Глава 6. Ряды наблюдаемых – источник данных для моделирования

При создании модели из «первых принципов» модельные величины наследуют смысл, заложенный в общих законах природы и производных от них формулировках типа уравнений Кирхгофа. При эмпирическом моделировании по временным реализациям выяснение взаимного соответствия параметров модели и объекта становится отдельной задачей. Далеко не всегда можно организовать измерение значений всех величин, включенных в модель, в принципе или с необходимой точностью (по различным техническим причинам или из-за специфики объекта). Приходится довольствоваться тем, что доступно, и учитывать это обстоятельство, проводя дополнительные преобразования данных при построении модели.

В этой главе мы остановимся на проблемах получения и обработки информативных сигналов с объекта с целью получения представлений о возможностях и специфике его моделирования – на том, что в схеме процесса моделирования на рис.5.1 значится как «получение и анализ временных рядов наблюдаемых» и «априорная информация».

6.1. Наблюдаемые и модельные величины

6.1.1. Наблюдения и измерения

Необходимую для моделирования информацию мы получаем, организовав *наблюдение* за объектом с позиций имеющихся представлений – априорных моделей (см. п. 1.1). Результат – качественные или количественные данные. К качественным выводам может привести просто *созерцание*, количественные данные добываются путями измерения и подсчета. *Подсчет* используют, например, для регистрации числа испускаемых частиц, особей в популяциях и в других подобных ситуациях, когда наблюдается совокупность или последовательность дискретных элементов. *Измерение* – сравнение с однородной величиной, принятой за единицу, носителем которой являются эталоны различных уровней. Далее, говоря о «наблюдении», «наблюдаемых величинах» и «результатах наблюдения», мы будем иметь в виду, соответственно, сами процессы измерения или подсчета, величины, которые этим процедурам подвергаются, и полученные количественные данные [125]. Следуя традиции, слово «величины» будем отбрасывать и говорить просто о *наблюдаемых*. Для их обозначения используем букву η, похожую на вопросительный знак, подчеркивая этим наличие непростого вопроса об их связи с модельными величинами.

Любая реальная система обладает неограниченной совокупностью

свойств, но для достижения целей моделирования (исследования) бывает достаточно ограничиться некоторым их конечным набором – *модельными величинами* – переменными $x_1, x_2, ..., x_D$ и параметрами $c_1, c_2, ..., c_P$. Число наблюдаемых величин и переменных модели, да и сами они (по своему содержанию, физическому смыслу) могут находиться в различном соответствии. Так, число k наблюдаемых $\eta_1, \eta_2, ..., \eta_k$ обычно меньше размерности модели D, а сами эти величины часто не те, что заложены в модель, а лишь некоторым образом связаны с ее переменными. Это может быть связано с условиями экспериментов, возможностями доступа к объекту, его формой или габаритами, отсутствием необходимой аппаратуры, несовершенством инструментов и многими другими объективными и субъективными факторами. Например, возможности электрических измерений внутри биологической клетки ограничены ее малыми размерами (рис.6.1,а). Колеблющаяся струна на гитаре большая, доступ к ней открыт и можно измерить, например, скорость или координату каждой точки этого механического резонатора, а доступ к внутреннему объему закрытого резонатора можно реализовать лишь через отверстия в его стенках, т.е. после частичного разрушения объекта (рис.6.1,б,в).

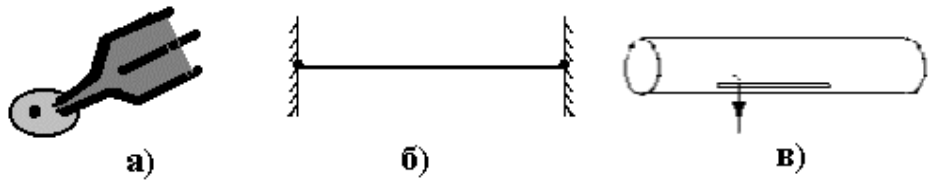

Рис.6.1. Электрическое подключение к клетке с помощью стеклянного капилляра (а); открытый (б) и закрытый (в) резонаторы

Связь наблюдаемых и модельных величин может оказаться очевидной, они могут даже совпадать. Но чаще простой однозначной зависимости между наблюдаемыми и модельными величинами, не бывает. В общем случае этот вопрос требует специального анализа, например:

1) при моделировании эволюции биологической популяции отображением, показывающим зависимость количества особей в последующий год x_{n+1} от их количества в предыдущий год x_n. Если экспериментальные данные η собираются наблюдателями путем непосредственного подсчета, то *модельная и наблюдаемая величины совпадают*: $x = \eta$. Но если наблюдению доступны только следы жизнедеятельности особей (следы на земле, отметины на деревьях, помет и т.п.) и по ним опосредованно делается вывод об их численности, то требуются какие-то формулы пересчета или методики для установления *зависимости между x и* η;

2) физикам хорошо знакома ситуация, складывающаяся при измерении электродвижущей силы (э.д.с.) E источника с внутренним сопротивлением r_i с помощью вольтметра (рис.6.2,а,б). Здесь объект – источник тока, модельная величина – э.д.с ($x = E$), наблюдаемая – напряжение U на клеммах источника ($\eta = U$). Для фиксации наблюдаемой вольтметр подключается к источнику, как это показано на рис.6.2,а. Сопротивление между входными клеммами реального вольтметра R_v (входное сопротивление) не может быть бесконечным, поэтому при подключении прибора цепь замкнется и потечет ток I. При этом напряжение на клеммах уменьшится по сравнению со случаем, когда прибора не было: $U = E - Ir_i = E\left(1 - r_i/\left(r_i + R_v\right)\right)$. Таким образом, наблюдаемая будет отличаться от величины, заложенной в модель, но *имеется однозначная функциональная связь* наблюдаемой и модельной величин $\eta = f(x)$;

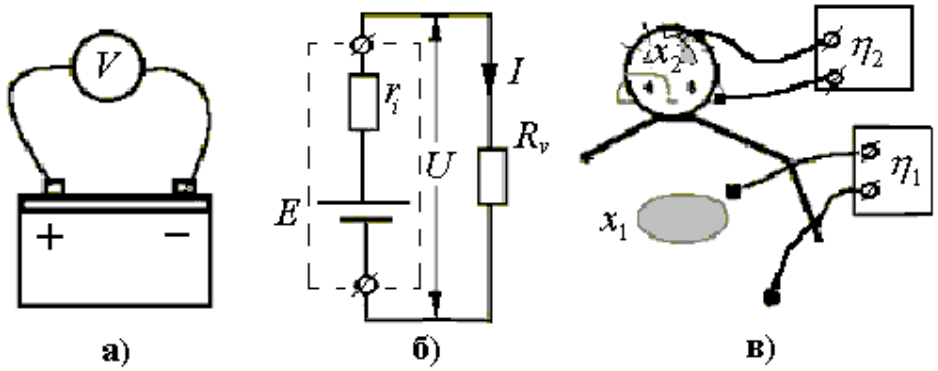

Рис.6.2. Примеры характеризующих величин и наблюдаемых: а) эксперимент с источником эдс, б) эквивалентная схема процесса измерения эдс; в) наблюдаемые величины при анализе активности сердца и мозга

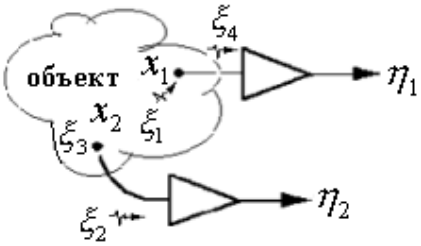

Рис.6.3. Схема получения данных об объекте наблюдения. x_i – величины, характеризующие моделируемое свойство объекта, η_i – наблюдаемые. Точки – датчики, с которых сигналы с шумами ξ_i поступают на измерительные

3) при записи электрических потенциалов на коже человека наблюдаемой η служит напряжение между двумя точками на поверхности тела, одну их которых принимают за опорную (см. рис.6.2,в). При съеме электроэнцефалограмм (ЭЭГ) это, как правило, напряжение между точками на коже головы и на ухе, для электрокардиограмм (ЭКГ) – между точкой на груди и на ноге. Не надо быть специалистом, чтобы согласиться с тем, что измеряемые напряжения являются лишь сильно трансформированными отголосками процессов в ансамблях клеток мозга или в

сердце. Поэтому связь наблюдаемых потенциалов η с величинами, характеризующими функционирование человеческого организма x, представляет предмет самостоятельного изучения. Но если моделируется сама запись потенциалов, а не их источники (аналогично кинематическому описанию в механике), то модельная и наблюдаемая величины совпадают.

При любом измерении на результат влияют особенности методики эксперимента и параметры приборов, внешние регулярные воздействия и шумы. На рис.6.3 изображена схема типичной процедуры получения экспериментальных наблюдаемых: внешней извилистой линией ограничен объект моделирования, черными точками обозначены датчики измеряемых величин, треугольники – приборы (усилители, ограничители и т.п.), преобразующие данные с датчиков в наблюдаемые η_i. Даже если доподлинно известно, как деформируется измеряемый сигнал приборами, и удается ввести соответствующие поправки,[1] на практике невозможно полностью избавиться от помех – влияния внешних регулярных и случайных воздействий[2]. Регулярные воздействия реально подавить, а нерегулярные, называемые шумами, – только ослабить. Шумы могут иметь как внешнее происхождение (их называют *измерительными*), так и быть присущими самому объекту – *внутренние* или *динамические* (т.к. влияют на динамику объекта). На рис.6.3 шумовые добавки представлены извилистыми линиями со стрелками. Измерительный шум может быть аддитивным – $\eta = f(x) + \xi$, мультипликативным – $\eta = f(x) \cdot \xi$, или входить в формулы для связи наблюдаемых и модельных величин более сложным образом.

Вид функции f, связывающей наблюдаемые и модельные величины, определяется и свойствами измерительных приборов и передающих цепей. Чтобы эти устройства не вносили искажений, они должны обладать

1) достаточно широким динамическим диапазоном $U_{\max} - U_{\min}$, позволяющим одинаково, не нарушая пропорций, передавать большие и малые сигналы. Например, в представленном на рис.6.4 примере динамический диапазон устройства недостаточен для передачи большого сигнала;

2) необходимым частотным диапазоном (полосой пропускания) – возможностью воспринимать и одинаково передавать все гармонические

---

[1] Например, в случае измерения эдс вольтметром можно ввести поправку на конечность R_v или использовать классический бестоковый компенсационный метод измерения с помощью потенциометра [84, с. 162-163], свободный от отмеченного недостатка.

[2] Так, опыт показывает, что и в случаях, показанных на рис.6.2, и аналогичных ситуациях измерения электрических сигналов чувствительный широкополосный измеритель (например, радиоприемник), подключенный параллельно вольтметру, кардиографу и т.п., дополнительно обнаружит шум, человеческую речь и музыку и т.п.

составляющие спектра мощности (п. 6.4.2.1) входного сигнала[3] и невосприимчивостью к сигналам на «чужих» частотах. Излишне широкая полоса частот, пропускаемых устройством, нежелательна из-за опасности проникновения в наблюдаемый сигнал помех и шумов, но и ее уменьшение чревато внесением искажений и ростом инерционности – временем τ_u реакции на короткий импульсный сигнал (см. рис.6.4,в). Узкую полосу пропускания преобразующего устройства часто используют для целенаправленной фильтрации сигнала. Пропускание только «низкочастотной» части спектра приводит к его сглаживанию. Фильтр высоких частот удаляет постоянную составляющую и *тренд* (медленный дрейф);

3) достаточной чувствительностью – уровень собственных шумов устройства должен быть достаточно мал, чтобы на их фоне можно было уверенно различать полезный входной сигнал на выходе.

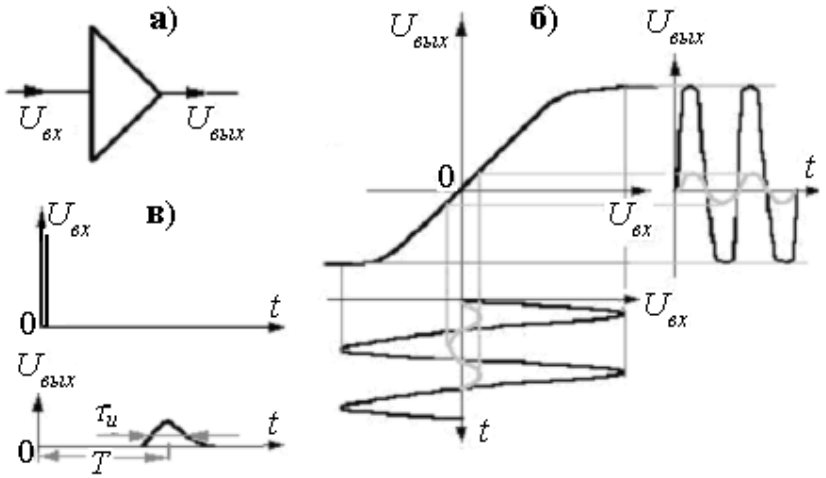

Рис.6.4. Устройство, передающее сигнал (а), и его передаточная характеристика $U_{вых} = f(U_{вх})$ (б). Малый входной гармонический сигнал (серая кривая) передается без искажений, а большой (черная кривая) ограничен, его верхушки срезаны. При передаче короткого импульса, поступившего на вход в момент $t = 0$, отклик появится на выходе с некоторой задержкой T расплывшимся из-за конечного времени инерционности устройства τ_u (в)

Сказанное о неидеальности измерительных устройств, о соответствии модельных величин и наблюдаемых имеет отношение и к статистическим данным, используемым при моделировании общественных процессов и гуманитарных проблем. Их искажение людьми и временем (в историческом смысле) может быть весьма существенно и неконтролируемо. При моделировании по таким данным надо быть тем более осторожным, если даже радиоизмерения с использованием

[3] Иметь одинаковый коэффициент передачи $U_{вых}/U_{вх}$ для всех компонент спектра.

достаточно «объективных» приборов и методов борьбы с шумами оставляют место сомнениям и заставляют быть внимательным при использовании данных наблюдения.

6.1.2. Методы увеличения и уменьшения числа характеризующих величин

Очень часто наблюдаемых величин не хватает, чтобы обеспечить динамическое описание объекта – задать его состояние **x** (D-мерный вектор), обеспечив однозначный прогноз будущих состояний по текущему. Для увеличения числа переменных (размерности вектора **x**) существует несколько подходов. Часть из них имеет теоретическое обоснование (см. теоремы Такенса в п.10.1.1), другие опираются на интуицию и т.п.

