке на рис.10.8,а. Разность между максимальным и минимальным

значениями z (одной из компонент вектора **z**) в пределах ячейки s_k назовем *локальным разбросом* ε_k. По величине максимального разброса $\varepsilon_{\max} = \max\limits_{1\le k\le M} \varepsilon_k$ и графику $\varepsilon_{\max}(\delta)$ оценивается пригодность величин **x** и z для глобального моделирования. Для построения глобальной модели переменные нужно выбирать так, чтобы график $\varepsilon_{\max}(\delta)$ стремился к началу координат плавно, без изломов (рис.10.8,б, жирная линия), для каждой из аппроксимируемых зависимостей $z_k(\mathbf{x})$, $k = 1,...,D$.

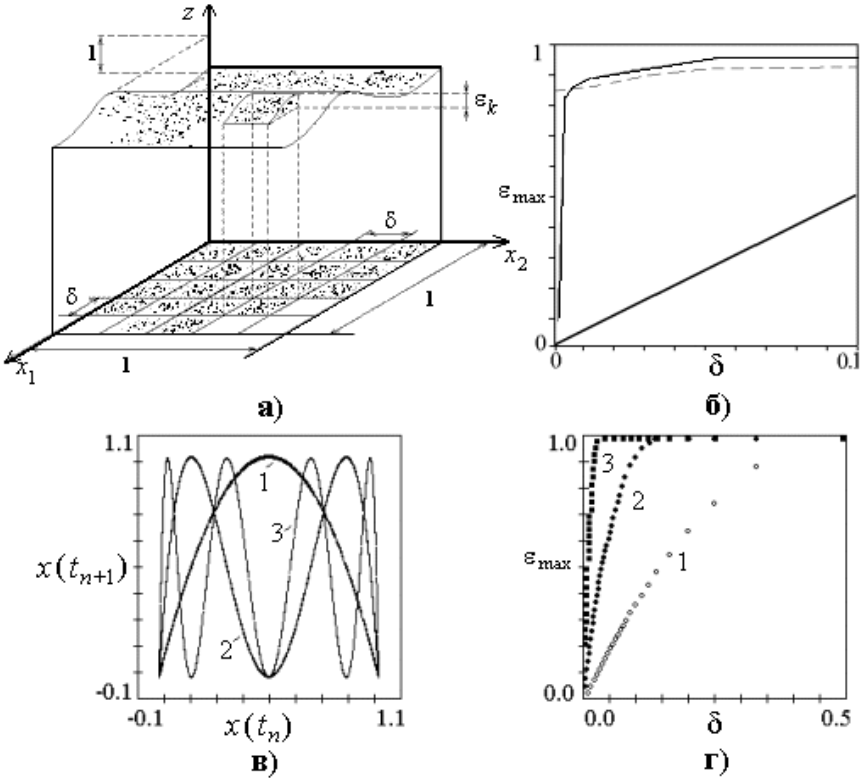

Рис.10.8. а) Иллюстрация методики проверки однозначности и непрерывности зависимости $z(\mathbf{x})$ при $D = 2$. б) Типичные графики $\varepsilon_{\max}(\delta)$ для различных переменных. Жирная линия – лучший случай, штриховая – худший (неоднозначность или разрывность), тонкая линия соответствует зависимости с областями быстрого изменения. в) Графики первой, второй и третьей итераций квадратичного отображения (помечены цифрами). г) Графики $\varepsilon_{\max}(\delta)$ для зависимостей $x(t_{n+1})$ от $x(t_n)$ в этих трех случаях

При этом желательно, чтобы наклон графика $\varepsilon_{\max}(\delta)$ был как можно меньшим, потому что для аппроксимации тогда достаточно использовать более простую модельную функцию, например, многочлен низкого порядка. На рис.10.8,в,г это проиллюстрировано на простом примере – аппроксимации зависимости следующего значения наблюдаемой от предыдущего, когда наблюдаемая генерируется первой, второй или третьей итерацией квадратичного отображения $x(t_{n+1}) = \lambda - x^2(t_n)$. График

первой итерации «наименее осциллирующий», а потому наклон $\varepsilon_{\max}(\delta)$ – самый маленький. Именно в этом случае проще всего получить «хорошую» модель (действительно, достаточно многочлена второго порядка, тогда как для описания третьей итерации нужен многочлен 8-го порядка). Особенно различаются эти три случая по трудности реконструкции при наличии шумов. Дополнительные подробности и примеры представлены в [314, 26].

10.2. Аппроксимация функций многих переменных

10.2.1. Модельные отображения

Векторы $[\eta_n, \eta_{n+\tau}, \eta_{n+2\tau}, ..., \eta_{n+(D-1)l}]$, полученные методом задержек, часто используются для построения многомерных модельных отображений

$$x_n = f(x_{n-1}, x_{n-2}, ..., x_{n-D}, \mathbf{c}), \tag{10.5}$$

где переменная x соответствует наблюдаемой, а время задержки принято равным $l = 1$ для простоты обозначений. Возможны различные варианты выбора вида функции f в (10.5). Говорят, что функция f, заданная в замкнутой форме (п. 3.5.1.2) во *всем* фазовом пространстве, обеспечивает *глобальную* аппроксимацию. В этом случае говорят также о глобальной модели и *глобальной реконструкции*. Используют и *локальную* аппроксимацию – функцию со своим набором параметров для каждой небольшой области фазового пространства.[18] В этом случае говорят о *локальной* модели.

Один из известных видов глобальной аппроксимации – алгебраические многочлены [36] – зачастую плохо работает на практике при аппроксимации функций уже двух переменных [206, 250, 193, 218, 134, 136]. Число их параметров и ошибки быстро растут с ростом размерности D. Подобные методы называют методами *слабой аппроксимации*. К ним относятся также тригонометрические многочлены и вейвлеты. При моделировании «черного ящика» нередко приходится использовать значения D, по меньшей мере, около 5-6. Поэтому алгебраические многочлены не нашли широкого практического применения.

Много усилий исследователей было потрачено на поиск методов *сильной* аппроксимации, т.е. относительно слабо чувствительных к росту размерности D. К ним относятся локальные методы низкого порядка [224, 206, 181, 260, 298, 303], радиальные, «цилиндрические» и «эллиптические» базисные функции [230, 317, 249-251, 308, 309] и искусственные нейронные сети [112, 200, 335]. Эти конструкции также содержат много

[18] Обычно используются функции простого вида: постоянная или линейная.

