

Ленинградский ордена Ленина
и ордена Трудового Красного Знамени
государственный университет имени А.А.Жданова

Т.Б.Яновская, Л.Н.Порохова

ОБРАТНЫЕ ЗАДАЧИ ГЕОФИЗИКИ

Учебное пособие

Ленинград
Издательство Ленинградского университета
1983

Печатается по постановлению
Редакционно-издательского совета
Ленинградского университета

УДК 550.3

Я н о в с к а я Т.Б., П о р о х о в а Л.И. Обратные задачи геофизики: Учеб. пособие. - Л.: Изд-во Ленингр. ун-та, 1983. 212 с.

Предлагаемое учебное пособие обобщает основные существующие методы решения обратных геофизических задач, описание которых до сих пор имелось только в специальной литературе. Авторы пытаются проследить связи между отдельными методами и определить место каждого из них в общей системе способов решения обратных задач. Обратная задача формулируется как определение модели среды по совокупности получаемых из наблюдений характеристик геофизического поля при условии, что известен оператор прямого преобразования модели среды в соответствующие характеристики поля. Рассмотрены методы, достаточно подробно освещенные в отечественной литературе (оптимизационные и метод регуляризации), а также методы, мало знакомые советскому читателю (метод Вэйкуса - Гильберта и метод псевдообращения). Строгое изложение материала иллюстрируется многочисленными примерами из области сейсмологии, сейсморазведки, геоэлектрики, гравиметрии.

Книга предназначена студентам старших курсов и аспирантам университетов по специальности геофизика, а также будет полезна геофизикам, занимающимся интерпретацией данных геофизических наблюдений.

Библиогр. 70 назв. Ил. 37. Табл. 2.

Р е ц е н з е н т ы: д-р физ.-мат. наук А.Л. Левшин (Ин-т физики Земли АН СССР), д-р физ.-мат. наук С.В. Гольдин (Ин-т геологии и геофизики СО АН СССР)

Я 1903010000-151 98-83
У76(02) - 83

© Издательство Ленинградского
университета, 1983 г.

Существовавшие до начала 60-х годов методы интерпретации геофизических наблюдений определялись спецификой используемых данных (метод характерных точек в гравия- и магниторазведке, метод Герглоца - Вихерта в сейсмологии и т.д.). Применение их для интерпретации относительно небольшого объема данных вполне могло осуществляться при отсутствии вычислительной техники, однако позволяло получать лишь грубую количественную информацию об изучаемом объекте.

В последние два десятилетия в связи с резким возрастанием объема геофизических наблюдений, обусловленным внедрением новой аппаратуры, расширением сети наблюдательных станций и интенсивным применением геофизических методов в практике разведки полезных ископаемых, изменились возможности и соответственно возросли требования к результатам. Одновременно появилась возможность обрабатывать все возрастающий поток данных на ЭВМ. Это привело к появлению новых методов интерпретации, использующих достижения современной вычислительной математики и математической статистики. Эти методы, в математической постановке сводящиеся к решению обратных задач, сейчас уже могут быть выделены в отдельный раздел геофизики.

На начальной стадии методы решения обратных задач развивались как методы перебора вариантов строения среды, в процессе которого отбирались такие, которые удовлетворяли бы некоторому критерию оптимизации. Поскольку возросли требования к качеству решения (точности, детальности), развиваются и совершенствуются сами критерии оптимизации, использующие выводы математической статистики (Л.А.Халфин, Ф.М.Гольцман, С.В.Гольдин, Т.Б.Калинина). Проблемы, связанные с неединственностью и

неустойчивостью решения обратных геофизических задач, привели к тому, что в геофизике стали широко применяться методы, разработанные в математике для решения некорректных задач (А.Н.Тихонов, М.М.Лаврентьев, В.Н.Страхов, В.Б.Гласко, В.И.Старостенко).

Важный этап в развитии теории обратных задач связан с именами Дж.Бэйкуса и Ф.Гильберта, разработавшими линейную теорию решения обратных задач при использовании ограниченного набора данных. Эти методы, начиная с середины 60-х годов, широко применяются за рубежом для решения различных обратных задач сейсмологии, гравиметрии и магнитовариационного зондирования. Метод Бэйкуса - Гильберта и оптимизационные методы, основанные на использовании метода наименьших квадратов в линейных задачах, оказалось возможным обобщить в рамках метода псевдообращения, разработанного в линейной алгебре на основе сингулярного разложения прямоугольных матриц.

Таким образом, теперь уже можно говорить о том, что методы решения обратных задач оформились в отдельный раздел геофизики, не связанный с каким-то определенным геофизическим методом. В то же время по сложившейся традиции методы решения обратных задач включаются в разные курсы по отдельным геофизическим дисциплинам (сейсмологии, геоэлектрики, гравитационной разведке), что приводит к ненужному параллелизму. При этом чем дальше, тем больше оказывается общих положений и методов, содержащихся в разных курсах. Поэтому сейчас назрела необходимость выделить теорию обратных задач в отдельную дисциплину, сохранив при этом, конечно, в каждом отдельном курсе методы, специфические для данной области геофизики, такие, как, например, методы аналитического продолжения потенциальных полей в задачах гравитационной и магниторазведки, τ -метод в сейсмологии, методы продолжения волновых полей в задачах сейсморазведочной интерпретации и т.д.

Отсутствие в отечественной литературе руководств, в которых были бы достаточно полно изложены и обобщены все созданные

ные к настоящему времени методы решения обратных геофизических задач, и привело к необходимости создания настоящей книги.

Данная книга создана на основе отдельных разделов курсов сейсмологии и геоэлектрики, читаемых авторами на кафедре физики Земли Ленинградского университета, и предназначена в качестве учебного пособия для студентов старших курсов и аспирантов по указанным геофизическим специализациям, а также по специализации грави- и магниторазведка. При этом авторы ставили перед собой цель изложить существующие методы независимо от конкретных задач, а затем показать их применимость при интерпретации данных, получаемых в разных областях геофизики.

При отборе материала, включенного в данное пособие, авторы исходили из необходимости отразить, с одной стороны, наиболее распространенные и хорошо разработанные методы, а с другой — также, которые хотя и не применяются широко, но представляются достаточно перспективными. Кроме того, авторам казалось важным изложить также методы, которые пока недостаточно освещены в геофизической литературе. Так, в настоящее время начинает входить в практику геофизической интерпретации метод псевдообращения, хорошо известный в линейной алгебре, но мало знакомый геофизикам. В связи с этим методу псевдообращения посвящена в настоящей книге отдельная глава (V). Подробно излагается (гл. IV) метод Бейкуса — Гильберта, широко применяемый за рубежом, но мало известный в нашей стране. Пока не получили распространения методы решения обратной задачи при наличии ограничений (§ 9, 10 гл. II), они слабо освещены в специальной геофизической литературе, однако есть все основания считать, что в недалеком будущем они найдут широкое применение. Изложенное позволяет надеяться, что данная книга может оказаться полезной не только студентам и аспирантам, но и специалистам-геофизикам, занимающимся интерпретацией геофизических наблюдений.

Материал в книге излагается с учетом того, что читатель знаком с элементами линейной алгебры, теорией вероятностей и математической статистики. Правда, поскольку самый большой раздел книги посвящен методам, основанным в значительной степени

на использовании выводов и следствий математической статистики (гл. II), для лучшего усвоения материала авторы изложили в § I этой главы основные положения теории статистических выводов.

Цитируемая литература содержит как оригинальные статьи, так и публикации обзорного характера, которые могут быть использованы в качестве дополнительной литературы, рекомендуемой при изучении материала данной книги. (Такая литература отмечена в списке звездочкой.)

Главы I, II-V написаны Т.Б.Яновской, глава VI - Л.Н.Пороховой при участии Т.Б.Яновской. Примеры, иллюстрирующие излагаемые методы, из области геоэлектрики подготовлены Л.Н.Пороховой, из области сейсмологии - Т.Б.Яновской. Для облегчения усвоения материала приводимые примеры максимально упрощены; читателю надо, конечно, иметь в виду, что практические задачи значительно сложнее.

Авторы благодарят докторов физ.-мат. наук А.Л.Левшина, С.В.Гольдина и И.А.Ибрагимова за критические замечания, полезные советы и дискуссии, способствовавшие улучшению содержания рукописи, а также М.М.Харламова и А.В.Юрина за помощь при выполнении расчетов.

§ I. Понятие об обратных задачах геофизики

Геофизика представляет комплекс физических наук, охватывающий широкий круг исследований — начиная с изучения физических свойств вещества Земли и их пространственного распределения, на основании которого делаются заключения о строении Земли, кончая изучением происходящих в Земле физических и физико-геологических процессов. При этом если даже ограничиться только той частью геофизики, которая занимается исследованием пространственного распределения физических свойств вещества, то мы столкнемся с многообразием задач, относящихся к разным, мало связанным разделам этой науки. Так, изучение распределения упругих свойств — предмет сейсмологии и сейсморазведки, электропроводности — предмет геоэлектрики, плотности — предмет гравиметрии и отчасти сейсмологии, магнитных свойств горных пород — предмет магнитометрии. Но все эти, казалось бы, мало связанные между собой задачи характеризуются общим методом исследования, который сводится к наблюдению на поверхности Земли (а также в скважинах и околоземном пространстве) тех или иных геофизических полей — сейсмических, электромагнитных, гравитационных, анализу структуры наблюдаемых полей и выводу заключений о глубинном распределении соответствующих физических параметров. Выводы о глубинном строении земных недр возможны благодаря тому, что наблюдаемые на поверхности Земли поля зависят от пространственного распределения свойств вещества в глубинных ее частях.

