

С.И.МИХЕЕВ

А.С.МИХЕЕВ

СЕЙСМОРАЗВЕДКА
ОСНОВЫ СПЕКТРАЛЬНЫХ И
СТАТИСТИЧЕСКИХ СПОСОБОВ ОБРАБОТКИ
СЕЙСМИЧЕСКИХ МАТЕРИАЛОВ

2007

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ

Саратовский ордена Трудового Красного Знамени
государственный университет им. Н.Г.Чернышевского

С.И.МИХЕЕВ
А.С.МИХЕЕВ

СЕЙСМОРАЗВЕДКА

ОСНОВЫ СПЕКТРАЛЬНЫХ И СТАТИСТИЧЕСКИХ СПОСОБОВ ОБРАБОТКИ СЕЙСМИЧЕСКИХ МАТЕРИАЛОВ

Учебное пособие для студентов геофизической
специальности

Издательство «Научная Книга»
2007

Михеев С.И., Михеев А.С.

Сейсморазведка. Основы спектральных и статистических способов обработки сейсмических материалов: Учебное пособие / Изд-во «Научная книга».- Саратов, 2007, -77 с.

ISBN 978-5-9758-0584-3

В пособии кратко освещены алгоритмы и методические аспекты спектральной и статистической обработки сейсмических материалов. Приводятся примеры практического применения этих методов. Акцентированное внимание уделяется новым способам спектрального и статистического анализа.

Предназначено для студентов, обучающихся по специальности 011200 «Геофизика». В процессе обучения магистров используются все разделы учебного пособия. При обучении бакалавров из рассмотрения исключаются разделы 1.11, 2.9.7, 2.10.

Пособие будет также полезно студентам геологических факультетов других специальностей.

Рекомендовано к изданию кафедрой геофизики Саратовского госуниверситета.

ВВЕДЕНИЕ

В настоящее время геофизические методы являются наиболее эффективным инструментом геологоразведочных работ. Методике полевых наблюдений, используемой измерительной аппаратуре, обработке и интерпретации геофизических материалов посвящено большое количество публикаций, в том ориентированных на использование в учебном процессе. Однако развитие теоретических основ способов обработки геофизической информации, в том числе спектральных и статистических происходит так стремительно, что приводимые в учебной литературе материалы быстро устаревают. Кроме того, большинство весьма содержательных и в наибольшей мере ориентированных на проблемы геофизики учебников и учебных пособий по вопросам спектрального и статистического анализа стало библиографической редкостью и практически недоступно для большинства студентов. Поэтому существует необходимость издания новых учебных пособий, оперативно освещающих состояние вопросов обработки и интерпретации геофизических материалов. Настоящее пособие предназначено для методического обеспечения практикума по курсу «Сейсморазведка». Разработано в ходе реализации инновационного образовательного проекта СГУ в 2007г.

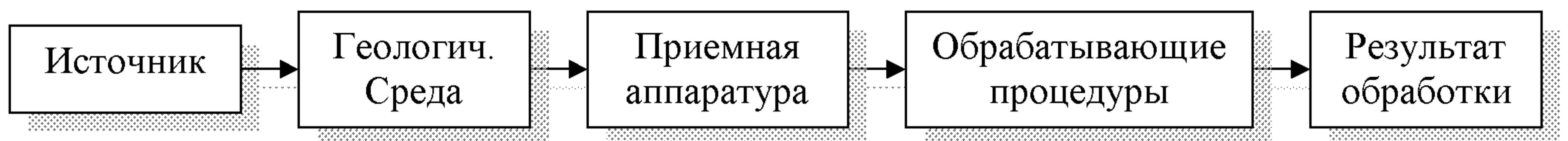
Подробно описать современные способы обработки геофизической информации в рамках одного учебного пособия практически невозможно. Это определило необходимость отбора имеющегося материала. При этом основное внимание уделено наиболее актуальным по нашему мнению вопросам, либо проблемам и методам, слабо освещенным в учебной литературе.

1. МЕТОДЫ СПЕКТРАЛЬНОГО АНАЛИЗА ГЕОФИЗИЧЕСКИХ ПОЛЕЙ

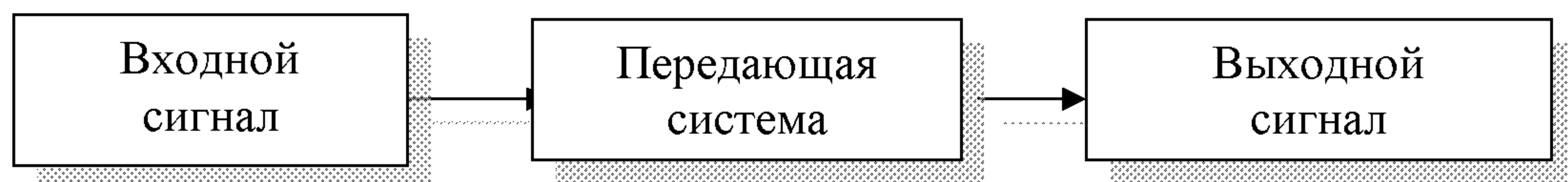
1.1. Линейные системы

1.1.1. Системы передачи сигналов в геофизике

Во многих случаях удобно изучать геофизические системы выделяя и анализируя в ней отдельные звенья (элементы). Так, зачастую целесообразно вводить в рассмотрение систему со следующими укрупненными элементами



Большинство задач геофизики сводится к анализу подсистем вида



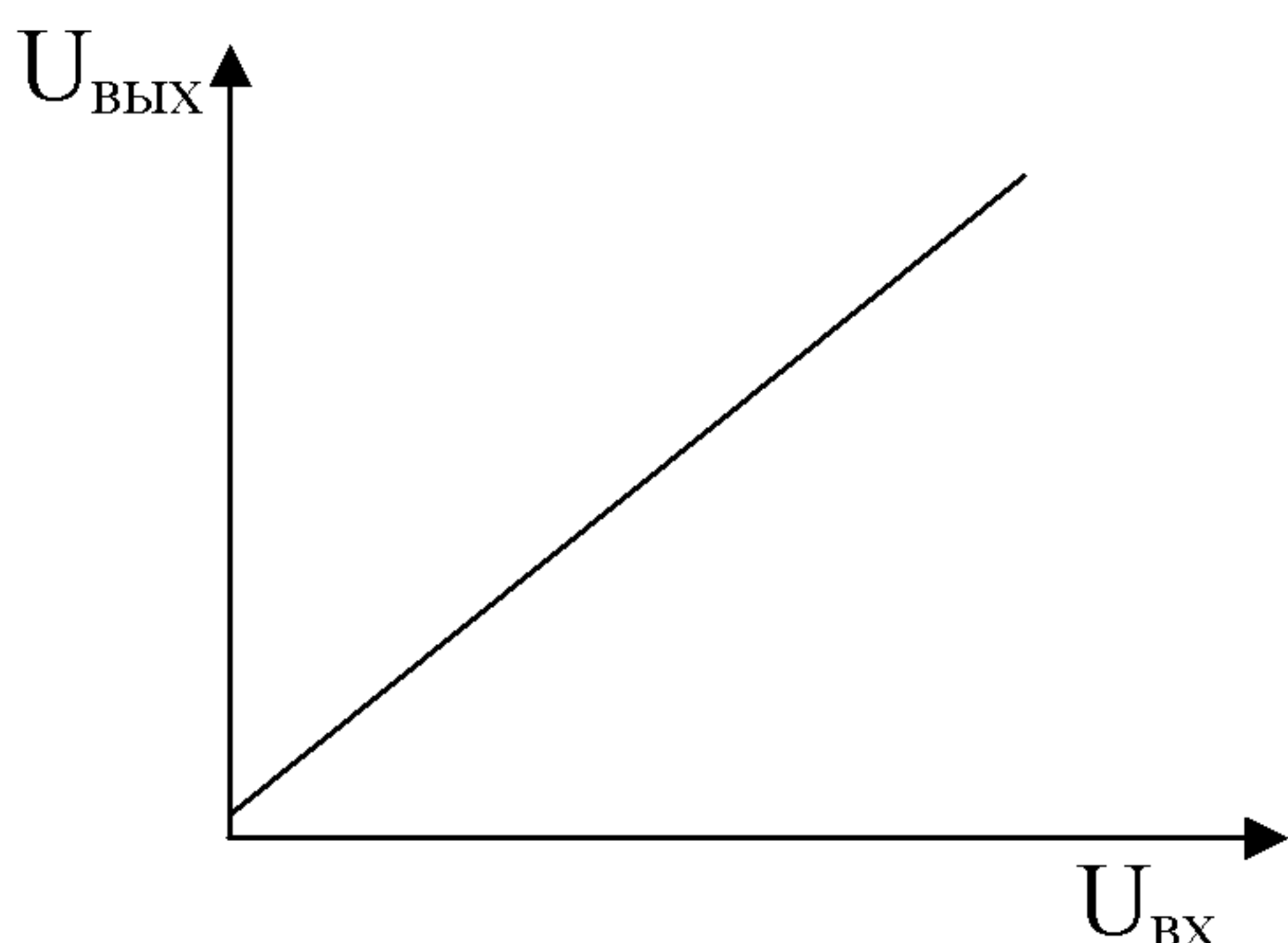
Анализ геофизических систем удобно выполнять, используя методы спектрального анализа.

1.1.2. Понятие о линейной системе

Чаще всего геофизические системы рассматриваются как линейные. Линейной системой передачи (преобразования) сигналов называется система, для которой амплитуда сигнала на выходе $U_{\text{ВЫХ}}$ пропорциональна амплитуде сигнала на входе $U_{\text{ВХ}}$.

$$U_{\text{ВЫХ}} = K * U_{\text{ВХ}}.$$

Зависимость $U_{\text{ВЫХ}}$ от $U_{\text{ВХ}}$ называется амплитудной характеристикой системы. Ее часто изображают в графической форме, представленной на рисунке 1.1.



Линейные системы бывают двух видов:

- а) с постоянными коэффициентами;
- б) с переменными коэффициентами.

В любом случае для линейных систем выполняется условие независимости коэффициента передачи системы K от входного сигнала : $K \neq K(U_{\text{ВХ}})$.

Рис. 1.1. Пример амплитудной характеристики системы.

1.1.3. Линейные преобразования

Преобразования сигналов в линейных системах называются линейными преобразованиями и обозначаются символом Λ . Пример записи линейного преобразования

$$U_{\text{ВЫХ}} = \Lambda\{U_{\text{ВХ}}\}.$$

Основное свойство линейных преобразований - аддитивность. Математически оно определяется соотношением

$$\Lambda\{X_1 + X_2\} = \Lambda\{X_1\} + \Lambda\{X_2\}.$$

Левую часть данного соотношения можно интерпретировать как реакцию системы на суммарный сигнал, а правую - как сумму реакций системы на отдельные сигналы.

Свойство аддитивности есть следствие линейности характеристики системы. При произвольном числе слагаемых имеем:

$$\Lambda\left\{\sum_{i=1}^n X_i\right\} = \sum_{i=1}^n \Lambda\{X_i\}.$$

При одинаковых X_i получаем свойство однородности:

$$\Lambda\{nX\} = n\Lambda\{X\}.$$

Исходя из свойства однородности, постоянный коэффициент можно выносить за знак линейного преобразования.

1.1.4. Сумма и произведение линейных преобразований

Суммой линейных преобразований Λ_1 и Λ_2 называется линейное преобразование Λ сопоставляющее величине X величину Y в соответствии с равенством

$$Y = \Lambda_1\{X\} + \Lambda_2\{X\}.$$

Таким образом

$$Y = \Lambda\{X\} = \Lambda_1\{X\} + \Lambda_2\{X\} = (\Lambda_1 + \Lambda_2)\{X\}.$$

Понятие суммы линейных преобразований используется в случае, когда отдельные линейные системы соединены параллельно. Например, если имеется две соединенные параллельно геофизические системы часто ставится и решается задача нахождения суммарного сигнала на выходе систем. Это соответствует нахождению суммы линейных преобразований.

Произведением линейных преобразований Λ_1 и Λ_2 называется преобразование Λ , которое получается в результате последовательного применения Λ_1 и Λ_2 к входному сигналу. Математически определяется выражением

$$Y = \{X\} = \Lambda_2 \{ \Lambda_1 \{X\} \} = (\Lambda_1 \Lambda_2) \{X\}.$$

В практике задача нахождения произведения линейных преобразований чаще всего решается в случае, когда требуется найти сигнал на выходе двух последовательно соединенных геофизических систем.

Доказано, что сумма и произведение линейных преобразований также является линейным преобразованием.

1.1.5. Методы исследования линейных систем

Как уже отмечалось большинство геофизических систем рассматриваются как линейные. Обычно полагают, что они характеризуются постоянными коэффициентами. Именно такие случаи и будут обсуждаться далее.

Исследование линейных систем может базироваться либо на временном, либо на спектральном методах.

Временной метод исследования

Временной метод основан на определении и изучении временных функций (например, напряжения $U(t)$, тока $I(t)$, смещения частиц среды $A(t)$ и др.). В этом методе составляют уравнения, в которые входят параметры системы, временные функции $U(t)$, $I(t)$, $A(t)$ и др., их производные и интегралы.

Отметим, что под независимой переменной t не обязательно понимается время. Это может быть, например, и пространственная координата.

Пример уравнения

$$a_1 \frac{dU}{dt} + a_2 U = U_0$$

Здесь U - неизвестное значение напряжения; U_0 - заданное значение напряжения; a_1 , a_2 - коэффициенты, характеризующие изучаемую систему.

Основная проблема практического применения временного метода - сложность решения интегральных и дифференциальных уравнений.

Спектральный метод

В основе спектрального метода лежит замена временных функций их спектрами. Под спектром обычно понимают совокупность слагающих его гармонических колебаний (гармоник). Метод исследования, базирующийся на

изучении спектров, называют спектральным. По сравнению с временным методом он характеризуется меньшей сложностью вычислений. Спектральный метод решения задач анализа систем обычно включает три этапа. Например:

1. Нахождение спектра сигнала ($U_{\text{вх}} \rightarrow \text{спектр}$).
2. Определение амплитуд (A) и фаз (φ) гармоник после прохождения сигнала через систему.
3. Вычисление по найденному на этапе 2 спектру сигнала на выходе системы ($U_{\text{вых}}$).

Содержание и последовательность реализации этапов определяется условиями поставленной задачи. На практике могут реализовываться не все перечисленные этапы (например, если ставится задача изучения искажений в системе спектра сигнала, а форма сигнала на выходе не интересует).

1.1.6 Принцип суперпозиции

Данный принцип заключается в независимости прохождения через линейную систему нескольких сигналов X_i одновременно. Пусть реакция системы на сигнал X_i равна Y_i , т.е. $Y_i = \Lambda\{X_i\}$. Исходя из свойства аддитивности результирующий сигнал будет равен

$$Y = \sum_{i=1}^n Y_i = \sum_{i=1}^n \Lambda\{X_i\} = \Lambda\left\{\sum_{i=1}^n X_i\right\}$$

Принцип суперпозиции имеет фундаментальное значение. Например, в спектральном анализе после разложения геофизического сигнала на гармонические или другие составляющие можно изучать воздействие системы на каждую из гармоник по отдельности. Спектр результирующего сигнала, исходя из принципа суперпозиции, можно найти просто просуммировав спектральные составляющие на выходе.

1.2. Нелинейные явления и системы в геофизике

До недавнего времени нелинейные явления и системы в геофизике вообще не рассматривались. Однако в ряде случаев наблюдались явления и процессы, которые не могли быть объяснены с позиций теории линейных систем. К ним относятся:

- сейсмическая эмиссия;
- изменение упругих свойств среды под воздействием сейсмических

вибраций;

- возникновение волн-гармоник и волн с комбинационными частотами при сейсмических работах с применением вибрационных источников;
- нарушение принципа суперпозиции (ПСС) при сейсмических исследованиях;
- сейсмoeлектрические эффекты и др.

Сейсмическая эмиссия заключается в возникновении высокочастотных шумов в деформируемых средах. Причины этого заключаются в следующем.

Реальные геологические тела в отличие от идеально упругих обладают способностью механически связывать в себе некоторое количество потенциальной энергии упругих деформаций и сохранять ее в “скрытом” виде. Среда со “скрытой” упругой энергией можно рассматривать как среда с собственными источниками энергии. Наличие “скрытой” энергии выражается в присутствии в твердых телах внутренних (структурных) напряжений. Под ними понимают систему напряжений, которые могут существовать в равновесии внутри тела, когда к его поверхности не приложены напряжения. Если же такие напряжения прикладываются, то может произойти процесс релаксации, в ходе которого геологическая среда вновь приходит к равновесному состоянию. Период времени, в течение которого осуществляется этот переход - период релаксации. Наиболее известный пример релаксации – землетрясения, когда незначительные “провоцирующие” напряжения вызывают высвобождение огромной потенциальной энергии, накопленной в земных недрах.

Другое характерное проявление нелинейности - изменение упругих свойств среды под воздействием сейсмических вибраций. Так, известен эффект возникновения постоянной составляющей поля смещений в рыхлых грунтах. Для них отмечается нелинейная связь напряжений и деформаций, следствием чего является неодинаковое искажение фаз сжатия и растяжения. Поэтому, например, при работе сейсмическими вибраторами на дневной поверхности имеет место преобладающее движение вверх, что приводит к возникновению постоянной составляющей смещений.

Еще одним примером проявления нелинейности могут служить волны-гармоники в сейсморазведке при использовании вибрационных источников. Так, при возбуждении вибратором в среде монохроматического сигнала с частотой f в среде и в системе источник-среда-приемник возникают сигналы с частотой $2f$, $3f$, $4f$ и т.д.

Волны комбинационных (суммарных и разностных) частот возникают при взаимодействии двух монохроматических колебаний. Регистрация волн

комбинационных частот противоречит принципу суперпозиции, утверждающего, что два малых колебания, приходящие в одну точку среды, накладываются друг на друга без нелинейного взаимодействия.

Сейсмоэлектрические эффекты заключаются во взаимном влиянии поля упругих волн и электрического поля друг на друга при одновременном их возбуждении. Эффект первого рода (J- эффект) проявляется в изменении силы тока I , проходящего через среду между двумя электродами под действием упругих колебаний. Эффект второго рода (E – эффект) заключается в возникновении разности потенциалов между точками, расположенными на различных расстояниях от источника излучения упругих волн.

На настоящий момент количество теоретических и экспериментальных работ, посвященных изучению нелинейных свойств реальных сред и объясняющих или подтверждающих наблюдаемые нелинейные явления уже достаточно велико. Например, первое математическое описание сейсмоэлектрического эффекта в горных породах было дано в работе Я. Френкеля, относящейся еще к 1944 году. Спустя 50 лет, С. Прайд более строго описал это явление с помощью системы уравнений, объединяющей уравнения Био и Максвелла. Впоследствии в работах Б.С. Светова, В.П. Губатенко и др. была предложена более простая формулировка этой задачи, сводящаяся к последовательному решению уравнений Био и Максвелла. Связующую роль между уравнениями Био и Максвелла при этом играет представление для стороннего тока. Величина стороннего тока может быть выражена через поверхностную плотность зарядов на границе твердой фазы или через потенциал порового раствора.