параметров и для них особенно актуален вопрос выбора размера модели и оптимизации ее структуры (пп. 7.2.3, 9.2).

10.2.1.1. Обобщенный многочлен. Для построения глобальной модели (10.5) выбирают вид f и рассчитывают параметры обычным МНК

$$S(\mathbf{c}) = \sum_{i=D+1}^{N} \left(\eta_i - f(\eta_{i-1}, \eta_{i-2}, ..., \eta_{i-D}, \mathbf{c})\right)^2 \to \min . \qquad (10.6)$$

Для простоты расчетов желательно выбрать функцию f, линейную по $\mathbf{c}$. Это имеет место для так называемого *обобщенного многочлена*

$$f(\mathbf{x}) = \sum_{k=1}^{P} c_k f_k(\mathbf{x}) \qquad (10.7)$$

по некоторой *системе базисных функций* $f_1, f_2, ..., f_P$. При этом задача (10.6) линейна и не возникает проблемы локальных минимумов. К такому способу относится использование алгебраического многочлена, порядок которого увеличивается, пока не будет найдена адекватная модель или пока не выполнится другое условие (о выборе размера модели см. п. 7.2.3). Подробности этого подхода представлены в лабораторной работе [36].

10.2.1.2. Радиальные базисные функции. Это функции вида $\phi_k(\mathbf{x}) = \phi\left(\|\mathbf{x} - \mathbf{a}_k\| / r_k\right)$, где $\|\cdot\|$ означает норму (длину) вектора, в качестве «материнской» функции ϕ обычно берется хорошо локализованная функция, например, «гауссиана» $\phi(y) = \exp(-y^2/2)$, величины $\mathbf{a}_k$ называют «центрами», а r_k – «радиусами». Модельная функция f представляет собой обобщенный многочлен по системе функций ϕ_k: $f(\mathbf{x}, \mathbf{c}) = \sum_k c_k \phi_k(\mathbf{x})$. Каждое слагаемое существенно отлично от нуля только на расстоянии, не большем r_k от центра $\mathbf{a}_k$ (рис.10.9). Интуитивно ясно, что с помощью такой суперпозиции можно приблизить очень сложный (но гладкий) рельеф. Радиальные базисные функции имеют много привлекательных свойств и часто используются в практике аппроксимации. Однако мы ограничимся сказанным и остановимся несколько подробнее на двух других, еще более распространенных методах.

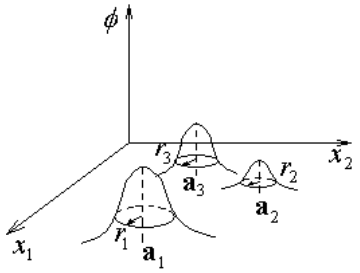

Рис.10.9. Графики радиальных базисных функций (качественно), зависящих от двух переменных: три «гауссовских холма»

10.2.1.3. Искусственные нейронные сети. Модели с ИНС (п. 3.8) широко и успешно используются для решения многих задач. Они представляют собой не сумму базисных функций, а *композицию* (см. с.96-97). В отличие от обобщенного многочлена, они обязательно нелинейно

зависят от оцениваемых параметров. Это «наиболее универсальный» способ аппроксимации функций многих переменных в том смысле, что он не только теоретически обоснован, но и нередко успешно работает на практике.

Введем здесь ИНС формально (в дополнение к п. 3.8) на примере многослойного персептрона. Пусть $\mathbf{x} = (x_1,...,x_D)$ – аргумент искомой функции *f*. Составим набор функций $f_j^{(1)}(\mathbf{x})$, $j = 1,...,K_1$:

$$f_j^{(1)}(\mathbf{x}) = \phi\left(\sum_{i=1}^{D} w_{j,i}^{(0)} \cdot x_i - \upsilon_j^{(0)}\right) \equiv \phi\left(\mathbf{x}^{\mathbf{T}} \cdot \mathbf{w}_j^{(0)} - \upsilon_j^{(0)}\right), \qquad (10.8)$$

где постоянные $w_{j,i}^{(0)}$ называют весами, $\upsilon_j^{(0)}$ – порогами, ϕ – функцией активации. Функция ϕ обычно нелинейна и имеет график ступенчатого вида. Часто используется так называемый классический сигмоид $\phi(x) = 1/\left(1 - e^{-x}\right)$. Будем говорить, что каждая функция $f_j^{(1)}$ представляет выход *стандартного формального нейрона* с номером *j*, на вход которого подан вектор **x** (живой нейрон суммирует внешние стимулы и реагирует на их совокупность пороговым образом, отсюда и свойства функции ϕ), см. рис.10.10,а. Система функций $f_1^{(1)},...,f_{K_1}^{(1)}$ – это набор нейронов первого слоя (рис.10.10,б). Назовем значения функций $f_j^{(1)}$ выходными значениями нейронов первого слоя и обозначим их совокупность вектором $\mathbf{y}^{(1)}$ с компонентами $y_j^{(1)} = f_j^{(1)}(\mathbf{x})$.

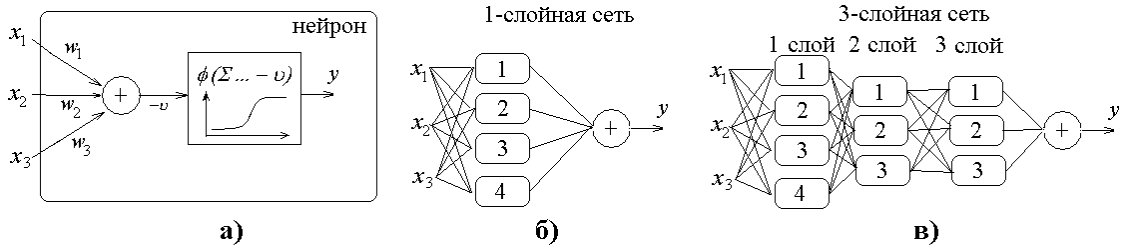

Рис.10.10. а) Стандартный формальный нейрон. б) Схема однослойной искусственной нейронной сети с одним выходом (один прямоугольник – один нейрон). в) Схема многослойной ИНС с одним выходом

Определив в качестве функции *f* линейную комбинацию $f_j^{(1)}$, получим так называемую *однослойную* ИНС:

$$f(\mathbf{x}) = \sum_{j=1}^{K_1} w_j^{(1)} y_j^{(1)} + \upsilon^{(1)} \equiv \sum_{j=1}^{K_1} w_j^{(1)} \phi\left(\sum_{i=1}^{D} w_{j,i}^{(0)} x_i + \upsilon_i^{(0)}\right) + \upsilon^{(1)}, \qquad (10.9)$$

где параметры $w_j^{(1)}, \upsilon^{(1)}$ – дополнительные веса и порог. Свободных параметров в такой модели $P = K_1(D+1) + (K_1+1)$ штук. Это

представление похоже на обобщенный многочлен (10.7), но ИНС нелинейно зависит от параметров $w_{j,i}^{(0)}$ и $\upsilon_j^{(0)}$.