По заданному распределению тех или иных физических свойств теоретически можно предсказать ожидаемые значения соответствующих физических полей. Такие задачи, называемые прямыми задачами геофизики, имеют свою специфику в зависимости

от природы полей. Построение же распределения физических параметров по значениям поля в некоторой части среды представляет собой обратную задачу геофизики. При этом в разных областях - в сейсмологии, геоэлектрике, гравитационной и магнитометрии - методы решения обратной задачи имеют много общего. Эта общность позволяет выделить обратные задачи в отдельную дисциплину.

На ранних этапах развития геофизики обратные задачи ставились и решались весьма примитивно. По сути они сводились к выбору одной из двух или из небольшого числа моделей среды, для которой рассчитанные значения характеристик поля лучше, чем для других, согласовывались с соответствующими наблюдаемыми значениями. Например, в течение долгого времени дискутировался вопрос, какой из моделей скоростного разреза Земли - Джеффриса или Гутенберга - следует отдать предпочтение. С этой целью для таких двух моделей (и, как правило, только для них) рассчитывались различные характеристики сейсмического поля - годографы, дисперсионные кривые поверхностных волн, периоды собственных колебаний Земли, которые и сравнивались с наблюдениями. Недостатки такого подхода очевидны. Прежде всего нет никакой гарантии, что среди альтернативных моделей среды окажется именно та, которая отражает истинное строение Земли. Далее остаются неразрешенными вопросы: достаточна ли разрешающая способность исходных данных, чтобы различить выбранные модели; позволяют ли наблюдения произвести большую детализацию структуры; в каком смысле понимать лучшее или худшее согласие теоретических данных с наблюдениями; как погрешность наблюдений скажется на погрешности искомой модели среды; как наиболее оптимальным образом организовать поиск модели, отвечающей наблюдениям, и т.д.

Необходимость ответа на эти и другие вопросы и привела к разработке методов решения обратных задач. Этому же способствовала возможность использования в геофизике быстродействующих ЭВМ, ибо решение обратных задач требует чрезвычайно большого объема вычислений. Развитие методов решения обратных задач происходило по разным направлениям в зависимости от того, какому из аспектов придавалось большее значение. К настоящему времени можно уже сформулировать общие принципы постановки об-

ратных задач и требования, предъявляемые к их решению, определить место каждого из существующих методов и указать их взаимосвязь.

§ 2. Математическая постановка обратной задачи

В наиболее общем виде постановка обратных задач геофизики может быть сформулирована следующим образом.

Даны: значения характеристик геофизических полей, полученные из наблюдений либо путем непосредственного измерения, либо в результате обработки наблюдаемых процессов. Эти значения имеют, как правило, дискретный характер и образуют набор данных y_l ($l = 1, 2, \dots, n$). Будем обозначать их n -мерным вектором y . В отдельных случаях наблюдения y являются непрерывной функцией некоторого параметра l в конечном интервале изменения этого параметра. Для общности можно считать, что $y(l)$ есть вектор в бесконечномерном пространстве.

Наблюдаемые величины всегда отягощены случайными ошибками. В некоторых случаях бывает известны какие-то характеристики распределения ошибок (дисперсия, ковариационная матрица), которые можно также рассматривать как исходные данные при решении обратной задачи.

Известна связь рассматриваемых полей с теми или иными физическими параметрами Земли. Поскольку свойства вещества Земли меняются от точки к точке, то физические параметры, характеризующие соответствующие свойства, являются функциями координат. Будем обозначать их $m(r)$, где r — координата точки, а m — в общем случае вектор, компонентами которого являются разные физические параметры среды, определяющие рассматриваемое геофизическое поле. Вектор $m(r)$ будем называть для общности моделью Земли. Разумеется, в разных задачах будут изучаться разные модели, описывающие те или иные физические свойства Земли. Например, в задачах магнитотеллурического зондирования (МТЗ) под моделью понимают функцию $c(z)$, описывающую изменение удельной проводимости с глубиной; в сейсмологии, в задачах, используемых в качестве исходного материала дисперсионные кривые релеевских волн, моделью является совокупность трех функций — па-

раметров Ламе $\lambda(z)$, $\mu(z)$ и плотности $\rho(z)$, образующих трехмерный вектор. В каждом случае известен оператор, позволяющий по заданной модели $m(r)$ вычислять значения характеристик поля:

$$y = \Phi(m). \quad (1)$$

Вид этого оператора может быть задан аналитически, но в ряде случаев он определяется только алгоритмически — путем задания последовательности вычислительных операций, позволяющих по m рассчитать y . Оператор Φ называют оператором решения прямой задачи.

Кроме того, считаются известными априорные ограничения, наложенные на модель. Они могут быть самыми разнообразными: может быть фиксирован класс моделей (например, рассматриваются модели в классе сферически-симметричных или кусочно-постоянных функций); может быть известна функция $m(r)$ в некоторой части исследуемой области; могут быть известны пределы, внутри которых может находиться значение рассматриваемого физического параметра. В случае, когда вид функции $m(r)$ фиксирован, под моделью следует понимать совокупность параметров, определяющих данную функцию. При этом связь характеристик поля с моделью определяется функцией этих параметров.

Требуется извлечь информацию о модели, содержащуюся в заданной выборке наблюдений y . Что под этим подразумевается? Очевидно, что в идеале было бы желательно просто определить $m(r)$, т.е. построить оператор, обратный (I), который бы позволял найти $m(r)$ по y . Однако поскольку, как правило, количество наблюдений ограничено, то определить полностью функцию $m(r)$ принципиально невозможно. Поэтому можно ставить задачу лишь об определении некоторых осредненных характеристик модели. А так как исходные данные всегда содержат случайные ошибки, то даже такие характеристики не удастся получить точно, можно найти только их оценки.

Приведем несколько примеров постановки обратных задач.

Пример I. Определение скоростного разреза по годографу сейсмической волны. Из наблюдений получают набор значений времен пробега волны t_i , соответствующих эпицентральному расстоянием Δ_i :

$$t_i = t(\Delta_i), \quad i = 1, 2, \dots, n, \quad \Delta = |\mathbf{r}_2 - \mathbf{r}_1|.$$

Время пробега волны определяется только распределением скорости соответствующей волны в среде $v(\mathbf{r})$. В общем случае связь $t(\mathbf{r}_2, \mathbf{r}_1)$ и $v(\mathbf{r})$ определяется дифференциальным уравнением луча. Однако если время пробега является только функцией расстояния между источником и точкой приема и не зависит от их координат $\mathbf{r}_1, \mathbf{r}_2$, то скорость зависит только от одной координаты: в случае сферической модели Земли это $v(r)$, в приближении уплощенной модели — $v(z)$. В этом состоит априорное ограничение на модель.

Для сферически-симметричной модели географ $t(\Delta)$ связан с распределением скорости $v(r)$ параметрически:

$$t(p) = 2 \int_{r_m}^R \frac{dr}{v(r) \sqrt{1 - p^2 v^2 / r^2}}, \quad \Delta(p) = 2p \int_{r_m}^R \frac{v(r) dr}{r^2 \sqrt{1 - p^2 v^2 / r^2}}, \quad (2)$$

где p — параметр луча. Параметрическое задание функции $t(\Delta)$ в виде (2) определяет оператор решения прямой задачи (1). Обратная задача состоит в том, чтобы оценить $v(r)$ по заданной выборке наблюдений t_i, Δ_i .

Пример 2. Нахождение распределения упругих параметров по дисперсии групповой скорости поверхностной волны. Непосредственно измерить зависимость групповой скорости от периода — $u(T)$ — нельзя. Для нахождения $u(T)$ используется либо ручная обработка сейсмограмм, в результате которой по измеренным значениям времен экстремумов колебаний t_i определяются скорости $u_i = \Delta / t_i$ и соответствующие этим скоростям периоды $T_i = t_{i+1} - t_{i-1}$, либо спектрально-временной анализ (СВАН) [26], преобразующий сейсмограмму в спектрально-временную диаграмму, по максимуму которой можно построить кривую $u(T)$ в некотором диапазоне периодов (рис.1). Таким образом, исходные наблюдения в обоих случаях подвергаются предварительной обработке. Но в первом случае получаемые данные представляют дискретный ряд значений u_i, T_i , а во втором — некоторый отрезок кривой $u(T)$. Фазовая скорость (c) поверхностной волны (для простоты будем считать, что это волна Лява) определяется в результате нахождения собственных значений следующей краевой задачи [26]:

$$\frac{d}{dz} \left(\mu(z) \frac{dV}{dz} \right) + (\omega^2 \rho(z) - \xi^2 \mu(z)) V = 0, \quad (3)$$

$$\mu = \frac{dV}{dz} \Big|_{z=0} = 0, \quad V \Big|_{z \rightarrow \infty} = 0.$$

Здесь $V(z)$ — распределение амплитуды волны с глубиной, ω — круговая частота $\omega = 2\pi/T$, ξ — волновое число, $\xi = \omega/c(\omega)$. При заданном волновом числе круговая частота определяется как собственное значение краевой задачи (3), откуда и находится фазовая скорость: $c(\xi) = \omega/\xi$, а по ней и групповая: $u = c + \xi \frac{dc}{d\xi}$ (рис.2).