С практической точки зрения весьма важно отметить, что результаты ряда исследований (например, Сибиряков Б.П., Корнеев В. и др.) в области нелинейной сейсмологии показывают, что наиболее ярко нелинейные сейсмические эффекты должны проявляться в микронеоднородных, многокомпонентных, трещиноватых горных породах. Тем самым, по сути, обосновываются возможности прогнозирования коллекторских свойств горных пород, а также прямого прогнозирования залежей углеводородов.

В связи с вышесказанным нелинейные явления все чаще кладутся в основу новых геофизических технологий. Наиболее активно развиваются технологии, основанные на сейсмоэлектрических явлениях и сейсмические технологии с использованием нелинейных волн.

1.3. Спектр геофизического сигнала

1.3.1. Общие представления о спектре

Зарегистрированные геофизические сигналы, как правило, имеют сложную форму. Это предопределяет трудности их изучения. Анализ сигналов можно упростить, если представить их в виде суммы более простых (элементарных) сигналов. Возможность разложения сложных сигналов на гармонические составляющие (синусоиды или косинусоиды) доказана, в частности, Шарлем Фурье (1807г.).

Спектром сигнала называют совокупность слагающих его колебаний простой формы (чаще всего синусоид и косинусоид).

Сами слагающие сигнал гармонические колебания называются спектральными составляющими (гармониками).

Вновь подчеркнём, что сигналы можно раскладывать не только на гармонические, но и на другие составляющие (например, единичные функции и др.).

1.3.2. Элементарные сигналы

Выбор элементарных сигналов

Ранее указывалось, что к линейным системам применим принцип суперпозиции. Поэтому вопрос о передаче системой сигнала сложной формы может решаться путем разложения входного сигнала на сумму простых (элементарных) сигналов и изучения того, как воздействует система на каждый из них. Выходной получается суммированием реакций системы на элементарные сигналы. Выбор элементарных сигналов определяется тем, что они должны быть достаточно "удобными" для теоретических и экспериментальных исследований. В частности, должен иметь достаточно простой и удобный математический аппарат для:

а) разложения сигнала сложной формы сумму выбранных элементарных сигналов;

б) вычисления реакции линейных систем на элементарный входной сигнал;

в) синтеза выходного сигнала путем суммирования реакций системы на элементарные сигналы.

С экспериментальной точки зрения необходимо иметь возможность достаточно просто генерировать элементарные сигналы. Перечисленным

условиям удовлетворяют непрерывные сигналы в виде гармонических функций и разрывные сигналы: единичная функция, прямоугольная функция и единичный импульс (импульс Дирака).

Гармонические элементарные сигналы

К гармоническим элементарным сигналам относятся синусоидальные и косинусоидальные функции, которые записываются в обычном виде:

$$\begin{aligned} U(t) &= A \sin(\omega t + \varphi_1) = A \cos(\omega t + \varphi_2) = \\ &= A \sin(2\pi f t + \varphi_1) = A \cos(2\pi f t + \varphi_2). \end{aligned}$$

В этих выражениях A — амплитуда гармонического колебания, $\omega = 2\pi f = \frac{2\pi}{T}$ — круговая частота колебания (число радианов в секунду), f — частота колебаний (число колебаний в секунду, герцы), T — период колебания, φ — фазовый сдвиг колебания. Последний определяет величину колебания в начальный момент времени, при $t = 0$

$$U(0) = A \sin \varphi_1 = A \cos \varphi_2.$$

Таким образом, гармонический сигнал определяется амплитудой, частотой и фазой. Взаимный переход между синусами и косинусами осуществляется путем изменения фазы колебания на $\frac{\pi}{2}$.

Чаще всего в геофизической практике рассматриваются колебания, происходящие во времени. Однако, в качестве аргумента гармонических функций могут выступать и другие показатели. Так, в сейсморазведке приходится иметь дело с процессами, которые носят колебательный характер не только во времени, но и в пространстве. Аргументом таких процессов является линейная комбинация функций времени и координат пространства. Для одномерного случая, когда плоская волна распространяется вдоль координаты x , этот аргумент имеет вид:

$$\mathcal{G} = \omega t - \nu x = \omega t - \frac{\omega}{\nu} x = 2\pi(ft - \chi x).$$

Здесь ν — скорость распространения волны, χ — волновое число, ν — круговое волновое число. В каждый заданный момент времени $t = t_1$ сигнал становится функцией только расстояния x и описывает отклонения величины U от нуля в различных точках x :

$$U(x) = A \sin(\nu x) = -A \sin(2\pi\chi x).$$

Волновые числа называют также пространственными частотами, т. к.

для колебания в пространстве волновое число χ и круговое волновое число ν являются аналогами частоты f и круговой частоты ω , а длина волны $\lambda = \frac{1}{\chi} = \frac{2\pi}{\nu}$ является аналогом периода T колебания. Указанная аналогия

позволяет рассматривать основные вопросы теории спектральных представлений в частотно-временной области, а затем переносить полученные результаты в пространственно-волновую область.

Разрывные элементарные сигналы

Единичная функция. Это разрывная функция, которая при определенном значении аргумента $t = \tau$ скачком меняет свое значение от 0 до некоторого значения y и дальше остается постоянной при произвольном увеличении аргумента. Значение y чаще всего принимается за единицу, вследствие чего эта функция и названа единичной. Эту функцию называют также функцией Хевисайда. Чтобы подчеркнуть, что амплитуда единичной функции равна единице, ее иногда обозначают $1(t-\tau)$. Аналитическая запись единичной функции:

$$\sigma(t - \tau) = \begin{cases} 0 & t < \tau \\ 1 & t > \tau \end{cases}$$

Вид единичной функции, возникающей в момент времени τ показан на рис. 1.2 а.

С помощью единичной функции можно ограничить функцию в момент $t = \tau$ “слева”, помножив ее на $\sigma(t - \tau)$, либо ограничить в момент $t = \tau$ “справа”, помножив функцию на $[1 - \sigma(t - \tau)]$. Поэтому единичную функцию иногда называют ограничивающей или обрезающей функцией.

Прямоугольный импульс. Прямоугольным импульсом длительностью Δt называется разрывная функция, которая мгновенно принимает единичное значение в какой-то момент времени $t = \tau_1$, остается равной единице в течение времени Δt и мгновенно принимает значение 0 при $t = \tau_2 = \tau_1 + \Delta t$. Прямоугольный импульс длительностью Δt можно представить как разность двух единичных функций, взаимно смещенных на Δt . Вид прямоугольного импульса показан на рис. 1.2 б. Прямоугольную функцию называют также вырезающей функцией, так как если из какой-либо

функции требуется вырезать интервал длительностью $\Delta\tau$, то эту функцию надо помножить на прямоугольную функцию заданной длительности.

Прямоугольная функция обозначается символом “rect”. При ее использовании удобно, чтобы при изменении длительности функции, площадь оставалась постоянной и равной единице. В таком случае амплитуда прямоугольной функции должна быть равна $1/\Delta\tau$. Соответственно ее аналитическое выражение записывается:

$$\begin{aligned} \text{rect}(\Delta\tau, t) &= \frac{1}{\Delta\tau} [\sigma(t - \tau_1) - \sigma(t - \tau_2)] = \\ &= \frac{1}{\Delta\tau} \left[\sigma\left(t - \tau + \frac{\Delta\tau}{2}\right) - \sigma\left(t - \tau - \frac{\Delta\tau}{2}\right) \right], \end{aligned}$$

где τ_1, τ_2, τ_3 — времена соответственно начала, конца и середины прямоугольного импульса.

Прямоугольный импульс единичной площади длительностью $\Delta\tau$ с центром на времени τ будем обозначать $\text{rect}(\Delta\tau, \tau)$.

Единичный импульс. Единичным импульсом называется разрывная функция, которая при значении аргумента, равном нулю, скачком принимает бесконечно большое значение, имеющее бесконечно малую протяженность, в то время как площадь этой функции равна условной единице. Единичный импульс обычно обозначают буквой δ и если он возникает в момент $t = \tau$ то его записывают $\delta(t - \tau)$. Единичный импульс можно определить следующим образом:

$$\begin{aligned} \delta(t) &= \begin{cases} \infty, & t = 0 \\ 0, & t \neq 0 \end{cases} \\ \int_{t_1}^{t_2} \delta(t) dt &= \begin{cases} 1, & t_1 < 0 < t_2 \\ 0, & t_1 \cdot t_2 > 0 \end{cases} \end{aligned}$$

Важно отметить, что единичный импульс может быть получен из прямоугольного импульса путем устремления длительности последнего к нулю. То есть

$$\delta(t - \tau) = \lim_{\Delta\tau \rightarrow 0} \text{rect}(\Delta\tau, \tau) = \lim_{\Delta\tau \rightarrow 0} \frac{\sigma\left(t - \tau + \frac{\Delta\tau}{2}\right) - \sigma\left(t - \tau - \frac{\Delta\tau}{2}\right)}{\Delta\tau}.$$

Из данного выражения следует, что единичный импульс является производной от единичной функции, т. е. можно записать

$$\delta(t - \tau) = \frac{d}{dt} \sigma(t - \tau).$$

Умножение некоторой функции $U(t)$ на единичный импульс $\delta(t - \tau)$ вырезает из этой функции одно ее значение $U(\tau)$. Свойство единичного импульса вырезать значение какой либо функции в момент его действия называется фильтрующим свойством δ - функции.

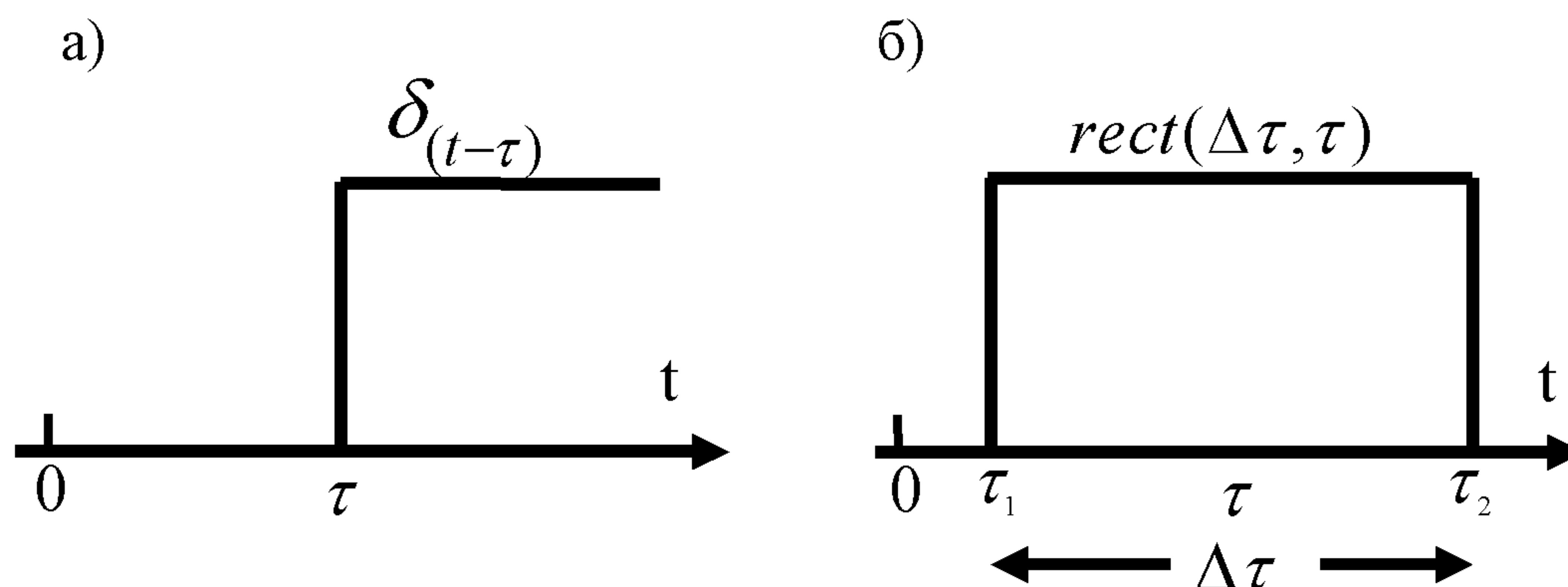


Рис. 1.2. Функции: единичная (а), прямоугольный импульс (б)

1.3.3. Ряд Фурье. Дискретный спектр сигнала

Ряд Фурье

Пусть имеется некоторый геофизический сигнал $U(t)$. Представим его в виде гармонических функций с частотами, увеличивающимися на величину $\Delta\omega = 2\pi\Delta f$. Пусть для определенности это будут косинусоиды. Тогда:

$$U(t) = C_{(0)}/2 + \sum_{-\infty}^{\infty} C_{(n)} \cos(2\pi n \Delta f t - \varphi_n),$$

где $n = 1, 2, 3 \dots$ порядковый номер слагаемого (гармоники). $C_{(n)}$ называют амплитудой гармоники, φ_n — фазовым сдвигом (фазой) гармоники. Случай $n = 1$ соответствует первой (основной) гармонике.

Разложение функции в ряд вышеприведенный вида называется разложением в ряд Фурье. Подразумевается, что коэффициенты $C_{(n)}$ вычислены таким образом, что ряд совпадает с исходной функцией $U(t)$ на интервале ее существования.

Важно указать, что ряд Фурье является периодической функцией с периодом $T = 1/\Delta f$.

Дискретные спектры

Совокупность амплитуд гармоник $C_{(n)}$ в ряде Фурье называют амплитудным спектром функции $U(t)$, совокупность φ_n — фазовым спектром. Амплитудный и фазовый спектры могут быть изображены в виде линий в системах координат $(C_{(n)}, f)$ и (φ_n, f) . Величина f принимает значения кратные Δf . Исходя из вида изображения (см. рис. 1.3) описанные спектры еще называют линейчатыми.

Помимо величин $C_{(n)}, \varphi_n$ для количественного описания спектров используют также характеристики a_n и b_n . Между парами указанных величин существует функциональная взаимосвязь

$$a_n = C_{(n)} \cos \varphi_n;$$

$$b_n = C_{(n)} \sin \varphi_n.$$

Из теории рядов Фурье известно, что a_n и b_n могут быть определены для периодической функции с периодом T по формулам

$$a_n = \frac{2}{T} \int_{T/2}^{T/2} U(t) \cos 2\pi n \Delta f t \, dt,$$

$$b_n = \frac{2}{T} \int_{T/2}^{T/2} U(t) \sin 2\pi n \Delta f t \, dt.$$

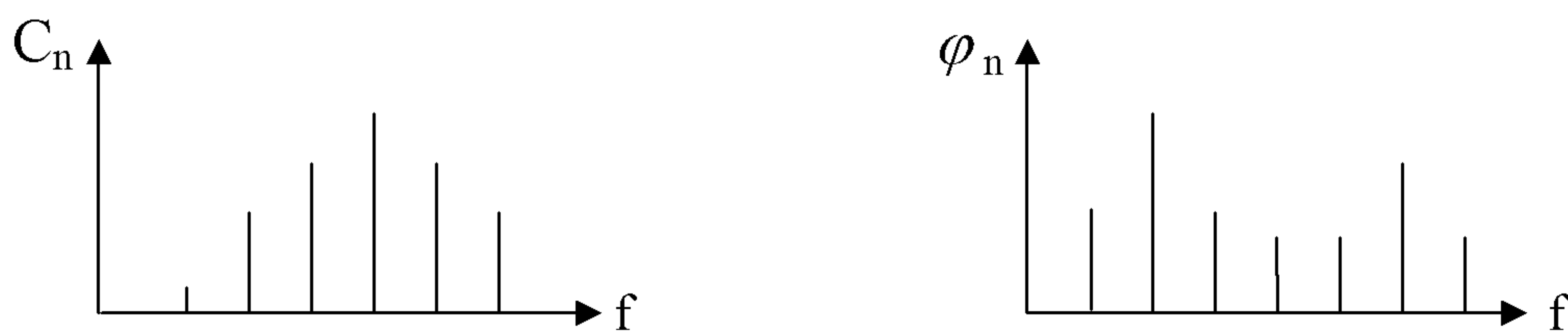


Рис. 1.3. Линейчатые спектры

Комплексная форма ряда Фурье

Действительная форма записи ряда Фурье может быть переписана в компактной комплексной форме. С этой целью синусы и косинусы заменяются показательными функциями с использованием формул Эйлера.

$$\cos(2\pi n \Delta f t - \varphi_n) = [e^{j(2\pi n \Delta f t - \varphi_n)} + e^{-j(2\pi n \Delta f t - \varphi_n)}] / 2,$$

где e — основание натурального логарифма; j — мнимая единица.

Используя формулы Эйлера действительная форма записи ряда Фурье может быть преобразована к виду

$$U(t) = \sum_{-\infty}^{\infty} C'_{(n)} e^{2\pi n \Delta f t}.$$

Здесь $C'_{(n)}$ комплексная величина

$$C'_{(n)} = (C_{(n)}/2) e^{-j \varphi_n}.$$

Согласно теории рядов Фурье

$$C'_{(n)} = 1/T \int_{T/2}^{T/2} U(t) e^{2\pi n \Delta f t} dt.$$

Комплексный частотный спектр сигнала

Величину $C'_{(n)}$ называют комплексным частотным спектром сигнала. Он может быть записан

$$C'_{(n)} = \frac{C_{(n)}}{2} e^{-j \varphi_n} = \frac{a_n - j b_n}{2}.$$

Модуль $|C'_{(n)}|$ называют амплитудным спектром, а аргумент $C'_{(n)}$ — фазовым частотным спектром. Для вычисления амплитудного и фазового спектров применяются формулы

$$|C'_{(n)}| = \frac{\sqrt{a_n^2 + b_n^2}}{2}, \quad \varphi_n = \operatorname{arctg}\left(\frac{b_n}{a_n}\right).$$

1.4. Спектры неустановившихся колебаний (интеграл Фурье)

Ряд Фурье удобен для анализа периодических колебаниях так как сам является периодической функцией. В геофизической практике значительно чаще приходится иметь дело с неустановившимися колебаниями, которые изменяют свои характеристики со временем и имеют ограниченную длительность. Ряд Фурье может применяться и в таких случаях, если рассматривать ограниченные во времени сигналы как периодические с неограниченно возрастающим периодом T (т.е. $T \rightarrow \infty$). При этом для дискретного спектра расстояние между спектральными линиями, т.е., $\Delta f = 1/T$ будет уменьшаться и при $T \rightarrow \infty$ становится бесконечно малым (далее обозначаем его df).

Дискретно меняющаяся в дискретном спектре частота $n \Delta f$ переходит в непрерывно изменяющуюся частоту f . С учетом вышесказанного комплексная амплитуда $C'_{(n)}$ запишется

$$\begin{aligned} \lim_{T \rightarrow \infty} C'_{(n)} &= \lim_{T \rightarrow \infty} \left[\frac{1}{T} \int_{-T/2}^{T/2} U(t) e^{-j2\pi n \Delta f t} dt \right] = \\ &= df \int_{-\infty}^{\infty} U(t) e^{-j2\pi f t} dt = S(f) df. \end{aligned}$$

Таким образом

$$S(f) = \int_{-\infty}^{\infty} U(t) e^{-j2\pi f t} dt.$$

Делая аналогичные преобразования для $U(t)$ (см. ряд Фурье в комплексной форме) получим

$$U(t) = \int_{-\infty}^{\infty} S(f) e^{j2\pi f t} df.$$

Приведенные выражения для вычисления $S(f)$ и $U(t)$ являются фундаментальными в теории разложения колебаний и составляют так называемую пару преобразований Фурье. Выражение для определения спектра $S(f)$ по сигналу $U(t)$ называют прямым преобразованием Фурье, а выражение, позволяющее вычислить сигнал во временной области по его спектру — обратным.