Рассмотрим теперь еще одну систему функций $f_j^{(2)}, j = 1,...,K_2$, имеющих вид (10.8), но зависящих от K_1-мерного аргумента. Это нейроны второго слоя, на вход которых подаются выходы нейронов первого слоя $\mathbf{y}^{(1)}$ (рис.10.10,в). Обозначим их выходные значения вектором $\mathbf{y}^{(2)}$ размерности K_2 и определим функцию f как линейную комбинацию выходных значений нейронов второго слоя:

$$f(\mathbf{x}) = \mathbf{w}^{(2)} \cdot \mathbf{y}^{(2)} + \upsilon^{(2)} \equiv \sum_{j_2=1}^{K_2} w_{j_2}^{(2)} \phi\left(\sum_{j_1=1}^{K_1} w_{j_2,j_1}^{(1)} \phi\left(\sum_{i=1}^{D} w_{j_1,i}^{(0)} x_i + \upsilon_{j_1}^{(0)} \right) + \upsilon_{j_2}^{(1)} \right) + \upsilon^{(2)}. \quad (10.10)$$

Это двухслойная ИНС, которая включает в себя уже композиции функций, что существенно отличает ее от псевдолинейной модели (10.7). Увеличение числа слоев достигается очевидным образом (см. рис.10.10,в).

Чаще всего для решения задач аппроксимации используют двухслойные ИНС (10.10), реже – трехслойные [115]. Увеличение числа слоев не приводит к существенному улучшению. Улучшения добиваются за счет увеличения числа нейронов в слоях K_1, K_2. Теоретическая основа использования ИНС – обобщенная аппроксимационная теорема (ее частным случаем является теорема Вейерштрасса), которая утверждает, что любая непрерывная функция может быть сколь угодно точно равномерно приближена с помощью ИНС, строгое изложение см. в [63].

Процедура расчета параметров ИНС путем минимизации (10.6) – ее «обучение» – это сложная задача многомерной нелинейной оптимизации, для решения которой развиты специальные «технологии»: алгоритм обратного распространения ошибки, обучение с расписанием, обучение с шумом, стохастическая оптимизация (генетические методы, метод имитации отжига), см. подробности и ссылки в [79]. ИНС может содержать очень много лишних элементов, и структуру этой модели (архитектуру сети) желательно сделать более компактной. Для этого нейроны, веса и пороги которых слабо меняются в процессе обучения, исключаются из сети.

Если имеется несколько альтернативных ИНС с разной архитектурой, полученных в результате обучения по тренировочному ряду, то лучшую из них обычно выбирают по наименьшей тестовой ошибке аппроксимации (п. 7.2.3.2). Для получения «честного» показателя прогностической эффективности модели используют еще один ряд (не тренировочный и не тестовый, т.к. оба они использовались для построения модели), который называют «экзаменационным».

Превосходство ИНС над другими конструкциями при решении задач

моделирования нелегко понять интуитивно [115]. Если получена хорошо работающая модель с ИНС, то, как правило, неясно, за счет чего она так хороша. Это проблема «прозрачности сети»: модель «черного ящика» и сама оказывается часто «черным ящиком». Тем не менее, модель можно исследовать численно и использовать для прогноза, см. п. 10.2.1.6.

10.2.1.4. Локальные модели. Локальные модели строятся из соображения минимизации суммы квадратов типа (10.6) в локальной области фазового пространства. Так, для прогноза значения η_{i+D}, которое последует за состоянием $\mathbf{x}_i = [\eta_i, \eta_{i+1}, ..., \eta_{i+D-1}]$,[19] используется следующая процедура. Из всех векторов тренировочного ряда отыскивают k *ближайших соседей* текущего вектора $\mathbf{x}_i$. Это векторы с временными индексами n_j, расстояния от которых до $\mathbf{x}_i$ меньше всего (*аналоги* $\mathbf{x}_i$, рис.10.11, 10.12):

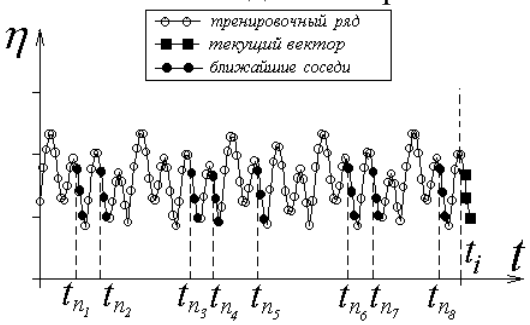

Рис.10.11. Иллюстрация для трехмерной локальной модели (D = 3): ближайшие соседи (черные кружки) вектора $\mathbf{x}_i$ (черные квадраты), найденные по тренировочному ряду

$$\left\| \mathbf{x}_{n_j} - \mathbf{x}_i \right\| \le \left\| \mathbf{x}_l - \mathbf{x}_i \right\|, \; j = 1,...,k, \; l \ne i, \; l \ne n_j. \quad (10.11)$$

Значения наблюдаемой, которые следовали в прошлом за соседями $\mathbf{x}_{n_j}$, известны. По ним можно построить модель (10.5). Обычно используют простую функцию $f(\mathbf{x}, \mathbf{c})$, параметры которой находят обычным МНК, хотя используют и более сложные методики оценки [260]. Полученную функцию $f(\mathbf{x}, \hat{\mathbf{c}}_i)$ используют для прогноза значения η_{i+D} по формуле $\hat{x}_{i+D} = f(\mathbf{x}_i, \hat{\mathbf{c}}_i)$, рис.10.12. Параметры $\hat{\mathbf{c}}_i$ снабжены нижним индексом i, т.к. соответствуют только окрестности текущего вектора $\mathbf{x}_i$. Согласно так называемому *итерационному* способу прогноза (см. п. 10.3), чтобы предсказать