Самый простой и популярный способ – метод временных задержек. В случае скалярной наблюдаемой, согласно этому подходу, в качестве компонент вектора $\mathbf{x}(t)$ берутся ее последовательные значения, разделенные некоторым интервалом τ (временем задержки): $x_1(t)=\eta(t)$, $x_2(t)=\eta(t+\tau)$, …, $x_D(t)=\eta(t+(D-1)\tau)$.

Метод последовательного дифференцирования состоит в использовании временных производных наблюдаемой в качестве динамических переменных: $x_1(t)=\eta(t)$, $x_2(t)=d\eta(t)/dt$, …, $x_D(t)=d^{D-1}\eta(t)/dt^{D-1}$. Его применение затруднительно при наличии измерительных шумов, что иллюстрирует, например, рис.7.8,а,б: чуть заметная «бахрома» на графике немного зашумленной гармонической зависимости $\eta(t)=\cos\ t+\xi(t)$ после дифференцирования может нарастать в такой степени, что не позволит даже заметить профиль ожидаемой синусоиды на графике $d\hat{x}/dt$.

Имеющиеся шумы частично подавляются при использовании для получения дополнительной переменной процедуры интегрирования – использовании интегралов с переменным верхним пределом от наблюдаемой: $x_2(t)=\int_0^t \eta(t)dt$, но при этом теряется часть информации о наблюдаемой. С помощью взвешенного суммирования можно составить переменную из последовательных значений наблюдаемой: $x_2(t)=a_0\eta(t)+a_1\eta(t-\Delta t)+a_2\eta(t-2\Delta t)+\ldots$, где a_k – весовые коэффициенты. Возможны комбинации всех упомянутых подходов и другие способы получения рядов модельных переменных (см. п.10.1.2).

Однако встречаются ситуации, когда структура уравнений динамической системы известна, но получить значения некоторых динамических переменных по наблюдаемым невозможно. При этом говорят, что часть переменных являются «скрытыми». В таком случае

определенные шансы на успех дают специальные приемы моделирования, описанные в п. 8.2.

На практике может возникать и необходимость сократить число наблюдаемых, если некоторые из них не несут дополнительной информации об объекте, полезной для моделирования. Это можно делать, например, анализируя по наблюдаемым данным взаимные зависимости между величинами и устраняя те величины, которые являются линейными комбинациями других. В некоторых ситуациях могут оказаться полезными *методы размерности и подобия* [156, 163]. Эффектный исторический пример перехода от величин, измеряемых непосредственно, к меньшему числу безразмерных комбинаций размерных величин (критериев подобия) можно найти в [23]; он связан с решением задачи о протекании жидкостей по трубам.

6.2. Аналого-цифровые преобразователи

При измерении величин различной природы, как постоянных, так и изменяющихся во времени, часто стараются преобразовать их в электрические напряжения и токи. Электрические сигналы легко передать с пространственно удаленных датчиков, для их обработки существует арсенал стандартных приборов и методик. Ранее измеряемые значения фиксировались отклонениями стрелок гальванометров, следами краски на лентах самописцев и свечением на экранах осциллографов. Но современные измерительные системы, как правило, представляют данные в цифровом виде, используя для этого специальные транзисторные устройства – *аналого-цифровые преобразователи* (АЦП). Задача аналого-цифрового преобразования состоит в преобразовании напряжения на его входе в момент измерения в пропорциональное ему число, а если напряжение меняется – в дискретную последовательность чисел [27]. При этом неискаженное представление формы сигнала требует соблюдения условий теоремы Котельникова: по дискретной числовой последовательности только тогда можно без погрешностей восстановить непрерывный исходный сигнал, когда частота выборки по меньшей мере вдвое превышает наибольшую частоту, присутствующую (соответствующую ненулевой компоненте) в спектре мощности исходного сигнала.

Познакомимся с принципом работы АЦП и причинами, ограничивающими точность получаемых с их помощью данных, на примере схемы рис.6.5. В ней реализован так называемый параллельный подход: входное напряжение одновременно сравнивают с n опорными (эталонными) напряжениями. Количество эталонных напряжений и интервал между ними определяется диапазоном измеряемых значений и требуемой точностью преобразования – числом двоичных разрядов

оцифрованного значения на выходе прибора. Для реализуемого в примере трехразрядного представления, позволяющего записать восемь различных чисел, включая нуль, требуется семь соответствующих эквидистантных эталонных напряжений. Они формируются с помощью резистивного делителя опорного напряжения $U_{\text{опор}}$, которое определяет верхний предел измеряемых напряжений (на схеме оно обозначено как $7U$).

Измеряемое напряжение $U_{\text{с}}$ сравнивается с эталонными уровнями с помощью семи компараторов k_i, напряжения на выходе которых принимают значения, считающиеся равными

- 1 (в двоичной системе счисления), когда напряжение на входе (+) превышает напряжение на входе (–),
- 0 – в противном случае.

Так, если измеряемое напряжение не выходит за пределы диапазона от $5U/2$ до $7U/2$, то компараторы с 1-го до 3-го устанавливаются в состояние «1», а компараторы с 4-го по 7-й – в состояние «0». Специальная логическая схема (приоритетный шифратор) преобразует эти состояния в двоичное число $z_1 z_2 z_3$ (в примере 011) или в соответствующее ему десятичное (в примере 3). Если же напряжение меняется во времени, то приоритетный шифратор нельзя подсоединять непосредственно к выходам компараторов (может быть получен ошибочный результат), поэтому с помощью схем выборки-хранения, обозначенных на схеме прямоугольниками, запоминают мгновенное[4] значение напряжения на выходе компараторов и удерживают их неизменными в течение времени измерения. Моменты, выбранные для измерения, диктуются специальными тактовыми сигналами; если они периодические, то на выходе шифратора окажется последовательность двоичных чисел – временной ряд.

[4] Реально время срабатывания этих элементов тоже конечно, но оно должно быть меньше интервала дискретизации.

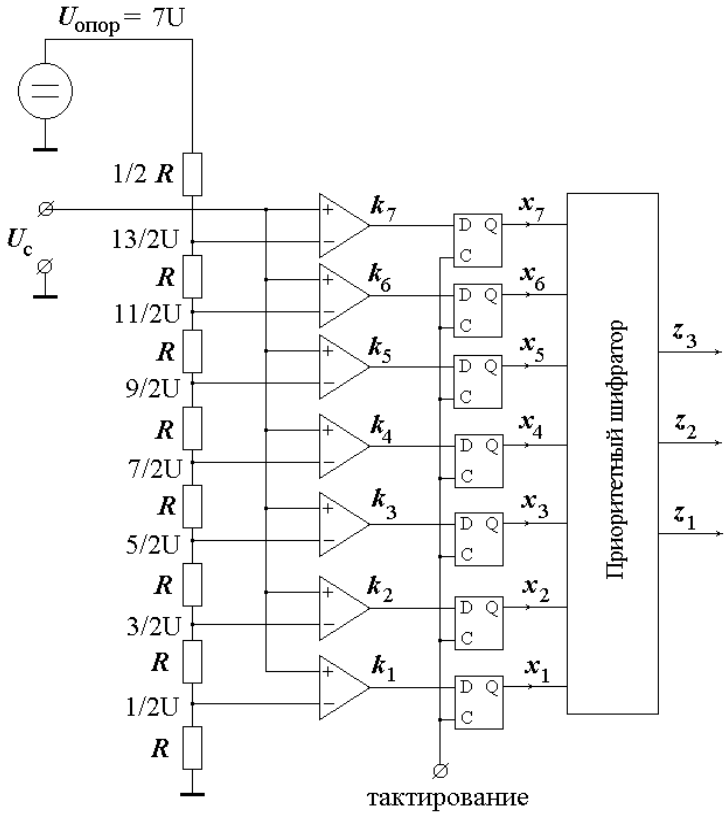

Рис.6.5. Схема 3-разрядного АЦП, реализующая параллельный подход

Описанное преобразование аналоговой величины в число с конечным количеством разрядов сопровождается «ошибкой квантования», составляющей величину, равную половине приращения входного напряжения, которое необходимо для изменения кода в младшем разряде. 8-разрядный АЦП имеет $2^8 = 256$ градаций: $\Delta x = x_{\max}/256$; 12-разрядный АЦП имеет $2^{12} = 4096$ градаций: $\Delta x = x_{\max}/4096$.[5] Наряду с этим существуют ошибки, обусловленные схемой, дрейфом параметров и их нелинейностью, так что погрешность значений наблюдаемых величин определяется комбинацией этих факторов и указывается в паспорте прибора.

Описанный параллельный метод реализации аналого-цифрового преобразования неэкономен – для каждого эталонного уровня необходимо свое сравнивающее устройство. То есть для диапазона измеряемых значений от 0 до 100 с единичным шагом необходимо 100 компараторов, а при большем разрешении и того больше. Поэтому были разработаны и широко используются более совершенные в этом отношении подходы,

[5] Если с помощью цифро-аналогового преобразователя произвести обратное преобразование полученного числа в напряжение, ошибка квантования проявляется в виде накладывающегося «шума».

например, *весовой, числовой*.[6] При *весовом методе* результат не может быть получен за один шаг, поскольку на каждом шаге определяется лишь один разряд двоичного числа. Сначала устанавливают, превышает ли входное напряжение опорное напряжение старшего разряда. Если оно выше, то старший разряд получает значение «1» и из входного напряжения вычитается опорное. Остаток сравнивают с соседним младшим разрядом и т.д. Очевидно, что для этого необходимо столько шагов сравнения, сколько разрядов в числе и сколько опорных напряжений. Простейший метод – *числовой.* В этом случае подсчитывается число суммирований опорного напряжения младшего разряда, необходимое для получения входного напряжения. Если максимальное число, которое может быть представлено, равно n, то необходимо, следовательно, максимум n шагов для получения результата. На практике часто используют комбинации различных способов.

6.3. Временные ряды

6.3.1. Термины

При использовании АЦП и во многих других ситуациях данные об исследуемом объекте (процессе) представляют собой *конечную последовательность значений наблюдаемой величины в различные моменты времени*, т.е. *временной ряд*: $\eta(t_1),\eta(t_2)...,\eta(t_N)$, где $t_1,t_2,...,t_N$ – моменты наблюдений, а их число N называется *длиной ряда*. Если в каждый момент времени t_i измеряется значение только одной скалярной величины, то временной ряд называют *скалярным*. Используется обозначение $\{\eta(t_i)\}_{i=1}^N$ или $\{\eta_i\}_{i=1}^N$, где $\eta_i = \eta(t_i)$. Если же одновременно в момент t_i измеряются значения k величин $\eta_1,...,\eta_k$ (рис.6.3), то ряд называют *векторным*[7], т.к. эти величины можно рассматривать как компоненты k-мерного вектора $\boldsymbol{\eta}$. Векторный ряд обозначают так же, как и скалярный, – $\{\boldsymbol{\eta}(t_i)\}_{i=1}^N$ или $\{\boldsymbol{\eta}_i\}_{i=1}^N$.

Элементы временного ряда (числа или векторы) называют также *точками*. Порядковый номер точки i называют *дискретным временем*. Если временные интервалы между последовательными моментами наблюдений t_i одинаковы – $t_i - t_{i-1} = \Delta t, i = 2,...,N$, то ряд называют *эквидистантным*, в противном случае – неэквидистантным. Говорят

[6] В отечественной литературе их часто называют методом поразрядного взвешивания и методом последовательного счета, соответственно.

[7] Обозначение η_i с одним нижним индексом используется нами, таким образом, в двух разных случаях: 1) скалярный временной ряд, i означает временной индекс; 2) векторная наблюдаемая, i означает номер координаты. Значение определяется контекстом.

также, что выборка произведена равномерно или неравномерно, соответственно. Интервал Δt между последовательными измерениями называют *выборочным интервалом* или интервалом дискретизации. В случае неэквидистантного ряда выборочный интервал есть величина переменная: $\Delta t_i = t_i - t_{i-1}$. На практике чаще встречаются и используются эквидистантные ряды.

С точки зрения моделирования полезно сразу заметить, что после того, как модель по временному ряду построена, для проверки ее качества (подробнее см. пп. 7.3, 8.2.1.4, 10.4) желательно иметь еще временные ряды от того же процесса, которые не использовались при построении модели. Поэтому, если позволяет объем экспериментальных данных, во временном ряде $\{\eta_i\}_{i=1}^{N}$ выделяют две части. Одну из них используют для построения (подгонки параметров) моделей и называют *тренировочным* временным рядом. А другую – используют для проверки модели и называют *тестовым* временным рядом.[8]

6.3.2. Примеры

Рассмотрим некоторые примеры временных рядов из различных областей практики и задачи, которые решаются с опорой на такие данные. Первая из этих задач – прогноз будущего развития процессов.

Метеорология (наука об атмосферных явлениях). На метеостанциях ежечасно измеряются различные показатели: температура воздуха и его влажность, атмосферное давление, скорость ветра (на различных высотах), количество осадков и т.д. На некоторых из станций записи ведутся более 100 лет. Изменения этих величин, характерные временные масштабы которых порядка 1 суток, относят к *погодным* процессам [124]. Прогноз погоды – это задача *синоптической метеорологии*.

Более медленные процессы, временные масштабы которых превышают 3-4 месяца, относят к *климатическим*. Они изучаются *климатологией*. Это температура поверхности океана (средняя температура верхнего перемешанного слоя, глубина которого – десятки метров, рис.6.6), уровень океана (в различных прибрежных пунктах), толщина ледового покрова, площадь льдов, растительный покров Земли, суммарное месячное количество осадков на данной территории, уровень бессточных озер и т.д. Кроме того, к климатическим процессам

[8] В области искусственных нейронных сетей принята несколько другая терминология. Тестовым рядом называется ряд, используемый для сравнения различных полученных моделей. Он позволяет выбрать наилучшую из них. Чтобы сравнить ее «честно» с объектом используется еще один ряд, который так и называют – ряд для проверки адекватности модели (validation time series). Тренировочный временной ряд часто называют обучающим. Мы придерживаемся варианта, указанного в основном тексте.