Спектр $S(f)$ в общем случае комплексная величина. Ее можно записать в показательной форме

$$S(f) = \Phi(f) e^{-j\varphi(f)},$$

и в алгебраической форме

$$S(f) = A(f) + jB(f).$$

Функцию $\Phi(f)$ называют амплитудно-частотным спектром неустановившихся колебаний, а функцию $\varphi(f)$ — фазо-частотным спектром.

Имеют место очевидные, часто используемые в практике соотношения

$$\Phi(f) = \sqrt{A(f)^2 + B(f)^2}; \quad \varphi(f) = \operatorname{arctg} \left[\frac{B(f)}{A(f)} \right]$$

1.5. Основные теоремы о преобразовании сигналов спектров

Теорема суперпозиции (теорема о линейности спектрального преобразования)

Пусть имеется несколько сигналов и соответствующих им спектров

$$U_{1(t)} \longleftrightarrow S_{1(f)}$$

$$U_{2(t)} \longleftrightarrow S_{2(f)}$$

.....

$$U_{n(t)} \longleftrightarrow S_{n(f)},$$

тогда спектр суммы этих сигналов (функций), умноженных на постоянные числа C_i , равен сумме спектров этих сигналов, умноженных на те же числа, т.е.

$$U_{\Sigma(t)} = \sum_{i=1}^n C_i U_{i(t)} \longleftrightarrow S_{\Sigma(f)} = \sum_{i=1}^n C_i S_{i(f)}.$$

Доказательство этого соотношения основано на применении прямого преобразования Фурье.

$$\begin{aligned} S_{(f)} &= \int_{-\infty}^{\infty} U_{\Sigma(t)} e^{-j2\pi ft} dt = \int_{-\infty}^{\infty} \left(\sum_{i=1}^n C_i U_{i(t)} \right) e^{-j2\pi ft} dt = \\ &= \int_{-\infty}^{\infty} C_1 U_{1(t)} e^{-j2\pi ft} dt + \dots + \int_{-\infty}^{\infty} C_n U_{n(t)} e^{-j2\pi ft} dt = \\ &= C_1 \int_{-\infty}^{\infty} U_{1(t)} e^{-j2\pi ft} dt + \dots + C_n \int_{-\infty}^{\infty} U_{n(t)} e^{-j2\pi ft} dt = \\ &= C_1 S_{1(f)} + \dots + C_n S_{n(f)} = \sum_{i=1}^n C_i S_{i(f)} = S_{\Sigma(f)}. \end{aligned}$$

Что и требовалось доказать.

Теорема о смещении сигнала во времени и спектра по частоте (теорема смещения)

Рассмотрим, как изменится спектр сигнала, если сместить его во времени, не изменяя форму. С этой целью запишем прямое преобразование Фурье, заменив переменную t в функции, описывающей сигнал, на переменную $t_1 = t - \tau$, где τ — величина смещения. Предварительно отметим, что $t = t_1 + \tau$ и $dt = dt_1$. Получим следующее выражение для спектра смещенного сигнала

$$\begin{aligned}
S_{\tau}(f) &= \int_{-\infty}^{\infty} U(t-\tau) e^{-j2\pi ft} dt = \int_{-\infty}^{\infty} U(t_1) e^{-j2\pi (t_1 + \tau)} dt_1 = \\
&= e^{-j2\pi f\tau} \int_{-\infty}^{\infty} U(t_1) e^{-j2\pi ft_1} dt_1 = e^{-j2\pi f\tau} S(f).
\end{aligned}$$

Обратим внимание на то, что подчеркнутый интеграл равен $S(f)$, т.е. спектру исходного сигнала, хотя интегрирование и происходит по переменной t_1 , а не по переменной t . Данное обстоятельство не имеет значения, так как интегрирование происходит в бесконечных пределах.

Таким образом, смещение сигнала во времени на величину $\pm \tau$ соответствует умножению спектра на функцию $e^{-j2\pi f\tau}$ (функцию смещения). Проанализируем полученный результат.

$$\begin{aligned}
S_{\pm \tau}(f) &= e^{\pm j2\pi f\tau} S(f) = e^{\pm j2\pi f\tau} |S(f)| e^{j\varphi(f)} = \\
&= |S(f)| e^{j(\varphi(f) \pm 2\pi f\tau)}
\end{aligned}$$

Отсюда амплитудный спектр смещенного во времени сигнала есть

$$|S_{\pm \tau}(f)| = |S(f)|,$$

а фазовый спектр

$$\varphi_{\pm \tau}(f) = \varphi(f) \pm 2\pi f\tau.$$

Вывод: смещение сигнала во времени не меняет амплитудный спектр, а влияет лишь на фазовый. На применении этой теоремы базируется, например, частотная теория направленного приема в сейсморазведке.

Теорема изменения масштабов сигнала и спектра (теорема масштабов)

Если у некоторого сигнала $U(t)$ заменить аргумент t на аргумент at ($a \neq 1$), то изменяется его горизонтальный масштаб. При $|a| > 1$ сигнал сжимается, при $|a| < 1$ сигнал растягивается.

Чтобы изменение горизонтального масштаба не меняло площади сигнала, будем далее рассматривать и анализировать спектр функции $U_{a(t)} = aU(at)$. Спектр такого сигнала обозначим $S_{a(f)}$.

$$S_{a(f)} = \int_{-\infty}^{\infty} aU(at) e^{-j2\pi ft} dt = \int_{-\infty}^{\infty} U(at) e^{-j2\pi (f/a)(at)} d(at) =$$

$$= \int_{-\infty}^{\infty} U(t_1) e^{-j2\pi (f/a) t_1} dt_1 = S(f/a).$$

По сути доказана теорема масштабов, которая формулируется следующим образом: при изменении горизонтального масштаба сигнала горизонтальный масштаб соответствующего ему спектра меняется в обратном направлении (при растяжении сигнала спектр сжимается, при сжатии — растягивается).

Теорема позволяет сделать качественный вывод: более длительным по времени сигналам соответствуют более узкие спектры и наоборот, более коротким сигналам соответствуют более широкие спектры. Такой вывод имеет большое практическое значение. Например в сейсморазведке все большее внимание уделяется повышению временной разрешающей способности метода, что нацеливает на применение минимальных по длительности сигналов. В частотной области это соответствует условию достижения максимально широкого спектра.

Теорема о произведении спектров и сигналов (теорема свертывания)

Пусть имеется два геофизических сигнала, описываемых функциями $U_{1(t)}$ и $U_{2(t)}$. Спектры этих сигналов обозначим через $S_{1(f)}$ и $S_{2(f)}$. Рассмотрим, что представляет из себя функция $U_{(t)}$, спектр которой $S_{(f)}$ равен произведению спектров $S_{1(f)}$ и $S_{2(f)}$.

$$S_{(f)} = S_{1(f)} S_{2(f)} .$$

Теорема: функция $U_{(t)}$ равна интегралу следующего

$$U_{(t)} = \int_{-\infty}^{\infty} U_1(\tau) U_2(t-\tau) d\tau .$$

Записанный интеграл называется “интегралом свертки” и реализует процедуру “свертки” или “конволюции” функций $U_{1(t)}$ и $U_{2(t)}$. Для обозначения конволюции используется обозначение $*$.

Таким образом, рассматриваемая теорема утверждает:

$$U_{1(t)} * U_{2(t)} = U_{(t)} \langle \text{-----} \rangle S_{1(f)} S_{2(f)} = S_{(f)}.$$

Доказательство:

$$S_{(f)} = S_{1(f)} S_{2(f)} = S_{2(f)} \int_{-\infty}^{\infty} U_1(\tau) e^{-j2\pi f\tau} d\tau =$$

(Примечание: в использованном прямом преобразовании Фурье заменили t на τ . Это не изменит результата, так как интегрирование происходит в бесконечных пределах)

$$\int_{-\infty}^{\infty} U_1(\tau) (S_{2(f)} e^{-j2\pi f\tau}) d\tau = \int_{-\infty}^{\infty} U_1(\tau) \left(\int_{-\infty}^{\infty} U_2(t-\tau) e^{-j2\pi ft} dt \right) d\tau =$$

(Примечание: внесли $S_{2(f)}$ под знак интеграла, а затем воспользовались теоремой смещения)

$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} U_1(\tau) U_2(t-\tau) d\tau \right) e^{-j2\pi ft} dt = \int_{-\infty}^{\infty} U(t) e^{-j2\pi ft} dt.$$

(Примечание: изменяли порядок интегрирования.)

Таким образом, получили

$$S_{(f)} \langle \text{-----} \rangle U_{(t)}.$$

Что и требовалось доказать.

Существует и обратная теорема: спектр сигнала, соответствующего произведению функций равен свертке спектров этих функций.

Свертка сопряженных функций

Функции с противоположными знаками аргумента называются сопряженными. Пусть имеется две функции $U_1(t)$ и $U_2(t)$ и соответствующие им спектры $S_{1(f)}$ и $S_{2(f)}$

$$U_{1(t)} \langle \text{-----} \rangle S_{1(f)} = |S_{1(f)}| e^{j\varphi_1(f)}$$

$$U_{2(t)} \langle \text{-----} \rangle S_{2(f)} = |S_{2(f)}| e^{j\varphi_2(f)}$$

Обозначим $S'_{2(f)}$ — спектр функции $U_{2(-t)}$. Из свойства четности амплитудно-частотного спектра и нечетности фазо-частотного следует

$$S'_{2(f)} = |S_{2(f)}| e^{-j\varphi_2(f)}$$

Запишем интеграл свертки, сменив знак у аргумента $U_{2(t)}$

$$U_{(t)} = \int_{-\infty}^{\infty} U_1(\tau) U_2(\tau - t) d\tau \langle \text{-----} \rangle S_{(f)} = |S_{(f)}| e^{j\varphi(f)} =$$

(Примечание: смотри теорему свертки)

$$S_{1(t)}S_{2(t)} = |S_{1(t)}| |S_{2(t)}| e^{j(\varphi_1 - \varphi_2)}$$

Таким образом,

$$|S_{(t)}| = |S_{1(t)}| |S_{2(t)}|$$

$$\varphi_{(t)} = \varphi_{1(t)} - \varphi_{2(t)}$$

Сформулировать вывод предлагается самостоятельно.

1.6. Функция автокорреляции и ее спектр

Функция автокорреляции

Рассмотрим важный для практики частный случай, когда сворачиваются сопряженные одинаковые по форме функции. Пусть $U_{2(t)} = U_{1(-t)}$. Из четности амплитудно-частотного спектра и нечетности фазо-частотного спектра следует, что рассматриваемые сигналы имеют одинаковые амплитудные спектры, но противоположные по знаку фазовые. Исходя из ранее рассмотренного случая свертки сопряженных функций, можно заключить, что полученная функция имеет спектр $S_{(t)}$ со свойствами:

$$|S_{(t)}| = |S_{1(t)}|^2; \quad \varphi_{(t)} = 0.$$

Определение: функцией автокорреляции (ФАК) сигнала $U_{1(t)}$ называется функция $U_{(t)}$, равная свертке этой функции на себя с противоположным знаком аргумента.

ФАК имеет нулевой фазовый спектр и является представителем некоторого класса сигналов, обладающих одинаковым амплитудным спектром. Их форма различна и определяется фазовым спектром.

Теорема Винера-Хинчина

Пусть имеется сигнал $f(t)$, определенный на интервале $-\frac{T}{2} + \frac{T}{2}$. Будем полагать, что при $|t| > \frac{T}{2}$ $f(t) = 0$.

Запишем для такого случая прямое и обратное преобразования Фурье:

$$F_{(\omega)} = \int_{-\infty}^{\infty} f_{(t)} e^{-j\omega t} dt; \quad f_{(t)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_{(\omega)} e^{j\omega t} d\omega.$$

Автокорреляционная функция $B_{nn(t)}$ сигнала $f(t)$ определяется выражением

$$f(t) * f(-t) = B_{nn(t)} = \int_{-T/2}^{T/2} f(\tau) f(\tau-t) dt.$$

(интегрирование выполняется в конечных пределах, так как за пределами интервала $-T/2 \div +T/2$ по условию $f(t)=0$).

Выразим в $f(\tau-t)$ через обратное преобразование Фурье.

$$B_{nn(t)} = \int_{-T/2}^{T/2} f(\tau) \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega(\tau-t)} d\omega d\tau.$$

Изменим далее порядок интегрирования и воспользуемся прямым преобразованием Фурье

$$B_{nn(t)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{-j\omega t} d\omega \int_{-T/2}^{T/2} f(\tau) e^{j\omega \tau} d\tau = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{-j\omega t} F(-\omega) d\omega.$$

Здесь

$$F(\omega) = |F(\omega)| e^{j\varphi t}; \quad F(-\omega) = |F(\omega)| e^{-j\varphi t}; \quad F(\omega)F(-\omega) = |f(\omega)|^2.$$

Величина $F(\omega)F(-\omega) = |f(\omega)|^2$ называется энергетическим спектром.

Обозначим его через $P(\omega)$. Тогда

$$B_{nn(t)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\omega) e^{-j\omega t} d\omega.$$

Таким образом, энергетический спектр $P(\omega)$ является Фурье-трансформантой функции автокорреляции и его можно найти по формуле

$$P(\omega) = \int_{-\infty}^{\infty} B_{nn(t)} e^{-j\omega t} dt.$$

Введем понятие интегральный спектр $\lambda(\omega)$

$$\lambda(\omega) = \int_{-\pi}^{\omega} P(\omega) d\omega.$$

Тогда энергетический спектр выражается через $\lambda(\omega)$ как $P(\omega) = d\lambda / d\omega$ и выражение для функции автокорреляции примет вид

$$B_{nn(t)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega t} d\lambda(\omega).$$

Данное соотношение известно как теорема Винера-Хинчина.

При решении геофизических задач важно, прежде всего, то, что энергетический спектр и функция автокорреляции связаны парой преобразований Фурье.

1.7. Минимально-фазовые операторы и сигналы

На основе обратного преобразования Фурье геофизический сигнал можно представить в виде суммы синусоид или косинусоид с различными частотами, амплитудами и фазами. Различным фазовым спектрам соответствуют отличающиеся по форме сигналы. Априори фазы гармоник неизвестны. Поэтому на практике часто делают те или иные предположения о свойствах фазо-частотного спектра. Очень часто реальные геофизические сигналы, регистрирующие и обрабатывающие комплексы и др. рассматриваются как минимально-фазовые сигналы или фильтры (операторы). В пользу такого предположения свидетельствуют многочисленные теоретические и экспериментальные данные.

Минимально-фазовым оператором (фильтром) называют такой оператор, который создает на выходе системы минимально-возможный фазовый сдвиг гармоник. Из всех сигналов с одинаковым амплитудным спектром минимально-фазовой характеристикой наибольшей скоростью нарастания энергии.

Интерес к минимально-фазовым сигналам и фильтрам объясняется тем обстоятельством, что вещественная и мнимая части комплексной частотной характеристики таких фильтров взаимосвязаны.

Пусть $U_{(\omega)}$ -вещественная, а $Q_{(\omega)}$ -мнимая части частотной характеристики:

$$H(\omega) = U(\omega) - jQ(\omega).$$

Можно показать, что

$$U_{(\omega)} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{Q_{(\nu)}}{\nu - \omega} d\nu; \quad Q_{(\omega)} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{U_{(\nu)}}{\nu - \omega} d\nu.$$

Эти интегралы называют преобразованиями Гильберта. Аналогичные зависимости можно установить между амплитудно-частотной и фазовой характеристиками.

$$\ln |H_{(\omega)}| = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\psi(\nu)}{\nu - \omega} d\nu; \quad \psi(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\ln |H_{(\nu)}|}{\nu - \omega} d\nu.$$

Данные соотношения можно использовать для нахождения фазовой характеристики минимально-фазового фильтра (сигнала) по его амплитудному спектру. В результате, имея только амплитудный спектр минимально-фазового сигнала можно полностью восстановить его форму во

временной области.

При обработке сейсмических материалов реальные импульсы иногда заменяются нуль - фазовыми эквивалентами. Фазовый спектр нуль – фазового импульса тождественно равен нулю. Такой импульс симметричен относительно основного максимума (или минимума) с наибольшей амплитудой.

1.8. Оценка формы сигнала

Оценка формы сигнала для случая минимально-фазовой модели

Из предыдущего раздела следует, что в рамках минимально-фазовой модели между амплитудным и фазовым спектрами сигнала существует функциональная взаимосвязь. Поэтому, имея информацию об амплитудном спектре можно определить фазовый спектр и восстановить сам сигнал.

Покажем, как можно вычислить амплитудный спектр исходного сигнала.

Пусть $S(t)$ -исходный сигнал; $J(t)$ -импульсная сейсмограмма; T -длительность сигнала.

Регистрируемую сейсмическую запись можно представить в виде свертки

$$U(t) = \int_{t-T}^t S(\tau) J(t-\tau) d\tau .$$

Исходя из данного выражения можно показать, что АКФ сейсмограммы B_{uu} может быть представлена через АКФ исходного сигнала

B_{ss} и АКФ импульсной сейсмограммы B_{jj}

$$B_{uu}(\theta) = \frac{1}{T} \int_{t-T-\theta}^{t-\theta} B_{ss}(\varphi) B_{jj}(\theta - \varphi) d\varphi .$$

Экспериментально показано, что B_{jj} , близка к единичному импульсу, т.е.

$$B_{jj}(\theta - \varphi) = a\delta(\theta - \varphi) ,$$

где $a = \text{const}$.

Подставляя данное выражение в формулу для вычисления B_{uu} и используя свойство δ -импульса вырезать значение функции в момент его действия получим:

$$B_{uu}(\theta) = \frac{a}{T} \int_{t-T-\theta}^{t-\theta} B_{ss}(\varphi) \delta(\theta - \varphi) d\varphi = \frac{a}{T} B_{ss}(\theta) \int_{t-T-\theta}^{t-\theta} d\varphi = aB_{ss}(\theta).$$

Таким образом, АКФ сейсмической трассы с точностью до постоянного множителя равен АКФ сигнала.

Зная B_{ss} , можно определить амплитудный спектр сигнала на основе теоремы Винера-Хинчина. Согласно теоремы, энергетический спектр сигнала вычисляется по формуле:

$$P_{(\omega)} = \int_{-\infty}^{\infty} B_{ss}(\theta) e^{j\omega\theta} d\theta.$$

Амплитудный спектр тогда равен:

$$|S_{(\omega)}| = \sqrt{P_{(\omega)}}$$

Полагая сигнал минимально-фазовым и применяя преобразования Гильберта, либо приравнивая фазовый спектр нулю, легко через обратное преобразование Фурье найти сам сигнал $S(t)$.