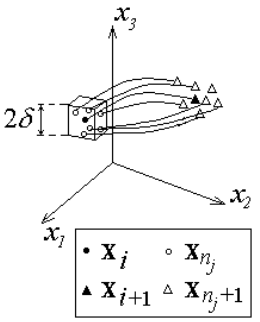

Рис.10.12. Ближайшие соседи (белые кружки) вектора $\mathbf{x}_i$ (черный кружок) и следующие за ними векторы (белые треугольники). По этим данным осуществляют прогноз следующего вектора $\mathbf{x}_{i+1}$ (черный

[19] В первую очередь обычно возникает задача – предсказать зн е последует сразу за последним значением η в тренировочном р м $\mathbf{x}_{N-D+1} = [\eta_{N-D+1}, \eta_{N-D+2}, ..., \eta_N]$.

следующее значение η_{i+D+1}, необходимо для нового вектора состояния модели $\hat{\mathbf{x}}_{i+1} = (\eta_{i+1},...,\eta_{i+D-1},\hat{x}_{i+D})$ повторить процедуру поиска соседей и оценки параметров. Таким образом, получают новый прогноз: $\hat{x}_{i+D+1} = f(\hat{\mathbf{x}}_{i+1},\hat{\mathbf{c}}_{i+1})$.

Опираясь на теорему Тейлора, используют следующие аппроксимирующие функции: константа $f(\mathbf{x},\mathbf{c}) = c_1$; линейная функция $f(\mathbf{x},\mathbf{c}) = c_1 + \sum_{j=1}^{D} c_{j+1} x_j$; многочлены более высоких порядков *K*. С одной стороны, погрешность аппроксимации тем меньше, чем ближе найденные соседи к текущему вектору. Поэтому она должна уменьшаться с ростом длины ряда, т.к. происходят все более близкие возвраты в окрестность каждой точки. С другой стороны, для борьбы с шумом следует использовать большее число соседей *k*. Возникает необходимость соблюдать компромисс: нельзя использовать слишком *далеких* «соседей», чтобы погрешность аппроксимации многочленом низкого порядка не стала слишком большой, но нельзя брать и слишком *малое число* близких соседей.

Локально-постоянные модели менее требовательны к числу данных и более устойчивы к действию шумов, т.к. они содержат всего один свободный параметр для каждой окрестности. При небольших шумах и достаточно больших длинах ряда (конкретные величины зависят от необходимой размерности модели *D*) преимущество имеют локально-линейные модели. Для их построения нужно использовать минимум $k = D + 1$ соседей, т.к. они содержат $D + 1$ параметров в каждой «ячейке». Погрешность аппроксимации для них в случае очень длинного и «чистого» ряда – порядка $\hat{\delta}^2$, где $\hat{\delta}$ – характерное расстояние между близкими векторами в тренировочном ряде. Локальные модели с многочленами более высокого порядка используются крайне редко, т.к. они имеют преимущества только в (почти нереалистичном) случае огромных объемов очень «чистых» данных.

В описанном подходе модельная функция *f*, как правило, разрывна, т.к. различные «куски» локальной аппроксимации не сшиваются между собой. Иногда это приводит к нежелательным особенностям динамики модели, которые не наблюдаются у исходной системы. Устранить разрывность можно с помощью триангуляции и других методов [308]. При этом модель приобретает некоторые глобальные свойства (*f* становится непрерывной), и ее тогда называют глобально-локальной. Но это сильно усложняет алгоритм и редко используется на практике.

Локальные модели часто используются на практике для прогноза. Есть разнообразные варианты алгоритмов их построения, учитывающие тонкие детали. Это современный вариант прогностического метода

«аналогов» (см. рис.10.11 и с. 261-262). Познакомиться подробнее с процедурой их построения можно в лабораторной работе [36].

10.2.1.5. Поиск близких соседей. Поиск соседей может занять очень много времени, если тренировочный ряд велик. Искать расстояния от каждого текущего вектора до каждого вектора в тренировочном ряде и выбирать наименьшие – очень долго (число операций порядка N^2). Здесь мы изложим эффективный алгоритм быстрого поиска [254], основанный на предварительном разбиении тренировочного множества на ячейки.

Описанные выше локальные модели – это *модели с фиксированным числом соседей*. Рассмотрим здесь другой (но очень близкий) вариант локальных моделей – *модели с фиксированным размером окрестности*. Согласно ему для текущего вектора $\mathbf{x}_i$ ищутся соседи, которые находятся на расстоянии от $\mathbf{x}_i$ не большем δ (рис.10.12):

$$\left\| \mathbf{x}_{n_j} - \mathbf{x}_i \right\| \le \delta . \qquad (10.12)$$

Число соседей может быть разным для разных $\mathbf{x}_i$. Главное, чтобы оно было не меньше $D+1$. Если соседей слишком мало, то надо увеличить размер окрестности. При увеличении уровня шума и размерности модели для фиксированной длины ряда N оптимальный размер окрестности сдвигается в сторону больших δ. Оптимальное значение δ подбирают методом проб и ошибок. Норму вектора в (10.11) или (10.12) можно использовать любую. Удобнее всего норма $\|\mathbf{x}\| = \max\{|x_1|, |x_2|, ..., |x_D|\}$, поскольку она вычисляется быстро. Для нее окрестность (10.12) – это гиперкуб со стороной 2δ.

Расчет расстояний от текущего вектора до всех векторов тренировочного ряда требует значительных затрат машинного времени. Желательно было бы не рассчитывать расстояния до тех векторов тренировочного ряда, которые заведомо не могут быть соседями. Для этого векторы тренировочного ряда предварительно *сортируют*. Для сортировки используют первую и последнюю координаты векторов. Пусть $\eta_{\min}$ и $\eta_{\max}$ – минимальное и максимальное значения наблюдаемой в тренировочном ряде. Тогда на плоскости (x_1, x_D) фазовая траектория лежит внутри квадрата со сторонами $x_1 = \eta_{\min}, x_1 = \eta_{\max}, x_D = \eta_{\min}, x_D = \eta_{\max}$ (см.