принадлежат изменения многих характеристик погоды, усредненных по значительному промежутку времени и/или области пространства: среднемесячное значение температуры воздуха в данном месте, среднее значение температуры в данный момент в пределах 5-градусной широтной зоны, среднегодовое значение температуры воздуха Северного полушария. Что касается усреднения по пространству, то в настоящее время его обеспечивает наличие глобальной сети метеостанций, которые покрывают большую часть поверхности всех материков. Расстояния между соседними станциями составляют примерно 10-100 км.

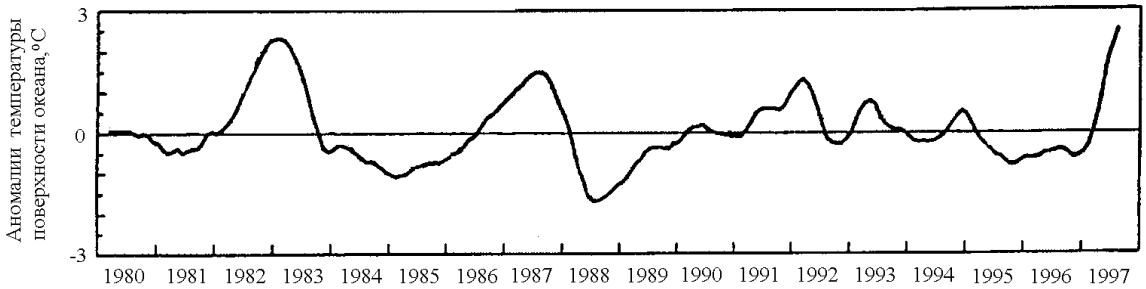

Рис.6.6. Аномалии температуры поверхности в восточной экваториальной зоне Тихого океана (средние значения за 5 месяцев). Интервал выборки Δt = 1 месяц [256]

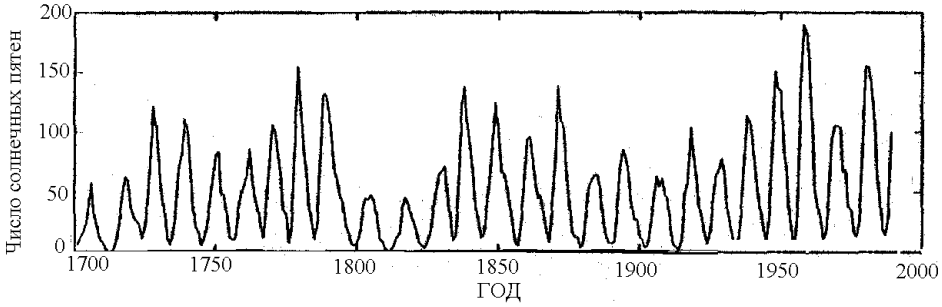

Рис.6.7. Числа солнечных пятен по Вольфу (средние за год) [170]. Интервал выборки Δt = 1 год. Заметен 11-летний цикл Солнечной активности

Физика Солнца. На протяжении долгого периода популярным объектом исследования является временной ряд годового количества пятен на Солнце (рис.6.7). Эта величина фиксируется со времени изобретения телескопа, точнее с 1610 г. (т.е. уже почти 400 лет). Этот процесс отражает магнитную активность Солнца, которая влияет, в частности, на энергию излучения и силу солнечного ветра [170, 250, 260, 338], что в свою очередь приводит к изменениям климата Земли.

Сейсмология. Землетрясения происходят в блочной, пересекаемой многочисленными разломами, сильно неоднородной, твердой оболочке Земли – литосфере – в результате ее деформирования под действием тектонических сил. Уравнения, которым подчиняется процесс, до сих пор не установлены [151]. Есть основания полагать, что на сейсмический режим влияют различные физические и химические процессы: тепловые, электрические и пр. На практике измеряются значения поля механических

напряжений в различных участках земной поверхности. Другая форма представления данных – временные интервалы между двумя последовательными сильными землетрясениями: $\Delta t_i = t_i - t_{i-1}$, где t_i – момент i-го толчка (что похоже на случай переменного выборочного интервала, если не принимать во внимание, что больше ничего не измеряется). Величины Δt_i сильно меняются в зависимости от региона и периода наблюдения.

Для изучения сейсмической активности используют и эквидистантные ряды. Например, измеряют величину, пропорциональную ускорению колебаний земной поверхности, с помощью сейсмографов [95].

Финансы. Для участников событий на рынках ценных бумаг важно уметь предвидеть изменения курсов валют, акций и т.п. На эти процессы влияет множество факторов, они весьма сложны и быстро изменяются. Временные ряды курса валют записываются с выборочным интервалом, например, в 1 час (и даже в 2 минуты), курс акций какой-либо компании фиксируется ежедневно (рис.6.8, 6.9) [45, 99, 112, 263, 277, 208].

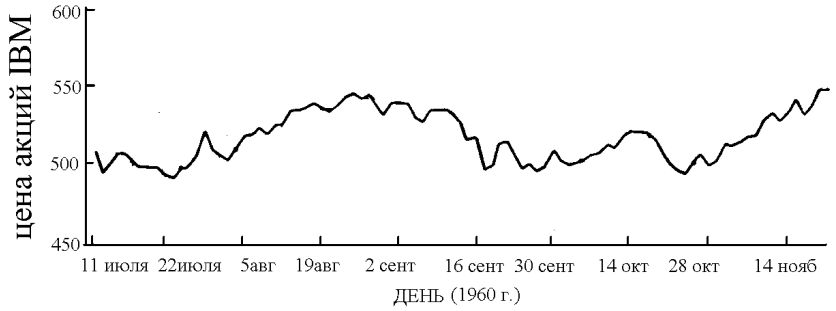

Рис.6.8. Курс акций IBM. Интервал выборки Δt = 1 день (значения на конец дня) [45]

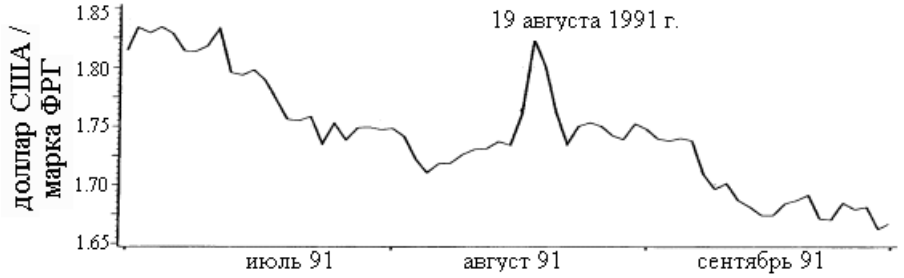

Рис.6.9. Курс доллара США к марке ФРГ в 1991 г. [263]. Интервал выборки Δt = 1 день

Физиология и медицина. Здесь чаще возникает задача не прогноза, а диагностики. Физиологические сигналы, отражающие деятельность сердца, мозга и т.д., – один из основных источников информации для врачей при постановке диагноза. Используются электрокардиограммы (разность потенциалов между различными участками на поверхности груди, рис.6.10), электроэнцефалограммы (потенциалы на коже головы), электрокортикограммы (внутрикорковые потенциалы мозга), электромиограммы (потенциалы внутри мышц и на коже), записи значений

ускорения при колебаниях пальцев вытянутой руки (физиологическом треморе), концентрации кислорода в крови, частоты пульса, объема груди (данные о дыхательной деятельности), и др. Типичный интервал выборки имеет порядок 1 мс.

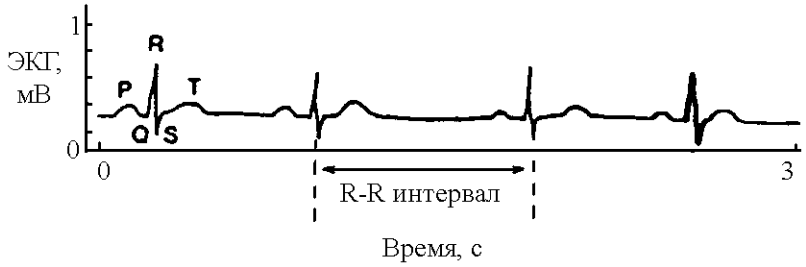

Рис.6.10. Электрокардиограмма [292]. Показан характерный PQRST-комплекс и так называемый R-R интервал. Интервал выборки $\Delta t = 1$ мс

Важно суметь обнаружить признаки заболевания на ранних стадиях. Для этого могут потребоваться весьма тонкие методы анализа сигналов.

Транспорт. В этой области, как и во многих технических приложениях, часто возникает и решается проблема автоматического управления. Например, записываются данные, фиксирующие одновременно курс корабля и угол поворота руля (рис.6.11). Располагая ими, можно спроектировать автоматическую систему, позволяющую повысить эффективность управления кораблем (сократить его «виляния») и снизить расход топлива [111].

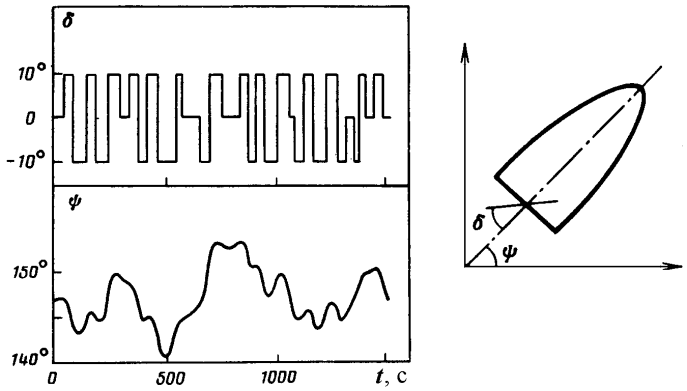

Рис.6.11. Одновременно записанные временные ряды – угол поворота руля (случайные повороты) и курс корабля [111]. Интервал выборки $\Delta t = 10$ с

Гидродинамика. Изучение турбулентного движения жидкости – одна из старейших (и наиболее сложных) задач нелинейной динамики. Хаотические режимы реализуются в экспериментах с жидкостью между двумя цилиндрами, вращающимися в различных направлениях (или в миксере, внутри которого вращается турбина с лопастями). Данные о процессе – это, например, временной ряд значений скорости в некоторой точке жидкости, измеряющийся с типичным интервалом выборки 1 мс [265].

Химия. Значительное внимание исследователей привлекает хаотическое поведение во многих химических реакциях. Данные об этих системах представляются в виде временных реализаций реагентов. Например, концентрация ионов Ce^{IV} в реакции Белоусова – Жаботинского [201, 267].

Лазерная физика. Наблюдаемое сложное поведение лазера при периодическом импульсном характере накачки (рис.6.12) можно использовать для вычисления некоторых параметров лазера и далее для изучения зависимости этих параметров от внешних условий (температуры и т.п.). Данные – временной ряд интенсивности излучения, измеряющейся с помощью фотодиода (в приведенном примере Δt = 1 мкс).

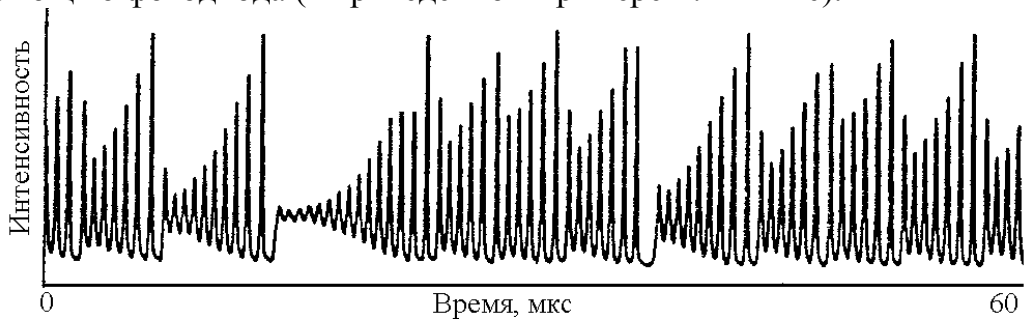

Рис.6.12. Данные с кольцевого лазера в хаотическом режиме [243]. Реализация интенсивности излучения. Интервал выборки Δt = 40 нс

Астрофизика. Фактически единственным источником информации об удаленных от нас объектах Вселенной – звездах, являются временные ряды интенсивности их излучения. Эти ряды получают с помощью радио- и оптических телескопов (рис.6.13).

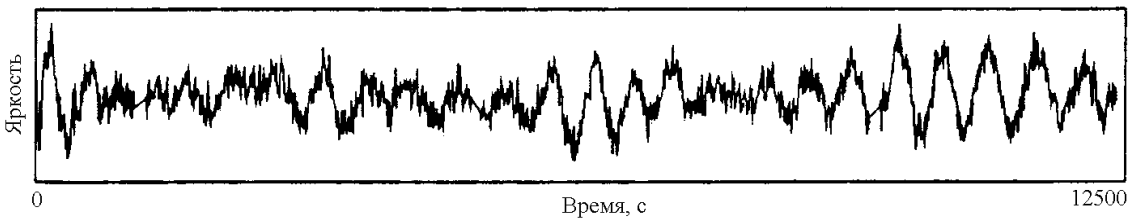

Рис.6.13. Изменения яркости излучения звезды – белого карлика PG-1159. Δt = 10 с [214]. Записывался ряд непрерывно в течение 231 часа

Перечень примеров, задач и объектов можно было бы продолжать. Их количество постоянно растет. Исследователю уже доступны АЦП с $\Delta t = 10^{-8} - 10^{-9}$ с и запоминающие устройства с объемами памяти в сотни гигабайт. В развитии техники получения, запоминания и обработки временных рядов еще отнюдь не наблюдается насыщение. Даже приведенных примеров достаточно, чтобы понять, что временные реализации движений в реальных системах, как правило, имеют очень сложный вид, являются нерегулярными. Тем не менее, сейчас, когда известно, что сложные хаотические решения могут демонстрировать

простые нелинейные динамические системы, задача анализа и моделирования по рядам не кажется безнадежной, хотя и требует разработки непростых технологий решения.

6.4. Элементы анализа временных рядов

6.4.1. «Визуальный» экспресс-анализ

Способности людей различать видимые изображения (картинки) развиты настолько, что при решении задач обработки образов мы можем конкурировать даже со специализированными компьютерами. Считается, что около семидесяти процентов информации человек получает через глаза. Визуальный анализ данных об объекте, если они представлены в графической форме, может оказаться весьма полезным и эффективным при моделировании – подсказать подходящий вид модельных функций, вид и размерность модельных уравнений. Наиболее естественна визуальная оценка изображений временных реализаций процесса – зависимостей величины η от времени t (рис.6.14).