Оценка формы смешанно-фазового сигнала

Функциональная связь между амплитудным и фазовым спектрами имеет место только для минимально-фазовой модели сигнала. Однако, большой практический интерес представляет оценка исходного сигнала произвольной формы. Такая оценка может базироваться на кепстральном преобразовании. В соответствии с этим преобразованием геофизический сигнал преобразуется в кепстр, компоненты которого связаны аддитивно. Например, сейсмическую трассу $U(t)$ можно представить как результат свертки импульсной сейсмограммы $J(t)$, полученной из последовательности коэффициентов отражения с сигналом $S(t)$. При переходе в частотную область свертке $J(t)$ и $S(t)$ соответствует умножение их спектров.

$$U_{(\omega)} = J_{(\omega)} S_{(\omega)}.$$

Задача оценки формы исходного сигнала сводится к исключению влияния импульсной трассы, т.е. устранения мешающего действия мультитактивной помехи $J_{(\omega)}$. Для решения этой задачи прологарифмируем выражение для $U_{(\omega)}$, после чего в правой части уравнения получим вместо произведения сумму:

$$\ln U_{(\omega)} = \ln J_{(\omega)} + \ln S_{(\omega)}.$$

Применим к этой сумме обратное преобразование Фурье, в результате чего получим:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \ln U_{(\omega)} e^{j\omega t} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \ln J_{(\omega)} e^{j\omega t} d\omega + \frac{1}{2\pi} \int_{-\infty}^{\infty} \ln S_{(\omega)} e^{j\omega t} d\omega.$$

В левой части данного выражения находится так называемый кепстр трассы $\hat{U}(t)$, в правой-сумма кепстра импульсной трассы $\hat{J}(t)$ и кепстра сигнала $\hat{S}(t)$.

$$\hat{U}(t) = \hat{J}(t) + \hat{S}(t).$$

Кепстр сигнала является непрерывной функцией, которая сосредоточена в относительно непротяженном временном интервале вблизи значения $t=0$ и быстро убывает с возрастанием t . Кепстр же импульсной трассы распределен в большем временном интервале. Поэтому с целью определения спектра сигнала выбирается ограниченное число первых отсчетов η кепстра трассы и пересчитывают их в частотную область. Величина η должна удовлетворять условию

$$\eta \Delta t \approx \tau,$$

где τ - длина кепстра сигнала.

1.9. Сглаживание трассы специальными функциями

При расчете спектров геофизических сигналов приходится оперировать с конечной последовательностью отсчетов, соответствующей определенному временному или пространственному интервалу геофизических наблюдений. Получение такой последовательности эквивалентно умножению бесконечной последовательности отсчетов на временную функцию прямоугольной формы. В частотной области умножение эквивалентно свертке Фурье-преобразования трассы $U(\omega)$ с Фурье-преобразованием прямоугольной весовой функции $F(\omega)$:

$$F(\omega) = 2L_m \frac{\sin \omega L_m}{\omega L_m}.$$

Таким образом, результирующая частотная характеристика анализируемого интервала наблюдений будет равна:

$$K(\omega) = U(\omega) * F(\omega).$$

Применение прямоугольной весовой функции может привести к появлению больших искажений особенно при анализе коротких анализируемых интервалов (например, небольших временных окон при анализе сейсмограмм). Для борьбы с указанными искажениями применяют

окна с неравномерным взвешиванием. Они позволяют постепенно уменьшать значения в начальной и конечной частях анализируемых последовательностей наблюдений. При выборе вида и параметров весовых функций учитывают то обстоятельство, что резкие их срезы во временной области приводят к сильным искажениям в области высоких частот. Однозначного способа определения наилучшей весовой функции не существует. Так как при ее выборе приходится основываться на компромиссе между различными требованиями, то оптимальную функцию обычно находят путем переборных различных вариантов.

1.10. Способы повышения точности и разрешенности спектральных оценок

Спектральные оценки по методу максимальной энтропии

При анализе геофизических наблюдений обычно полагают, что за пределами анализируемого временного или пространственного окна отсчеты равны нулю. Такой прием по сути является одним из путей преодоления неоднозначности. Другой из таких методов - метод максимальной энтропии (МЭ) в геофизике предложил Дж. Бург (1967). Данный метод в ряде случаев позволяет повысить разрешенность спектров сигналов, в связи с чем получил широкое распространение.

Само понятие энтропии было введено К.Шенноном в 1948 году. Энтропия характеризует состояние и статистические свойства системы в плане упорядоченности ее элементов и может быть вычислена по формуле:

$$H = - \sum_{i=1}^n P_i \log_e P_i.$$

Здесь P_i - вероятность нахождения элементов системы в определенном (i -м) состоянии, а $\sum_{i=1}^n P_i = I$. Основание логарифма e в принципе может быть любым числом, но обычно его полагают равным 2.

Максимум упорядоченности элементов системы соответствует случаю, когда вероятность K -го состоянию возрастает до единицы ($P_K = I$), а всех остальных равна нулю. При таких условиях энтропия равна нулю. С уменьшением упорядоченности энтропия увеличивается. Доказано, что энтропия достигает максимума при достижении состояния равновесия, для

которого $P_1 = P_2 = \dots = P_n = I/n$.

Бург на примере анализа сейсмических трасс показал, как можно найти энергетический спектр, образуя с помощью специальной фильтрации исходных отсчетов новое множество с максимальной энтропией. Требование максимальной энтропии эквивалентно требованию того, чтобы отсчеты трассы были максимально «случайными» вне анализируемого временного окна. Спектры, рассчитанные по такому множеству имеют высокую разрешенность.

Спектральные оценки по методу максимального правдоподобия

Метод максимального правдоподобия (ММП) создан в 1969 году Дж. Кейпоном. В ММП вместо фиксированных временной и частотной весовых функций используются весовые функции, зависящие соответственно от волнового числа или от частоты. Весовые функции рассчитываются из условия, чтобы максимально подавить все гармоники спектра за исключением одной, полезной.

Пусть сигнал на входе некоторой геофизической системы является синусоидой с амплитудой A . В ММП выдвигается требование, чтобы сигнал на выходе системы X_k был равен сигналу на входе. В нашем случае имеем:

$$Ae^{j2\pi f k \Delta t} = \sum_{n=1}^N W_n A e^{j2\pi f \Delta t (k+1-n)}.$$

Здесь W - импульсная реакция системы, Δt - шаг дискретизации; K -номер отсчета; f -частота.

Деля обе части на $Ae^{j2\pi f \Delta t (k+1-n)}$ получим:

$$1 = \sum_{n=1}^N W_n e^{j2\pi f \Delta t (1-n)}.$$

Из теоретических соображений система должна пропускать и комплексно-сопряженный сигнал.

$$Ae^{-j2\pi f \Delta t}.$$

Если на входе одновременно действует аддитивная помеха E с нулевым математическим ожиданием, то она вызовет на выходе системы реакцию (уравнение свертки):

$$\sum_{n=1}^N W_n E_{k+1-n}.$$

Так как математическое ожидание помехи равно 0, то ее дисперсия на выходе с учетом сигнала будет

$$\delta^2 = E\left(\sum_{n=1}^N W_n E_{k+1-n}\right)^2,$$

где $E=A^2$.

ММП предусматривает минимизацию δ^2 .

Метод ММП обычно используется, когда известно, что сигнал хорошо аппроксимируется синусоидой, а фон отвечает гауссовскому процессу.

1.11. Z-преобразование

Z - преобразование последовательностей дискретных чисел, в частности цифровых сейсмических записей, позволяет давать весьма удобные и компактные описания процессов обработки. Z-преобразованием последовательности $y_t=y_1, y_2, y_3, \dots, y_n$ называется полином следующего вида

$$y(z)=y_0+y_1z+y_2z^2+\dots+y_{n-1}z^{n-1}.$$

Пусть, например, геофизический сигнал описывается последовательностью $y_t=3, -1, -2$. Тогда Z-преобразование такого сигнала будет

$$y(z)=3z^0 + (-1)z^1 + (-2)z^2=3 - z - 2z^2.$$

Исходя из определения по заданному Z-преобразованию можно легко восстановить исходную временную последовательность. Например, если $Y(z)=1-z^2$, то исходная временная последовательность будет равна $Y_{(t)}=1, 0, -1$.

Анализируя выражения для Z - преобразования, можно отметить, что символ z, являющийся сомножителем члена полинома y_1z , показывает, что элемент временной последовательности y_1 смещен на один шаг временного квантования записей по отношению к началу отсчета процесса; символ z^2 означает, что элемент y_2 смещен на два шага, символ z^p — что элемент y_p смещен на p шагов, и т.д. Это частично раскрывает смысл Z-преобразования. Чтобы сделать это с большей полнотой обратимся к прямому преобразованию Фурье:

$$Y_{(\omega)}=1/2\pi n \sum_{t=0}^{n-1} y_t e^{-j\omega t}.$$

Перепишем эту формулу, заменив t на $i\Delta t$ ($i=0,1,\dots,n-1$), положив $\Delta t=1$ и отбросив масштабный коэффициент $2\pi/n$:

$$Y_{(\omega)} = \sum_{t=0}^{n-1} y_t e^{-j\omega t} .$$

Придавая i значения $0,1,\dots$, представим это выражение следующим образом:

$$Y_{(\omega)} = y_0 + y_1 e^{-j\omega} + y_2 e^{-2j\omega} + y_3 e^{-3j\omega} + \dots$$

Z-преобразование той же временной последовательности Y имеет вид:

$$Y(z) = y_0 + y_1 z^{-1} + y_2 z^{-2} + y_3 z^{-3} \dots$$

Из сопоставления этих двух выражений вытекает, что если под символом единичного сдвига z понимать комплексную величину $e^{-j\omega}$, то Z-преобразование некоторой временной последовательности есть своеобразное представление комплексного спектра этой последовательности. Это представление очень удобно, так как наглядно отображает структуру последовательности y_t во временной области и, в то же время, обладает преимуществами спектральных представлений. Z-преобразования можно достаточно просто записать и оперировать с ними, как с обычными полиномами. Например, пусть требуется определить во временной области результат фильтрации сигнала $f_t = (1, -1, 1/2)$ фильтром с временной характеристикой $g_t = (1, 1/2, -1/2)$. Их Z-преобразования равны $F(z) = 1 - z + 1/2 z^2$ и $G(z) = 1 + 1/2 z^{-1} - 1/2 z^{-2}$. Тогда Z-преобразование результата фильтрации, описываемый сверткой функций f_t и g_t , запишется

$$f_t * g_t \leftrightarrow F(z)G(z) = (1 - z + 1/2 z^2)(1 + 1/2 z^{-1} - 1/2 z^{-2}) = (1 - 1/2 z^{-1} - 1/2 z^{-2} + 3/4 z^{-3} - 1/4 z^{-4}).$$

Отсюда профильтрованный сигнал $f_t * g_t = (1, -1/2, -1/2, 3/4, -1/4)$.

1.12. Вейвлет-анализ

Общие положения и задачи вейвлет-анализа

Прямое преобразование Фурье позволяет представить сложный геофизический сигнал в виде суммы простых (синусоиды либо косинусоиды) элементарных сигналов. Однако выделяемые при этом гармоники не локализованы во времени и характеризуют весь интервал наблюдений. Это и обуславливает применимость Фурье-преобразования, прежде всего, к анализу стационарных сигналов. Большинство же реальных геофизических наблюдений являются сложными функциями от времени и пространственных координат и относятся к нестационарным, то есть их статистические характеристики изменяются в пределах изучаемых интервалов. Для анализа таких наблюдений необходим метод, способный обеспечить локализацию различных составляющих анализируемых числовых данных. Есть два подхода к анализу данных такого типа. Первый - локальное преобразование Фурье (short-time Fourier transform). Следуя по этому пути, нестационарным сигнал предварительно разбивается на сегменты (фреймы), в пределах которых свойства сигнала существенно не меняются. Анализируя результаты специального анализа для различных сегментов можно получить представление о том, как меняется спектр сигнала во времени или в пространстве. Вторым подходом основан на вейвлет-анализе.

Вейвлет-анализ заключается в исследовании сигнала $f(x)$ при помощи вычисления системы функций, зависящих от двух непрерывных параметров. Эта система получается из некоторой фиксированной функции $\psi(x)$ путем сжатий, растяжений и сдвигов.

Вейвлет-преобразование стремительно завоевывает популярность в различных областях науки. Благодаря хорошей приспособленности к анализу нестационарных сигналов оно стало мощной альтернативой преобразованию Фурье и в ряде геофизических приложений. Наиболее известны применения вейвлет-анализа для решения следующих задач:

- обработка нестационарных сигналов;
- подавление шума, повышение отношения сигнал/помеха;
- определение характеристик сигналов и помех;
- сжатие наблюдаемых данных;
- обработка изображений;
- локализация сигналов в частотной и временной областях и др.

Вейвлеты и их основные свойства

Функцию $\psi(x)$ чаще всего называют вейвлетом (по-английски – wavelet). Непосредственный перевод слова wavelet звучит как *маленькая волна* или *волнушка*. Некоторые авторы используют слово *всплеск*.

Проблема выбора конкретного вейвлета обычно решается с использованием в качестве критерия внешнего сходства вида исследуемой функции и функции преобразования. Выбранный анализирующий вейвлет должен обладать следующими свойствами:

1. Быстро сходиться к нулю при увеличении абсолютного значения аргумента.
2. Быть непрерывным.
3. $\int_{-\infty}^{\infty} \psi(x) dx = 0$

В практике широко применяются такие анализирующие вейвлеты как вейвлет “сомбреро”, вейвлет Морле, вейвлет Шеннона и др. На рис. 1.4 в качестве примера приведены графики первых двух из указанных вейвлетов.

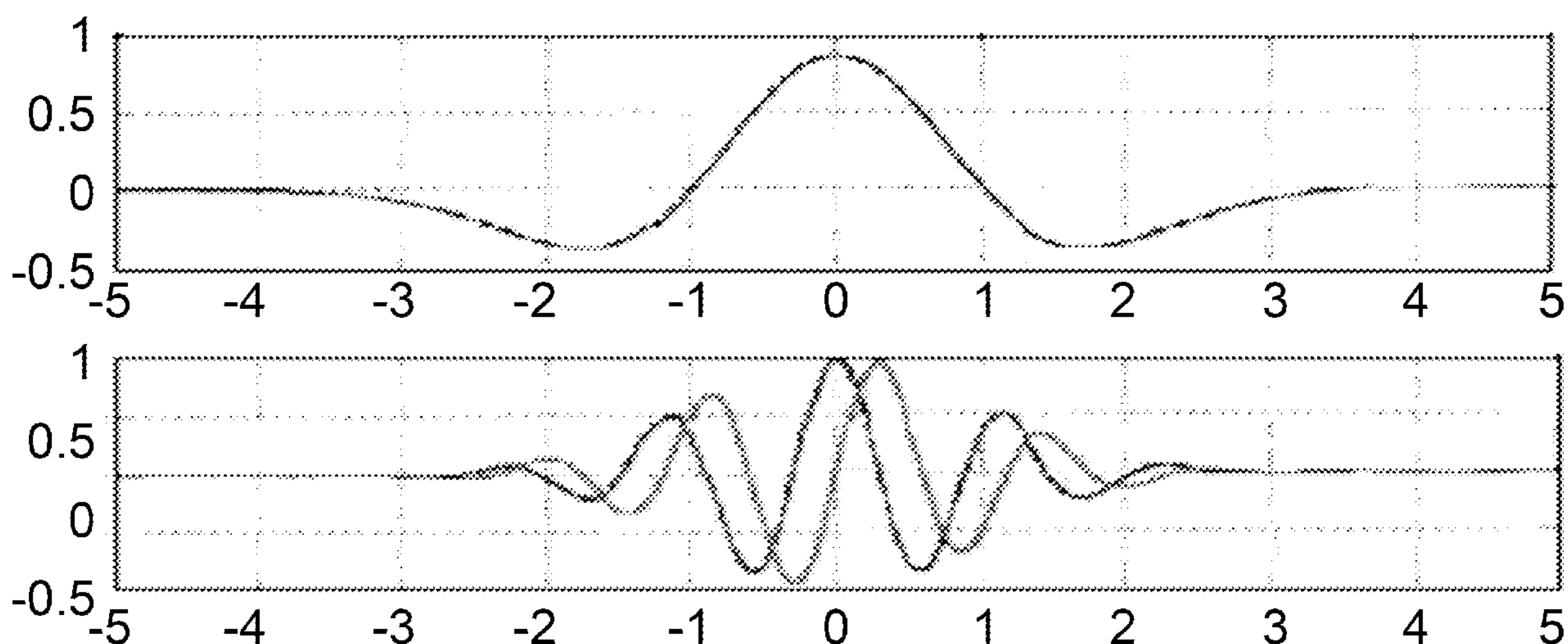


Рис. 1.4. Примеры вейвлетов. Сверху изображен вейвлет "сомбреро" (Mexican Hat), снизу - вейвлет Морле (вещественная и мнимая составляющие).

Прямое и обратное вейвлет-преобразования

Прямое и обратное вейвлет-преобразования переводят исследуемую функцию $f(x)$ в набор вейвлет-коэффициентов $W_\psi(a,b)f$ и обратно соответственно. Различают непрерывное и дискретное преобразования. Далее мы ограничимся рассмотрением непрерывного варианта.

Прямое вейвлет-преобразование осуществляется согласно правилу

$$W_\psi(a,b)f = \frac{1}{\sqrt{C_\psi}} \int \frac{1}{\sqrt{|a|}} \psi\left(\frac{x-b}{a}\right) f(x) dx,$$

где a и b — параметры, определяющие соответственно масштаб и смещение функции ψ (анализирующий вейвлет), C_ψ — нормировочный множитель. Интегрирование ведут по всей числовой оси.

Имея известный набор коэффициентов $W_\psi(a,b)f$, можно восстановить исходный вид функции $f(x)$. Для этого осуществляется обратное вейвлет-преобразование:

$$f(x) = \frac{1}{\sqrt{C_\psi}} \iint \frac{1}{\sqrt{|a|}} \psi\left(\frac{x-b}{a}\right) [W_\psi(a,b)f] \frac{da db}{a^2}$$

Отметим, что в отличие от преобразования Фурье, вейвлет-преобразование определено неоднозначно так как зависят от выбранной для анализа функции ψ . При этом каждому вейвлету соответствует свое преобразование.

Обратим также внимание на отсутствие в вейвлет-преобразовании явного частотного параметра. Вместо него имеется параметр масштаба, который можно определить как величину обратную частоте.

Число используемых при разложении сигнала вейвлетов задаёт уровень декомпозиции сигнала. При этом за нулевой уровень декомпозиции часто принимается сам сигнал, а последующие уровни декомпозиции образуют ниспадающее вейвлет-дерево. Точность представления сигнала по мере перехода на более низкие уровни декомпозиции снижается, но зато появляется возможность вейвлет-фильтрации сигналов, удаления из сигналов шумов и сжатия сигналов, т.е. становится возможной вейвлет-обработка сигналов.

2. СТАТИСТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ И ИНТЕРПРЕТАЦИИ ГЕОФИЗИЧЕСКИХ ДАННЫХ

2.1. Детерминированный и вероятностно-статистический подходы к обработке и интерпретации геофизических наблюдений

При детерминированном подходе решение тех или иных задач при этом ищется как единственно возможное, т.е. в виде определенной функциональной зависимости или в виде определенного числа.

Основой детерминированного подхода является применение аналитических методов теории потенциала (в грави-, магнито-, терморазведке, электроразведке постоянным током), уравнений Максвелла (в электроразведке переменным током) и теории упругости (в сейсморазведке).

Применение детерминированного подхода считается оправданным, если производится преобразование и анализ интенсивных аномалий от геологических объектов с заметной дифференциацией физических свойств, благоприятными для решения поставленных задач размерами, формой и глубиной залегания.