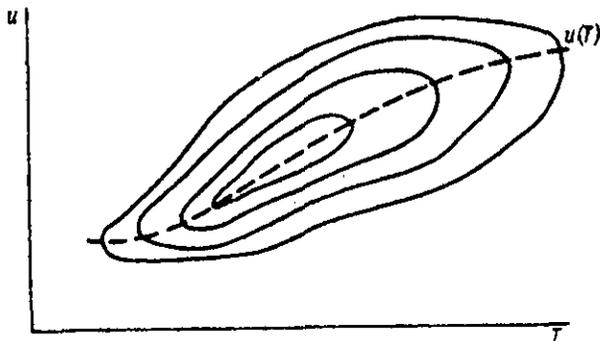


Рис. 1.

Из формулы (3) видно, что скорость зависит от распределения с глубиной модуля сдвига μ и плотности ρ . Таким образом, в данном случае модель — это набор двух функций $\mu(z)$ и $\rho(z)$, и обратная задача состоит в оценке этих функций по наблюдаемым значениям u_i , T_i или отрезку кривой $u(T)$.

При решении этой задачи часто используют априорное ограничение на модель: среду считают слоисто-однородной, содержащей фиксирован-

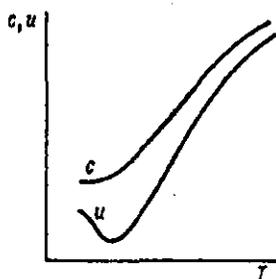


Рис.2.

ное число слов. В этом случае нахождение неизвестных распределений $\mu(x)$ и $\rho(x)$ сводится к оценке мощностей слов и значений μ_i и ρ_i в этих словах.

Пример 3. Построение отражающей границы по годографу отраженной волны. На основе сейсмических исследований построен продольный годограф отраженной волны $l(x)$. Пусть априори известно, что профиль наблюдений направлен по падению отражающей границы, а скорость распространения волны в среде можно считать постоянной. Уравнение годографа $l(x)$ в этом случае имеет вид

$$v^2 l^2 = (2h \sin \alpha + x)^2 + 4h^2 \cos^2 \alpha.$$

Неизвестными элементами среды (или моделью) в этом случае будут скорость v и положение отражающей границы, определяемые глубиной под источником (h) и углом падения (α).

Пример 4. Определение удельного сопротивления по данным магнитотеллурического зондирования. Магнитотеллурические наблюдения позволяют определять импеданс $Z(\omega)$ путем измерения компонент естественного электромагнитного поля $-E_x$, E_y , H_x , H_y в некоторой точке дневной поверхности. Для нахождения значений импеданса проводится обработка магнитотеллурической программы методом видимых амплитуд либо путем цифровой фильтрации на ЭВМ. Это дает возможность определить амплитуды и фазы компонент поля на разных частотах. Если среду в окрестности точки наблюдения можно считать горизонтально однородной, то входной импеданс вычисляется как отношение взаимно перпендикулярных компонент электрического и магнитного полей:

$$Z = E_x / H_y = -E_y / H_x.$$

Для решения обратной задачи используются амплитуда и фаза комплексного импеданса $Z(\omega) = |Z(\omega)| e^{i\varphi(\omega)}$ в некотором конечном интервале частот.

Импеданс $Z(\omega, z)$ в полупространстве $z > 0$, характеризующемся распределением проводимости $\sigma(z)$, определяется из решения уравнения

$$Z'(\omega, z) = Z^2(\omega, z) \sigma(z) + i\omega \mu_0 \quad (4)$$

с граничными условиями непрерывности импеданса на границах

разрыва $G(z)$ и предельным условием $Z \rightarrow 0$ при $z \rightarrow \infty$. Уравнение (4) является уравнением Риккати и в общем случае - для произвольной функции $G(z)$ - для него не может быть получено решение в явном виде. Но для случая кусочно-постоянной функции $G(z)$, т.е. для слоисто-однородной структуры, решение $Z(\omega, z)$ может быть легко построено, и соответственно может быть получено выражение для входного импеданса $Z = Z(\omega, 0)$. Если параметрами структуры являются мощности слоев h_s ($s = 1, 2, \dots, N-1$), и удельное сопротивление в них $\rho_s = G_s^{-1}$ ($s = 1, 2, \dots, N$), то зависимость входного импеданса от параметров структуры определяется соотношением

$$Z = Z_1 R_N,$$

где $Z_1 = i\omega\mu_0/k_1$ - входной импеданс однородной среды с удельным сопротивлением ρ_1 ; μ_0 - магнитная проницаемость среды; k_1 - волновое число; R_N определяется при помощи следующего рекуррентного соотношения [22]:

$$R_{N-s} = \text{cth} \{ k_{s+1} h_{s+1} + \text{arcth} \sqrt{\rho_{s+1}/\rho_s} R_{N-(s+1)} \}, \quad (5)$$

в котором $R_1 = 1$, $k_s = 2\pi/(\sqrt{10\rho_s T}) \exp(-i\pi/4)$. Обратная задача МГЗ заключается в оценке параметров среды по значениям $Z(\omega)$.

§ 3. Линеаризация оператора прямой задачи

Из приведенных в предыдущем параграфе примеров видно, что формы оператора решения прямых задач геофизики весьма разнообразны. Большинство же методов решения обратных задач, как мы увидим далее, разработано для случая, когда оператор прямой задачи Φ линеен. Однако в геофизике линейность оператора Φ имеет место только в задачах гравиметрии относительно плотности и в магнитометрии относительно компонент вектора намагниченности. Действительно, связь ускорения свободного падения $g(\mathbf{x})$ с распределением плотности в среде $\rho(\mathbf{r})$ определяется интегралом

$$g(\mathbf{x}) = \gamma \iiint \frac{\rho(\mathbf{r}) \cos(\mathbf{r} - \mathbf{x}, \mathbf{e}_z)}{|\mathbf{r} - \mathbf{x}|^2} dV,$$

где γ — гравитационная постоянная. В случае, когда среда условно разбита на блоки с постоянными значениями плотности в них,

$$g(x) = \sum_i G_i(x) \rho_i.$$

Но если модель распределения плотности характеризовать положением границы $z(x)$, разделяющей слой разной плотности (а такая постановка задачи довольно распространена в практике интерпретации гравиметрических данных), то относительно $z(x)$ оператор прямой задачи уже не будет линейным.

Если же обратную задачу формулировать не как поиск неизвестных параметров m , относительно которых оператор решения прямой задачи нелинеен (он может быть даже вообще не представим в аналитической форме), а как поиск поправок Δm к некоторому начальному приближению $m^{(0)}$, то связь Δm и отклонения наблюдений y от решения, соответствующего модели $m^{(0)}$, приближенно будет линейной. Приближение это тем лучше, чем ближе начальная модель к истинному решению, т.е. чем меньше Δm . Для нахождения решения m с заданной точностью можно использовать метод последовательных приближений: вначале определить $\Delta m^{(1)}$, по которому находится первое приближение $m^{(1)} = m^{(0)} + \Delta m^{(1)}$, затем полученное приближение $m^{(1)}$ принять в качестве начального и определять поправку к нему и т.д. до сходимости процесса, т.е. пока поправка $\Delta m^{(n)}$ не станет достаточно малой. При этом на каждом шаге процесса должна решаться линейная обратная задача

$$A \Delta m^{(s)} = \Delta y^{(s-1)} = y - \Phi(m^{(s-1)}). \quad (6)$$

В случае, когда пространство моделей M является конечномерным, A в формуле (6) представляет собой матрицу с элементами $A_{ik} = \left(\frac{\partial \Phi_i}{\partial m_k} \right)_{m = m^{(s-1)}}$, а если M является функциональным пространством, то (6) есть интегральное преобразование вида $\int G_i(r) \Delta m(r) dr = \Delta y_i$, $i = 1, 2, \dots, n$.

Описанный подход аналогичен решению трансцендентных уравнений $F(x) = 0$ итеративным методом Ньютона, в котором решение на j -й итерации определяется через решение на предыдущей итерации по формуле

$$x^{(j)} = x^{(j-1)} - \frac{F(x^{(j-1)})}{F'(x^{(j-1)})}.$$

Сведение оператора прямой задачи произвольного вида к линейному оператору (6) называется линеаризацией задачи.

Как уже говорилось, основное преимущество линеаризованной обратной задачи заключается в том, что для линейных задач существуют хорошо разработанные методы и алгоритмы. Но, кроме того, оказывается, что в тех случаях, для которых оператор решения прямой задачи относительно исходной модели не может быть описан только алгоритмически, можно построить оператор преобразования поправок к модели Δw в поправки к наблюдениям Δy в виде явной функциональной зависимости.

Покажем, как производится линеаризация оператора прямой задачи на примерах, рассмотренных в § 2.