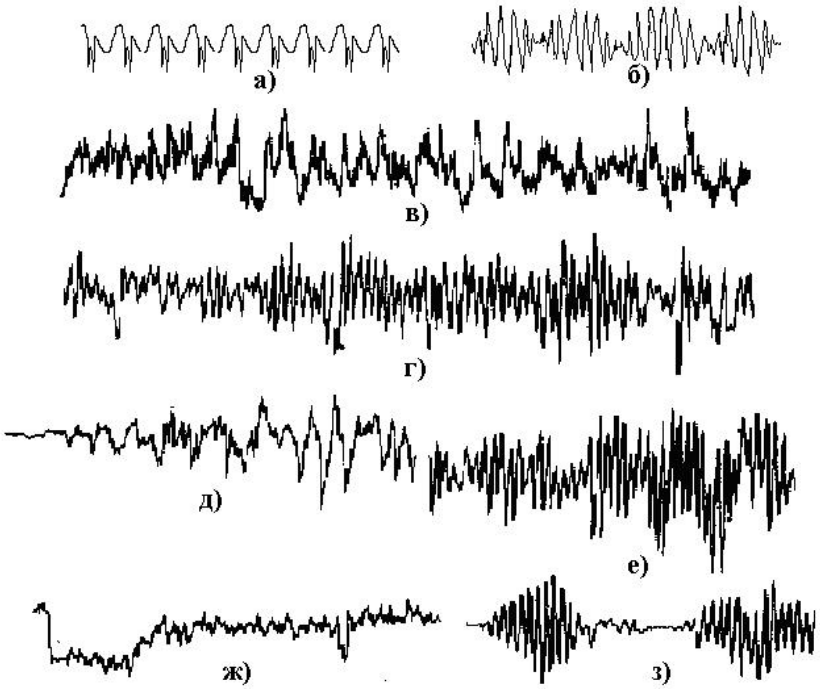

Рис.6.14. Временные реализации регулярных (а,б) и нерегулярных (в-з) сигналов: а) периодический; б) квазипериодический; в) широкополосный стационарный; г) узкополосный стационарный; д) широкополосный, нестационарный по дисперсии; е) узкополосный, нестационарный по математическому ожиданию; ж) широкополосный, нестационарный по математическому ожиданию; з) узкополосный периодически нестационарный по дисперсии

Необходимо рассматривать временные реализации достаточно большой протяженности, чтобы в ней появились особенности движения,

позволяющие его идентифицировать. Так, для *периодических* движений (рис.6.14,а) такой особенностью является полная повторяемость через период *T*. Движение *квазипериодично*, если имеется два или более характерных временных масштабов (периодов гармонических составляющих, суперпозицией которых можно представить данное движение, рис.6.14,б), для которых отношение T_i/T_j иррационально. Периодические и квазипериодические движения называют *регулярными* в отличие от вариантов, чьи реализации представлены на рис.6.14,в-з, – хаотических, без видимой закономерности, *нерегулярных*.[9]

Верхнюю половину рис.6.14 составляют временные реализации *стационарных* процессов, которые качественно и количественно не меняются с течением времени – варианты (а-г), нижнюю – *нестационарных*, варианты (д-з), более подробно см. п.п. 4.1.3 и 6.4.4. Нестационарные движения распознаются (в простых случаях визуально) по наличию во временной реализации качественно или количественно неоднородных этапов.

Если цена деления на оси времени известна, по расстоянию между экстремумами временной реализации и крутизне участков можно оценить характерный период и частоту, и даже диапазон частот гармонических составляющих сигнала (п. 6.4.2.1). Например, расстояние между максимумами в процессе рис.6.14,г почти постоянно вдоль всей реализации, а на рис.6.14,в меняется в широких пределах, при этом на кривой имеются и очень крутые и пологие участки, следовательно, спектр процесса покрывает большую полосу частот, процесс рис.6.14,в широкополосен.

Другой распространенный способ визуальной оценки опирается на процедуру *восстановления фазовой траектории* – по координатным осям откладываются полученные по наблюдаемой динамические переменные (см. пп. 6.1.2, 10.1.2) и наносятся точки, изображающие состояние в отдельные моменты времени (рис.6.15). Возможности визуального анализа фазовых портретов весьма ограниченны. Без специальных средств существует возможность рассматривать только двухмерные проекции фазового портрета на плоском экране (рис.6.15,б). Циклы легко идентифицируются на изображении по малой толщине линий (на рисунке слева). Изображение тора на плоскости (вверху) выглядит, как лента с четкими границами, чем отличается от более «размазанного» изображения проекции хаотического аттрактора (внизу, справа). Более информативны для различения хаотического и квазипериодического движений сечения

[9] В отличие от книги по экспресс-анализу [78], где эти процессы называют случайными, мы подчеркиваем только отсутствие видимой закономерности, т.к. такие реализации могут быть порождены маломерной нелинейной динамической системой (см. главу 3).

фазового портрета (стробоскопические или с помощью выделения экстремумов на временных реализациях переменных). Такое сечение для тора представляет собой замкнутую линию, для цикла – точку или несколько точек, для хаотического множества – сложно структурированный набор линий (рис.6.15,б). Для идентификации сложных непериодических движений анализ фазовых портретов более продуктивен, чем спектральный анализ сигналов.

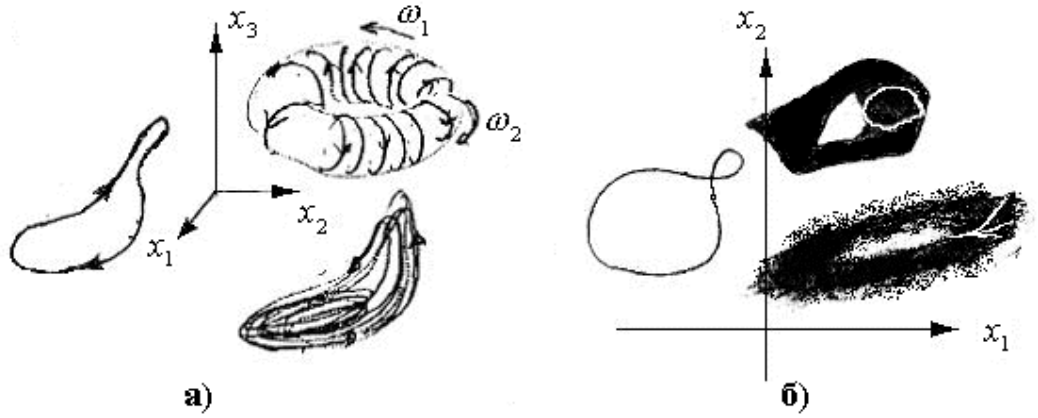

Рис.6.15. Качественный вид возможных фазовых траекторий в трехмерном пространстве, восстановленных по наблюдаемому ряду (а) и их проекции на плоскость (типичный вид на экране осциллографа) (б). Белые линии на темном фоне – сечения портрета

6.4.2. Методы спектрального анализа (фурье-анализ и вейвлеты)

Чаще всего термин «спектральный анализ» относят к использованию преобразования Фурье (разложения сигнала на гармоники). Но в обобщенном смысле спектральным анализом можно называть любое представление наблюдаемого сигнала в виде суперпозиции некоторых *базисных* функций, т.к. используется набор (спектр) функций-компонент. Здесь мы кратко рассмотрим традиционный фурье-анализ, а также более новый и «модный» инструмент – вейвлет-анализ, который очень полезен, в частности для исследования нестационарных сигналов.

6.4.2.1. Преобразования Фурье и спектр мощности. Эта тема является предметом рассмотрения огромного числа книг и статей и лежит в основе целых отраслей прикладной математики – спектрального анализа [74] и цифровой фильтрации [147]. Коснемся кратко некоторых моментов.

Напомним, прежде всего, что согласно теореме Вейерштрасса с помощью тригонометрического многочлена можно сколь угодно точно представить любую непрерывную на отрезке [*a*,*b*] функцию $\eta = F(t)$ при $F(a) = F(b)$. Обратимся к практике применения этой идеи в случае временного ряда. Будем считать, что ряд эквидистантный с четным числом точек *N*, интервал выборки Δt, $a = t_1$, $b = a + N \cdot \Delta t = t_N + \Delta t$. Можно

показать, что исходный сигнал $\{\eta(t_i)\}_{i=1}^{N}$ может быть однозначно представлен во все моменты наблюдений в виде суммы

$$\eta(t_i) = a_0 + \sum_{k=1}^{N/2} a_k \cos(k\omega t_i) + \sum_{k=1}^{N/2-1} b_k \sin(k\omega t_i), \quad i = 1,...,N. \tag{6.1}$$

где основная частота есть $\omega = \dfrac{2\pi}{b-a} = \dfrac{2\pi}{N\Delta t}$. Коэффициенты тригонометрического многочлена (6.1) находятся по простым формулам

$$a_0 = \frac{1}{N}\sum_{i=1}^{N}\eta_i\,,\ a_{N/2} = \frac{1}{N}\sum_{i=1}^{N}(-1)^i\eta_i\,, \tag{6.2}$$

$$a_k = \frac{2}{N}\sum_{i=1}^{N}\eta_i \cos(k\omega t_i)\,,\ k = 1,...,N/2-1, \tag{6.3}$$

$$b_k = \frac{2}{N}\sum_{i=1}^{N}\eta_i \sin(k\omega t_i)\,,\ k = 1,...,N/2-1. \tag{6.4}$$

Формулы (6.2)-(6.4), переводящие значения η_i в коэффициенты a_k и b_k, называются прямым дискретным преобразованием Фурье (ДПФ). Формула (6.1), обеспечивающая расчет η_i по a_k и b_k, называется обратным ДПФ.

Опираясь на эти преобразования, можно построить приближенное описание исходного сигнала (аппроксимацию) или провести его сглаживание. Например, высокие частоты (соответствующие большим k) часто отвечают помехам и шумам и от них желательно избавиться. Этого можно достичь в простейшем варианте, приравняв нулю соответствующие коэффициенты a_k и b_k. Выполнив теперь обратное преобразование Фурье (6.1), получим более плавно меняющийся («более гладкий») сигнал. Это – вариант фильтра низких частот. Аналогично можно реализовать фильтр высоких частот (обнулив коэффициенты, соответствующие малым k) или полосовой фильтр (сохранив ненулевыми только коэффициенты из некоторой полосы частот). Мы упомянули далеко не самые эффективные варианты цифровых фильтров, подробнее см., например, [147].

С помощью тригонометрических многочленов часто можно получить достаточно хорошую аппроксимацию непрерывной на отрезке функции $F(t)$. Причем можно использовать даже тригонометрические многочлены очень высокого порядка, тогда как использование алгебраических многочленов ведет к существенным трудностям в таком случае (см. п.7.2.3). Если функция $F(t)$ периодическая, то ее можно хорошо приблизить тригонометрическим многочленом и на всей числовой оси.

Подчеркнем, что тригонометрическая система особенно полезна и не имеет «конкурентов» при аппроксимации периодических функций.

Физическая интерпретация – спектр мощности. Можно показать, что средний квадрат наблюдаемой (он имеет смысл мощности, если наблюдаемая – это напряжение на участке электрической цепи или сила тока) равен сумме средних квадратов слагаемых в правой части (6.1). Другими словами, мощность сигнала распределяется *по частотам*:

$$\frac{1}{N}\sum_{i=1}^{N}\eta_i^2 = \sum_{k=0}^{N/2} S_k\,, \tag{6.5}$$

где S_k есть мощность гармонической составляющей с частотой $k\omega$: $S_k = (a_k^2 + b_k^2)/2$ при $1 \le k < N/2$, $S_k = a_k^2$ при $k = 0, N/2$. Физически, наблюдаемый сигнал часто представляет собой суперпозицию сигналов от нескольких источников. Если эти источники демонстрируют гармонические колебания каждый со своей частотой, то в наблюдаемом сигнале их интенсивность будет отражена величиной S_k на соответствующей частоте. Коэффициенты S_k, помимо решения задачи фильтрации сигнала, позволяют выявить присутствие и значимость различных источников колебаний. Если каждую значимую частоту связать с колебаниями одного гармонического осциллятора, то число таких существенных частот равно *числу степеней свободы*, вовлеченных в процесс. Поскольку полная мощность сигнала, согласно (6.5), представляется в виде набора (спектра) нескольких составляющих, то этот набор S_k называют *спектром мощности* процесса $\eta(t)$. Строго говоря, это лишь оценка спектра мощности (см. ниже).

Понятие спектра мощности так легко вводится только для детерминированной периодической функции $\eta(t)$ с периодом $2\pi/\omega$, поскольку она однозначно представляется в виде тригонометрического ряда Фурье

$$\eta(t) = a_0 + \sum_{k=1}^{\infty}[a_k\cos(k\omega t) + b_k\sin(k\omega t)], \tag{6.6}$$

все коэффициенты которого в общем случае ненулевые и выражаются через интегралы от исходной функции $\eta(t)$.

Но и здесь следует учесть, что полученный с помощью ДПФ по временному ряду конечный набор коэффициентов (6.1) является только приближением для теоретического спектра. Если основная мощность содержится только в низких частотах, то приближение достаточно точно. Известно явление мимикрии (маскировки) частот, которое состоит в следующем. В модель (6.1) входит максимальная частота $\omega N/2 = \pi/\Delta t$,

которую называют частотой Найквиста. Период соответствующей компоненты равен двум интервалам выборки. Составляющие с частотами $k\omega$, большими частоты Найквиста, в точках временного ряда являются линейной комбинацией базисных функций, входящих в (6.1): если их добавлять в модель, то более низкие частоты нужно исключать, чтобы обеспечить линейную независимость базисных функций. Другими словами, составляющие с более высокими частотами нельзя отличить от низкочастотных составляющих, например, $\cos\left(\left(N/2+k\right)j\omega\Delta t\right)=\cos(\pi j+k\omega j\Delta t)=\left(-1\right)^j\cos(jk\omega\Delta t)$, где $k>0$, $t_j=j\Delta t$. Одни как бы «маскируются» под другие.