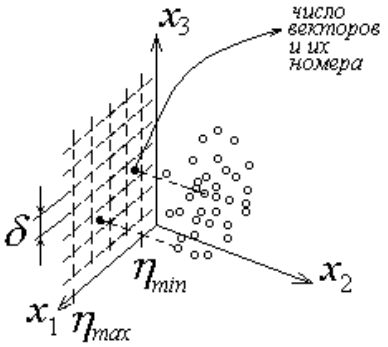

Рис.10.13. Векторы тренировочного ряда сортируются по значениям первой и последней координат: создается квадратный массив, в элементы которого записывается информация о количестве векторов в

рис.10.13). Этот квадрат разбивается на квадратные ячейки размером δ. Определяется, какой вектор в какую ячейку попадает. Создается массив, элементы которого соответствуют ячейкам разбиения. В каждый элемент записываются временные индексы векторов, попавших в соответствующую ячейку. Теперь, чтобы отыскать ближайших соседей вектора **x**, нужно определить, какой ячейке принадлежит этот вектор, и рассчитать расстояния от него только до тех векторов, которые принадлежат той же ячейке или имеющим с ней общую вершину (всего нужно проверить максимум 9 ячеек). Такой алгоритм существенно ускоряет процесс поиска соседей и требует всего порядка N операций при отсутствии слишком густо и слишком редко «населенных» областей в восстановленном пространстве.

10.2.1.6. Практический пример прогноза. Наблюдаемый процесс – хаотическая динамика лазера (рис.10.14) – был предложен организаторами конференции в Санта-Фе (США) в 1993 году всем желающим посоревноваться в области прогноза сложных сигналов [326]. Авторы должны были представить продолжение ряда, которое не было им заранее известно (100 следующих точек по имеющимся 1000 точкам). Конкурс выиграл аспирант Э. Ван, который использовал ИНС прямого распространения [335].

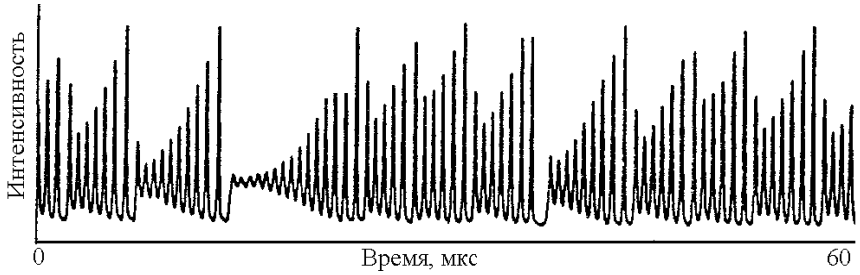

Рис.10.14. Данные с кольцевого лазера в хаотическом режиме [243], $\Delta t = 40$ нс

На рис.10.15,а представлены графики наблюдаемого ряда (пунктир) и прогноза (сплошная линия) с помощью модели с ИНС для различных стартовых моментов. Близкую точность прогноза, а по некоторым параметрам и лучшее воспроизведение динамики [326], обеспечили локально-линейные модели [298], см. рис.10.15,б. Локально-линейная модель «работает» лишь чуть хуже ИНС для стартовых моментов 1000 и 2180 и лучше для других трех случаев. Относительный успех модельного прогноза на разных отрезках зависит от того, насколько точно они предсказывают момент переключения колебаний с высокоамплитудных на низкоамплитудные. Кроме того, оказывается, что локальная линейная модель лучше воспроизводит долговременную динамику, что можно увидеть на верхней панели рис.10.15,б, где представлен итерационный прогноз на 400 точек вперед и видно хорошее соответствие с

экспериментом на всем интервале, для ИНС результаты несколько хуже [298, 335]. Таким образом, в данном случае модельные отображения с ИНС и локальной линейной аппроксимацией примерно одинаково хороши. Их примерно одинаковая эффективность наблюдается и во многих других случаях.

Ряд примеров применения локальных моделей для прогноза можно найти в [224, 254, 260]. Модели с ИНС применяются все же несколько чаще, т.к. они менее требовательны к длине ряда и уровню шумов. Есть примеры их успешного применения для прогноза в геофизике и области финансов (курсы валют) [99, 112]. Успешные результаты построения моделей в виде (10.5) см. также в [251, 308].