При вероятностно-статистическом подходе полагают, что экспериментальный материал имеет случайный характер. Как уже отмечалось выше, характерной особенностью детерминированного подхода является то, что решение находится в форме единственно возможного, т. е. достоверного. Между тем такая постановка обратных задач принципиально неверна. Кроме неоднозначности решения обратных задач значительную роль играют и другие обстоятельства. Так, на практике мы не можем знать геофизические поля с абсолютной точностью. Действительно, результаты геофизических съемок всегда отягощены инструментально-методическими погрешностями измерений и обработкой данных, а также искажаются под влиянием многообразных технических и геологических помех. В результате обычно можно лишь предсказать область возможных значений поля, вероятность появления того или иного значения. Предположение о случайности наблюдений обуславливает случайность суждений об изучаемых объектах и явлениях.

Случайным полагается также расположение изучаемых геологических объектов, точек, площадей наблюдения и др.

Таким образом, в практике обработки и интерпретации геофизических полей исследователь имеет дело с данными, которые описываются случайными величинами и процессами. Изучение этих величин и процессов требует использования математического аппарата теории вероятностей и ее

приложений. Разработанные на их основе алгоритмы решения геофизических задач позволяют получить для одних и тех же исходных данных разные решения, согласующиеся с имеющимися геофизическими и геологическими материалами. Это расширяет возможности геологической интерпретации геофизических наблюдений.

В наибольшей мере возможности вероятностно-статистического подхода раскрываются при комплексной интерпретации геофизических данных, полученных несколькими методами.

2.2. Основные понятия теории вероятностей и математической статистики в задачах обработки геофизических данных

Теория вероятностей – наука, изучающая закономерности случайных событий во времени и пространстве и приемы их количественного описания.

Наблюдаемые при проведении экспериментов события делятся на достоверные, невозможные и случайные. В качестве события в геофизике выступают: появление конкретного значения геофизического поля, появление аномалии и т.д.

Если событие может как произойти, так и не произойти, то его называют случайным.

Если событие в ходе эксперимента происходит обязательно, то его называют достоверным.

Невозможным называют такое событие, которое не может произойти.

Два события называются несовместимыми, если появление одного из них исключает появление другого.

Суммой событий называется событие, состоящее в появлении хотя бы одного из этих событий.

Произведением событий называется событие, состоящее в совместном появлении всех этих событий.

События A_1, A_2, \dots, A_n образуют полную группу, если они попарно несовместимы, а в сумме образуют достоверное событие (т.е. какое либо из них обязательно происходит, причем только одно).

Противоположными событиями называются два несовместимых события, образующих полную группу (например, наличие и отсутствие залежи).

Вероятность события – количественная мера степени объективной возможности того или иного события A . Обозначается $P(A)$:

$$P(A) = m / n,$$

где n – общее число равновозможных исходов экспериментов;

m – число исходов, благоприятствующих событию A .

Это классическое определение вероятности. В геофизической практике чаще используют статистическое определение вероятности – вероятность события, это относительная частота его появления при многократном воспроизведении комплекса условий эксперимента. При большом числе опытов частота события A стремится к вероятности $P(A)$ в ее классическом определении. (В целом недостаток классического определения в том, что оно сводит понятие вероятность к равновозможным событиям. Это тривиальный случай.)

Условной вероятностью события A называют вероятность, вычисленную при условии, что произошло событие B . Обозначается $P(A/B)$

События называются независимыми, если появление одного из них не меняет вероятности появления другого. Выполняется $P(A/B)=P(A)$.

Для зависимых $P(A/B) \neq P(A)$.

Априорная вероятность – вероятность, известная (или заданная) до проведения эксперимента.

Апостериорная вероятность – вероятность, установленная в результате эксперимента.

2.3. Законы распределения случайных величин

Закон, устанавливающий связь между возможными значениями случайной величины и соответствующими им вероятностями называют законом распределения.

Для описания геофизических показаний (физических полей, физических свойств пород и др.) наибольшее применение получили:

- 1) нормальный закон распределения;
- 2) равномерный закон распределения;
- 3) закон Пуассона;
- 4) распределение Стьюдента.

Нормальный закон характерен для большинства наблюдаемых в геологии и геофизике явлений и процессов. Это является отражением так называемой центральной предельной теоремы, которая гласит, что распределение случайной величины, являющейся суммой n независимых величин при больших n приближается к нормальному. Поэтому нормальный закон распределения рассматривается в геофизической практике чаще всего. Он имеет функцию распределения

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left[-(t - \bar{x})^2 / 2\sigma^2\right] dt.$$

Функция $F(x)$ определяет для каждого значения x вероятность того, что случайная величина X примет значение меньше x , т.е.

$$F(x) = P(X < x).$$

σ – среднее квадратичное отклонение (стандарт) случайной величины X ;
 \bar{x} - математическое ожидание величины X (среднее значение).

$$\bar{x} = \sum_{i=1}^n x_i p_i$$

Отметим

$$\sigma = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 p_i}.$$

Важный частный случай имеет место при $\bar{x}=0$ и $\sigma=1$:

$$F(x) = \Phi(x) = \frac{1}{2\pi} \int_{-\infty}^x \exp(-t^2 / 2) dt.$$

Имея произвольный набор геофизических данных легко их преобразовать к случаю, когда $\bar{x}=0$ и $\sigma=1$. Значения $\Phi(x)$ приведены практически во всех справочниках и могут привлекаться для анализа. Например, $\Phi(0) = 0,5$; $\Phi(-1) \approx 0,16$; $\Phi(1) \approx 0,84$.

Так как практически все значения случайной величины (99,7 %) укладываются в интервал 3σ . Способ оценки диапазона возможных значений случайной величины носит название «правило трех сигм».

В геофизической практике часто встречаются ситуации, когда по нормальному закону распределена не сама анализируемая случайная величина, а логарифмы ее значений. В таком случае говорят, что случайная величина имеет логнормальный закон распределения.

Помимо функции распределения $F(x)$ в практике часто применяют функцию плотности распределения $f(x)$. Она является производной функции $F(x)$ и характеризует плотность, с которой распределены значения случайной величины в данной точке.

Равномерный закон имеет функцию распределения

$$F(x) = \begin{cases} (x - a) / (b - a), & a < x < b, \\ 0, & x < a \\ 1, & x > b \end{cases}$$

Данный закон используется при априорном задании распределений параметров объектов (глубины, мощности и т.д.) или параметров аномалии (амплитуды, протяженности).

Закон Пуассона справедлив для дискретной случайной величины, принимающей последовательно значения $X=0, 1, 2, \dots, m$ с вероятностью

$$P_m = \frac{a^m \exp(-a)}{m!}.$$

Ее функция распределения

$$F_m = \sum_{m=0}^{\infty} \frac{a^m \exp(-a)}{m!}$$

здесь a – параметр распределения.

Закон Пуассона используется при анализе результатов измерений, в процессе которых производится расчет суммы событий (например, суммы импульсов, фиксирующих радиоактивный распад какого-либо элемента за время t).

Распределение Стьюдента (t -распределение). Оно похоже на нормальное, но характеризуется более широкой областью значений. Форма кривой распределения меняется в зависимости от числа наблюдений в выборке. Когда число наблюдений в выборке бесконечно, то t -распределение совпадает с нормальным.

На распределении Стьюдента основаны так называемые t -критерии, которые полезны для проверки гипотезы о том, что данная выборка геофизических наблюдений извлечена из совокупности с заданными характеристиками или же для проверки гипотезы об однородности двух выборок.

2.4. Параметры и статистики

Распределения имеют ряд характеристик. Эти характеристики называются параметрами, если они описывают совокупности наблюдений в целом, и статистиками, если они относятся к выборкам.

Статистики можно использовать для оценки параметров исходных совокупностей и для проверки гипотез, сформированных относительно этих совокупностей.

Чаще всего применяют следующие статистики:

- среднее значение

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i ;$$

- мода (наиболее часто встречающееся в совокупности значение признака);
- медиана (средняя точка распределения);
- дисперсия (мера разброса отдельных значений относительно среднего)

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n};$$

- стандартное отклонение (также мера разброса). Равно $\sqrt{\sigma^2}$;
- квантиль распределения (значение случайной величины, для которого функция распределения принимает конкретно заданное значение);
- центральный момент k -го порядка

$$\mu_k = \sum_{i=1}^n (x_i - \bar{x})^k P_i$$

(P_i – вероятность появления x_i);

- асимметрия распределения A

$$A = \mu_3 / \sigma^3;$$

- эксцесс распределения

$$E = \mu_4 / \sigma^4 - 3.$$

2.5. Проверка статистических гипотез. Ошибки первого и второго рода

Необходимость проверки статистических гипотез возникает в практике геологоразведочных работ достаточно часто. В статистической гипотезе речь может идти о наличии или отсутствии залежей полезных ископаемых, о предполагаемой величине параметров известного распределения, о равенстве параметров двух или нескольких распределений и т. п.

Охарактеризуем кратко процедуру проверки статистической гипотезы.

Гипотезу, которую необходимо проверить, называют нулевой, или основной. Ее обозначают H_0 . Наряду с выдвинутой H_0 рассматривают и противоречащую ей альтернативную гипотезу H_1 . Например, нулевая

гипотеза состоит в предположении, что математическое ожидание $M[x]$ пористости пород равно 12%. Тогда конкурирующая гипотеза может состоять в том, что математическое ожидание пористости не равно 12%. Это записывается так: $H_0: M[x]=12$; $H_1: M[x] \neq 12$.

Различают гипотезы, которые содержат одно предположение или более. Гипотеза, содержащая только одно предположение, называется простой. Поэтому, в рассматриваемом случае гипотеза $H_0: M[x]=12$ — простая.

Проверяя нулевую гипотезу, мы можем ее принять, отвергнуть или прийти к выводу, что данных для принятия решения недостаточно. При этом могут быть допущены ошибки двух родов. Гипотеза может быть отвергнута как неправильная, хотя она верна, — это *ошибка первого рода*. Гипотеза может быть принята как правильная, хотя она неверна, — это *ошибка второго рода*.

Вероятность совершить ошибку первого рода обычно обозначается через α . Ее называют уровнем значимости. Величина α выбирается, как правило, небольшой (практически 0,001; 0,01 или 0,05). Выбор ее основывается на так называемом принципе практической уверенности, или принципе практической невозможности маловероятных событий. Принцип практической уверенности формулируют следующим образом: если вероятность некоторого события в одних и тех же условиях весьма мала, то можно быть практически уверенным, что при однократном наблюдении это событие не произойдет.

Суть проверки статистических гипотез состоит в вычислении и анализе так называемых статистических критериев K . Значение критерия вычисляется по той же выборке, на основе которой формулируются проверяемая и альтернативная гипотезы. Статистический критерий выбирается в зависимости от содержания нулевой гипотезы и от условий, в которых осуществляется проверка. К условиям относятся знание или незнание закона распределения, которому подчиняется изучаемая случайная величина, а также ее дисперсии, объем выборки и т. п. После того как критерий выбран, множество всех его возможных значений разбивают на два непересекающихся подмножества: одно из них содержит те значения критерия, при которых проверяемая гипотеза отвергается, другое — при которых она принимается.

Совокупность значений критерия, при которых нулевая гипотеза отвергается, называется критической областью гипотезы. Совокупность

значений, при которых гипотеза принимается, — областью принятия гипотезы или областью допустимых значений критерия.

Основной принцип проверки статистических гипотез состоит в следующем. По данным выборки вычисляется значение критерия K_H (наблюдаемое). Это значение сравнивается с критическим $k_{кр}$, которое находится по специальной таблице. Если наблюдаемое значение критерия принадлежит критической области, проверяемую (нулевую) гипотезу отвергают, если области допустимых значений — ее принимают.

Следует подчеркнуть, что наблюдаемое значение критерия может попасть в критическую область не только потому, что неверна проверяемая гипотеза, но и из-за недостаточного числа наблюдений, из-за непредставительности выборки и др. Вместе с тем, если проверяемая гипотеза принята, не следует быть уверенным, что она тем самым окончательно доказана. Один пример, подтверждающий справедливость некоторого общего утверждения, еще не доказывает его правильность. Поэтому правильно говорить не о том, что нулевая гипотеза правильна, а о том, что данные наблюдений согласуются с нулевой гипотезой и не дают оснований ее отвергнуть.

2.6. Статистические оценки геофизических характеристик

Точечные и интервальные оценки

На практике исследователь работает с выборками и задача сводится к оценке истинного значения того или иного параметра по выборочным значениям.

Точечная оценка – оценка, выраженная одним числом. Ее недостаток заключается в отсутствии информации о точности получаемого результата. В практике значения точечной оценки могут быть далеки от истинной величины оцениваемого параметра.

Интервальная оценка – это некоторый интервал значений, который с заданной вероятностью включает истинное значение изучаемого параметра.

Задаваемая вероятность обычно выбирается близкой к единице.

Пусть истинное значение параметр θ , а его точечная оценка $\hat{\theta}$. Тогда интервальная оценка записывается как интервал $(\hat{\theta} - \lambda, \hat{\theta} + \lambda)$. Интервальные оценки называют еще доверительными интервалами.

Критериями качества любых оценок являются:

- несмещенность, т.е. совпадение в среднем оценки параметра с истинным значением;
- состоятельность (оценка параметра стремится к истинному значению с ростом мощности выборки);
- эффективность (оценка параметра обладает наименьшей степенью случайных отклонений от истинного значения);
- достаточность (отсутствие необходимости привлечения других оценок).

Пример интервальной оценки

Часто в геофизической практике приходится оценивать средние значения той или иной величины. С этой целью обычно используется следующая формула для интервальной оценки неизвестного среднего

$$P\left(\bar{x} - t_{n,\alpha} \frac{S}{\sqrt{n}} \leq \theta \leq \bar{x} + t_{n,\alpha} \frac{S}{\sqrt{n}}\right) \approx 1 - \alpha,$$

где $t_{n,\alpha}$ – допустимые значения критерия Стьюдента.

Данная функция получена с учетом четности дифференциальной функции распределения (плотности вероятности) для распределения Стьюдента.

Пусть, например, произведено 10 замеров амплитуды сейсмических колебаний: 120, 122, 121, 118, 120, 116, 124, 122, 121, 120. Требуется дать интервальную оценку истинной амплитуды с доверительной вероятностью 0,99. Вычисляем вначале среднее арифметическое выборочное

$$\bar{A} = \frac{1}{10} (A_1 + A_2 + \dots + A_{10}) = \frac{1}{10} (120 + 122 + \dots + 120) = 120,4.$$

Далее определяем среднее квадратическое отклонение

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{t=1}^{10} (x_t - 120,4)^2}{10 - 1}} = \sqrt{\frac{0,16 + 2,56 + \dots + 0,16}{9}} \approx 2,22.$$

По данным $n=10$ и $P=0,99$ определяем по таблице допустимых значений критерия Стьюдента значение $t_{\alpha} = 3,17$.

Находим доверительные границы

$$\bar{x} - t_{n,\alpha} \frac{S}{\sqrt{n}} = 120,4 - 3,17 \frac{2,22}{3,16} = 118,17$$

$$\bar{x} + t_{n,\alpha} \frac{S}{\sqrt{n}} = \dots = 122,63.$$

Таким образом, с надежностью 0,99 истинное значение амплитуды находится в интервале $118,17 \leq A \leq 122,63$.

Условия построения интервальных оценок

При доверительном оценивании геофизических характеристик исходят из предположения о независимости и стационарности наблюдений.

Независимость означает, что результаты наблюдений в одних точках не зависят от результатов наблюдений в других. Зависимость наблюдений опасна возможностью пристрастного выбора данных. Например, если все наблюдения разместить в пределах небольшого интервала залежи, то выборку уже нельзя назвать случайной. Измерения тех или иных геофизических характеристик будут использовать изменения монотонного характера с закономерным изменением расположения соседних точек.

Стационарным случайным процессом называется такой случайный процесс, для которого все числовые характеристики не изменяются при сдвиге аргумента t . На практике условие стационарности зачастую не выполняется. Так, может наблюдаться нестационарность по математическому ожиданию, нестационарность по дисперсии и нестационарность по преобладающему периоду и др.

Проверка стационарности процесса

Практически при доказательстве стационарности процесса анализируют математическое ожидание, дисперсию, функцию автокорреляции. Должно выполняться

$$MX(t) = const \quad S^2 X(t) = const$$

$$R_{X(t_i, t_j)} = R_{X(t_i, t_i + \tau)}, \quad (*)$$

где $t_j = t_i + \tau$, т.е. АКФ зависит только от расстояния между t_i и t_j и является, таким образом, функцией лишь одного аргумента τ .

Условие (*) является основным для выполнения стационарности, т.к. АКФ с одной стороны определяется через математическое ожидание, а с другой – при $t_i = t_j = t$ совпадает с дисперсией процесса.

Проверка независимости наблюдений

Для проверки независимости наблюдений чаще всего вычисляют и анализируют радиус корреляции. Под радиусом корреляции понимается критическое расстояние между точками наблюдений, при котором значения случайной величины можно рассматривать как независимые.

Пусть радиус корреляции r_K , тогда при $r > r_K$ наблюдения независимы, при $r < r_K$ – зависимы (r – расстояние между точками наблюдений).

Наиболее просто радиус корреляции можно определить с помощью функции автокорреляции. Для непрерывной функции $U_1(t)$ она записывается:

$$U(t) = U_1(t) * U_1(-t) = \int_{-\infty}^{\infty} U_1(\tau) \cdot U_1(t - \tau) d\tau.$$

АКФ симметрична относительно точки $t = 0$.

Радиус корреляции соответствует точке, для которой $U(t) = 0,1 \div 0,3U(0)$.

Отметим, что с помощью АКФ можно оценивать радиусы корреляции, как полезных сигналов, так и помех. Например, для аддитивной модели сейсмической трассы $S(t)$

$$S(t) = a(t) + n(t),$$

где $a(t)$ – полный сигнал, $n(t)$ – помеха, можно выбрать участки, где преобладают сигнал или помеха и проанализировать их по отдельности.

2.7. Исключение грубых погрешностей измерений

В процессе обработки экспериментальных данных следует исключать грубые ошибки, так как они могут весьма значительно сказаться на достоверности получаемых результатов. Наиболее простой способ исключения таких ошибок основан на правиле трех сигм: разброс случайных величин от среднего не должен превышать значений

$$x_{\max, \min} = \bar{x} \pm 3\sigma,$$

где σ – среднеквадратическое отклонение.

Более достоверными являются методы, базируемые на использовании доверительных интервалов. Если выборка подчиняется нормальному закону распределения и относительно невелика ($n \leq 50$), то можно использовать следующие критерии

$$\beta_1 = (x_{\max} - \bar{x}) / \sigma \sqrt{(n-1)/n};$$

$$\beta_2 = (\bar{x} - x_{\min}) / \sigma \sqrt{(n-1)/n}.$$

Вычислив β_1, β_2 , задаются доверительной вероятностью и по таблицам (критерий появления грубых ошибок) определяют β_{\max} . В таблице указано n и доверительная вероятность γ .

Если $\beta_1 > \beta_{\max}$, то значение x_{\max} следует исключить.

При $\beta_2 < \beta_{\max}$ исключается значение x_{\min} .

Далее процесс повторяется (n – уменьшается).

Известны и другие способы исключения грубых ошибок.