Еще сложнее ситуация в случае непериодической функции $\eta(t)$. Такая функция не может быть представлена точно в виде ряда (6.6) на всей оси. Но для нее при некоторых условиях (интегрируемость на всей оси, гладкость) можно записать аналогичное представление в виде интеграла Фурье, т.е. заменить дискретные значения частоты $k\omega$ в (6.6) на непрерывный ряд значений:

$$\eta(t)=\int_0^\infty A(\omega)\cos(\omega t)d\omega+\int_0^\infty B(\omega)\sin(\omega t)d\omega, \tag{6.7}$$

$$A(\omega)=\frac{1}{\pi}\int_{-\infty}^{\infty}\eta(t)\cos(\omega t)dt,\ B(\omega)=\frac{1}{\pi}\int_{-\infty}^{\infty}\eta(t)\sin(\omega t)dt. \tag{6.8}$$

Преобразования (6.7) и (6.8) называют непрерывными преобразованиями Фурье (обратным и прямым, соответственно). Дискретные преобразования – их аналоги. Энергия[10] сигнала $\eta(t)$ раскладывается в непрерывный *энергетический спектр* $\int_{-\infty}^{\infty}\eta^2(t)dt=\int_0^\infty E(\omega)d\omega$, $E(\omega)=A^2(\omega)+B^2(\omega)$.

Наконец, рассмотрим случай, когда $\eta(t)$ – реализация стационарного случайного процесса. Почти всегда она не периодическая. При этом не существует интегралов (6.8), т.е. нельзя определить A и B даже как случайные величины. Спектральный состав процесса характеризуют, используя *финитное* преобразование Фурье (см., например, [172]), т.е. величины

$$A_T(\omega)=\frac{1}{\pi}\int_{-T/2}^{T/2}\eta(t)\cos(\omega t)dt,\ B_T(\omega)=\frac{1}{\pi}\int_{-T/2}^{T/2}\eta(t)\sin(\omega t)dt. \tag{6.9}$$

[10] Не мощность: средняя мощность равна нулю, т.к. сигнал затухает на бесконечности.

Далее рассчитывают математическое ожидание A_T, B_T и определяют *спектр мощности* как

$$S(\omega) = \lim_{T\to\infty} \frac{M[A_T^2(\omega) + B_T^2(\omega)]}{T}. \qquad (6.10)$$

Это именно мощность, т.к. в знаменателе стоит величина временного интервала *T*. В этой постановке значения S_k, полученные с помощью ДПФ, – это случайные величины. Их набор является грубой оценкой спектра мощности. Ее называют периодограммой (если умножить все компоненты на *N*, т.е. перейти от мощности к энергии). Чтобы получить оценку с лучшими статистическими свойствами, желательно усреднить S_k по нескольким реализациям процесса или «сгладить» отдельную периодограмму.

Важность понятия спектра мощности связана с тем, что поведение многих реальных систем в режимах небольших амплитуд колебаний адекватно описывается с помощью гармонических функций. Такие режимы поведения хорошо изучены и поняты, они повсеместно реализуются на практике и широко используются в технике, например, в системах связи. Линейные системы (фильтры, усилители и т.д.) описываются в терминах передаточных функций, т.е. их основная характеристика – это способ преобразования спектра мощности входного сигнала. Зачастую важен и фазовый спектр – набор начальных фаз гармонических составляющих в (6.1). Многочисленные особенности оценки спектров на практике и методов фильтрации отражены в литературе, см. работы [74, 147, 172] и ссылки в них.

Пример. Сигнал с меняющимся частотным составом и оконное ДПФ. В заключение обсудим распространенную ситуацию, когда исследуемый сигнал представляет собой последовательность двух фрагментов синусоиды с различными частотами:

$$\eta(t) = \begin{cases} \sin 2t, -\pi \le t < 0, \\ \sin 4t, \quad 0 \le t < \pi. \end{cases} \qquad (6.11)$$

Анализ проводится на отрезке [-π, π] по ряду длиной 20 точек с интервалом выборки $\Delta t = \pi/10$ и начальным моментом $t_1 = -\pi$ (рис.6.16,а).

Этот сигнал можно хорошо описать с помощью тригонометрического многочлена (6.1), который содержит много значимых компонент (рис.6.16,а). Но более полезную информацию можно получить, если разбить сигнал на два фрагмента (окна) и по каждому из них построить свой тригонометрический многочлен (рис.6.16,б). Это так называемое оконное преобразование Фурье. В каждом окне получим спектр из единственной значимой компоненты, что облегчает физическую

интерпретацию результатов. При оконном ДПФ видно, что частотный состав сигнала меняется во времени, что не может быть выявлено с помощью единого многочлена (6.1). Подобные нестационарные сигналы часто встречаются на практике. Для их анализа можно использовать оконное ДПФ, но существует и гораздо более удобный современный инструмент – вейвлет-преобразование.

6.4.2.2. Вейвлет-преобразование и вейвлет-спектры. Для анализа функций $\eta = F(t)$, имеющих импульсный характер, разрывы, изломы и прочие особенности, весьма эффективным оказывается использование базиса, состоящего из функций $\phi_k(t)$, локализованных по времени и по частоте, – вейвлетов. В последние 20 лет они приобрели огромную популярность, см., например, обзорные работы [17, 94] и ссылки в них.

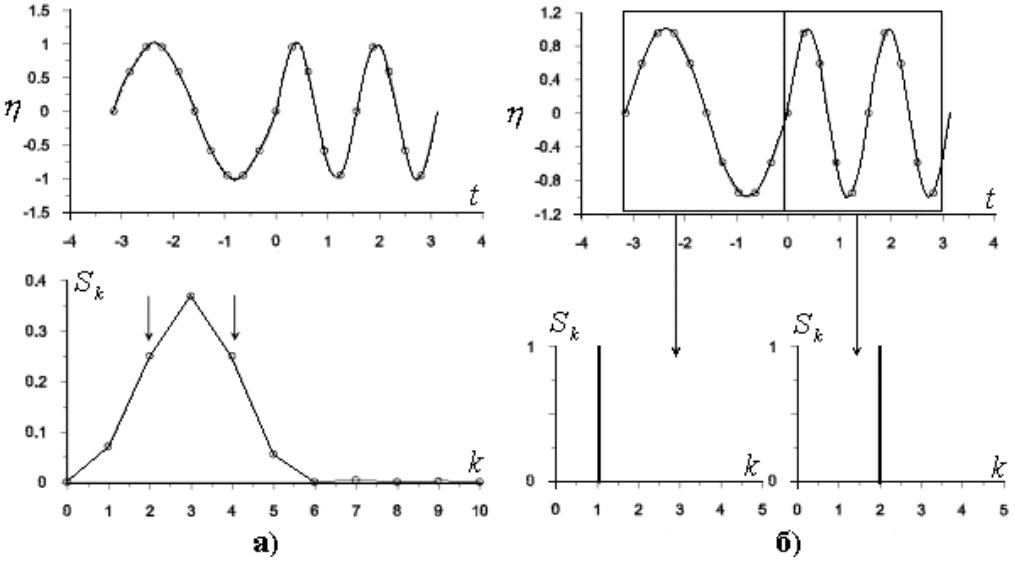

Рис.6.16. а) Два фрагмента синусоиды и наложенный на них график аппроксимирующего тригонометрического многочлена (6.1). Спектр мощности содержит 5 значимых компонент, стрелками показаны частоты фрагментов (снизу). б) Аналогичные графики для оконных ДПФ, в спектрах по одной частоте. Они соответствуют числам *k*, в два раза уменьшенным по сравнению с левым рисунком, т.к. число точек в каждом из окон в два раза меньше числа точек исходного ряда

«Многие исследователи называют вейвлет-анализ “математическим микроскопом”» [17]. По-английски «wavelet» (термин введен в 1984 г.) означает приблизительно «маленькая волна», но уже и в русскоязычной литературе принято использовать термин «вейвлет». Не вдаваясь в строгость математических определений, ограничимся пояснением основных моментов. Вейвлетом называют функцию $\phi(t)$, которая

1) хорошо локализована по времени *t* (т.е. быстро стремится к нулю с ростом $|t|$) и по частоте (ее Фурье-образ тоже хорошо локализован);

2) имеет нулевое среднее $\int_{-\infty}^{\infty} \phi(t)dt = 0$;

3) удовлетворяет условию автомодельности (при изменении масштаба по времени число осцилляций не меняется).

Пример представлен на рис.6.17 – это так называемый DOG-вейвлет[11]:

$$\phi(t) = e^{-t^2/2} - 0.5e^{-t^2/8}. \qquad (6.12)$$

На основе вейвлета можно построить систему функций путем его сдвигов и масштабных преобразований (сжатий и растяжений вдоль оси t). Полученные функции удобно обозначать двумя индексами:

$$\phi_{j,k}(t) = 2^{j/2}\phi\left(2^j t - k\right), \; -\infty < j,k < \infty, \qquad (6.13)$$

где j и k – целые числа. Увеличение индекса j на единицу дает изменение масштаба по оси времени в 2 раза («сжатие» графика функции), а увеличение индекса k на единицу дает сдвиг графика функции $\phi_{j,k}$ на интервал $k/2^j$ вправо вдоль оси t (рис.6.17). Нормировочный множитель $2^{j/2}$ введен только для удобства, чтобы сохранять неизменной норму (интеграл от квадрата функции) $\phi_{j,k}$: $\|\phi_{j,k}\|^2 = \|\phi\|^2$.

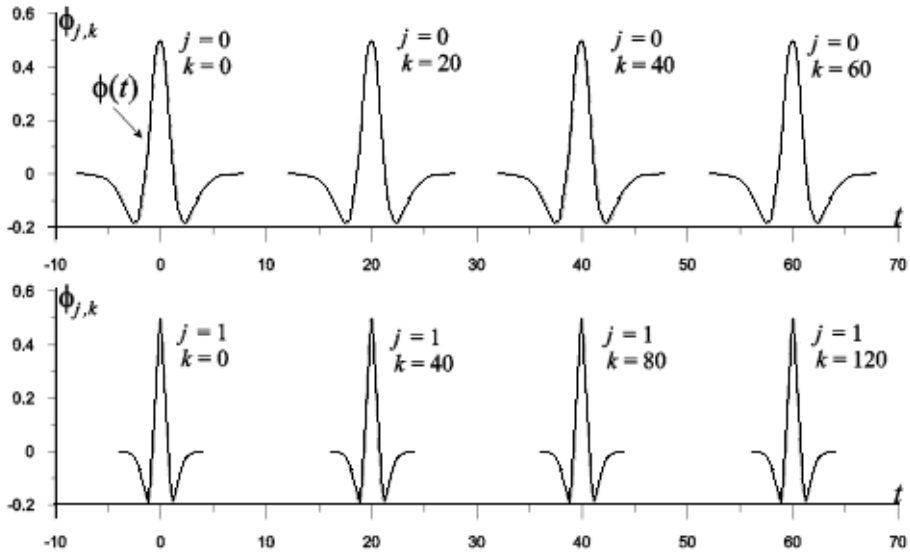

Рис.6.17. Пример вейвлета и конструирования на его основе системы базисных функций путем сдвигов (слева направо) и сжатий (сверху вниз) без учета нормировки

Сконструированная система локализованных функций $\phi_{j,k}$ за счет сдвигов и растяжений покрывает всю ось t. При определенном выборе функции $\phi(t)$ она является базисом в пространстве функций, суммируемых

[11] От Difference of Gaussians – разность «гауссиан» (т.е. функций Гаусса).

с квадратом на всей оси (только тогда $\phi(t)$, строго говоря, и называют вейвлетом [17]). Это выполняется для широкого класса функций, например, для представленной на рис.6.17. Базисные функции $\phi_{j,k}$ часто называют вейвлет-функциями. Поскольку все они получены преобразованиями функции $\phi(t)$, то последнюю часто называют «материнской функцией».

Примеры материнских функций очень разнообразны, см. рис.6.18. Библиотеки насчитывают сотни таких функций [350]. Отметим часто используемый комплексный вейвлет Морле

$$\phi(t) = \pi^{-1/4} e^{-t^2/2} (e^{-i\omega_0 t} - e^{-\omega_0^2/2}), \tag{6.14}$$

где ω_0 определяет количество осцилляций, а второй член в скобках введен для обеспечения нулевого среднего. Действительная часть вейвлета Морле представлена на рис.6.18,в для $\omega_0 = 6$.

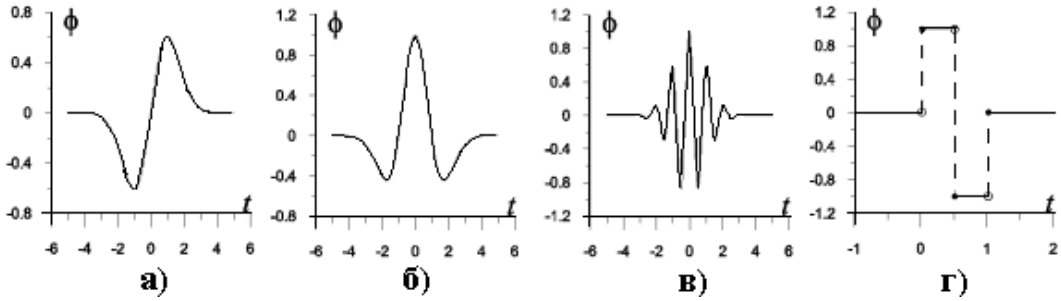

Рис.6.18. Примеры вейвлетов: а) WAVE-вейвлет; б) «мексиканская шляпа»; в) вейвлет Морле (действительная часть); г) HAAR-вейвлет

При построении аппроксимирующей функции по временному ряду нужно ограничить бесконечную систему функций $\phi_{j,k}$ конечным числом составляющих. На примере HAAR-вейвлета (рис.6.18,г) проиллюстрируем аппроксимацию вейвлетами [276]. Пусть имеется эквидистантный временной ряд длиной $N = 2^m$, m – положительное целое число, интервал выборки $\Delta t = 1/N$, $t_1 = 0$. В качестве базисных функций удобно взять из общей системы $\phi_{j,k}$ следующие, дополнив ее функцией-константой:

$$\begin{gathered}
1, \\
\phi_{0,0}(t) \equiv \phi(t), \\
\phi_{1,0}(t) \equiv \phi(2t), \qquad \phi_{1,1}(t) \equiv \phi(2t-1) \\
\phi_{2,0}(t) \equiv \phi(4t), \phi_{2,1}(t) \equiv \phi(4t-1), \phi_{2,2}(t) \equiv \phi(4t-2), \phi_{2,3}(t) \equiv \phi(4t-3), \\
\ldots, \\
\phi_{m-1,0}(t) \equiv \phi(2^{m-1}t), \qquad \ldots, \qquad \phi_{m-1,2^{m-1}-1}(t) \equiv \phi(2^{m-1}t - 2^{m-1} + 1).
\end{gathered} \tag{6.15}$$