Годограф отраженной волны. Для этого примера линеаризация выполняется наиболее просто. Наблюдаемая величина — время прихода отраженной волны в точку x — имеет простое аналитическое выражение через основные параметры h , v , α . Пусть их начальные значения будут h_0 , v_0 , α_0 , а годограф для модели с такими параметрами будет $l_0(x)$, причем $\Delta h = h - h_0$, $\Delta v = v - v_0$, $\Delta \alpha = \alpha - \alpha_0$. Тогда

$$\Delta l(x) = l(x) - l_0(x) = a \Delta h + b \Delta v + c \Delta \alpha,$$

$$\text{где } a = \frac{4h_0 + 2x \sin \alpha_0}{v_0 \sqrt{(2h_0 \sin \alpha_0 + x)^2 + 4h_0^2 \cos^2 \alpha_0}},$$

$$b = -\sqrt{(2h_0 \sin \alpha_0 + x)^2 + 4h_0^2 \cos^2 \alpha_0} / v_0^2,$$

$$c = \frac{2h_0 x \cos \alpha_0}{v_0 \sqrt{(2h_0 \sin \alpha_0 + x)^2 + 4h_0^2 \cos^2 \alpha_0}}.$$

Фазовая скорость волны Лява. Ограничимся только рассмотрением случая, когда наблюдаемой величиной является фазовая скорость волны Лява. Для групповой скорости вывод выражения, связывающего вариацию скорости и вариацию параметров модели, значительно более сложен.

Как было показано в § 2, задача об определении фазовой скорости волны Лява $s(\omega) = \omega / \xi(\omega)$, соответствующей заданным

распределениям модуля сдвига $\mu(z)$ и плотности $\rho(z)$ сводится к нахождению собственных значений следующей краевой задачи:

$$\frac{d}{dz} \left(\mu(z) \frac{dV}{dz} \right) + (\rho \omega^2 - \mu \xi^2) V = 0, \quad (7)$$

$$\mu \frac{dV}{dz} \Big|_{z=0} = 0, \quad V \Big|_{z \rightarrow \infty} = 0.$$

Нахождение собственных значений ω по заданному значению волнового числа ξ — сложная вычислительная задача. Она является типичным примером такой, где связь между наблюдениями и моделью $\mu(z)$, $\rho(z)$ не может быть представлена в виде явной функциональной зависимости. В то же время зависимость поправки к скорости от малых вариаций упругих параметров $\Delta\mu(z)$ и $\Delta\rho(z)$ определяется в явном виде.

Из самосопряженности оператора (7) вытекает следующее интегральное соотношение [26]:

$$\omega^2 \int_0^{\infty} \rho V^2 dz = \frac{\omega^2}{c^2} \int_0^{\infty} \mu V^2 dz + \int_0^{\infty} \mu \left(\frac{dV}{dz} \right)^2 dz. \quad (8)$$

Рассматривая вариации левой и правой части выражения (8), нетрудно прийти к следующему соотношению, связывающему вариацию скорости Δc с вариациями модели:

$$2\Delta c(\omega) \int_0^{\infty} \mu_0 V_0^2 dz = c_0 \int_0^{\infty} \Delta\mu \left[\frac{c_0^2}{\omega^2} \left(\frac{dV_0}{dz} \right)^2 + V_0^2 \right] dz - c_0^3 \int_0^{\infty} \Delta\rho V_0^2 dz. \quad (9)$$

Здесь собственная функция краевой задачи $V_0(z)$ и фазовая скорость волны Лява c_0 соответствуют начальному распределению параметров $\mu_0(z)$, $\rho_0(z)$. Иначе формулу (9) можно записать в виде

$$\Delta c = \int_0^{\infty} G^{(1)}(z) \Delta\mu(z) dz + \int_0^{\infty} G^{(2)}(z) \Delta\rho(z) dz,$$

где

$$G^{(1)}(z) = c_0 \left[\frac{c_0^2}{\omega^2} \left(\frac{dV_0}{dz} \right)^2 + V_0^2 \right] \left(2 \int_0^{\infty} \mu_0 V_0^2 dz \right)^{-1},$$

$$G^{(2)}(z) = -V_0^2 c_0^3 \left(\int_0^{\infty} \mu_0 V_0^2 dz \right)^{-1}.$$

Годограф рефрагированной волны. Время пробега сейсмической волны определяется функционалом Ферма $l(P, P_0) = \int_{P_0}^P ds/v(r)$, где интегрирование выполняется вдоль экстремали (луча). Если $v = v(r)$, а точки P и P_0 находятся на поверхности Земли ($r = R$) на расстоянии Δ , то зависимость $l(\Delta)$ (годограф) определяется в параметрической форме (2).

Пусть распределение скорости в среде $v(r)$ отличается от некоторого заданного $v_0(r)$ на малую величину $\Delta v(r)$. Тогда время пробега волны между точками P и P_0 будет отличаться от того, которое соответствует распределению $v_0(r)$, на величину, определяемую интегралом

$$\Delta l = - \int_{P_0}^P \frac{\Delta v(r)}{v_0^2(r)} ds, \quad (10)$$

в котором интегрирование выполняется вдоль луча, соответствующего $v_0(r)$. Вариация самого луча, обусловленная отличием $v(r)$ от $v_0(r)$, дает вклад в поправку к значению времени пробега второго порядка малости по сравнению с интегралом (10). Таким образом, поправка к годографу $l_0(\Delta)$, соответствующему скоростному разрезу $v_0(r)$, определяется выражением

$$\Delta l(\Delta) = -2 \int_{r_m}^R \frac{\Delta v(r) dr}{v_0^2(r) \sqrt{1 - p_0^2 v^2/r^2}}.$$

В этом выражении p_0 — параметр луча, выходящего на эллиптическом расстоянии Δ в среде, характеризующейся скоростным разрезом $v_0(r)$. Таким образом, поправка к годографу является линейным функционалом $\Delta v(r)$. Ядро этого функционала в данном случае разрывное и имеет сингулярность в точке $r = r_m$:

$$G(r) = \begin{cases} -2(v_0^2 \sqrt{1 - p_0^2 v^2/r^2})^{-1}, & r_m < r \leq R, \\ 0, & r < r_m. \end{cases}$$

Задача магнитотеллурического зондирования. Как уже упоминалось в § 2, импеданс $Z(\omega, z)$ является решением дифференциального уравнения Риккати, к в общем случае для него нельзя

получить явное выражение. Но если известно решение $Z_0(\omega, z)$ для некоторого начального распределения проводимости $\sigma_0(z)$, то для структуры $\sigma(z)$, отличающейся от $\sigma_0(z)$ малой поправкой $\Delta\sigma(z)$, можно определить входной импеданс в виде

$$Z(\omega) = Z_0(\omega, 0) + \Delta Z(\omega),$$

так что $\Delta Z(\omega)$ будет выражаться линейно через $\Delta\sigma(z)$. Покажем это.

Уравнение для $Z_0(\omega, z)$ имеет вид [7]

$$Z_0' = Z_0^2 \sigma_0 + i\omega\mu_0, \quad (11)$$

а для $Z = Z_0 + \Delta Z$

$$Z_0' + \Delta Z' = (Z_0 + \Delta Z)^2 (\sigma_0 + \Delta\sigma) + i\omega\mu_0. \quad (12)$$

Вычитая (11) из (12) и ограничиваясь членами первого порядка малости, получим дифференциальное уравнение относительно $\Delta Z'$:

$$\Delta Z' + p(z)\Delta Z + Q(z) = 0,$$

$$\Delta Z \rightarrow 0 \text{ при } y \rightarrow \infty,$$

где $p = -2Z_0(\omega, z)\sigma_0(z)$, $Q = -Z_0^2(\omega, z)\Delta\sigma(z)$.

Решение этого уравнения имеет вид

$$\Delta Z(\omega, z) = -\exp\left[-\int_{\infty}^z p(\eta) d\eta\right] \int_{\infty}^z Q(\eta) \exp\left(\int_{\infty}^{\eta} p(\xi) d\xi\right) d\eta. \quad (13)$$

Поправка к входному импедансу $\Delta Z(\omega)$ получается, если в выражении (13) положить $z=0$:

$$\Delta Z(\omega) = \int_0^{\infty} Q(x) \exp\left(\int_0^x p(\xi) d\xi\right) dx = -\int_0^{\infty} Z_0^2 \exp\left(-2\int_0^x Z_0 \sigma_0 d\xi\right) \Delta\sigma(x) dx,$$

т.е. она записывается в виде интегрального преобразования $\Delta\sigma(x)$ с ядром

$$G(\omega, x) = -Z_0^2(\omega, x) \exp\left(-2\int_0^x Z_0(\omega, \xi) \sigma_0(\xi) d\xi\right).$$

§ 4. Некорректность обратных задач

Наблюдения y можно рассматривать как элемент пространства \mathcal{Y} , а модель m — как элемент пространства \mathcal{M} (в общем слу-

чае эти пространства могут быть бесконечномерными). Оператор Φ преобразует пространство M в пространство Y . Такое преобразование является однозначным: какую бы ни взяли модель, ей можно сопоставить вполне определенное геофизическое поле (при условии, конечно, что модель достаточна для построения поля). Обратная задача может рассматриваться как построение обратного оператора Φ^{-1} , преобразующего пространство Y в пространство M . Будет ли это преобразование однозначным? Иначе говоря, будет ли заданному геофизическому полю отвечать только одна определенная модель? Известно, что в общем случае это не так. Например, одному и тому же полю силы тяжести, измеренному на поверхности Земли, могут соответствовать разные распределения плотности вещества внутри Земли, тогда как прямая задача — нахождение гравитационного поля по заданному распределению плотности — является однозначной. Другой пример — восстановление скоростного разреза по годографу при наличии на нем зоны тени, обусловленной существованием в среде волновода. Внутри волновода и ниже его распределение скорости восстанавливается по годографу неоднозначно.