2.8. Проверка на воспроизводимость результатов

Ответственные геофизические эксперименты должны быть проверены и на воспроизводимость результатов, т.е. на их повторяемость в определенных пределах измерений при заданной доверительной вероятности.

Пусть выполнено m серий опытов. В каждой серии количество измерений n . Тогда можно применить следующую методику проверки воспроизводимости результатов:

- 1) вычисление для каждой серии среднеарифметического значения \bar{x}_i ;
- 2) вычисление дисперсии D_i ;
- 3) определение критерия Кохрена (расчетного):

$$K_{\text{КПТ}} = \max D_i / \sum_{i=1}^m D_i,$$

где $\max D_i$ - наибольшее значение дисперсии в сериях экспериментов;

- 4) определение табличного критерия Кохрена (для определения используется число степеней свободы $q = n - 1$, доверительная вероятность и количество серий m).

Сопоставление расчетного и табличного $K_{\text{КПТ}}$ значений критерия Кохрена.

Опыты считаем воспроизводимыми, если

$$K_{кр} \leq K_{КПТ}.$$

Например: Кол-во серий $m = 3$, кол-во измерений $n = 5$.

Тогда $q = 5 - 1 = 4$. Зададимся $\gamma = 0,95$. При таких условиях $K_{КПТ} = 0,74$.

Если $K_{кр} = 0,55$, то $K_{кр} < K_{КПТ}$ и эксперименты следует считать воспроизводимыми. В противном случае необходимо было бы увеличить число серий m или число измерений n .

2.9. Многомерный статистический анализ геофизических данных

Общие сведения

Многомерным статистическим анализом именуют группу методов обработки данных, являющихся многомерными в том смысле, что каждый объект наблюдения характеризуется значениями многих наблюдаемых переменных. Ранее уже отмечалось, что вероятностно-статистический подход особенно эффективен при обработке и комплексной интерпретации геофизических данных, полученных несколькими методами. На настоящий момент разработано большое число методов многомерного статистического анализа. Они различаются в зависимости от того: 1) являются ли анализируемые признаки количественными или качественными; 2) постулируется ли коррелированность или независимость случайных переменных в исследуемом наборе признаков; 3) каковы теоретические представления о линейности или нелинейности взаимосвязей характеристик изучаемого явления; 4) каковы характеристики функций распределения переменных и т. д. Безусловно, это далеко не исчерпывающая и весьма условная схема систематизации методов многомерного анализа.

Каждый из методов многомерного статистического анализа ориентирован на решение определенных задач, устанавливая ту или иную особенность изучаемых объектов. В этой связи целесообразно применять несколько статистических методов одновременно, что позволяет системно проанализировать имеющиеся экспериментальные данные, с наибольшей полнотой изучить различные свойства и проявления объектов исследования. Обсуждая возможности использования методов многомерной статистики следует указать на то обстоятельство, что статистические критерии и

процедуры большинства из них разработаны лишь при очень сильных ограничениях. Вид этих критериев и их поведение при более слабых допущениях, которые обычно используются при решении большинства реальных задач, изучены плохо. В этой связи практически всегда результаты многомерного статистического анализа геолого-геофизических данных могут быть подвергнуты серьезной критике. Однако, познавательная ценность обсуждаемых методов очень велика, что и определило их широкое внедрение в практику.

Приведем далее краткое описание концептуальных идей и теоретических основ наиболее распространенных в геофизической практике вероятностно-статистических методов.

2.9.1. Корреляционно-регрессионный анализ

На основе корреляционно-регрессионного анализа изучаются статистические (стохастические) связи (зависимости) между различными величинами. В качестве таких величин могут выступать данные наблюдений различных геофизических методов, физические характеристики пород, глубины границ раздела в осадочной толще пород и т.д.

Статистической называют зависимость, при которой изменение одной величины вызывает изменение распределения другой.

В ходе изучения статистических зависимостей устанавливают корреляцию и регрессию.

Корреляция определяет существование взаимосвязи между случайными величинами, а также силу (тесноту) этой связи.

Регрессия определяет характер (форму) зависимости между величинами, а также возможность оценки одних величин по другим.

Оценки тесноты корреляционной связи для нормального закона

Пусть имеются две случайные величины X и Y (физические характеристики, геофизические поля и др.), распределенные по нормальному закону. В качестве характеристики линейной связи между ними чаще всего используется коэффициент корреляции

$$r_{XY} = \frac{1}{n\sigma_X\sigma_Y} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

где $\bar{x}, \bar{y}, \sigma_X, \sigma_Y$ - оценки средних значений и среднеквадратических отклонений X и Y .

На практике удобнее использовать формулу

$$r_{XY} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \cdot \sqrt{n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2}}.$$

Свойства коэффициента корреляции состоят в следующем:

- 1) $-1 \leq r_{XY} \leq 1$;
- 2) если $r_{XY} = 0$, величины X и Y называют некоррелированными;
- 3) при возрастании $|r_{XY}|$ корреляционная зависимость становится все более тесной и при $|\hat{r}_{XY}| = 1$ переходит в функциональную;
- 4) r_{XY} оценивает тесноту лишь линейной связи между X и Y .

Если $r > 0$, то говорят о положительной корреляции Y и X , если $r < 0$, то об отрицательной.

После вычисления r_{XY} следует его проанализировать. Так, исходя из теоретических соображений доверительный интервал для r_{XY} в предположении нормальности распределения X и Y (при доверительной вероятности γ) будет

$$r_{XY} \pm t_{(\gamma+1)/2} (1 - r_{XY}^2) / \sqrt{n},$$

где $t_{(\gamma+1)/2}$ - γ -квантиль нормального распределения;

$(1 - r_{XY}^2) / \sqrt{n}$ - среднеквадратическое отклонение для r_{XY} .

Если r_{XY} выходит за пределы доверительного интервала, то гипотеза об отсутствии линейной связи отвергается с вероятностью γ .

Практически удобно пользоваться критерием В.И. Романовского: если $|r_{XY}| \geq 3 \sigma_r$ (σ_r - среднеквадратическое отклонение r_{XY}), то связь X и Y значимая.

Пример.

Пусть количество измерений 1434, а выборочный коэффициент корреляции $r_{XY} = 0,648$. Тогда

$$\sigma_r = (1 - r_{XY}^2) / \sqrt{n} = (1 - 0,648^2) / \sqrt{1434} = 0,015.$$

$$|r_{XY}| = 0,648 > 3 \sigma_r = 0,045.$$

Таким образом, можно констатировать наличие значимой связи.

Во многих справочниках по математической статистике можно найти таблицы с критическими значениями коэффициента корреляции.

При практическом применении коэффициента корреляции следует обратить внимание на условие нормальности распределения анализируемых

величин. На практике это условие вследствие большой трудоемкости вычислений чаще всего не проверяется, что снижает достоверность выводов о наличии и тесноте корреляционных связей. Такие же отрицательные последствия имеет отклонение корреляционных связей от линейных. Проверка статистической гипотезы о виде корреляционной связи также представляет значительные вычислительные трудности и обычно не делается. Качественный вывод о виде корреляционных зависимостей можно сделать, построив и проанализировав графики анализируемых величин.

Оценка тесноты корреляционной связи в случае произвольного закона распределения

Большинство статистических методов, используемых в настоящее время в практике геологических и геофизических работ, относятся к методам параметрической статистики. Свое название она получила из-за того, что основной объект ее изучения - это выборки из распределений, описываемых одним или небольшим числом параметров. Однако, в подавляющем большинстве реальных ситуаций точного соответствия наблюдаемых распределений тем или иным теоретическим моделям (параметрическим семействам) нет. Неоднократно публиковались экспериментальные данные, свидетельствующие о том, что распределения реально наблюдаемых случайных величин, в частности, ошибок измерения, в подавляющем большинстве случаев отличны от нормальных (гауссовских). Для преодоления возникающих в результате проблем и ошибок применяются непараметрические методы. С их помощью можно решать практически тот же круг статистических задач, что и с помощью параметрических. Так, если случайные величины X и Y распределены не по нормальному закону, а произвольно, то зависимость между ними можно оценить с помощью коэффициента ранговой корреляции Спирмена ρ (непараметрический критерий).

Пусть имеется n пар измерений X и Y . Рассмотрим наиболее простой случай, когда в выборках нет совпадающих значений. Тогда ранг будет равен порядковым номерам измерений, расположенным в порядке возрастания и значение ρ определяется как

$$\rho = 1 - \frac{6 \sum d_i^2}{n^3 - n},$$

где d_i – разность рангов.

Пример.

Пусть $n = 8$, X меняется от 1 до 8, Y меняется от 0,1 до 2,1. Для оценки тесноты корреляционной связи вначале составим следующую таблицу:

| № т. набл. | Знач. X | Знач. Y | Ранг X | Ранг Y | d |
|------------|-----------|-----------|----------|----------|----------|
| 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 1 | 1,5 | 1 | 7 | -6 |
| 2 | 2 | 2,1 | 2 | 8 | -6 |
| 3 | 3 | 0,3 | 3 | 2 | 1 |
| 4 | 4 | 0,9 | 4 | 6 | -2 |
| 5 | 5 | 0,5 | 5 | 4 | 1 |
| 1 | 2 | 3 | 4 | 5 | 6 |
| 6 | 6 | 0,6 | 6 | 5 | 1 |
| 7 | 7 | 0,4 | 7 | 3 | 4 |
| 8 | 8 | 0,1 | 8 | 1 | 7 |

Подставляя значения в формулу для вычисления ρ , получим:

$$\rho = 1 - \frac{6(36 + 36 + 1 + 4 + 1 + 1 + 16 + 49)}{8^3 - 8} = -0,71.$$

Для проверки гипотезы о наличии связи между X и Y следует далее сравнить ρ с его доверительным интервалом, равным

$$\rho \pm t_{(1+\gamma)/2} \sqrt{1/(n-1)}.$$

В нашем случае примем $\gamma = 0,95$.

$t_{(1+\gamma)/2}$ - квантиль распределения Стьюдента с $(n - 2)$ степенями свободы.

Из таблиц математической статистики:

$$t_{(1+\gamma)/2} = 1,96, \text{ тогда } t_{(1+\gamma)/2} \sqrt{1/(8-1)} \approx 1,96 \cdot 0,38 = 0,74.$$

Т.к. $|\rho| = 0,71 < 0,74$, то гипотеза о наличии корреляционной связи между X и Y не выполняется.

Можно воспользоваться готовой таблицей (таблица критических значений выборочного коэффициента корреляции рангов). Для рассмотренного случая выбираем значения из таблицы:

$$\rho_{кр} = 0,88 \quad (\gamma = 0,99)$$

$$\rho_{кр} = 0,72 \quad (\gamma = 0,95).$$

Т.о. полученный выше результат подтвердился.

Регрессионный анализ

Основные задачи регрессионного анализа

Регрессионным анализом называют совокупность статистических методов, ориентированных на исследование стохастической зависимости одной переменной Y от набора других переменных $\{X_j\} = (X_1, X_2, \dots, X_P)$. Отличительной особенностью регрессионного анализа является выделение двух видов переменных – зависимых и независимых. Пусть имеются две переменные Y на X . Тогда, в принципе, можно определить два уравнения регрессии: $y_x = \varphi(x)$ - уравнение регрессии Y на X и $x_y = f(y)$ - уравнение регрессии X на Y . Разделение переменных на зависимую и независимые в регрессионном анализе производится исследователем на основе содержательных соображений. Если число независимых переменных два и больше, то регрессионные уравнения называют уравнениями множественной регрессии, если используется единственная независимая переменная, то — уравнениями парной регрессии. В практике используется ещё одно название регрессии- тренд.

В общем случае основными задачами регрессионного анализа являются:

1. Установление формы зависимости Y от $\{X_j\}$.
2. Определение вида уравнения регрессии.
3. Прогнозирование значений результирующей (прогнозируемой) переменной Y , носящей название отклика по известным значениям переменных X_1, X_2, \dots, X_P , которые часто называют регрессорами либо аргументами.

Математические модели регрессионного анализа

Пусть y_i -экспериментальные данные, $\varphi(x_i)$ - сглаживающая их функция. Представим $\varphi(x)$ в виде функции от аргумента x и неизвестных коэффициентов $\beta_0, \beta_1, \beta_2, \dots$, т.е. $\varphi(x, \beta) = \varphi(x, \beta_0, \beta_1, \beta_2, \dots)$. Согласно метода наименьших квадратов необходимо выполнить требование, чтобы сумма квадратов отклонений экспериментальных данных y_i от сглаживающей функции $\varphi(x_i)$ обращалась в минимум, т.е.

$$\sum_{i=1}^n [y_i - \varphi(x_i)]^2 \Rightarrow \min .$$

Для нахождения неизвестных коэффициентов $\beta_0, \beta_1, \beta_2 \dots$ данное выражение дифференцируется каждому из них и производные приравниваются к нулю. В результате получают, а затем решают относительно коэффициентов так называемую систему нормальных уравнений.

В общем виде при анализе геофизических данных необходимо учитывать погрешности измерений и помехи различной природы. Поэтому наблюдаемая величина y_i может быть записана

$$y_i = \varphi(x, \beta) + e_i.$$

Здесь e обозначает ошибку. Обычно полагают, что ошибки в точках наблюдений e_i имеют нормальное распределение с нулевым математическим ожиданием и одинаковой дисперсией, не коррелированы между собой и не зависят от φ . Данная модель также приводит к системе нормальных уравнений, решая которую и определяют искомые коэффициенты.

Следует указать, что искомые зависимости между геолого-геофизическими параметрами чаще всего аппроксимируются линейными моделями первого порядка. В противном случае используются полиномиальные модели второго (вида $y = \beta_0 + \beta_1x + \beta_2x^2$) и более высоких порядков, функции гиперболического, степенного и логарифмического типов. Обычно составляют модели не выше второго порядка.

Описанный выше подход не дает возможности определить наилучшую форму уравнения регрессии. Поэтому, после расчета конкретного уравнения необходимо проверить статистическую гипотезу о соответствии данной регрессионной модели анализируемой совокупности. Такая проверка может быть осуществлена с использованием критерия Фишера. Роль регрессии по сравнению с влиянием случайных ошибок существенна, если $D_{рег}/D_{отк} < F_{табл}$. Здесь $D_{рег}$ дисперсия самого уравнения регрессии, $D_{отк}$ -оценка дисперсии отклонения исходных данных от регрессии, $F_{табл}$ – табличное значение критерия Фишера.

Качество регрессий обычно оценивается по среднеквадратическим отклонениям исходных данных от линии регрессии (σ) и по значениям коэффициентов детерминации (R^2). Последний равен квадрату обычного коэффициента корреляции R и является отношением суммы квадратов,

обусловленных регрессией, к общей сумме квадратов откликов, скорректированной на среднее. Он измеряется в процентах и показывает долю общего разброса Y относительно среднего значения, объясняемую регрессией. Чем ближе значение R^2 к ста процентам, тем лучше подобранная модель описывает данные эксперимента.

Одновременное использование двух и более оценок качества регрессий имеет важное значение, так как каждая из оценок в отдельности не дает полного представления о возможности и целесообразности решения поставленных задач методом регрессионного анализа. Например, регрессионная модель анализируемой совокупности геофизических данных может быть статистически значимой, но точность прогноирования целевой характеристики с использованием регрессионного уравнения окажется неудовлетворительной. С другой стороны, очень маленькое значение среднеквадратического отклонения исходных данных от линии регрессии (σ) может оказаться просто следствием малой изменчивости зависимой характеристики.

С целью повышения точности и достоверности результатов в настоящее время часто применяется метод пошаговой регрессии, позволяющей автоматически выбрать те переменные, которые существенно влияют на результаты регрессионного анализа. Данный метод состоит в том, что на каждом шаге в модель включается либо исключается какая-либо независимая переменная, после чего полученный результат анализируется на предмет статистической значимости.

Применение регрессионного анализа в практике геологоразведочных работ

Регрессионный анализ нашел широкое применение при решении большого числа геолого-геофизических задач: при комплексной интерпретации промыслово-геофизических данных, при геологической интерпретации гравитационных и магнитных аномалий, при прогнозировании структурных планов глубоких горизонтов и др.

Регрессии часто используются при определении значений одного физического параметра по известным значениям другого. В частности, плотность горных пород $\sigma_{\text{п}}$, необходимую для интерпретации результатов гравиметрических наблюдений, можно вычислить по скорости упругих волн, полученную по результатам сейсмических работ. Впервые получившая

широкое распространение в практике оценка взаимосвязи скорости и плотности была получена Н.Н. Пузыревым. Она имеет вид $v = 6\sigma_{\Pi} - 11$.

Оценка глубины залегания фундамента H (или другой геологической границы) по данным гравиразведки (или магниторазведки) $H = a_1\Delta g + a_0$ может служить примером использования линейной регрессии. Но более широкое применение в практике получила методика КОМР (Шрайбман В.И., Жданов М.С., Витвицкий О.В. и др.) Сущность метода КОМР состоит в следующем.

В пределах исследуемой территории выделяются эталонные прогнозные пространства, характеризующиеся сходством геолого-геофизического строения. В каждой точке эталонного пространства заданы значения изучаемой глубины H и геофизических параметров, прогнозирующих характеристику.

Из суммарного прогнозирующего поля Δg выделяется полезная компонента Δg_0 , наиболее тесно корреляционно связанная с H путем вычитания из Δg фоновой составляющей Δg_{ϕ} .

$$\Delta g_0 = \Delta g - \Delta g_{\phi}$$

Последняя аппроксимируется многочленом степени от координат пространства X и Y :

$$\Delta g_{\phi}(x,y) = \sum_{P=0}^A \sum_{S=0}^{B-P} \lambda_{ps} X^P Y^S \quad (P + S > 0),$$

где λ_{ps} - коэффициенты фонового многочлена, $n = \max(A, B)$ - порядок фоновой составляющей, изменяющейся от 1 до 4.

Коэффициенты λ_{ps} определяются из условия максимума модуля коэффициента корреляции $r(\Delta g_0, H)$.

Поле Δg помимо влияния изучаемой границы отражает влияние других геолого-геофизических факторов, которые при разделении следует рассматривать как помехи. Поэтому в процедуре КОМР предусматривается возможность включения в обработку некоторого дополнительного корректирующего параметра j с тем, чтобы максимально исключить вносимые им искажения.

Выбор оптимальной остаточной составляющей Δg_0 разделяемого поля представляет большую трудность. В настоящее время разработан ряд критериев выбора оптимального порядка фоновой компоненты. Например, критерием может являться минимум значения коэффициента корреляции между фоновой составляющей Δg_{ϕ} и прогнозируемой характеристикой H .

При выполнении этого условия в остаточной составляющей будет в наибольшей степени учтена вся информация об изучаемой границе, содержащаяся в наблюдаемом поле Δg .

Обсуждая возможности применения регрессионного анализа важно подчеркнуть, что при обработке одних и тех же данных можно получить несколько видов уравнений регрессий, практически равноценных по ошибке прогноза. Это обстоятельство особенно сильно проявляется при наличии взаимно коррелируемых признаков. Поэтому коэффициентам уравнения регрессии обычно нельзя придать определенный физический смысл, а регрессионному уравнению поставить в соответствие ту или иную физико-геологическую модель.