Исходный временной ряд представляется в точности в виде суммы

$$\eta_i = c_0 + \sum_{j=0}^{m-1} \sum_{k=0}^{2^j-1} c_{j,k} \phi_{j,k} \ , \tag{6.16}$$

Коэффициенты при функциях с $j = m-1$ зависят только от разности значений наблюдаемой в соседних точках, т.е. от колебаний с временным масштабом $2\Delta t$. Поскольку эти функции описывают только самые мелкомасштабные изменения (изменения с частотой Найквиста), то значения составляющей исходного сигнала, равной сумме этих функций:

$$D_1(t_i) = \sum_{k=0}^{2^{m-1}-1} c_{m-1,k} \phi_{m-1,k}(t_i) \, , \tag{6.17}$$

называют *деталями первого уровня*. А оставшуюся составляющую

$$A_1(t_i) = \eta_i - D_1(t_i) = c_0 + \sum_{j=0}^{m-2} \sum_{k=0}^{2^j-1} c_{j,k} \phi_{j,k} \ , \tag{6.18}$$

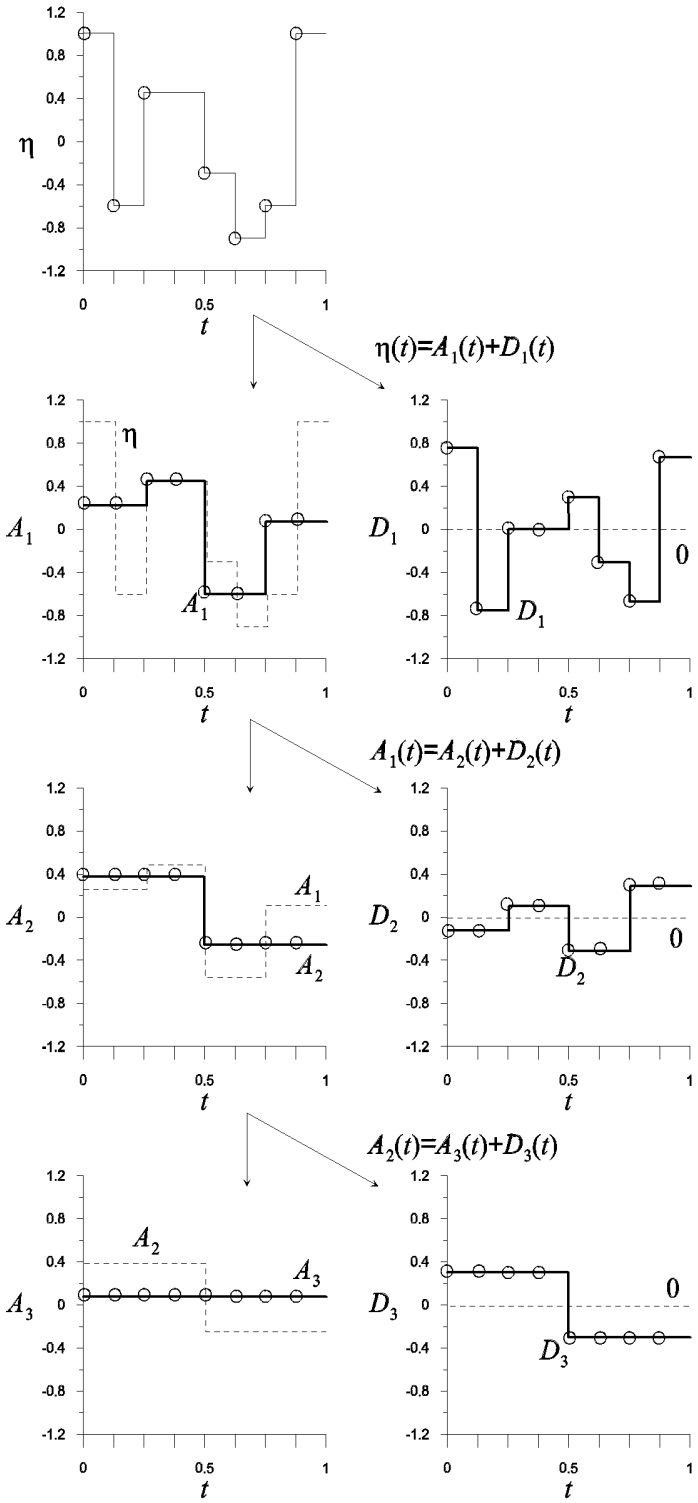

Рис.6.19. Вейвлет-преобразование: аппроксимации и детали разных уровней

называют *аппроксимацией первого уровня*. Аппроксимация первого уровня уже не содержит изменений с масштабом $2\Delta t$ (рис.6.19). Аналогично, дальше можно выделять детали в аппроксимации первого уровня и т.д., что и осуществляется базисными функциями с меньшими j (рис.6.19).

Общее определение деталей *n*-го уровня и аппроксимации *n*-го уровня дается аналогично (6.17) и (6.18): детали D_n есть мелкомасштабные составляющие сигнала, аппроксимации $A_n(t_i)$ – крупномасштабные. Аппроксимация последнего *m*-го уровня есть просто среднее значение сигнала. В итоге исйходный сигнал равен сумме среднего значения и деталей всех уровней: $\eta_i = A_m(t_i) + D_m(t_i) + D_{m-1}(t_i) + ... + D_1(t_i)$.

Величина *j* определяет масштаб рассмотрения (чем она больше, тем меньше этот масштаб), а *k* – временную точку, около которой проводится рассмотрение. Если провести дальше аналогию вейвлета с микроскопом, то *k* – точка его фокусировки, *j* определяет увеличение, а вид вейвлета (форма материнской функции) – оптические свойства.

Если оказались значимыми только несколько вейвлет-коэффицентов $c_{j,k}$, то остальные можно отбросить (обнулить) и оставшиеся слагаемые дадут компактное и достаточно точное приближение исходного сигнала. В этом случае говорят также, что вейвлет обеспечивает сжатие информации (множество измеренных значений ряда можно заменить несколькими вейвлет-коэффициентами и хранить только их, при необходимости по коэффициентам можно восстановить исходный сигнал с малой погрешностью). Вейвлеты эффективны для «сжатия» сигналов различного характера, особенно импульсного. Для сжатия сигналов определенного вида годятся алгебраические или тригонометрические многочлены, но область приложений вейвлетов на практике значительно шире [276, 170, 350].

Вейвлет-анализ. Вейвлет-коэффициенты также можно анализировать и делать по ним нетривиальные выводы о характере сигнала. Это составляет содержание вейвлет-анализа в отличие от аппроксимации или восстановления значений сигнала, что относится к задачам синтеза. В представленной постановке говорят о *дискретном вейвлет-анализе*, т.к. индексы *j*, *k* системы вейвлет-функций меняются дискретно. Все больший интерес привлекает *непрерывный вейвлет-анализ*, когда переходят к непрерывнозначным «индексам» сдвига и масштаба [17, 94]. Интегральное (или непрерывное) вейвлет-преобразование сигнала *F*(*t*) определяется выражением:

$$W(s,k) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} \eta(t) \cdot \phi\left(\frac{t-k}{s}\right) dt \equiv \int_{-\infty}^{\infty} \eta(t) \cdot \phi_{s,k}(t) dt, \qquad (6.19)$$

где *s*, *k* – вещественные числа (непрерывнозначные параметры масштаба и сдвига), вейвлет-функции обозначены $\phi_{s,k}(t) = \left(1/\sqrt{s}\right)\phi\left((t-k)/s\right)$, параметр *s* есть аналог величины 2^{-j}. Функция двух переменных $W(s,k)$ называется *вейвлет-спектром* сигнала $\eta(t)$. Она имеет наглядный физический смысл.

Большое значение $|W(s_1,k_1)|$ указывает на то, что в окрестности момента времени k_1 интенсивны изменения сигнала с временным масштабом s_1. Грубо говоря, значения $|W(s,k_1)|$ при фиксированном k_1 показывают частотный состав сигнала в окрестности момента k_1. Если велики значения с маленькими *s*, то присутствуют мелкомасштабные – высокочастотные – компоненты. Значения $|W(s_1,k)|$ при фиксированном s_1 показывают, как меняется во времени интенсивность компоненты сигнала с временным масштабом изменения s_1. Таким образом, вейвлет-спектр содержит информацию и о частотном составе сигнала, и о его временной локализации (в отличие от фурье-спектра мощности, который дает информацию только о частотном составе без локализации во времени), поэтому его еще называют частотно-временным спектром.

Для вейвлет-спектра выполняется «энергетическое условие», которое позволяет связать его с разложением энергии по времени и «частоте»:

$$\int_{-\infty}^{\infty}\eta^2(t)dt=\frac{1}{C_\phi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}W^2(s,k)\frac{dsdk}{s^2}, \qquad (6.20)$$

где C_ϕ – нормировочный коэффициент. Если величину W^2 проинтегрировать по времени *k*, то получим функцию только временного масштаба, которую называют *глобальным энергетическим спектром* или *скалограммой*:

$$E_W(s)=\int_{-\infty}^{\infty}W^2(s,k)dk\,. \qquad (6.21)$$

Ее можно использовать для глобальной характеристики частотного состава сигнала наряду с периодограммой. Скалограмма является более точной оценкой спектра мощности, она близка к сглаженной периодограмме [17].

При практическом использовании визуализировать вейвлет-спектр можно как поверхность в трехмерном пространстве. Но чаще используют картину линий уровня $|W(s,k)|$ или двухмерную карту ее значений на плоскости (*k*, *s*) в серых полутонах, где белый цвет означает большие, а черный – нулевые значения. Возможен только приближенный расчет интеграла (6.19) по временному ряду. Для этого нужно определить характер поведения сигнала и за пределами наблюдаемого отрезка [*a*,*b*] (часто полагают его равным нулю), что вносит свои искусственные особенности – краевые эффекты. Проиллюстрируем «работу» вейвлет-преобразования на простом примере сигнала с переменным частотным составом – двух фрагментов синусоиды с разной частотой (рис.6.20,а) [172, 94]. Используем DOG-вейвлет, временные ряды длиной 1000 точек на

отрезке [-π,π] и дополнение сигнала нулем вне отрезка. Глобальный фурье-анализ не позволял выявить «структуру нестационарности», тогда как на вейвлет-спектре хорошо видны характерные временные масштабы (рис.6.20,б). Он позволяет выделить колебания с низкой частотой в начале ряда и скачок частоты (белые пятна – большие абсолютные значения W – соответствуют положению экстремумов синусоиды и ее временному масштабу).

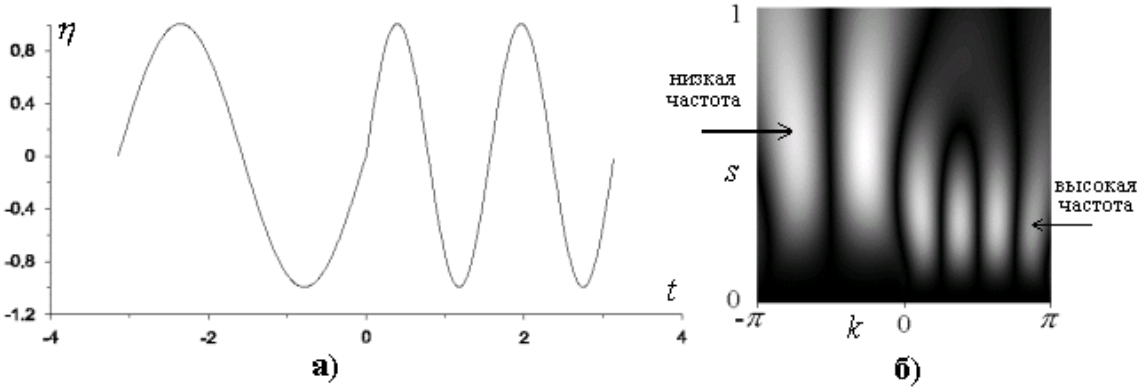

Рис.6.20. а) Сигнал с переменным частотным составом – два фрагмента синусоиды с разными частотами; б) вейвлет-спектр, полученный с помощью DOG-вейвлета

Вейвлет-анализ исключительно полезен для исследования нестационарных сигналов, содержащих сегменты с различным характером поведения. Кроме того, он весьма эффективен для существенно неоднородных сигналов (импульсного типа и т.д.), сигналов с особенностями (разрывами, изломами, разрывами производных более высокого порядка), поскольку позволяет локализовать особенности и выявить их характер. Вейвлет-спектр обнаруживает характерную регулярную форму для фрактальных (грубо говоря, сильно изрезанных и самоподобных) сигналов. Такие признаки демонстрируют многие реальные сигналы. При этом можно рассматривать не только зависимость переменной от времени, но и, например, от пространственной координаты. Так, в сечении рельеф поверхности Луны выглядит, как показано на рис.6.21 [276], т.е. имеет очень сложную форму с различными масштабами, которые связывают с бомбардировкой Луны метеоритами различного размера. Вейвлеты применяются для анализа данных в геофизике, биологии, медицине, астрофизике, системах обработки информации, для распознавания и синтеза речи, сжатия изображений и т.д. Много интересных материалов и библиографию по

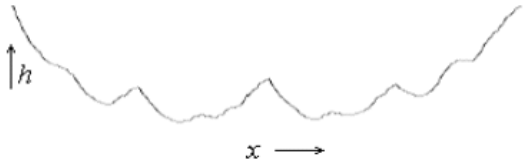

Рис.6.21. Рельеф поверхности Луны в сечении (x – координата вдоль поверхности, h – высота) – имитация

вейвлетам можно найти в [17, 94] и на сайте [350], где представлен вейвлет-дайджест, выходящий с 1992 года.