В большинстве случаев неединственность решения обратной задачи связана с неполнотой реальных геофизических данных: измерения обычно проводятся в дискретных точках, на ограниченной площади или в конечном интервале изменения некоторого параметра. Например, известно, что обратная задача геоэлектрики имеет единственное решение, если исходные данные — значения кажущегося сопротивления в зависимости от периода волны — определены в интервале периодов от 0 до ∞ . Но наблюдения всегда удается получить лишь в некотором конечном интервале периодов. А это приводит к неединственности решения.

Если сказанное изобразить схематически, то точке m из пространства M соответствует одна единственная точка из пространства Y , а точке y соответствует некоторая область M из пространства M , причем в общем случае она может даже не быть связанной (рис. 3). Как же подходить к решению обратной задачи при такой ситуации? Для геофизических построений важно получить область решений, соответствующих заданным наблюдениям. Но как описать компактно эту область? Очевидно, что желательнее

найти такие характеристики, которые определяли бы все модели, относящиеся к области решения. Сделать это можно двояко. Один способ заключается в априорном осреднении модели, т.е. вместо истинного распределения $m(r)$ заранее определяются некоторые его осредненные свойства. Обычно это осуществляется путем разделения среды на слои и рассмотрения средних значений искоемых физических параметров в каждом слое. В грави- и магниторазведке вместо поиска распределения аномальных масс часто ограничиваются нахождением некоторой "средней" массы, совмещенной с центром тяжести аномального тела. При таком подходе в принципе можно добиться взаимно-однозначного соответствия между моделью и наблюдениями, но это не так просто. Дело в том, что при недостаточном осреднении мы будем продолжать оставаться в рамках описанной ситуации, когда одной точке из пространства Y соответствует область из пространства M . А при слишком сильном осреднении может возникнуть другая опасность: заданной точке из пространства Y не будет соответствовать вообще ни одной точки из пространства M . Примером может быть интерпретация голографа преломленной волны на криволинейной границе в рамках модели среды с плоской границей. Для такой модели голограф является прямой линией, в то время, как голограф, полученный из наблюдений, обязательно криволинеен. Пространство наблюдений Y в этом случае образовано совокупностью времен вступлений волн t_i ($i = 1, 2, \dots, n$) в последовательных точках профиля. При этом точки, соответствующие любой модели из выбранного класса, должны быть расположены на некоторой плоскости, принадлежащей пространству Y , а точка, отвечающая наблюдениям, будет находиться вне этой плоскости. Таким образом, при решении обратной задачи мы можем столкнуться как с неединственностью решения, так и с отсутствием точного решения.

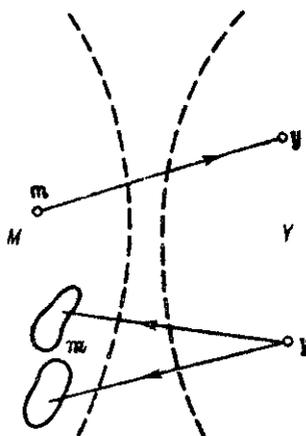


Рис. 3.

осреднении мы будем продолжать оставаться в рамках описанной ситуации, когда одной точке из пространства Y соответствует область из пространства M . А при слишком сильном осреднении может возникнуть другая опасность: заданной точке из пространства Y не будет соответствовать вообще ни одной точки из пространства M . Примером может быть интерпретация голографа преломленной волны на криволинейной границе в рамках модели среды с плоской границей. Для такой модели голограф является прямой линией, в то время, как голограф, полученный из наблюдений, обязательно криволинеен. Пространство наблюдений Y в этом случае образовано совокупностью времен вступлений волн t_i ($i = 1, 2, \dots, n$) в последовательных точках профиля. При этом точки, соответствующие любой модели из выбранного класса, должны быть расположены на некоторой плоскости, принадлежащей пространству Y , а точка, отвечающая наблюдениям, будет находиться вне этой плоскости. Таким образом, при решении обратной задачи мы можем столкнуться как с неединственностью решения, так и с отсутствием точного решения.

Более рациональным представляется выбрать способ определения характеристик модели непосредственно на основе используемых наблюдений. Поскольку каждая совокупность наблюдений определяет соответствующую ей область в пространстве моделей, то и способ осреднения моделей должен определяться исходными наблюдениями.

Оба указанных способа осреднения будут подробно изложены в последующих главах.

Перейдем теперь к рассмотрению устойчивости решений обратной задачи. Для этого необходимо определить метрику пространства. До сих пор мы говорили о пространстве наблюдений и о пространстве моделей как о некотором множестве элементов, не определяя "близость" этих элементов между собой. Если введено понятие расстояния $\rho(\mathbf{y}_1, \mathbf{y}_2)$ (метрика) между элементами $\mathbf{y}_1, \mathbf{y}_2$ до такого пространства называется метрическим. Часто используют квадратичную метрику (L_2), в которой расстояние ρ между элементами* \mathbf{y}_1 и \mathbf{y}_2 определяется как $\rho^2(\mathbf{y}_1, \mathbf{y}_2) = \sum_i (y_{2i} - y_{1i})^2$ или, если V - функциональное пространство, то

$$\rho^2(\mathbf{y}_1, \mathbf{y}_2) = \int_{x_1}^{x_2} (y_2(x) - y_1(x))^2 dx.$$

Другой пример метрического пространства - пространство с равнмерной метрикой (C), в котором расстояние между элементами определяется по формуле

$$\rho(\mathbf{y}_1, \mathbf{y}_2) = \max_i |y_{2i} - y_{1i}|$$

$$\text{или } \rho(\mathbf{y}_1, \mathbf{y}_2) = \max_{x \in [x_1, x_2]} |y_2(x) - y_1(x)|.$$

Будем обозначать расстояние между элементами метрического пространства через $\|\mathbf{y}_2 - \mathbf{y}_1\|$.

* Здесь и далее предполагается, что компоненты вектора имеют одинаковую размерность и одинаковый масштаб; в § 8 гл. IV описан способ приведения к такому виду векторов $\mathbf{y} \in V$ и $\mathbf{t} \in M$.

Оператор преобразования одного пространства в другое ($X \rightarrow Y$) называется устойчивым, если малые изменения в исходном пространстве X (в смысле расстояния между элементами) приводят к малым изменениям в пространстве Y . Строго понятие устойчивости оператора Φ , преобразующего пространство X в пространство Y , определяется так: по любому сколь угодно малому значению $\epsilon > 0$ всегда найдется такое $\delta > 0$, что для любых элементов из пространства X , удовлетворяющих условию $\|x_2 - x_1\| < \delta$, выполняется условие $\|y_2 - y_1\| < \epsilon$, где $y_i = \Phi(x_i)$.

Оператор решения прямой задачи, как правило, устойчив. Это обусловлено тем, что все рассматриваемые геофизические поля — сейсмическое, гравитационное, электромагнитное — устойчивы по отношению к малым флуктуациям параметров среды. Однако существуют и исключения. Они могут быть вызваны двумя причинами. Во-первых, мы обычно рассматриваем не полностью то или иное поле, а некоторые его характеристики. А характеристики могут быть выбраны так, что при некотором соотношении параметров среды они становятся неустойчивыми. Например, годограф сейсмической волны, отвечающий скоростному разрезу $v(x)$, оказывается неустойчивым в тех случаях, когда в какой-то точке $\frac{dv}{dx} = 0$, в этом случае на годографе возникает бесконечно протяженная петля. В то же время сейсмическое поле в целом, отвечающее такому разрезу, остается устойчивым, поскольку амплитуды волн, соответствующих этой петле, будут бесконечно малы. Вторая причина неустойчивости оператора решения прямой задачи связана с тем, что для построения такого оператора часто используются приближенные или асимптотические методы, которые в некоторых областях могут быть вообще неприменимы либо дают малую точность. Например, если для расчета амплитудной кривой сейсмической волны используется лучевое приближение, то в окрестности каустики или зоны тени при решении прямой задачи мы получим разрывы амплитуды, которые в действительности не имеют места.

Эти особенности оператора решения прямой задачи всегда надо иметь в виду при решении обратных задач и исключать из рассмотрения области неустойчивости. В дальнейшем мы будем предполагать, что оператор решения прямой задачи всегда устойчив.

Оператор же решения обратной задачи Φ^{-1} , как правило, оказывается неустойчивым. Это значит, что малые изменения в исходных данных \mathbf{y} могут приводить к сколь угодно большому изменению в модели \mathbf{m} . Но, как уже упоминалось при описании постановки обратной задачи, наблюдения \mathbf{y} всегда имеют некоторую ошибку, так что "истинные" значения характеристик поля всегда находятся в некоторой окрестности наблюдаемых.

Таким образом, наблюдения в действительности определяют не точку в пространстве Y , а некоторую область. Поэтому даже в случае, когда обратная задача имеет единственное решение, наблюдения, осложненные ошибками, будут определять область в пространстве M . Если оператор решения обратной задачи неустойчив, то эта область может включать в себя сильно различающиеся модели. При этом решение, отвечающее наблюдениям с погрешностью, может весьма существенно отличаться от истинного. Это схематически иллюстрирует рис. 4. Построим в пространстве Y в окрестности точки \mathbf{y} , отвечающей наблюдениям, малую область, размеры которой определяются погрешностью наблюдений.