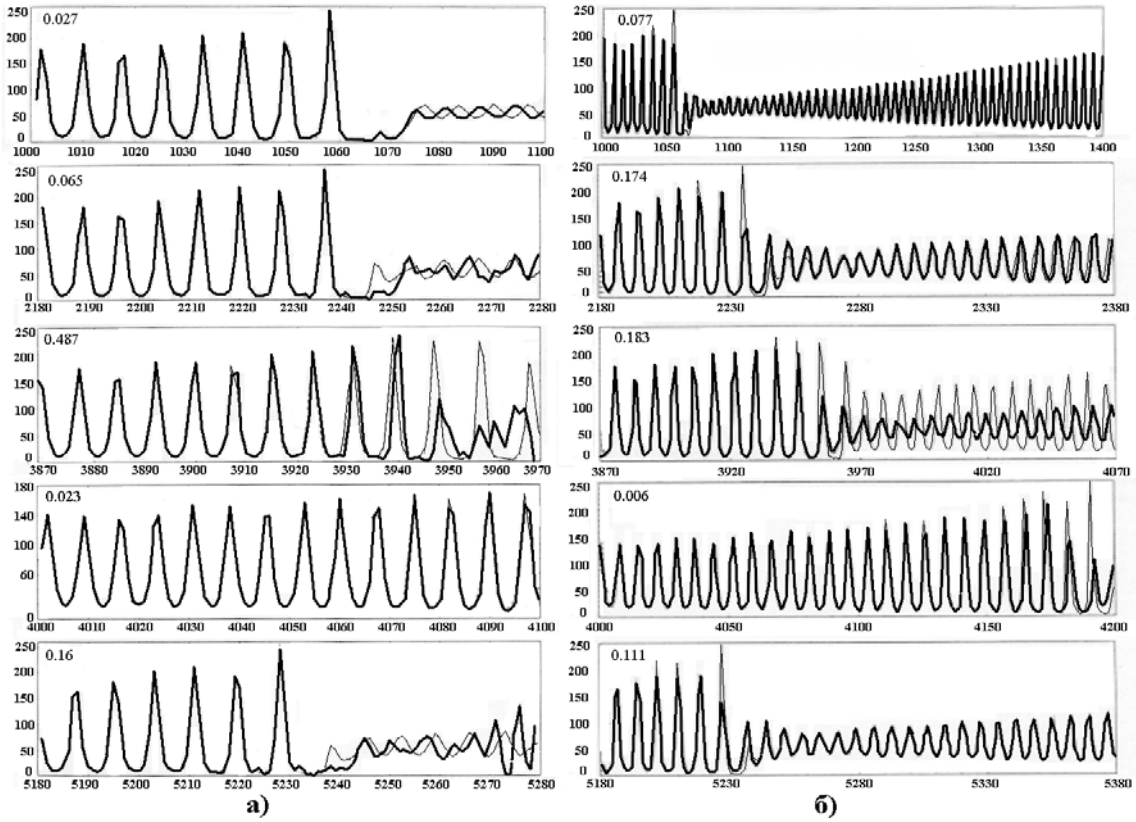

Рис.10.15. Прогноз интенсивности излучения лазера: а) модель с ИНС [335], б) локально-линейная модель [298]. По горизонтальной оси – время в единицах интервала выборки. Тонкие линии – измеренные значения, жирные линии – прогноз. Разные панели соответствуют прогнозу с различных стартовых моментов: 1000, 2180, 3870, 4000, 5180. Число в левом верхнем углу каждой панели – средний квадрат ошибки прогноза на первых 100 точках представленного фрагмента. Верхние панели показывают фрагмент, который требовалось предсказать в соревновании 1993 года (точки № 1001-1100)

10.2.2. Модельные дифференциальные уравнения

По скалярному временному ряду модельные ОДУ часто строят, восстанавливая векторы состояния методом последовательных

производных $[\eta, d\eta/dt, ..., d^{D-1}\eta/dt^{D-1}]$ в так называемой стандартной форме (п. 3.5.3):

$$d^D x/dt^D = f(x, dx/dt, ..., d^{D-1}x/dt^{D-1}, \mathbf{c}), \qquad (10.13)$$

где наблюдаемая $\eta = x$. О выборе аппроксимирующей функции здесь можно сказать все то же самое, что и выше для модели (10.5). Но здесь чаще наблюдаются «плавные» аппроксимируемые зависимости и используются алгебраические многочлены

$$f(x_1, x_2, ..., x_D, \mathbf{c}) = \sum_{l_1, l_2, ..., l_D = 0}^{K} c_{l_1, l_2, ..., l_D} \prod_{j=1}^{D} x_j^{l_j}, \ \sum_{j=1}^{D} l_j \le K. \qquad (10.14)$$

Почти не используют модельные ОДУ (10.13) с ИНС и т.п. [309].

Некоторые системы могут быть приведены к стандартному виду (10.13) даже аналитически. Так, эталонная хаотическая система – система Ресслера – имеет вид:

$$\begin{aligned} dx/dt &= -y - z, \\ dy/dt &= x + C_1 y, \\ dz/dt &= C_2 - C_3 z + xz. \end{aligned} \qquad (10.15)$$

Можно показать, что ее можно свести к трехмерной же системе с последовательными производными координаты y и многочленом второго порядка в правой части:

$$\begin{aligned} dx_1/dt &= x_2, \\ dx_2/dt &= x_3, \\ dx_3/dt &= -C_2 - C_3 x_1 + (C_1 C_3 - 1)x_2 + (C_1 - C_3)x_3 - C_1 x_1^2 + \\ &\quad + (C_1^2 + 1)x_1 x_2 - C_1 x_1 x_3 - C_1 x_2^2 + x_2 x_3, \end{aligned} \qquad (10.16)$$

где $x_1 = y$. Для последовательных производных координат x и z также можно получить уравнения (10.16), но с дробно-рациональными функциями в правой части [232, 37].

Стандартные модели часто использовались на практике [65, 66, 135, 178, 231-233, 265-267], но успехов остается считанное число. Очень часто структура (10.13) с алгебраическим многочленом (10.14) в правой части приводит к громоздким уравнениям.

При построении ОДУ по векторному ряду нужно аппроксимировать D скалярных функций, а не одну (см. п. 9.1). Практические примеры построения модельных ОДУ представлены в лабораторных работах [37].

10.3. Прогноз с помощью различных видов моделей

Бросая общий взгляд на проблему прогноза временных рядов и сопоставляя различные подходы, следует сказать, что новые методы, развитые в рамках нелинейной динамики и обсуждавшиеся выше, нередко оказываются наиболее эффективными для прогноза сложных реальных процессов. Это имеет место в тех случаях, когда достаточно ограничиться моделью невысокой размерности. Многочисленные «нелинейно-динамические» подходы можно отличать по целому ряду признаков: итерационный прогноз – прямой прогноз; модельные отображения различного вида (например, глобальные модели – локальные модели); модельные отображения – модельные ОДУ. Далее мы кратко обсудим преимущества и недостатки различных вариантов. Забегая вперед, отметим, что наиболее эффективным инструментом прогноза, как правило, являются модельные отображения (глобальные или локальные, в зависимости от объема данных и необходимой размерности модели) при использовании итерационного, прямого или комбинированного способа прогноза (в зависимости от необходимого упреждения). Но сначала вспомним и о более «старых» подходах.

Методы, не использующие нелинейную динамику. Для очень простых сигналов задача прогноза может успешно решаться даже с помощью явных функций времени (глава 7). Для нерегулярных стационарных сигналов без признаков нелинейности наиболее уместны линейные модели авторегрессии – скользящего среднего (пп. 4.4, 8.1, лабораторные работы в [39]), хотя эти возможности весьма ограничены. Можно показать, что прогноз с помощью линейной АРСС-модели может быть более или менее точным только на интервале порядка времени корреляции процесса τ_{cor} [12] (это время спадания автокорреляционной функции, см. п. 2.3, рис.2.8).

Для хаотического ряда время корреляции τ_{cor} может быть очень мало, а потому и дальность прогноза с помощью АРСС-модели будет мала. Хотя хаотический процесс в принципе нельзя предсказать очень далеко, для нелинейных моделей дальность прогноза может быть значительно больше τ_{cor}. Напомним, что эту дальность прогноза можно оценить из простых (но не всегда верных, подробнее см. п. 2.4) соображений [96, 318] по формуле (2.34): $\tau_{\text{pred}} = (1/2\Lambda_1)\ln\left(\sigma_x^2 / \sigma_\nu^2 + \sigma_\mu^2 + \sigma_{\Delta M}^2\right)$. Если шумы и погрешности модели не велики, то τ_{pred} может значительно превышать время корреляции, которое может быть грубо оценено как $\tau_{cor} \sim 1/\Lambda_1$.