6.4.3. Фаза сигнала и разложение на эмпирические моды

В самых разных задачах оказывается очень полезным рассмотрение динамики фазы сигнала. Современное понятие фазы мы рассмотрим ниже. Грубо говоря, это – переменная, характеризующая повторяемость в сигнале, она нарастает на 2π на интервале между двумя последовательными максимумами. Особенное внимание к этой переменной связано с тем, что она обычно очень чувствительна даже к слабым внешним воздействиям на систему. Для изменения амплитуды колебаний может потребоваться значительная энергия, тогда как фазу легко изменить слабым «толчком».[12]

Термин «фаза» – синоним слов «состояние», «этап» [133]. Мы уже использовали его в п. 2.1.3, говоря о векторе состояния и пространстве состояний динамической системы, называли «фазовой» траекторию, которую описывает вектор состояния системы. В теории обработки сигналов значения этого термина несколько изменилось. Так, фазой гармонического сигнала $x(t) = A\cos(\omega t + \phi_0)$ называют аргумент косинуса $\phi = \omega t + \phi_0$. Он отчасти определяет значение косинуса, но для полного задания состояния системы нужна еще и амплитуда A, т.е. фаза – не полная характеристика состояния. Кроме фазы $\phi = \omega t + \phi_0$, называемой «развернутой» (рис.6.22,б, верхняя панель), используют и «свернутую» фазу: $\phi(t) = (\omega t + \phi_0) \bmod 2\pi$, которая определена только в интервале $[0,2\pi)$ (рис.6.22,б, нижняя панель). Последнее имеет смысл, т.к. значения развернутой фазы, отличающиеся на 2π, соответствуют одинаковым *состояниям* (значениям косинуса).

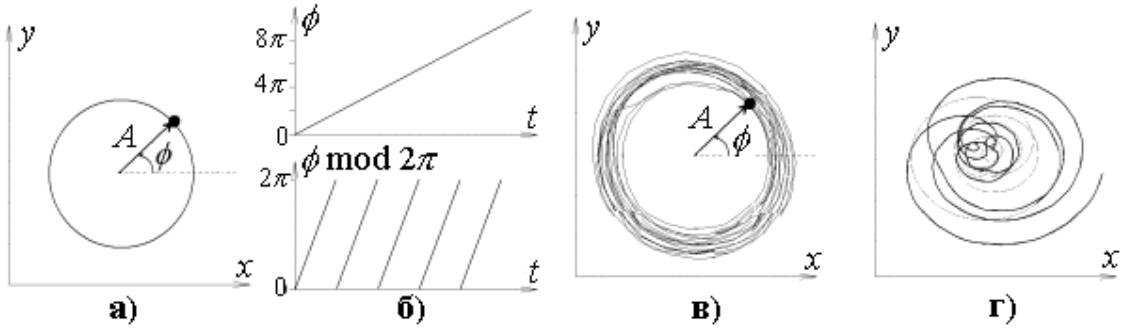

Рис.6.22. а) Траектория на комплексной плоскости для гармонического сигнала, его амплитуда и фаза. б) Развернутая и свернутая фазы гармонического сигнала в зависимости от времени. в) Определение амплитуды и фазы негармонического

[12] Так, фазу колебаний маятника (рис.3.5,а,в) можно изменить, придержав его в точке максимального отклонения от положения равновесия, не затрачивая на это энергии.

узкополосного сигнала с помощью преобразования Гильберта, г) то же самое для широкополосного сигнала (фаза не является хорошо определенной величиной)

Наглядное геометрическое толкование введенного понятия фазы возможно, если представить сигнал $x(t) = A\cos(\omega t + \phi_0)$ как вещественную часть $\mathrm{Re}\, z(t)$ комплексного сигнала $z(t) = Ae^{i(\omega t + \phi_0)}$. Тогда вектор $z(t)$ на плоскости (x,y), где $x(t) = \mathrm{Re}\, z(t)$, а $y(t) = \mathrm{Im}\, z(t)$, равномерно вращается с частотой ω по окружности радиуса A с центром в начале координат (рис.6.22,а). Фаза $\phi = \omega t + \phi_0$ есть угол поворота $z(t)$ относительно положительного направления оси x. При расчете «развернутой» фазы учитывается число полных оборотов, для чего при каждом обороте фаза увеличивается на 2π. Для гармонического сигнала «развернутая» фаза растет во времени линейно со скоростью, равной угловой частоте колебаний, график «свернутой» фазы – кусочно-линейная функция (пила), рис.6.22,б.

Исходно введенное только для гармонического сигнала понятие фазы затем было обобщено и на более сложные ситуации. Основной путь такого обобщения – определить фазу как угол поворота вектора на координатной плоскости, где по осям отложены значения исходного сигнала и сопряженного ему (в некотором смысле). Наиболее известно и распространено обобщение фазы на случай негармонических сигналов с помощью построения аналитического сигнала [228]. Аналитическим называется комплексный сигнал, в комплексном фурье-образе которого отличны от нуля только компоненты с положительными частотами. По исходному сигналу $x(t)$ строится аналитический сигнал $z(t) = x(t) + iy(t)$, где $y(t)$ – сопряженный по Гильберту для $x(t)$:

$$y(t) = \mathrm{P.V.} \int_{-\infty}^{\infty} \frac{x(t')dt'}{\pi(t - t')}, \tag{6.22}$$

P.V. означает главное значение несобственного интеграла. Фаза определяется как аргумент z, т.е. как угол поворота радиуса-вектора на комплексной плоскости (x, y). Этот подход широко используется в радиофизике и электротехнике [44, 137]. В случае гармонического сигнала $x(t) = A\cos(\omega t + \phi_0)$ получаем сопряженный сигнал $y(t) = \sin(\omega t + \phi_0)$ и фазу, совпадающую с первоначальным определением. Для более сложного (но «похожего» на гармонический) сигнала наблюдается вращение вектора z не по окружности, а «почти» по окружности (рис.6.22,в). Фаза нарастает в среднем со скоростью, равной средней угловой частоте колебаний.

Ясный физический смысл фаза имеет для сигналов с ярко выраженным основным ритмом колебаний, подробнее см. [137, 7]. В случае сложных нерегулярных сигналов может оказаться, что построение

аналитического сигнала через преобразование Гильберта не обнаружит вращения около четко определенного центра (рис.6.22,г). Тогда такое формальное определение фазы, как правило, бесполезно. Но если наблюдаемый сложный сигнал является комбинацией относительно простого сигнала от исследуемой системы и наложенных помех, то можно попытаться выделить более простой сигнал и определить его фазу, как описано выше. Рассмотрим два основных подхода к выделению простого сигнала – полосовую фильтрацию и разложение на эмпирические моды.

Полосовая фильтрация. Самый простой способ – использовать полосовой фильтр (п. 6.4.2.1), пропускающий только частоты в небольшой окрестности некоторой выделенной частоты. Если ширина полосы фильтра невелика, получим сигнал с явно выделенным основным ритмом, для которого фаза легко определяется описанным выше способом, см. рис.6.22,в. Но какую полосу фильтрации выбрать? Связан ли отфильтрованный сигнал с исследуемым процессом или является искусственной конструкцией? На эти вопросы можно ответить, только привлекая дополнительную информацию об исследуемой системе. Полоса фильтрации не должна быть слишком узкой (в предельном случае можно отфильтровать единственную гармонику, фаза которой хорошо определена, но не несет интересной информации) и слишком широкой (иначе не будет вращения вектора z вокруг одного центра – «повторяемости», с которой связана фаза).

Другим распространенным вариантом получения аналитического сигнала из исходного является комплексное вейвлет-преобразование [262]

$$z(t) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} x(t')\phi^*\left(\frac{t'-t}{s}\right)dt' \qquad (6.23)$$

при фиксированном масштабе s, например, с помощью вейвлета Морле (6.14): $\phi(t) = \pi^{-1/4}[\exp(-i\omega_0 t) - \exp(-\omega_0^2/2)]\exp(-t^2/2)$. Это эквивалентно полосовой фильтрации исходного сигнала с последующим применением преобразования Гильберта. А именно, при $\omega_0 = 6$ это полоса шириной $\Delta f/f = 1/4$ около частоты $f \approx 1/s$. При использовании вейвлет-преобразования (6.23) меньше проявляются краевые эффекты.

Разложение на эмпирические моды. Кроме линейной фильтрации возможны и другие подходы. Все большую популярность в последние годы приобретает недавно предложенный в [242] метод разложения сигнала на так называемые «эмпирические моды» (это вариант адаптивной нелинейной фильтрации). Для каждой из выделенных «мод» фаза хорошо определяется через преобразование Гильберта, т.к. обеспечено вращение траектории вокруг четко выделенного центра на комплексной плоскости.

Для этого каждая «мода» должна представлять собой сигнал с нулевым средним, максимумы которого положительны, а минимумы отрицательны, т.е. между каждым максимумом и минимумом сигнала $x(t)$ обязательно имеется пересечение его графика с прямой $x = 0$. Метод прост в реализации и требует совсем небольшого объема расчетов. Вот общий алгоритм.

1) Находят все экстремумы в сигнале $x(t)$.

2) Проводят интерполяцию между минимумами и получают нижнюю огибающую $e_{\min}(t)$. Например, соседние минимумы соединяются отрезками прямых (линейная интерполяция). Аналогично получают верхнюю огибающую $e_{\max}(t)$ по максимумам сигнала.

3) Рассчитывают среднюю величину $m(t) = \left(e_{\max}(t) + e_{\min}(t)\right)/2$.

4) Вычисляют так называемые *детали* (смысл термина аналогичен используемому в вейвлет-анализе, п. 6.4.2.2) $d(t) = x(t) - m(t)$. Величину $m(t)$ называют *остатком*.

5) Продолжают шаги 1-4 для выделенных деталей $d(t)$ (так называемая *процедура отсеивания*) и получают новые детали $d(t)$ и новый остаток $m(t)$. Продолжают процесс с новыми деталями и т.д. Итерации продолжаются до тех пор, пока детали не станут удовлетворять двум условиям. Текущий остаток $m(t)$ близок к нулю по сравнению с $d(t)$ и число экстремумов $d(t)$ отличается от числа пересечений $d(t)$ с осью $d = 0$ не более, чем на 1. При выполнении этих условий итерации прекращают и детали $d(t)$ называют эмпирической модой $f(t)$ (внутренней модовой функцией).

6) Рассчитывают остаток – разность между сигналом и эмпирической модой: $m(t) = x(t) - f(t)$. Шаги 1-5 приводят с остатком $m(t)$.

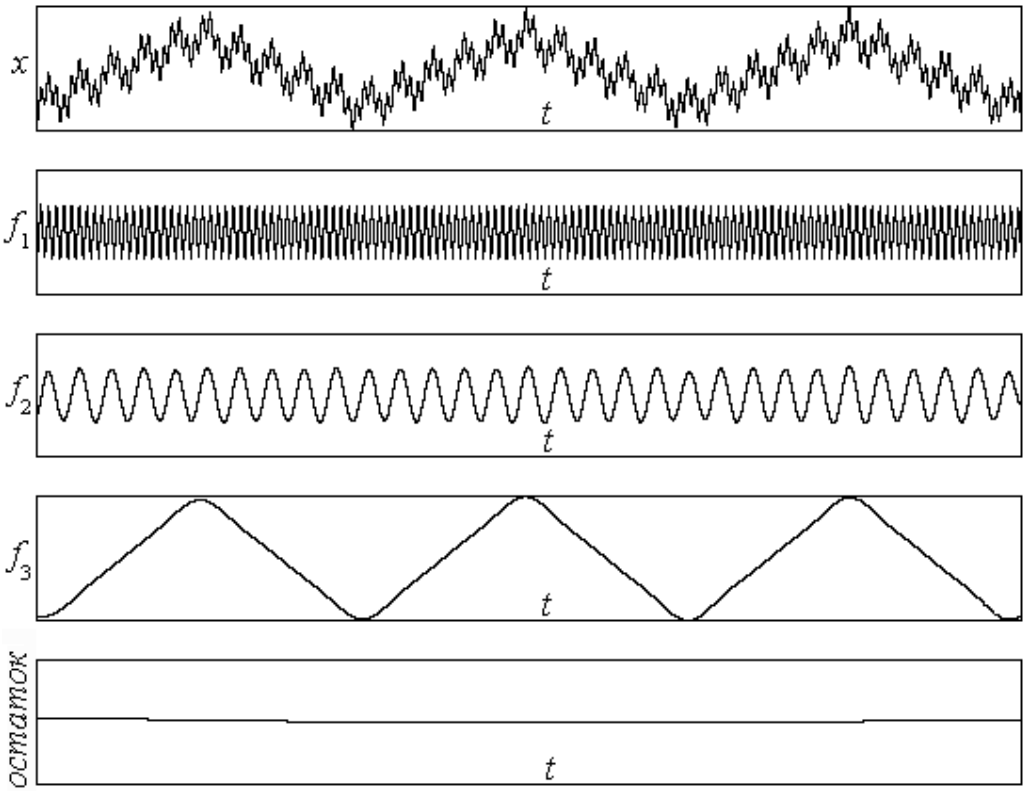

Рис.6.23. Разложение на эмпирические моды [242]. Сверху – исходный сигнал. Далее – выделенные моды. Первая – периодический сигнал треугольной формы с малым периодом, вторая – синусоида, третья – треугольная волна с большим периодом. Последний график – итоговый остаток разложения, который содержит всего один экстремум

Всю процедуру прекращают, когда в реализации $m(t)$ становится мало экстремумов. Процесс проиллюстрирован на рис.6.23 на примере сигнала, являющегося суммой синусоиды и двух периодических сигналов с треугольным профилем, период одного из которых больше, а другого – меньше, чем период синусоиды. В результате исходный сигнал представляется как сумма трех компонент (эмпирических мод) и остатка, близкого к нулю. Преимуществом данного метода перед полосовой фильтрацией является его адаптивный характер: он выделяет моды в зависимости от свойств сигнала без опоры на *предварительно заданную* полосу частот. В частности, он эффективнее при работе с нестационарными сигналами.

6.4.4. Анализ на стационарность/нестационарность

В общем случае стационарность процесса относительно некоторого свойства означает неизменность этого свойства во времени. Определения стационарности в узком и широком смысле представлены в п. 4.1.3. Кроме такой статистической стационарности, связанной с неизменностью законов

распределения или их моментов, выделяют динамическую стационарность, означающую неизменность оператора эволюции (п. 11.1).

В основе большинства методов анализа временных рядов лежит предположение о стационарности исследуемого процесса. Однако большинство реальных сигналов в природе (физиологических, финансовых и пр.) нестационарны. К этому приводит наличие в системе процессов, характерный масштаб которых превышает время наблюдения, внешние события, приводящие к перестройке динамики (например, адаптация в биологических системах). Многие характеристики, рассчитанные по нестационарному ряду, оказываются бессмысленными или недостоверными.