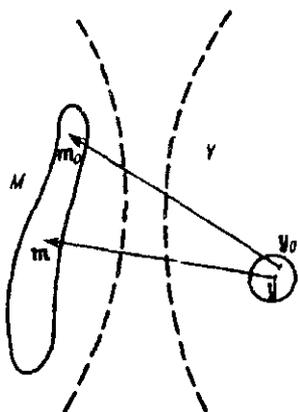


Рис. 4.

она включает в себя точку \mathbf{y}_0 , соответствующую точным данным без погрешности. Эта область преобразуется обратным оператором в достаточно протяженную область в пространстве M , и при этом точка \mathbf{m} , соответствующая наблюдениям \mathbf{y} , будет сильно смещена относительно точки \mathbf{m}_0 , отвечающей истинной модели. Приблизиться к истинному решению можно, только введя определенные априорные ограничения на модель.

Перечисленные особенности оператора решения обратной задачи (неединственность решения, возможное отсутствие точного решения, неустойчивость) относят обратные задачи к классу некорректных.

§ 5. Параметризация модели

В § 2, 3 уже упоминалось о том, что неоднозначность решения обратной задачи приводит к необходимости определять не искомую модель $m(\mathbf{r})$ полностью, а только некоторые ее осредненные характеристики. Степень осреднения модели зависит от того, какие исходные данные взяты для решения обратной задачи и какими они характеризуются погрешностями. Например, если исходными данными для определения скоростей упругих волн в слоях, разделенных горизонтальными плоскопараллельными границами, являются времена пробега волн, отраженных от границ слоев при нормальном падении, то очевидно, что такие данные определяют только средние скорости в слоях $v_i^{cp} = h_i \left(\int_0^{h_i} dh/v_i(h) \right)^{-1}$ (при условии, что мощности слоев h_i известны), но не несут никакой информации о распределении скорости с глубиной $v_i(h)$ в каждом слое. Таким образом, эти данные позволяют ставить задачу оценки лишь средних значений скорости. Другой пример — определение структуры земной коры по дисперсионной кривой поверхностных волн Лява. Из вида оператора решения прямой задачи (3) следует, что разным распределениям $\mu(x)$, $\rho(x)$ будут соответствовать разные дисперсионные кривые групповой скорости $u(T)$. Однако $u(T)$ определяется из наблюдений со значительной погрешностью, что не позволяет выявлять тонкие детали строения коры. Обычно при решении этой задачи ограничиваются рассмотрением модели коры, представляющей однородный слой, лежащий на однородном полупространстве. В такой модели искомыми параметрами будут модули сдвига μ_1, μ_2 и плотности ρ_1, ρ_2 соответственно в слое и полупространстве, а также толщина слоя h . Тогда поиск неизвестных распределений $\mu(x)$, $\rho(x)$ заменится поиском пяти неизвестных параметров. Такая замена оказывается возможной благодаря тому, что вариации упругих постоянных в коре и в мантии в допустимых пределах приводят к вариациям дисперсионной кривой в пределах погрешности наблюдений. Более того, часто пределы изменения μ и ρ в изучаемом регионе бывают известны из других исследований и они оказываются настолько узкими, что изменения μ и ρ внутри заданных пределов также приводят к вариациям дисперсионной кривой в пределах погрешности.

Тогда единственным определяемым параметром оказывается мощность кори h , т.е. обратная задача сводится к оценке всего одной величины.

Такая замена искомой модели $m(r)$ упрощенной моделью, полученной из $m(r)$ путем некоторого осреднения, называется параметризацией модели.

Можно сформулировать следующие общие принципы параметризации [1, 62]. Прежде всего определяется более или менее узкий класс функций $m(r)$, внутри которого производится поиск модели. В описанных примерах неизвестные распределения $v(z)$ (или $\mu(z)$, $\rho(z)$) рассматривались в классе кусочно-постоянных функций. В зависимости от специфики задачи могут выбираться различные функции: кусочно-линейные, экспоненциальные и т.д. Выбранные функции характеризуются некоторыми параметрами, так что задача поиска модели в классе этих функций сводится к поиску значений этих параметров. Таким образом, в результате параметризации искомая модель определяется уже не в функциональном, а в конечномерном пространстве. Размерность пространства моделей в результате параметризации будет равна числу определяемых параметров.

На основе каких же соображений следует параметризовать модель? Иначе говоря, исходя из чего следует выбирать параметры? Как уже говорилось, такой выбор диктуется используемыми данными. Обычно на основании результатов математического моделирования — расчетов поля для разных моделей среды — выявляется зависимость тех или иных характеристик поля от вариации модели. Например, путем анализа теоретических расчетов определяется, насколько существенно будет влиять на значения в заданной выборке данных замена однослойной модели двухслойной, трехслойной и т.д. Это позволяет оценить, какие свойства модели существенно влияют на поле, а какие — менее существенно, какие вариации в строении среды приводят к заметным различиям в значениях характеристик поля.

Параметризация должна быть оптимальной, т.е. желательно выбрать параметры так, чтобы их было не слишком много (не больше, чем позволит разрешающая способность исходных данных) и в то же время не слишком мало, чтобы не пропустить в решении

какие-то особенности модели, которые проявляются в используемых наблюдениях. Априорно выбрать оптимальную параметризацию довольно сложно и далеко не всегда удается. Но, как будет показано, если параметризация произведена неудачно, ее удается скорректировать в ходе решения.

§ 6. Классификация методов решения обратных задач

Сделаем краткий обзор основных направлений развития методов решения обратных задач. Подробный анализ методов с примерами их применения в разных геофизических задачах будет дан в последующих главах.

Как уже упоминалось в § 1, различия существующих методов решения обратных задач определяются тем, каким именно особенностям исходных данных, априорным представлениям о модели и требованиям к решению придается большее значение. В то же время, несмотря на кажущиеся различия, между всеми методами можно проследить определенную связь.

Мы будем в дальнейшем рассматривать только так называемые машинные методы, которые сводятся к последовательности математических операций и могут быть реализованы практически только на ЭВМ. Но нельзя совсем исключать и ручные методы, ибо они, хотя и отошли на задний план, но все еще сохраняют значение в практике геофизической интерпретации. С их помощью нельзя, конечно, построить решение, оптимальное в том смысле, чтобы оно, с одной стороны, отражало среду с той степенью детальности, которую разрешают используемые данные, а с другой — имело бы не слишком большую неопределенность. Но они дают возможность получить качественные и даже грубые количественные представления о модели, которые в дальнейшем могут быть использованы для уточнения машинными методами.

Сущность таких методов, как уже говорилось, сводится к сравнению данных наблюдений с теоретическими расчетами для небольшого числа моделей и выбору такой модели, которая лучше других согласуется с наблюдениями. Это простейший вариант ручной интерпретации, и его, собственно говоря, даже нельзя назвать методом, ибо он сильно зависит от искусства и интуиции

интерпретатора. К разряду ручных методов, хотя и самых простых, можно отнести палеточный метод, который позволяет оценивать некоторые параметры модели. Этот метод получил широкое распространение в практике электроразведки. Он применим в случае, когда исследуемое поле подчиняется некоторому принципу подобия. Сущность его сводится к следующему.

Пусть измеряется характеристика поля y в зависимости от некоторого параметра l . При этом среди неизвестных параметров модели имеются такие параметры α и β , которые в выражение зависимости $y(l)$ входят в виде масштабных коэффициентов, т.е. $y(l) = \beta \varphi(l/\alpha)$. В этом случае вместо того, чтобы рассчитывать множество кривых $y(l)$ для разных значений α и β , достаточно рассчитать одну кривую $\eta = \varphi(\xi)$, где $\eta = y/\beta$, $\xi = l/\alpha$. Эту кривую строят в логарифмическом масштабе, и также в логарифмическом масштабе представляют данные $y(l)$ (рис.5). Эти две зависимости накладывают одну на другую и сдвигают координатные системы одну относительно другой так, чтобы наблюдения хорошо

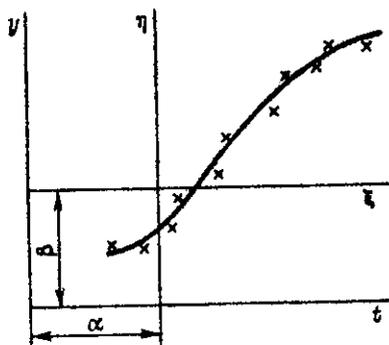


Рис. 5.

совпали с теоретической кривой. Тогда оси η, ξ отсекут на осях l, y величины, равные значениям неизвестных параметров α и β .

В случае, когда модель определяется не только параметрами, имеющими характер масштабных коэффициентов, но и некоторыми другими ($\theta_1, \theta_2, \dots, \theta_k$), рассчитывается семейство кривых $\varphi(\xi; \theta_1, \theta_2, \dots, \theta_k)$, соответствующих сетке значений

параметров θ_i . Наблюдения сопоставляются со всеми кривыми этого семейства, и из них выбирается такая, которая по форме лучше других соответствует наблюдениям.

Хотя в этом параграфе мы приводим лишь общий обзор методов, а подробное их изложение, иллюстрируемое примерами, будет дано в последующих главах, но в связи с тем, что на палеточ-

ном методе мы больше останавливаться не будем, здесь мы приведем некоторые примеры его применения при решении обратных геофизических задач.