Итерационный, прямой и комбинированный способы прогноза. Предсказать значения наблюдаемой, следующие за последним в ряде значением η_N, с помощью (нелинейной) модели (10.5) можно уже упомянутым (п. 10.2.1.4) *итерационным* способом:

1) получают прогноз на один шаг вперед

$$\hat{x}_{N+1} = f(\mathbf{x}_{N-D+1}, \hat{\mathbf{c}}) = f(\eta_{N-D+1}, \eta_{N-D+2}, ..., \eta_N, \hat{\mathbf{c}}), \qquad (10.17)$$

2) принимают предсказанное значение $\hat{x}_{N+1}$ в качестве последней координаты нового вектора состояния $\hat{\mathbf{x}}_{N-D+2} = (\eta_{N-D+2}, \eta_{N-D+3}, ..., \eta_N, \hat{x}_{N+1})$,

3) используют вектор $\hat{\mathbf{x}}_{N-D+2}$ как аргумент функции *f*, получают новый прогноз $\hat{x}_{N+2} = f(\hat{\mathbf{x}}_{N-D+2}, \hat{\mathbf{c}})$, и так далее.

В итоге получим прогноз $\hat{x}_{N+l}$ с любым упреждением *l* (число шагов)

Есть и альтернатива: для прогноза с упреждением *l* можно вместо *l*-кратного итерирования формулы (10.5) строить модель, аппроксимирующую *непосредственно* (прямо) зависимость η_{i+l} от $\mathbf{x}_i$. Форма этой зависимости при большом *l* может быть очень сложной в случае хаотической динамики из-за чувствительной зависимости будущего поведения η_{i+l} от начальных условий $\mathbf{x}_i$. В результате для очень большого *l* будет получена модельная функция примерно равная среднему значению наблюдаемой $f \approx \langle \eta \rangle$, т.е. совсем невысокая точность прогноза.[20] Однако для умеренных *l* прямой метод может иметь преимущества.

Как зависит дальность прогноза для обоих способов от длины тренировочного ряда *N* и других факторов? На этот вопрос можно ответить теоретически для локальных моделей с многочленом порядка *K*. По оценкам [224, 206] ошибка прогноза при итерационном методе растет с упреждением *l* как $\sigma_{\Delta M} \cdot e^{\Lambda_1 l \Delta t}$, а при прямом методе – как $\sigma_{\Delta M} \cdot e^{(K+1) H l \Delta t}$, где *H* –сумма положительных ляпуновских показателей. Отсюда видно, что скорость роста ошибки для прямого метода больше (причина указана: трудно аппроксимировать зависимость *далекого* будущего от настоящего). Однако следующий отсюда вывод о преимуществе итерационного способа прогноза справедлив не всегда, а только при некоторых условиях, а именно, модель должна давать одношаговый прогноз с большой точностью, для чего нужен, как правило, очень длинный тренировочный временной ряд и отсутствие шумов. Если же эти условия нарушаются, то, как показывает практика, для прогноза с упреждением *l*, превышающим один шаг, но меньшим характерного времени разбегания близких траекторий, лучше использовать прямой метод [251]. Это объясняется тем, что «одношаговая» эмпирическая модель (10.5) может содержать систематические ошибки (например, из-за неудачного выбора вида функций для аппроксимации или недостаточной размерности модели),

[20] Нетрудно увидеть, что для *l*, превышающих длину тренировочного временного ряда, прямой метод использовать невозможно из-за отсутствия нужных данных.

накопление которых при итерациях может дать больший вклад в ошибку прогноза, чем ошибка аппроксимации при прямом способе.

Для улучшения прогнозов с умеренным упреждением предложен комбинированный подход «предиктор – корректор» [251], который состоит в следующем. Делают прогноз итерационным или прямым путем с помощью имеющейся модели, назовем ее «базовым предиктором». Затем «подправляют» предсказанные значения с помощью дополнительного модельного отображения – так называемого «корректора», – которое строится по тренировочному ряду и связывает ошибки прогнозов базового предиктора для упреждения l с самими прогнозами. Для корректора используют значительно более простую структуру, чем для базового предиктора. Комбинация «предиктор – корректор» может дать существенно более точный прогноз по сравнению с «чистыми» прямым и итерационным способами.

Наконец, отметим одно существенное, с точки зрения прогноза, различие между динамическими моделями типа (10.5) и явными функциями времени (п. 7.4.1.3). В отличие от явной *экстраполяции* временной зависимости динамическая модель (10.5) опирается на *интерполяцию в фазовом пространстве* и потому оказывается намного эффективнее. Действительно, значение вектора состояния $\mathbf{x}_i$, с которого нужно начинать прогноз, лежит обычно «между» многими векторами тренировочного ряда, которые используются для построения модели (см., например, рис.10.12). Но если при итерациях модельного отображения значение $\mathbf{x}$ выйдет из области, в которой лежат векторы тренировочного ряда, то дальнейшее использование модели для прогноза означает *экстраполяцию в фазовом пространстве*. Тогда надежность прогноза резко снизится, а траектория модели может вести себя без всякого сходства с наблюдаемым процессом, например, уйти на бесконечность. Последнее особенно часто имеет место при использовании алгебраических многочленов, которые очень плохо экстраполируют.

Различные виды модельных отображений. Сопоставим прогностические возможности моделей (10.5) с различными видами функции f.

Алгебраические многочлены умеренного порядка K очень эффективны для аппроксимации функций одного переменного, причем плавно меняющихся – без скачков и изломов. Лучше них в этом случае только сплайны порядка 3 и выше [85, 153]. Чем больше необходимая размерность модели D и порядок многочлена K, тем меньше вероятность успеха.

Дробно-рациональные функции эффективны в тех же случаях, но могут лучше описывать зависимости с областями быстрого изменения.

Тригонометрические многочлены и вейвлеты (п. 6.4.2) тоже являются способами слабой аппроксимации и хороши для аппроксимации функций с определенными свойствами (п. 7.2.4).