Много усилий исследователей было посвящено проблеме проверки стационарности. Ранее при обнаружении нестационарности временной ряд отвергался как непригодный для детального анализа или делился на сегменты, достаточно короткие, чтобы рассматривать их как стационарные. Позднее авторы начали использовать информацию о характере нестационарности как полезную для исследования процесса. Возможны ситуации, когда изменения свойств процесса представляют собой наиболее интересную часть временного ряда. Например, целью анализа электроэнцефалограмм часто является обнаружение изменений состояния мозга. Такие изменения происходят между различными стадиями сна, между эпилептическим припадком и нормальной мозговой активностью, и т.д.

Чтобы более или менее надежно проверить стационарность процесса по временному ряду, длительность последнего должна существенно превышать все интересующие исследователя временные масштабы колебаний системы. При наличии составляющих с характерными временами порядка длины ряда, процесс признается нестационарным. Но зачастую процесс можно считать стационарным и в том случае, когда длина временного ряда существенно меньше характерных времен более медленных процессов в системе. Например, сердечные сокращения у человека в расслабленном состоянии, как правило, гомогенны на интервалах порядка нескольких минут. Более длинные фрагменты, однако, обнаруживают новые особенности, появляющиеся из-за медленных биологических циклов. Поскольку обычные 24-часовые записи электрокардиограмм покрывают только один цикл циркадного (суточного) ритма, их гораздо труднее рассматривать как стационарные, чем более длинные или более короткие записи.

Для обнаружения нестационарности используют следующие способы:

1) вычисление какой-либо характеристики в скользящем окне (фрагменте временного ряда, смещающемся вдоль оси времени). Если она меняется слабо и без видимых трендов, то ряд считается стационарным, в

противном случае – нет. Примеры нестационарности относительно среднего и дисперсии представлены на рис.6.14,д-з. В статистике разработаны методы проверки стационарности относительно среднего (*t*-критерий Стьюдента, непараметрический критерий сдвига, критерий инверсий), дисперсии (критерий Фишера, критерий Кокрена, критерий рассеяния), одномерных функций распределения (критерий Вилкоксона) [87, 145]. На основе теории проверки гипотез можно опровергнуть утверждение о стационарности относительно этих величин с заданным уровнем значимости;

2) сравнение характеристик в различных временных окнах. Используют различные меры из статистики (критерии χ^2, Крамера – Мизеса, Колмогорова – Смирнова [87]) и нелинейной динамики (кросскорреляционный интеграл [211], взаимная ошибка предсказания [301,302], расстояние между векторами коэффициентов динамических моделей [235]). Иллюстрации подходов этого типа представлены в п. 11.1.

6.4.5. Анализ взаимной зависимости (корреляция, когерентность, синхронизация)

Выше были рассмотрены варианты анализа скалярных временных рядов. Если же наблюдается векторный ряд, например, одновременные измерения двух величин $x(t)$ и $y(t)$, то возникают новые вопросы и возможности анализа. Во многих случаях важно выяснить взаимосвязь между $x(t)$ и $y(t)$ и сделать отсюда выводы о наличии и характере связи между источниками этих сигналов. Эта информация может быть полезна и для моделирования. Так, если существует однозначная связь между $x(t)$ и $y(t)$, то достаточно измерять только одну из величин, а другая не несет новой информации. Если же есть некоторая взаимная зависимость, но неоднозначная, то целесообразно строить модель, учитывающую наличие взаимодействия между источниками двух сигналов, и т.д.

Существуют различные подходы для анализа взаимозависимостей. Исторически первыми инструментами были взаимная корреляционная функция и функция когерентности. Эти характеристики развиты в рамках математической статистики. Они относятся к так называемым линейным методам анализа временных рядов. *Взаимная ковариационная функция* определяется как ковариация значений x и y в моменты времени t_1 и t_2:

$$K_{x,y}(t_1,t_2) = M\left[\left(x(t_1) - M[x(t_1)]\right)\left(y(t_2) - M[y(t_2)]\right)\right]. \qquad (6.24)$$

Для стационарных процессов она зависит только от разницы между рассматриваемыми моментами времени:

$$K_{x,y}(\tau) = M\left[\left(x(t) - M[x]\right)\left(y(t+\tau) - M[y]\right)\right]. \qquad (6.25)$$

Согласно одному из вариантов терминологии (принятому в математике) нормированную взаимную ковариационную функцию называют взаимной корреляционной функцией (ВКФ). Это коэффициент корреляции между x и y. Для стационарных процессов ВКФ имеет вид

$$k_{x,y}(\tau) = \frac{M\left[\left(x(t) - M[x]\right)\left(y(t+\tau) - M[y]\right)\right]}{\sqrt{\mathrm{D}[x]\mathrm{D}[y]}}, \qquad (6.26)$$

где $\mathrm{D}[x]$ и $\mathrm{D}[y]$ – дисперсии процессов; $-1 \le k_{x,y}(\tau) \le 1$. Абсолютная величина $k_{x,y}(\tau)$ достигает 1 в случае детерминированной линейной зависимости: $y(t+\tau) = \alpha\, x(t) + \beta$. Для статистически независимых процессов $k_{x,y}(\tau) = 0$. Однако, если процессы связаны однозначно, но не линейно, то ВКФ может быть нулевой (например, для $y(t) = x^2(t)$ и симметричного распределения x относительно нуля), т.е. не выявит наличия зависимости. Поэтому говорят, что ВКФ *характеризует линейную зависимость* между сигналами. Для оценки ВКФ используют обычную формулу для эмпирических моментов (п. 2.2.1.4).

Развиты многочисленные модификации и обобщения ВКФ. Так, для того чтобы охарактеризовать зависимость между компонентами сигналов только на какой-то выделенной частоте, а не интегрально, используются взаимная спектральная плотность и функция когерентности. Чтобы выявить нелинейные зависимости используют обобщения коэффициента корреляции: коэффициент ранговой корреляции Спирмена [87], корреляционное отношение [1], функцию взаимной информации [227] и т.д.

В нелинейной динамике эти подходы получили развитие в двух направлениях. Первая идея – анализ взаимных (возможно, нелинейных) зависимостей между векторами состояний $\mathbf{x}(t)$ и $\mathbf{y}(t)$, восстановленными по каждому из рядов методом временных задержек или каким-либо другим (пп. 6.1.2 и 10.1.2). Методики опираются на поиск ближайших соседей в пространствах состояний [186, 285] или на построение взаимных прогностических моделей [301, 226]. Если между одновременными векторами состояния систем имеется однозначная зависимость $\mathbf{x}(t) = \mathbf{F}(\mathbf{y}(t))$, говорят об *обобщенной синхронизации* [297, 285, 195].

Вторая идея состоит в анализе *только фаз* исследуемых сигналов. Фаза является очень чувствительной переменной, поэтому наличие взаимодействия между системами при довольно общих условиях в первую очередь проявляется в наличии зависимостей между их фазами, тогда как амплитуды могут быть некоррелированными.

Если разность (развернутых) фаз постоянна $\phi_x(t) - \phi_y(t) = const$, то говорят, что имеет место *фазовая синхронизация*. Это эффект

беспороговый, т.е. может проявляться при сколь угодно слабой связи между системами, если их основные частоты очень близки друг к другу [21, 137]. При наличии даже слабых шумов строгого постоянства разности фаз быть не может, поэтому на практике рассматривают несколько смягченное условие $\left|\phi_x(t)-\phi_y(t)-const\right|<2\pi$. Это определение синхронизации $1:1$ или синхронизации на основном тоне. Может наблюдаться и синхронизация более высокого порядка – $m:n$. Она определяется условием $\left|m\phi_x(t)-n\phi_y(t)-const\right|<2\pi$. При наличии существенных шумов даже смягченное требование ограниченности разности фаз может выполняться зачастую только на конечном интервале времени. Тогда говорят об эффективной синхронизации, если этот интервал времени значительно превышает характерный период колебаний каждой из систем.

Вводятся различные численные характеристики взаимной зависимости между фазами, которые называют коэффициентами фазовой синхронизации и т.д. Наиболее распространен так называемый коэффициент средней фазовой когерентности (он имеет и много других названий):

$$R_{m,n}=\sqrt{\left\langle\cos\left(m\phi_x(t)-n\phi_y(t)\right)\right\rangle^2+\left\langle\sin\left(m\phi_x(t)-n\phi_y(t)\right)\right\rangle^2}\,, \qquad (6.27)$$

где угловые скобки означают усреднение по времени. Он равен единице, когда разность фаз остается постоянной (фазовая синхронизация), и нулю, когда каждая из систем демонстрирует колебания со своей частотой, не зависимые от другой системы. В случае не строгого захвата фаз (из-за шума и т.д.) $R_{m,n}$ может принимать промежуточное значение, характеризуя «степень зависимости» между фазами колебаний систем. Пример эффективного использования анализа фаз колебаний для решения задачи медицинской диагностики можно найти в [286].

В п. 11.3 изложен метод, основанный на моделировании фазовой динамики и позволяющий выявить еще более слабые связи и их направление.

6.5. Экспериментальный пример вместо заключения

Данная глава посвящена методикам и проблемам, возникающим на начальном этапе моделирования, получению и предварительному анализу временных рядов. Вместо заключения, в котором можно было бы сказать, что вероятность успеха при моделировании тем больше, чем объемнее и точнее априорные сведения об объекте, рассмотрим реальный пример получения наблюдаемых и построения моделей. Объект – знакомая цепь, к которой мы уже обращались: источник постоянной э.д.с. с

подключенными к нему сопротивлениями (рис.6.2,а,б). Только вместо резистора R_v включим в нее образец полупроводникового материала (InSb, антимонид индия).[13] При комнатной температуре InSb – проводник, поэтому эксперименты с ним проводят в жидком азоте, находящемся при температуре кипения – приблизительно $-77\,^{o}$C (рис.6.24). Мы выбрали его для иллюстрации, несмотря на видимую простоту (во всяком случае, по сравнению с живыми системами), из-за разнообразия демонстрируемых движений, видов математического аппарата, требуемого для их описания, и трудностей получения рядов наблюдаемых в зависимости от целей моделирования.

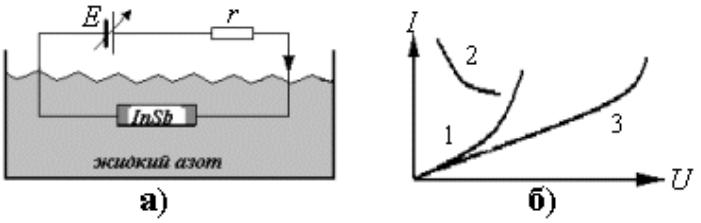

Рис.6.24. Схема эксперимента (а) и зависимость среднего тока через образец I от среднего напряжения на образце U (б)

Казалось бы, какой сложности можно ожидать от продолговатого образца, к концам которого двумя проводами подведено напряжение от источника с постоянной э.д.с.? Однако, в зависимости от условий проведения эксперимента и используемых приборов, здесь можно увидеть многообразие процессов от тривиального протекания постоянного тока до колебаний на сверхвысоких частотах и даже излучения в миллиметровом диапазоне длин волн. Регистрация процессов с характерными частотами от 1 Гц до 10^{12} Гц требует привлечения различных приборов и аналого-цифровых преобразователей. Более того, начиная с частот порядка нескольких ГГц, оцифровка и, следовательно, моделирование по временным рядам пока технически невозможны. Причем математическое моделирование наблюдаемых феноменов требует использования различных уравнений – от алгебраических до дифференциальных в частных производных.

При малых напряжениях U на концах образца он ведет себя как обычный резистор – через него течет постоянный ток силой I, а процессы моделируются алгебраическим соотношением – законом Ома $I = U/R$, где R – параметр (сопротивление образца). Характеризующие величины U и I легко измеряются (это наблюдаемые). С увеличением U линейность зависимости нарушается из-за нагрева образца, проводимость которого

[13] В настоящее время с этим узкозонным материалом, характеризующимся большой подвижностью носителей заряда, связывают надежды на увеличение быстродействия полупроводниковых устройств.

при этом возрастает, см. ветвь 1 характеристики на рис.6.24,б.

С увеличением тока I, а, следовательно, и интенсивности нагрева режим кипения жидкости переходит от пузырькового к пленочному[14]. На зависимости среднего тока от среднего напряжения это отражается в переходе рабочей точки системы на ветвь 2. Это сопровождается переходом системы «образец – источник тока – охлаждающая жидкость» в колебательный режим. Его характерные частоты – от долей до единиц Гц и в радиодиапазоне (десятки кГц – единицы МГц) – позволяют использовать АЦП и записать временные ряды. Более низкочастотные колебания определяются образованием большого числа пузырьков на поверхности, а более высокочастотные – реактивностью проводов, к которым подключено отрицательное сопротивление образца. Эти явления могут моделироваться с помощью стохастических ДУ. При этом в модель входят величины, напрямую с наблюдаемыми не связанные: тепловой поток с образца, температурная зависимость проводимости и реактивности цепи питания.

Еще более сложная для наблюдения и моделирования ситуация складывается, если для уменьшения влияния нагрева перейти от непрерывного питания к импульсному, когда напряжение на образец подается только на короткое время, а до следующего включения он успевает хорошо охладиться. При этом без теплового разрушения образца и изменения режима кипения можно достичь напряжений (ветвь 3 на рис.6.24,б), достаточных, чтобы в локальных областях образца начался ударный пробой и сформировавшиеся «куски» плазмы стали источником СВЧ-излучения. Для наблюдения этих процессов требуются специальная оснастка и СВЧ-приемники, а использование в качестве наблюдаемых тока $I(t)$ и напряжения $U(t)$ становится нерациональным. Причина в том, что не ясно, как характеризующие величины, входящие в дифференциальную модель этого колебательного механизма, связаны с такими наблюдаемыми. Более адекватным задаче становится выбор в качестве переменных модели напряженности поля в образце и концентрации носителей, учет зависимости дрейфовой скорости от напряженности поля, и т.п. Пример дифференциальной модели этого режима функционирования представлен в [27]. Еще более строгие модели рассматриваемой с виду простой системы требуют использования нелинейных уравнений в частных производных.

Этот пример нельзя считать исключением из правил на практике.

---

[14] Жидкий азот в термосе находится при температуре кипения. При малом тепловом потоке из образца на поверхности образуются отдельные пузырьки, которые при еще более интенсивном нагреве охватывают всю поверхность. Образуется паровая пленка, изолирующая образец от охлаждающей жидкости.