П р и м е р 1. Интерпретация кривых вертикального электрического зондирования (ВЗЗ) [32]. В практике электроразведочных работ по методу ВЗЗ измеряется кажущееся сопротивление ρ_k в зависимости от расстояния r между питающими и приемными электродами. Задачей интерпретации кривых $\rho_k(r)$ является получение информации о распределении с глубиной удельного сопротивления ρ . Интерпретация проводится в предположении, что среда является слоисто-однородной, так что оценке подлежат мощности слоев h_i ($i = 1, 2, \dots, N-1$) и значения удельного сопротивления в них ρ_i ($i = 1, 2, \dots, N$), при этом число слоев N считается фиксированным.

В случае, когда модель может быть представлена одним слоем на полупространстве, выражение для кажущегося сопротивления, определяемого с помощью трех- или четырехэлектродной симметричной установки, имеет вид

$$\rho_k(r) = \rho_1 \left\{ 1 + 2 \sum_{n=1}^{\infty} \frac{K_{12}^n r^3}{[r^2 + (2nh_1)^2]^{3/2}} \right\}, \quad (14)$$

где $K_{12} = (\rho_2 - \rho_1) / (\rho_2 + \rho_1)$. Выражение (14) может быть представлено так:

$$\xi = \frac{\rho_k}{\rho_1} = f\left(\Lambda, \frac{\rho_2}{\rho_1}\right), \quad \text{где } \Lambda = r/h_1.$$

Таким образом, достаточно рассчитать семейство кривых $\xi(\Lambda)$ при разных значениях параметра ρ_2/ρ_1 . Подбирая из этого семейства кривую, наиболее совпадающую по форме с наблюдаемой кривой $\rho_k(r)$, можно оценить ρ_2/ρ_1 , а по относительному смещению координатных осей оцениваются значения h_1 и ρ_1 .

П р и м е р 2. Оценка мощности слоя и скорости поперечной волны в нем по дисперсионной кривой групповой скорости волны Лява [25]. Если изучаемую модель аппроксимировать однородным слоем на однородном полупространстве так, что модель будет характеризоваться параметрами $\sigma_1, \sigma_2, \rho_1, \rho_2, h$ ($\sigma = \sqrt{\mu/\rho}$), то зависимость $\omega(\tau)$ может быть записана в форме

$$u = \sigma_1 \varphi \left(\frac{T\theta_1}{h}; \frac{\sigma_2}{\sigma_1}, \frac{\rho_2}{\rho_1} \right).$$

Для простоты примем, что отношения скоростей поперечных волн σ_2/σ_1 и плотностей ρ_2/ρ_1 известны. Тогда достаточно рассчитать кривую $\eta = \varphi(\xi)$, где $\xi = T\theta_1/h$, $\eta = u/\sigma_1$. Выполнив сопоставление наблюдений с теоретически рассчитанной кривой описанным способом, можно оценить скорость поперечной волны в слое θ_1 и величину h/σ_1 , на основании которых и определяется толщина слоя h .

Поскольку при ручных способах интерпретации сопоставление теоретических и наблюдаемых данных проводится визуально, то выбранное решение не отвечает формализованному критерию, а зависит от произвола и искусства интерпретатора. Оценка ошибки решения также оказывается в значительной степени субъективной. Теоретическую кривую можно сдвигать вверх и вниз, вправо и влево в таких пределах, пока с точки зрения интерпретатора совпадение еще остается удовлетворительным. На рис. 6 изображены два крайних положения теоретической кривой 1 и 2, которые можно считать еще удовлетворяющими исходным данным. Кривая 1 соответствует координатным осям (ξ_1, η_1) , а кривая 2 — осям (ξ_2, η_2) . Из этого рисунка видно, что пределы, в которых производится перемещение теоретической кривой вдоль оси абсцисс (от η_1 до η_2), дают оценку погрешности параметра α , а вдоль оси ординат (от ξ_1 до ξ_2) —

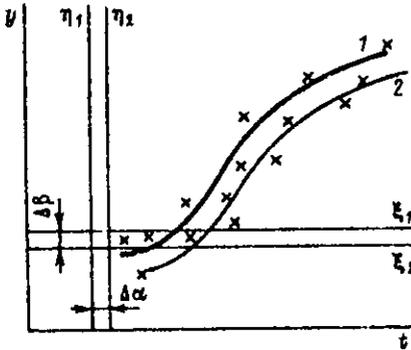


Рис. 6.

оценку погрешности параметра β .

Все сказанное относится и к выбору параметров, определяющих модель. Формализовать критерий оптимальной параметризации в таких методах невозможно. Обычно стараются выбрать класс

моделей предельно простым, чтобы по возможности уменьшить число альтернативных моделей. В этом случае решение получается чересчур сглаженным, не соответствующим разрешающей способности данных.

Мы довольно подробно остановились на анализе ручных (неформализованных) методов решения обратных задач, так как это позволит лучше уяснить особенности машинных (формализованных) методов.

Основными проблемами, возникающими при решении обратных задач, как уже упоминалось, являются: а) параметризация модели – выбор пространства M ; б) сопоставление наблюдений с теоретически рассчитанными характеристиками поля для заданной модели – выбор критерия близости точек в пространстве Y ; в) оценка качества решения – определение области допустимых решений; г) сужение области решения в случае неединственности или неустойчивости решения обратной задачи. Существующие методы имеет смысл рассматривать и классифицировать с точки зрения того, как они подходят к решению указанных проблем.

Остановимся на вопросе сопоставления теоретических и наблюдаемых данных, предполагая, что класс моделей, в котором отыскивается решение, из каких-то соображений фиксирован. Эта проблема возникает тогда, когда точного решения не существует, т.е. ни одна из точек в пространстве M не удовлетворяет условию $\Phi(m) = y$. Такая ситуация возможна в двух случаях: либо модель описывается слишком малым числом параметров, так что ни при каких значениях этих параметров нельзя получить все особенности имеющихся наблюдений, либо параметров выбрано достаточно, но наблюдения содержат погрешности. В этом случае в качестве приближенного решения принимается такая точка \hat{m} , которой в пространстве Y соответствовала бы точка \hat{y} , наиболее близкая к точке y , отвечающей реальным наблюдениям. При этом способе построения решения следует определить метрику пространства Y . Выбор метрики производится на основе анализа характера данных наблюдений. Чаще всего используется квадратичная метрика, но в некоторых случаях (например, когда недопустимы большие отклонения) может быть использована равномерная метрика.

Если расстояние $\|\hat{y} - y\|$ минимизируется в метрике L_2 , то построение решения производится по методу наименьших квадратов. В случае, если наблюдения имеют разную размерность, что имеет место при комплексировании разнородных наблюдений, или если просто наблюдения имеют разный вес, то минимизируется взвешенная сумма квадратов отклонений:

$$\lambda(m) = \sum_i \frac{(y_i - \Phi_i(m))^2}{\sigma_i^2}$$

В общем случае построение приближенного решения производится путем минимизации некоторой так называемой функции отклика или целевой функции. В частности, если пространство наблюдений имеет метрику L_2 , функция отклика представляет собой просто сумму квадратов отклонений, но в общем случае ее вид зависит от свойств исходных наблюдений. Характер исходных наблюдений учитывается наиболее полно при трактовке экспериментального материала с теоретико-вероятностных позиций. При таком подходе к данным пространство наблюдений рассматривается как выборочное пространство, а точка y , соответствующая наблюдениям, — как элементарное случайное событие. В этом пространстве задана вероятность появления того или иного события y для фиксированной модели m . Эта вероятность определяется путем анализа статистических свойств ошибок наблюдений. Для нахождения модели m по заданной точке в выборочном пространстве Y используются те или иные методы теории статистических решений. Такой подход к трактовке наблюдений как случайных величин, характеризующихся определенным распределением вероятности, лежит в основе статистических методов решения обратных задач. Статистические методы получили широкое развитие в работах ленинградской школы геофизиков (Ф.М. Гольцман, Т.Б. Калинин, Л.А. Халфин [14, 15, 17-19, 37, 46]).

И статистические методы, и метод наименьших квадратов сводятся к минимизации некоторой функции отклика в пространстве параметров модели. Поэтому их иногда объединяют под общим названием оптимизационных методов. В случаях, когда функция отклика строится с учетом статистических свойств ошибок наблюдений, качество решения характеризуется оценками дисперсий и ковариаций параметров, знание которых позволяет строить

доверительные области для параметров модели. Вопросы параметризации при использовании оптимизационных методов отодвигаются на второй план, ибо в их рамках трудно найти конструктивный способ оптимальной параметризации.

Как уже говорилось, методы оптимизации используются в тех случаях, когда по той или иной причине нельзя построить точного решения: либо из-за наличия погрешностей в исходных данных, либо из-за неадекватности модели среды.

Теперь рассмотрим другую ситуацию, когда наблюдения достаточно точны, а размерность пространства параметров слишком велика, так что эти наблюдения в рамках выбранного класса сред отвечает бесконечное множество решений. В этом случае решение не может быть построено, если не наложить на модель какие-то априорные ограничения. Простейший способ наложения априорных ограничений - это закрепление некоторых параметров на основании результатов каких-то других исследований. Иногда удается ввести определенные связи между параметрами. В том же другом случаях из исходного пространства выделяется некоторое подпространство, в котором и отыскивается решение. Есть и иной способ наложения ограничений: в качестве решения принимается такое, которое наиболее близко к некоторой заданной точке m_0 . Этот способ основан на том, что всегда из предыдущих исследований бывает известно некоторое приближение к искомой модели. Поэтому наиболее естественно строить решение так, чтобы оно было бы ближе других к этому начальному приближению. При таком подходе оказывается уже необходимым определить метрику пространства моделей M . В этом случае задача ставится так: из области всех возможных решений выбрать такое, которое минимизирует расстояние $\|m - m_0\|$. Указанный подход к построению единственного решения в случае, когда наблюдения y_i являются линейными функционалами от $m(r)$ (или соответственно линейными функциями параметров, описывающих модель), был разработан Дж. Бэйкусом и Ф. Гильбертом и получил название метода Бэйкуса - Гильберта.