Радиальные базисные функции [250] несравнимо эффективнее упомянутых подходов при большой размерности модели (ориентировочно, больше 3). ИНС имеют примерно такие же свойства, а по некоторым указаниям [206, 335] ИНС аппроксимируют сложные зависимости даже лучше. Требования к объему данных и уровню шума для всех перечисленных моделей относительно не строгие, т.к. это глобальные модели.

Локально-линейные модели очень эффективны при умеренных размерностях вектора состояния (больше 1, но меньше некоторого небольшого числа, зависящего от длины тренировочного ряда), длинных рядах (чтобы найти значительное число близких соседей каждого вектора) и малом измерительном шуме. Требования к объему данных и уровню шума очень строгие. Локально-постоянные модели лучше, чем локально-линейные при большем шуме и более коротком ряде.

Все подходы страдают от уже упоминавшегося «проклятия размерности». Очень высокоразмерные системы (ориентировочно, размерности около 10 и выше) на настоящем этапе не могут быть сколько-нибудь успешно описаны такими эмпирическими моделями.

Сопоставление модельных отображений и дифференциальных уравнений. В общем случае отображения дают лучший прогноз с умеренным упреждением, чем ОДУ [309]. Это можно понять по аналогии с тем, что итерационный прогноз хуже прямого при значительных погрешностях модели (10.5), обеспечивающей одношаговое предсказание. Модельные ОДУ строят так, чтобы точнее аппроксимировать скорость изменения вектора состояния $d\mathbf{x}(t_i)/dt$ в зависимости от $\mathbf{x}$ (9.3) и, следовательно, дать наилучший прогноз на ближайший момент времени: $\mathbf{x}(t_{i+1}) \approx \left(d\mathbf{x}(t_i)/dt\right)\Delta t$. Если использовать ОДУ для прогноза более далекого будущего, это похоже на итерационный прогноз. Он может оказаться совсем не точным при наличии систематической погрешности аппроксимации ОДУ.

Для долговременного описания динамики эмпирические модельные отображения тоже зачастую оказываются лучше [309]. Кроме того, их проще строить и использовать – не нужно численного дифференцирования рядов и численного интегрирования уравнений.

Модельные ОДУ могут быть хороши, если они «родные» для объекта, т.е. его динамика действительно почти точно подчиняется системе ОДУ со структурой, используемой при моделировании. Но это более типично для постановки «прозрачный» или «серый ящик» и очень маловероятно без априорной информации.

Тот факт, что многие авторы все-таки занимаются построением модельных ОДУ для «черного ящика», отчасти связан с проблемой «прозрачности». После получения «хорошей» модели желательно понять, как она «работает» и физически интерпретировать ее переменные и параметры. Для ОДУ с многочленами в правой части есть надежда на такое физическое истолкование, т.к. асимптотические модели многих реальных процессов имеют вид ОДУ с многочленом в правой части, например, в уравнения химической кинетики и динамики лазеров. Потому же часто используют ОДУ с последовательными производными (10.13), а не векторами задержек: последовательным производным можно придать смысл скорости, ускорения и т.д. Но обычно надежда на физические интерпретации не оправдывается: если в структуру ОДУ заранее не заложены физические соображения [38, 194], то «вытащить» физический смысл из алгебраического многочлена (10.14) или подобной универсальной структуры невозможно.

10.4. Диагностическая проверка модели

Хотя в предположении о наличии динамического шума уместно проводить *анализ остатков* (проверка их некоррелированности и нормальности, п. 7.3), при проверке адекватности динамических моделей обычно опираются на расчет и сопоставление с экспериментом тех характеристик модели, которые популярны в теории динамических систем. Перечислим основные возможности.

1) Для детерминированной модели *дальность прогноза* можно оценить по формуле $\tau_{\text{pred}} = (1/2\Lambda_1)\ln\left(\sigma_x^2/\sigma_\nu^2 + \sigma_\mu^2 + \sigma_{\Delta M}^2\right)$. Эта оценка совпадает с экспериментально полученной величиной для адекватной модели.

2) *Качественное сходство проекций траекторий на плоскости различных переменных.* Это – субъективный критерий, хотя и очень важный. Он направлен на оценку сходства *существенных особенностей динамики* объекта и модели. Разные варианты придания ему более точного количественного выражения приводят к разным способам проверки адекватности модели и указаны ниже, следуя обзору [233].

3) *Сравнение инвариантных мер* (плотностей распределения векторов в пространстве состояний) или их проекций – плотностей распределения одной из переменных. Подход применим и для стохастических моделей.

4) *Сравнение старшего ляпуновского показателя* модели с оценкой, полученной по наблюдаемому временному ряду.

5) *Сравнение фрактальных размерностей и энтропий* аттрактора модели с оценками, полученными по наблюдаемому временному ряду.

6) *Сравнение топологических свойств.* Это тонкий подход, основанный на поиске и анализе неустойчивых периодических орбит,

вложенных в аттрактор, и на определении их взаимного расположения в фазовом пространстве. Он применим, строго говоря, только для детерминированных систем размерности не выше 3 и представляет собой очень строгий тест для модели. Если она воспроизводит большую часть неустойчивых орбит, найденных по наблюдаемому временному ряду, это уже существенное свидетельство в пользу ее адекватности.

7) *Сравнение отображений Пуанкаре*. Это легко сделать для одномерных отображений Пуанкаре. Как правило, анализируется зависимость следующего максимума наблюдаемой величины от предыдущего. Подход имеет отношение к анализу топологических свойств аттракторов и часто используется в связке с ним.

8) *Синхронизация модели сигналом от объекта*. Модель признается адекватной, если она синхронизуется (с заданной точностью) наблюдаемым временным рядом при умеренной интенсивности воздействия [201].

9) Предлагалось также выяснять, имеет ли модель то же самое количество аттракторов того же типа, что и объект; расположены ли эти аттракторы в соответствующих областях фазового пространства; совпадают ли бассейны их притяжения. Это очень строгие требования к модели и ни одна эмпирическая модель в типичном случае их не выдержит.

В заключение главы 10 отметим, что мы почти обошли вопросы моделирования распределенных систем в виде дифференциальных уравнений в частных производных и других конструкций, хотя этой задаче в последнее время уделяется пристальное внимание [189, 284, 306, 330]. Мы только кратко коснулись построения стохастических модельных уравнений [329, 327, 307]. Разнообразную полезную информацию по этим и смежным вопросам можно найти в статьях, помещенных на сайты [341, 349, 348, 342].