2.9.2. Факторный анализ

Математические модели факторного анализа

При исследовании сложных объектов и систем, часто не удается непосредственно измерить величины, определяющие свойства этих объектов (так называемые факторы). В тоже время для измерений могут быть доступны иные величины, тем или иным способом зависящие от этих факторов. Общее число факторов может быть гораздо меньше, чем число измеряемых переменных. Хотя различные процедуры факторного анализа (ФА) существенно различаются по целям и лежащим в их основе математическим моделям, все они имеют одно общее средство: представление о наличии нескольких гипотетических переменных, которые сохраняют всю существенную информацию, содержащуюся в большом количестве наблюдаемых переменных. В качестве такой гипотетической переменной может, например, выступить фактор нефтегазоносности разреза, а в качестве наблюдаемых переменных – геофизические характеристики.

Наиболее употребительная математическая модель метод ФА в терминах наблюдаемых переменных выражается уравнением:

$$X_{(N*P)} = F_{(N*K)} A_{(K*P)} + E_{(N*P)},$$

где X - матрица данных,

F - матрица значений факторов,

A - матрица факторных нагрузок,

E - матрица остатков или ошибок,

P -число переменных,

N -число наблюдений,

K -скаляр,обозначающий число используемых факторов.

Для произвольного элемента матрицы данных, например расположенного в n -ой строке i -м столбце, из модели вытекает следующее соотношение в скалярных обозначениях:

$$x_{ni} = \sum_{j=1}^k f_{nj} a_{ij} + e_{ni}.$$

Для вектора-столбца, получаем:

$$X = Af + e.$$

Последнее уравнение является фундаментальным для всех видов факторного анализа в R -модификации. Оно показывает, что каждая наблюдаемая переменная равна взвешенной сумме факторов и члена, содержащего ошибки. Произведение Af представляет собой векторную оценку X , вектор e - разность между этой оценкой и наблюдаемым вектором. Предполагается, что остатки коррелируются с факторами. Вектор-столбец содержит факторы, а элементы матрицы A - это вклады каждой переменной в факторы.

В общем случае наличия k факторов можно записать

$$x_i = a_{i1}f_1 + a_{i2}f_2 + \dots + a_{ik}f_k + e_i,$$

где линейная комбинация

$$c_i = a_{i1}f_1 + a_{i2}f_2 + \dots + a_{ik}f_k,$$

является наилучшей линейной оценкой переменной x_i и называется ее общей частью. Остаточный член e_i , обычно, считается как сумма двух некоррелируемых частей:

$$e_i = s_i + \varepsilon_i,$$

где s_i - специфическая часть x_i , ε_i - погрешность измерений.

Известны две основные модификации факторного анализа: метод главных компонент (МГК) и собственно факторный анализ. В МГК факторы определяются в соответствии с критерием максимизации их вклада в суммарную дисперсию (по сути изменчивость) всех переменных и делается предположение, что e_i малы. В ФА факторы находятся по принципу максимизации связей между переменными.

Особенности применения факторного анализа для решения геологических и геофизических задач

При решении геофизических задач МГК обычно используется для уменьшения количества анализируемых характеристик без существенных потерь содержащихся в них информации. Проще говоря, МГК позволяет получить минимальное число новых переменных, являющихся линейными комбинациями исходных, причем эти новые переменные содержат то же количество информации. Практически количество характеристик обычно уменьшается до двух-пяти новых комплексных показателей – главных компонент (ГК). Как и для исходных геофизических характеристик при анализе ГК можно строить соответствующие графики и карты, выявлять зоны аномальных значений и др. При реализации собственно факторного анализа основное внимание уделяется содержательному истолкованию факторов. Это дает возможность проверять или формулировать различные гипотезы, например, о влиянии того или иного геологического процесса на формирование залежей углеводородов.

Положительных результатов от применения метода ФА следует ожидать только в том случае, когда подлежащие анализу исходные данные удовлетворяют определенным требованиям. Опишем основные из них.

Факторный анализ базируется на предположении, что слагающие исследуемые выборки признаки коррелируют между собой. Поэтому необходимо перед факторным экспериментом оценивать корреляционные связи между исходными характеристиками. При наличии таких связей возможны два случая: в первом из них один или несколько признаков определяют остальные, во втором - поведение признаков обусловлено воздействием других, не включенных в исследуемый набор переменных. Большой практический интерес представляет второй случай, т.к. ФА позволяет изучить закономерности проявления причин - факторов и соответствующих им природных закономерностей. При этом должен исключаться первый случай, т.е. должна отсутствовать прямая причинно-следственная связь между исходными переменными. Исходя из содержания этого требования недопустимо увеличение количества признаков путем арифметических действий с имеющимися.

Одним из обязательных условий является однозначность в определении признаков и их сопоставимости. Содержание этого требования заключа-

ется в том, что вкладываемый в переменные смысл не должен изменяться при переходе от одного объекта к другому.

Необходимо соблюдать также требования репрезентативности выборки: объем ее должен удовлетворять условию получения надежных оценок элементов корреляционной матрицы. Надежными можно считать элементы, значения которых по модулю не меньше дисперсии коэффициента корреляции, определяемой по формуле:

$$\sigma r = (1 - r^2) / n ,$$

где n - количество точек наблюдения;
 r - коэффициент корреляции.

Между подлежащими анализу факторным методом характеристиками не должно существовать нелинейных корреляционных связей. Качественные суждения о характере связи можно сделать по виду эмпирических линий регрессий. При заметном отклонении их от прямой линии гипотезу о виде корреляционной связи необходимо проверить аналитически, например, по критерию Фишера. Отметим, что нелинейность корреляционной зависимости зачастую вызывается асимметричностью статистических распределений параметра. При этом близкие к линейным связи можно получить, если воспользоваться логарифмическим масштабом.

Для получения объективных результатов факторного эксперимента следует также избегать неоднородностей массивов данных. Неоднородность, обычно, является следствием составления единой выборки по результатам наблюдений на различных профилях (участках) территории исследований.

Выполнение перечисленных требований в значительной степени уменьшают опасность появления "ложных" факторов, связанных не с природными объектами, а с особенностями вычислительных процедур ФА.

Формальное применение метода, без учета его специфических особенностей, может привести к ложным выводам. Поэтому важно сформулировать ряд моментов, которые необходимо учитывать при анализе геолого-геофизических материалов:

1. Знаки факторов и факторных нагрузок следует понимать в относительном смысле: если один из факторов положительно связан с признаком А и отрицательно с Б, то его нужно объяснять как фактор, оказывающий на А и Б действие противоположного характера.

2. Различают детерминированный и стохастический случаи применения ФА. Первый имеет место при анализе конкретной выборки и никаких

выводов о статистической совокупности в целом не делается. Во втором случае набор объектов рассматривается как случайная выборка из некоторой совокупности.

3. Для облегчения интерпретации факторов можно применять процедуру их вращения, которая приводит к тому, что средних факторных нагрузок становится меньше, а больших и малых - больше.

4. Вычисляемые факторы и факторные нагрузки должны проверяться на значимость. Значимыми факторами, обычно, считают те, которые в сумме выбирают от 75 до 95% суммарной дисперсии выборки.

5. С увеличением числа анализируемых факторов появляется возможность учета все более "тонких" свойств объектов исследования. Однако, интерпретация факторов с большими номерами может оказаться не достоверной из-за погрешностей в наблюдениях.

6. Решение конкретных геолого-геофизических задач может обеспечиваться анализом как отдельных факторов, так и их комбинации.

7. При описании и интерпретации вычисляемых факторов удобно пользоваться так называемыми тестовыми структурами. Они содержат номера факторов, выбираемую ими долю суммарной дисперсии, перечень имеющих значимые нагрузки переменных (приводятся в скобках) и величины этих нагрузок (показываются как индекс переменной). При этом, если переменные имеют положительные нагрузки, то они записываются в числителе, а если отрицательные – то в знаменателе формулы тестовой структуры. Пример записи тестовой структуры фактора второго фактора, выбирающего 33% суммарной дисперсии анализируемой выборки геофизических характеристик $X_1 \dots \dots X_6$, приведен ниже

$$II - 33 \frac{(X_1)_{71} (X_2)_{42} (X_3)_{38}}{(X_4)_{72} (X_5)_{39} (X_6)_{44}}.$$

8. Интерпретация факторов, в случае неоднозначного содержания слагающих векторов наблюдения характеристик, представляет значительные трудности. Для их преодоления можно либо привлекать эталонную информацию, либо вводить в признаковое пространство параметры, описывающие конкретные объекты (свойства) разреза. Такие признаки называются маркирующими.

2.9.3. Распознавание образов

Общие сведения о методах распознавания образов

Важное место в работе геофизика занимает поиск сходства среди множества встречающихся ситуаций и объектов. Например, проанализировав ситуацию в пределах хорошо изученных известных месторождений полезных ископаемых, можно найти похожую ситуацию в пределах других участков территории и на этой основе спрогнозировать там наличие новых месторождений. Решение такой задачи может основываться на использовании способов распознавания образов.

Основная идея способов распознавания образов - это поиск аналогий. На принципах теории распознавания базируется большое число приемов комплексной интерпретации геофизических данных. Как правило, использование таких приемов основано на использовании "обучающего" материала и задача распознавания формулируется следующим образом: статистическими методами нужно разделить пункты геофизических наблюдений на два и большее число классов при априорно заданных вероятностях появления классов и распределения вероятностей значений каждого признака. Таким образом, модель геофизических полей должна предусматривать наличие следующей априорной информации:

- количество классов, на которые требуется разбить объекты исследований. Чаще всего рассматривается два класса: нефтенасыщенная и водонасыщенная порода; точка наблюдения в пределах залежи и точка наблюдения вне залежи и т.д.;
- вероятности встречи этих классов;
- плотности распределения вероятностей значений каждого признака для тех же классов.

К настоящему времени разработан целый ряд алгоритмов решения задачи распознавания образов. В любом случае основой распознавания является анализ совокупности признаков, присущих определенным ситуациям. В этом смысле каждая ситуация может быть представлена некоторым k -мерным вектором, которому в эвклидовом пространстве соответствует точка с k координатами. Множество наблюдаемых ситуаций образует в пространстве R^k множество точек. Предполагается, что некоторому классу ситуаций в этом пространстве соответствует некоторая ограниченная область, а фиксированному множеству классов - такое же

количество непересекающихся областей. Задача заключается в том, чтобы найти границы областей соответствующих классов.

Общая схема аппарата распознавания образов такова. Строится некоторая функция $F(x_1, x_2, \dots, x_k)$, аргументами которой являются значения признаков в точках изучаемого профиля (площади). Функция F должна принимать максимально различные значения при подстановке аргументов, соответствующих объектам различных классов. На эталонных объектах значения, принимаемые F в различных точках являются критерием классификации и используются при отнесении наблюдаемых объектов к какому либо их классов ситуаций.

Распознавание образов на основе теории статистических решений

В практике геофизических работ наиболее широкое применение нашли методики распознавания образов, основанные на теории статистических решений. Базирующиеся на этой теории алгоритмы распознавания обладают преимуществами, обусловленными, в первую очередь высоким уровнем развития теории, особенно в отношении критериев принятия решений, в частности критериев максимального правдоподобия и максимума апостериорной вероятности.

Математическая постановка задачи распознавания образов сводится к следующему. Пусть наблюдается L экспериментальных полей x_1, x_2, \dots, x_L , тогда математическая модель исходных данных будет иметь вид

$$x_l = \begin{cases} s_{1l} + n_{1l} \Rightarrow H_0, \\ s_{2l} + n_{2l} \Rightarrow H_1, \\ \dots\dots\dots \\ s_{Ml} + n_{Ml} \Rightarrow H_M, \end{cases}$$

где s_{kl} — неслучайная компонента полей; n_{kl} — случайная компонента полей (помехи); индекс $k = 1, \dots, M$, характеризует конкурирующие (качественные) состояния объектов; M — общее число состояний или число классов объектов. Каждому такому состоянию отвечает гипотеза H_k . Когда записанная модель предполагает наличие априорной информации о количестве классов, на которые требуется разбить объекты исследований; вероятностях встречи этих классов; плотностях распределения вероятностей значений каждого признака для

тех же классов, то говорят о распознавании образов с обучением. Указанную априорную информацию получают на основе анализа данных предшествовавших исследований на эталонных объектах для каждого класса в отдельности.

Априорные вероятности классов, если о них не имеется никаких сведений, в первом приближении полагаются равными. Таким образом, задача заключается в следующем: статистическими методами нужно разделить пункты геофизических наблюдений на два класса (например, нефтеносные и «пустые» структуры) или на большее число классов при априорно заданных вероятностях появления классов и распределениях вероятностей значений каждого признака.

Рассмотрим решение задачи распознавания с обучением на основе использования коэффициент правдоподобия для случая наличия двух классов H_0 и H_1 .

Для независимых признаков x_1, x_2, \dots, x_L коэффициент правдоподобия равен отношению функций правдоподобия $P(x_i H_k)$, вычисляемых для каждого l -го признака и по каждому классу:

$$\Lambda(\vec{X}) = \frac{P(x_1 H_1) P(x_2 H_1) \dots P(x_L H_1)}{P(x_1 H_2) P(x_2 H_2) \dots P(x_L H_2)}$$

Распознавание исследуемых объектов сводится к нахождению в i -й точке (ячейке) значений $P(x_i H_0)$ и $P(x_i H_1)$, которые снимаются с гистограмм, полученных на эталонных объектах. После этого рассчитываются величины коэффициента правдоподобия. С практической точки зрения важно отметить, что ограничения на законы распределения признаков при этом отсутствуют. Решение о принадлежности i -го искомого объекта к классу H_0 принимается согласно критерию максимального правдоподобия при $\Lambda(\vec{X}_i) > 1$. Если $\Lambda(\vec{X}_i) < 1$, справедлива гипотеза H_1 , т. е. объект принадлежит ко второму классу.

Частные коэффициенты правдоподобия для каждого l - то признака $\Lambda(\vec{X}_i) = P(x_l H_0) / P(x_l H_1)$ характеризуют относительный вклад признака в величину общего (вычисленного по всем признакам) коэффициента правдоподобия.

Основной проблемой при применении способов распознавания образов является сложность организации процесса обучения и формирования обучающей последовательности. Например, при классификации выявленных геофизическими методами структур на пустые и продуктивные возникают

проблемы с формированием эталонных данных по пустым объектам. Последние, из очевидных причин изучаются со значительно меньшей детальностью, чем продуктивные.

В завершение отметим, что эмпирическую оценку качества решающего правила можно произвести по специально формируемой независимой выборке, которая называется экзаменационной последовательностью.

Дискриминантный анализ

Дискриминантным анализом называют раздел многомерной статистики, включающий методы классификации многомерных наблюдений на основе вычисления и анализа так называемых дискриминантных функций. В этом виде анализа предполагается, что в распоряжении исследователя имеется обучающая выборка. Классический вариант дискриминантного анализа основан на нахождении по обучающей выборке некоторой линейной комбинации исходных параметров, характеризующих две различные группы объектов, которая показывает максимальное различие между этими группами. Такая функция (линейная дискриминантная функция - ЛДФ) может быть использована для отнесения новых объектов в ту или иную группу. Один из наиболее распространенных методов нахождения ЛДФ - построение уравнений регрессии, где зависимыми переменными являются разности между многомерными средними двух групп. В матричном виде уравнение регрессии записывается:

$$[S^2_p] * [\lambda] = [D],$$

где $[S^2_p]$ – матрица дисперсий и ковариаций объединенной выборки; $[\lambda]$ – вектор столбец неизвестных коэффициентов дискриминантной функции; $[D]$ – вектор-столбец разностей между средними значениями двух групп.

Записанное выше уравнение регрессии решается относительно неизвестных коэффициентов $[\lambda]$ с помощью операций обращения и умножения матриц:

$$[\lambda] = [S^2_p]^{-1} * [D].$$

Вычисленный набор коэффициентов используется для построения дискриминантной функции вида

$$R = \lambda_1 \varphi_1 + \lambda_2 \varphi_2 + \dots \dots \dots + \lambda_m \varphi_m.$$

Суммируя слагаемые функции R получают число, называемое дискриминантной меткой. Данное число сравнивается с так называемым дискриминантным индексом R_0 (пороговым значением ЛДФ), которое вычисляется путем подстановки в уравнение дискриминантной функции среднего арифметического, полученного из средних для двух обучающих выборок. Отметим, что R_0 соответствует точке разделяющей прямой, которая лежит в точности посередине между центрами заданных в обучающей выборке групп.

Правило принятия решения при распознавании объектов с помощью сформированной ЛДФ заключается в том, что если полученная дискриминантная метка больше R_0 , то объект относится к первому классу, если меньше R_0 – ко второму.

2.9.4. Классификация с одной обучающей выборкой (алгоритм "Хоттелинг")

Алгоритмы классификации при наличии одной обучающей выборки можно построить, сформулировав распознавание как задачу проверки гипотез. Пусть имеется экзаменуемый вектор Y и независимая выборка x_1, \dots, x_N из нормального распределения $N(\mu, \Sigma)$, где μ и Σ - неизвестные значения среднего и ковариационной матрицы. Проверяется гипотеза $H_0: Y \in N(\mu, \Sigma)$ против гипотезы $H_1: Y \in N(\mu_1, \Sigma), \mu \neq \mu_1$. Можно показать, что критерий максимального правдоподобия для проверки гипотезы H_0 против H_1 эквивалентен критерию $T^2 \geq T_0^2(\alpha)$, где $T_0^2(\alpha) = \frac{(N-1)r}{N-r} F_{r, n-r}(\alpha)$; $F_{r, n-r}(\alpha) - 100\alpha\%$ - ная точка центрального F -распределения со степенями свободы r и $n-r$, а $T^2 = (Y - \bar{X})' S^{-1} (Y - \bar{X})$, где \bar{X} , S - оценки среднего и ковариационной матрицы, сделанные по обучающей выборке X_1, \dots, X_N ; T^2 - статистика Хоттелинга. Таким образом, для выбранного уровня значимости α можно построить классификационное правило, основанное на сравнения модифицированной статистики Хоттелинга T^2 с порогом $T_0^2(\alpha)$. Алгоритм "Хоттелинг" вычисляет для каждого экзаменуемого вектора вероятность ошибки I-го рода, что и является дискриминантной функцией, которая сравнивается с соответствующим пороговым значением (выбранным уровнем значимости).

2.9.5. Классификация без обучения

При поисках полезных ископаемых часто возникает проблема классифицирования объектов изучения в условиях отсутствия информации по эталонным объектам или существенном ограничении априорных данных.

Классификация геологических объектов без обучения приобретает особое значение при обработке геолого-геофизических данных в сложно построенных в геологическом отношении районах, когда сложный и непостоянный характер геофизических полей обычно исключает возможность использования эталонных объектов. Решить подобную проблему можно с помощью кластерного анализа (КА), корреляционного алгоритма классификации и др.

Кластерный анализ

С помощью КА совокупность объектов, например наблюдений на профиле, разбивается на группы однородных ("схожих") объектов, называемых кластерами. Подчеркнем, что в КА классификация объектов и наблюдений осуществляется на основе применения формальных математических методов.

Математически понятие однородности объектов определяется заданием правила вычисления величины ρ_{ij} , характеризующей расстояние $d(X_i, X_j)$ между объектами X_i и X_j из исследуемой совокупности. Если задана функция $d(X_i, X_j)$, то близкие в смысле этой метрики объекты считаются однородными, принадлежащими к одному классу. Естественно, при этом необходимо сопоставление $d(X_i, X_j)$ с некоторым пороговым значением.

Выбор расстояния или коэффициента сходства является одним из основных моментов исследования, который определяет понятие однородности объектов. В качестве меры сходства может, например, использоваться коэффициент корреляции. Заметим, что результаты кластеризации зависят от выбранного метода.

Большинство методов кластеризации являются объединительными - они начинают с создания элементарных кластеров, каждый из которых состоит ровно из одного исходного наблюдения (одной точки), а на каждом последующем шаге происходит объединение двух наиболее близких кластеров в один. Момент остановки процесса объединения может задаваться исследователем (например, указанием требуемого числа кластеров), либо выбираться автоматически. Графическое изображение процесса объединения кластеров может быть получено с помощью дендрограммы - дерева объединения кластеров. При этом на горизонтальной оси показываются объекты, скомбинированные в соответствии с выполненным анализом. На вертикальной оси откладывается расстояние для каждого шага работы иерархического алгоритма кластеризации.

Обсуждая возможности КА при решении геофизических задач можно отметить с одной стороны простоту реализации этого метода, с другой сложности, возникающие при содержательном истолковании выделенных кластеров. Последнее обстоятельство определяется применением исключительно формальных критериев различия кластеров. В результате по данным КА можно лишь констатировать, что объекты наблюдения отличаются друг от друга. Но, чем по сути определяются эти различия или сходство без привлечения дополнительной информации как правило установить не удастся. Поэтому результаты кластеризации часто служат лишь подспорьем для содержательного анализа. Иногда результаты кластеризации можно интерпретировать, используя другие, помимо кластерного анализа, статистические методы.

Корреляционный алгоритм классификации

Данный алгоритм позволяет провести классификацию геологических объектов по зависимым друг от друга признакам. Принцип классификации в нем для случая двух классов формулируется следующим образом: i -й объект относится к классу H_1 , если выполняется условие

$$\sum_{i=1}^L h_l x_{il} = y_i > y_{nop}$$

где y_i —комплексный параметр, равный сумме взвешенных значений каждого признака для i -го объекта; h_l —весовой коэффициент l -го признака; y_{nop} — порог принятия решения.

Если веса h_l выбраны так, что обеспечивают различные значения y_i для этих классов, то вычисление h_l можно свести к процедуре нахождения собственного вектора корреляционной матрицы исходных признаков, соответствующего ее максимальному собственному числу λ_{max} .

Порог для принятия решения о наличии объекта искомого класса определяется исходя из условия нормальности распределения y_i и определяется заданием вероятности

$$\gamma = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y_{nop}} e^{-x^2/2} dx,$$

$$\text{где } x = (y_{nop} - \bar{y}) / \sigma_y \sqrt{2}$$

Например, для вероятности $\gamma = 95\%$ $y_{пор}$ равно 1,65. Так как дисперсия величины y_i равна λ_{max} , то для вероятности 95 % порог будет $y_{пор} = 1,65 \sqrt{\lambda_{max}}$. Для вероятности $\gamma = 99\%$ имеем $y_{пор} = 2,4 \sqrt{\lambda_{max}}$.

Описанный алгоритм отличается простотой и наглядностью результатов. Он пригоден и для случая, когда имеется три и более классов. В таких ситуациях корреляционный алгоритм применяется многократно, что позволяет выполнить классификацию объектов внутри первоначально выделенных двух классов.

2.9.6. Повышение надежности прогнозирования геологических объектов за счет комплексирования геофизических методов

При прогнозировании целевых объектов в практике геологоразведочных работ часто допускаются ошибки как первого, так и второго рода. Часто, например, отнесенные по геофизическим данным к продуктивной структуры оказываются пустыми или вообще не подтверждаются. Повысить надежность прогнозирования объектов можно комплексировав геофизические методы. Предположим, что сейсморазведкой выявлена предположительно нефтегазоносная структура, относительно которой можно предположить две гипотезы. Одна из них H_0 - структура реально существует, другая H_1 - структура не существует. Естественно возникает вопрос, какая из гипотез верна. По сути возникает проблема классификации геофизических наблюдений. Обозначим априорные вероятности гипотез через $P(H_0)$ и $P(H_1)$. Предположим далее, что в результате обработки данных гравirazведки выявлен некоторый дополнительный поисковый признак Π (например, наличие положительной аномалии), также свидетельствующий о наличии структуры.

Вероятности гипотез $P(H_0)$ и $P(H_1)$ в связи с появлением нового поискового признака изменятся, причем они будут теперь условными, т.е. $P(H_0/\Pi)$ и $P(H_1/\Pi)$. Их вычисление, в частности $P(H_0/\Pi)$, производится по формуле Байеса.

$$P\left(\frac{H_0}{\Pi}\right) = \frac{P(H_0) \cdot P(\Pi/H_0)}{P(H_0) \cdot P(\Pi/H_0) + P(H_1) \cdot P(\Pi/H_1)}$$

Для практического использования данной формулы необходимо знать условные вероятности $P(H_0/\Pi)$ и $P(H_1/\Pi)$, называемые ещё функциями

правдоподобия. Они характеризуют частоту проявления признака Π при наступлении гипотез H_0 и H_1 . Величина $\lambda = \frac{P(\Pi/H_0)}{P(\Pi/H_1)}$ показывает прирост надежности информации за счет комплексирования методов. Если $\lambda = 1$, то прироста информации не будет, если $\lambda < 1$, то информация о нефтегазоперспективной структуре, наоборот, уменьшится, при $\lambda > 1$ информация увеличится. С учетом этого коэффициента можно записать

$$P\left(\frac{H_0}{\Pi}\right) = \frac{1}{1 + \frac{P(H_1)}{P(H_0)} \cdot \frac{1}{\lambda}}$$

При наличии m признаков $A_{1,2\dots m}$ о структуре, полученных при применении m отдельных геофизических методов имеет место следующая обобщенная формула

$$P(H_0 / \Pi_{1,2\dots m}) = \frac{1}{1 + \frac{P(H_1)}{P(H_0)} \prod_{k=1}^m \frac{1}{\lambda_k}}$$

Данная формула позволяют оценить прирост надежности классификации объектов наблюдения для тех или иных конкретных условий. Пусть в отдельном районе на основании опыта установлено, что в среднем 1/3 прогнозируемых сейморазведкой структур подтверждаются и продуктивны. Тогда $P(H_0) = 1/3$, $P(H_1) = 2/3$. Предположим, что опытным путем установлено, что $\lambda = 2$. Тогда для $m = 1$, $P(H_0/\Pi_1) = 1/2$. Иными словами, вероятность наличия нефтегазоносной структуры при учете лишь одного дополнительного (гравитационного) признака возросла до 1/2 по сравнению с исходной вероятностью $P(H_0) = 1/3$. Привлекая другие геофизические методы можно повысить достоверность прогнозирования нефтегазоносной структуры еще в большей мере.

2.9.7. Методы нечетких множеств

Теория нечетких множеств появилась сравнительно недавно, но все больше привлекает внимание геологов и геофизиков. Основные ее достоинства связаны с тем, что она позволяет эффективно решать различные, в том числе и геологоразведочные задачи в условиях недостаточной формализованности последних, неполноты и невысокой достоверности исходной информации, неопределенности в выборе оптимальных вариантов. Так, зачастую мы не можем безоговорочно отнести эталонный объект к тому или иному классу, но можем определить степень его сходства с объектами

того или иного класса. Соответствующая такому случаю теория множеств и критерии классификации получили название нечетких (расплывчатых, размытых).

Нечеткие множества не имеют четких границ, и характеризуются постепенным переходом от принадлежности к непринужденности.

Целесообразность подразделения геологических объектов на классы не имеющих четких границ появляются при решении широкого круга задач. Например, среди коллекторов нефти и газа есть переходные разности. Аналогичная ситуация имеет место при анализе литологического состава пород.

Пусть объекты описываются вектором параметров $\bar{x} = (x_1, \dots, x_n)$. Нечеткое множество A определяется заданием для каждого элемента x степени его принадлежности множеству A , которая обозначается $\mu_{A(x)}$ и называется функцией принадлежности элемента x множеству A . Равенство $\mu_{A(x)}=1$ означает полную принадлежность элемента x множеству A . $\mu_{A(x)}=0$ – полную непринадлежность x множеству A . $0 < \mu_{A(x)} < 1$ – промежуточные степени принадлежности.

Например, породы, характеризующиеся высокими фильтрационно-емкостными свойствами со степенью, близкой к 1, принадлежат к нечеткому классу коллекторов; неколлекторы имеют $\mu_{A(x)}$ близкие к нулю. Породы с пониженными фильтрационно-емкостными свойствами принадлежат к нечеткому классу коллекторов со степенями, лежащими в пределах от 0 до 1.

На настоящий момент предложено большое количество алгоритмов и программ для вычисления функций принадлежности, которые с успехом могут применяться при решении геологических и геофизических задач.

2.10. Технологии нейронных сетей как альтернатива статистическим методам

Описанные выше и другие статистические методы позволяют решать широкий круг различных геологических и геофизических задач. Практически все эти задачи, например распознавания образов, прогнозирование физических свойств пород и месторождений полезных ископаемых, могут быть решены также путем реализации технологий искусственных нейронных сетей (НС). Они особенно бурно развиваются в последние десятилетия и весьма эффективны в типичных для практики

геологоразведочных работ ситуациях, когда постановка задач неформальна, строгий алгоритмический подход невозможен, а решение неоднозначно. Сравнивая возможности НС с традиционными методами статистического анализа следует указать, что большинство из последних связано с построением моделей, основанных на тех или иных предположениях и теоретических выводах (например, что наблюдаемая геофизическая характеристика имеет нормальное распределение). Нейросетевой подход не связан с такими предположениями и в этом смысле предпочтителен.

Нейронные сети - основные понятия и принцип работы

В основу искусственных НС положены черты живых:

- в качестве простого обрабатывающего элемента выступает нейрон;
- в обработке информации участвует большое число нейронов;
- каждый нейрон связан с большим числом других нейронов (глобальные связи);
- связи между нейронами изменяются по весу;
- информация обрабатывается параллельно.

Прототипом для создания нейрона в искусственных НС послужил биологический нейрон головного мозга. Биологический нейрон имеет тело, отростки - дендриты, по которым в нейрон поступают входные сигналы, а также отросток - аксон, передающий выходной сигнал нейрона. Точка соединения дендрита и аксона называется синапсом. В общих чертах функционирование нейрона можно представить следующим образом:

- 1) нейрон получает от дендритов входные сигналы;
- 2) нейрон оценивает суммарное значение входных сигналов. При этом каждый вход характеризуется некоторым весовым коэффициентом, определяющим важность поступающей информации. С точки зрения математики нейрон вычисляет скалярное произведение вектора входных сигналов и вектора весовых коэффициентов;
- 3) нейрон формирует выходной сигнал, интенсивность которого зависит от значения вычисленного скалярного произведения. Если это произведение не превышает некоторого заданного порога, то выходной сигнал не формируется (нейрон "не срабатывает");
- 4) выходной сигнал поступает на аксон и передается дендритам других нейронов.

Поведение искусственной НС зависит как от весовых коэффициентов, так и от функции возбуждения нейронов (здесь и далее под нейроном

подразумевается искусственный нейрон, то есть ячейка НС). Известны три основных вида такой функции: пороговая, линейная и сигмоидальная. Для пороговых элементов выходной сигнал устанавливается на одном из двух уровней в зависимости от того, превышает или нет суммарный сигнал на входе нейрона пороговое значение. Для линейных элементов сигнал на выходе пропорционален суммарному взвешенному входу нейрона. Для сигмоидальных элементов в зависимости от входного сигнала, выход варьируется непрерывно, но не линейно, по мере изменения входных сигналов.

По структуре связей НС делятся на два класса: однослойные и многослойные. Многослойная сеть имеет входной, выходной и скрытые слои, на входной подается информация, с выходного снимается ответ, скрытые слои участвуют в обработке.

При создании НС с целью решения конкретной задачи, необходимо выбрать, каким образом следует соединять нейроны друг с другом, и соответствующим образом подобрать значения весовых параметров на этих связях. В зависимости от установленных соединений элементы НС могут влиять друг на друга, либо не влиять. Степень влияния определяется весами соединений. Выбор структуры НС осуществляется в соответствии с особенностями и сложностью задачи. На настоящий момент предложено и практически применяется большое число различных моделей НС: модель Маккалоха, модель Розенблата, модель Хопфилда и др. Если же решаемая задача не может быть сведена ни к одному из известных типов, разработчику приходится решать сложную проблему синтеза новой конфигурации. При этом следует руководствоваться следующими принципами: возможности сети возрастают с увеличением числа ее ячеек, плотности связей между ними и числом слоев; введением обратных связей; сложности алгоритмов функционирования сети (например, введение нескольких типов синапсов – возбуждающих, тормозящих и др.).

Задачи, решаемые на основе нейронных сетей

Наибольшее применение НС нашли при решении задач в следующих ситуациях:

- отсутствует алгоритм или не известны принципы решения задач, но имеется большое число примеров реализаций связанных с этими задачами ситуаций;

- проблема характеризуется большими объемами входной информации;

- имеющиеся данные неполны или избыточны, зашумлены, частично противоречивы.

НС, в частности, хорошо подходят для распознавания образов и решения задач классификации, оптимизации и прогнозирования. Из литературы известны следующие примеры их успешного применения:

- анализ геологической информации;
- разведка залежей минералов по данным аэрофотосъемок;
- анализ составов примесей;
- обработка и интерпретация наблюдаемых сигналов (разделение, идентификация, локализация, устранение шума);
- обобщение информации;
- сжатие и восстановление изображения.

Применение нейронных сетей

Для того, чтобы с помощью НС решить поставленную задачу, например задачу классификации объектов наблюдения необходимо вначале выполнить процедуру ее обучения. От того как сеть построена и как она обучается зависит класс задач, которые можно с ее помощью решать.

Существует два подхода к решению задачи обучения НС. Один из них, исторически более ранний, состоит в постепенной модификации весовых коэффициентов в процессе обучения, другой – в использовании проекционного алгоритма.

Подходы к обучению однослойных и многослойных сетей различны. Обучение многослойных сетей состоит в том, что на основе набора имеющихся примеров постепенно подбираются веса всех связей так, чтобы каждое входное состояние вызывало соответствующее условию задачи выходное. Например, определенные аномалии геофизических полей должны соответствовать наличию месторождений полезных ископаемых. Обучающие алгоритмы представляют собою итерационные процедуры с постепенным приближением к окончательным значениям весов связей. Этот подход получил дальнейшее развитие в алгоритмах типа обратного распространения.

Для однослойных НС в ряде случаев удастся выразить веса связей на основе заданного описания функционирования нейронной сети как "черного ящика". Если сеть должна реализовать заданную функцию, ее рассматривают как набор элементов пороговой логики и задача сводится к

кусочно-линейной аппроксимации этой зависимости и синтезу соответствующего автомата.

В общем случае, когда описание поведения сети задано в виде набора векторов возможных состояний, поиск синаптических весов можно свести к решению соответствующей системы нелинейных уравнений. Такое решение было впервые найдено Хопфилдом и привело к разработке проекционного алгоритма, позволяющего сократить затраты времени на обучение. Недостатком данного алгоритма является то, что при обучении необходимо сначала сформировать эталоны распознаваемых образов.

После завершения процесса обучения ИС и проверки эффективности обучения на контрольных примерах она может применяться для решения широкого круга геолого-геофизических задач. Например, подавая на вход наблюдаемые значения геофизических полей в какой либо точке наблюдения можно получить ответ на вопрос, находится ли данная точка в пределах месторождения или нет.

Таким образом, ИС можно применять в ситуациях, когда имеется некоторая хорошо известная информация (аналог эталонной информации в статистических методах) и необходимо на ее основе извлечь из всего объема имеющихся данных новую информацию.

Следует указать, что ИС относятся к наиболее «мощным» инструментам анализа экспериментальных данных. В тоже время методы нейронных сетей уступают традиционным статистическим методам по степени «открытости» и, соответственно, контролируемости. Кроме того, в зависимости от конкретных реализаций ИС, а также процедур и качества обучения полученные на основе их применения результаты могут существенно отличаться.

СОДЕРЖАНИЕ

| | |
|---|----|
| ВВЕДЕНИЕ | 3 |
| 1. МЕТОДЫ СПЕКТРАЛЬНОГО АНАЛИЗА ГЕОФИЗИЧЕСКИХ ПОЛЕЙ | 4 |
| 1.1. Линейные системы | 4 |
| 1.1.1. Системы передачи сигналов в геофизике | 4 |
| 1.1.2. Понятие о линейной системе | 4 |
| 1.1.3. Линейные преобразования | 5 |
| 1.1.4. Сумма и произведение линейных преобразований | 5 |
| 1.1.5. Методы исследования линейных систем | 6 |
| 1.1.6. Принцип суперпозиции | 7 |
| 1.2. Нелинейные явления и системы в геофизике | 7 |
| 1.3. Спектр геофизического сигнала | 10 |
| 1.3.1. Общие представления о спектре | 10 |
| 1.3.2. Элементарные сигналы | 10 |
| 1.3.3. Ряд Фурье. Дискретный спектр сигнала | 14 |
| 1.4. Спектры неустановившихся колебаний (интеграл Фурье) | 16 |
| 1.5. Основные теоремы о преобразовании сигналов и спектров | 18 |
| 1.6. Функция автокорреляции и ее спектр | 22 |
| 1.7. Минимально-фазовые операторы и сигналы | 24 |
| 1.8. Оценка формы сигнала | 25 |
| 1.9. Сглаживание трассы специальными функциями | 27 |
| 1.10. Способы повышения точности и разрешенности спектральных оценок | 28 |
| 1.11. Z-преобразование | 30 |
| 1.12. Вейвлет - анализ | 32 |

| | |
|--|-----------|
| 2. СТАТИСТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ И ИНТЕРПРЕТАЦИИ ГЕОФИЗИЧЕСКИХ ДАННЫХ | 35 |
| 2.1. Детерминированный и вероятностно-статистический подходы к обработке и интерпретации геофизических наблюдений | 35 |
| 2.2. Основные понятия теории вероятностей и математической статистики в задачах обработки геофизических данных | 36 |
| 2.3. Законы распределения случайных величин | 37 |
| 2.4. Параметры и статистики | 39 |
| 2.5. Проверка статистических гипотез. Ошибки первого и второго рода ... | 40 |
| 2.6. Статистические оценки геофизических характеристик | 42 |
| 2.7. Исключение грубых погрешностей измерений | 45 |
| 2.8. Проверка на воспроизводимость результатов | 46 |
| 2.9. Многомерный статистический анализ геофизических данных | 47 |
| 2.9.1. Корреляционно-регрессионный анализ | 48 |
| 2.9.2. Факторный анализ | 56 |
| 2.9.3. Распознавание образов | 61 |
| 2.9.4. Классификация с одной обучающей выборкой (алгоритм "Хоттединг"). | 65 |
| 2.9.5. Классификации без обучения | 65 |
| 2.9.6. Повышение надежности прогнозирования геологических объектов за счет комплексирования геофизических методов | 68 |
| 2.9.7. Методы нечетких множеств | 69 |
| 2.10. Технологии нейронных сетей как альтернатива статистическим методам | 70 |
| СПИСОК ЛИТЕРАТУРЫ | 75 |