В рамках этого метода решается и вопрос о разрешающей способности используемых наблюдений, что важно для выбора оптимальной параметризации. Метод Бэйкуса - Гильберта позволяет

построить также осредненные характеристики модели, которые по данной совокупности наблюдений определяются однозначно.

Решение обратной задачи, получаемое как оптимизационными методами, так и по методу Бэйкуса - Гильберта, может оказаться неустойчивым. Такое решение с точки зрения цели геофизической интерпретации оказывается бессмысленным, так как область допустимых решений, определяемая погрешностью наблюдений, становится весьма протяженной. В этих случаях используется метод регуляризации, разработанный А.Н.Тихоновым для построения устойчивых приближений при решении некорректных задач. Сущность регуляризации заключается в наложении некоторых априорных ограничений на модель, благодаря которым решение становится устойчивым.

Регуляризация приобретает вполне определенный смысл при статистическом подходе к построению решения. Если не только пространство Y , но и пространство моделей M рассматривать как пространство случайных событий, то наиболее распространенный тип ограничений (близость к заданной точке в метрике L_2), можно трактовать как априорную вероятность модели, понимая при этом под вероятностью меру неопределенности наших представлений о модели.

Основное различие между оптимизационными методами и методом Бэйкуса - Гильберта определяется тем, что они предназначены для решения задач с разным соотношением между числом параметров и числом наблюдений: оптимизационные методы требуют, чтобы число наблюдений было больше числа параметров, а метод Бэйкуса - Гильберта применим в случае, когда число параметров больше числа наблюдений (пространство параметров может быть и бесконечномерным). Очевидно, что хорошо было бы иметь универсальный метод, пригодный как в одном, так и в другом случае. Такой метод существует, но только для линейаризованных задач, т.е. таких, где зависимость наблюдений y_i от параметров модели линейная. Это так называемый метод псевдообращения (обобщенного обращения) или, как он часто в литературе называется, метод сингулярных разложений. Этот метод был разработан в линейной алгебре для решения систем линейных уравнений с произвольным соотношением между числом уравнений и числом

неизвестных. Он позволяет получать решение и в том случае, если точное решение отсутствует, а приближенных решений, определяемых по методу наименьших квадратов, оказывается бесконечное множество. При неустойчивости решения, получаемого этим методом, могут быть использованы те или иные способы регуляризации.

Основные вычислительные задачи, возникающие при реализации упомянутых методов, — это поиск экстремума многомерной функции и задачи линейной алгебры: решение систем линейных уравнений, нахождение собственных значений и собственных векторов матриц. Вычислительные трудности возникают из-за неустойчивости оператора обратной задачи и проявляются в том, что в широкой окрестности искомого экстремума функция отклика меняется слабо, а если задача построения решения сводится к решению линейной системы, то система оказывается плохо обусловленной. В этих случаях иногда пытаются повышать устойчивость решения применением определенных вычислительных алгоритмов. Но в конечном счете все эти способы и алгоритмы могут быть интерпретированы с точки зрения наложения априорных ограничений на искомую модель. Поэтому при использовании тех или иных вычислительных методов для решения обратной задачи следует исходить из анализа сущности самой задачи: характера исходных данных, априорных представлений об искомой модели и требований, предъявляемых к решению.

В последующих главах будут описаны принципиальные основы перечисленных выше методов.

Оптимизационные методы заключаются в нахождении оценки параметров среды m путем нахождения экстремума (оптимизации) некоторой целевой функции $\lambda(m, y)$ при заданном значении вектора наблюдений y . Целевую функцию следует выбрать на основе представлений о статистических свойствах погрешности наблюдений и с учетом требований, предъявляемых к качеству оценки. Оценки, получаемые путем оптимизации выбранной так целевой функции, наиболее полно отражают особенности экспериментального материала и являются в определенном смысле оптимальными. Строго говоря, такие методы следует называть статистическими, так как они сводятся к задаче статистического оценивания.

Однако очень часто целевая функция выбирается на основе эвристических соображений из-за полного отсутствия информации о характеристиках распределения ошибок наблюдений. Кроме того, во многих случаях построение статистических оценок теряет смысл из-за малого количества наблюдений. Но несмотря на это большую часть этой главы мы посвятим именно статистическому оцениванию параметров среды и трактовке решения, получаемого по методу оптимизации, с позиций математической статистики.

§ 1. Общие принципы статистического оценивания

В гл. I было отмечено, что данные геофизических наблюдений, используемые для решения обратной задачи, всегда осложнены случайными ошибками, и потому сами являются случайными величинами, имеющими некоторое распределение, характеристики которого зависят от параметров среды. При этом задачу определения параметров среды можно свести к задаче статистического оценивания параметров распределения случайных величин по дан-

ным ограниченного числа измерений. В этой главе мы рассмотрим методы решения обратных задач геофизики, основанные именно на таком подходе к трактовке экспериментальных данных. Для облегчения понимания дальнейшего изложения приведем основные сведения и понятия из математической статистики [8, 10, 21, 23, 33, 44].

Пусть производится наблюдение над некоторой n -мерной $X(X_1, X_2, \dots, X_n)$. Каждая компонента вектора X содержит случайную ошибку и потому в некоторой области принимает определенное значение в соответствии с функцией распределения вероятностей $F(x_i) = P(X_i \leq x_i)$. Вся совокупность данных принимает значения x_1, x_2, \dots, x_n в пространстве X в соответствии с функцией распределения $F(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$. Тогда X называют выборочным пространством, а случайный вектор X - выборкой объема n из генеральной совокупности. Функция

$$p(x) = \frac{\partial F(x_1, \dots, x_n)}{\partial x_1, \dots, \partial x_n}, \text{ (если } F(x) \text{ дифференцируема по всем}$$

x_i) называется общей плотностью вероятности случайных величин X_1, X_2, \dots, X_n . При независимых наблюдениях

$$p(x) = \prod_{i=1}^n p(x_i). \quad (15)$$

Если плотность вероятности $p(x)$ имеет известную математическую форму, но содержит некоторое число неизвестных параметров $\theta_1, \theta_2, \dots, \theta_k \in \Theta$ то задача статистического оценивания сводится к оценке неизвестных параметров распределения по выборочным значениям X_1, X_2, \dots, X_n . В общем виде эта проблема может быть сформулирована следующим образом: рассматривается некоторая случайная величина, для которой либо распределение имеет известную математическую форму, но содержит некоторое число неизвестных параметров, либо форма распределения неизвестна, но известны некоторые моменты распределения, выражающиеся также через неизвестные параметры. Задана конечная выборка наблюдаемых значений этой величины и требуется, используя выборочные значения $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$, получить оценки для вектора параметров $\Theta(\theta_1, \theta_2, \dots, \theta_k)$ и определить их точность.

Прежде чем перейти к рассмотрению методов оценивания параметров, укажем некоторые свойства функции $p(x; \Theta)$.

1. Если $p(x; \Theta)$ - непрерывная и дважды дифференцируемая функция параметров Θ и при этом область изменения x не зави-

сит от θ , то функция $p(x; \theta)$ называется регулярной. Примером нерегулярной функции является плотность равномерного распределения, в котором среднее значение зависит от параметра θ :

$$p(x; \theta) = [H(x - f(\theta) + \epsilon) - H(x - f(\theta) - \epsilon)] / (2\epsilon),$$

где $H(x)$ - функция Хэвисайда.

2. Для регулярной функции распределения справедливо следующее равенство:

$$-E_x \left(\frac{\partial^2 \log p(x; \theta)}{\partial \theta_i \partial \theta_j} \right) = E_x \left(\frac{\partial \log p(x; \theta)}{\partial \theta_i} \frac{\partial \log p(x; \theta)}{\partial \theta_j} \right), \quad (16)$$

$$i, j = 1, 2, \dots, k.$$

Здесь $E_x(\dots)$ - математическое ожидание, вычисленное в предположении, что θ - вектор истинных параметров. Действительно, пусть $p(x; \theta)$ есть плотность распределения случайной величины X и зависит от одного параметра θ . Тогда справедливо следующее равенство:

$$\int_{-\infty}^{\infty} p(x; \theta) dx_1 dx_2 \dots dx_n = 1.$$

Из этой формулы следует, что

$$\int_{-\infty}^{\infty} \frac{\partial p(x; \theta)}{\partial \theta} dx_1 \dots dx_n = \int_{-\infty}^{\infty} \frac{\partial \log p(x; \theta)}{\partial \theta} p(x; \theta) dx_1 \dots dx_n = E_x \left(\frac{\partial \log p(x; \theta)}{\partial \theta} \right) = 0.$$

Продифференцируем последнее выражение по θ :

$$\int_{-\infty}^{\infty} \frac{\partial^2 \log p(x; \theta)}{\partial \theta^2} p(x; \theta) dx_1 \dots dx_n + \int_{-\infty}^{\infty} \frac{\partial \log p(x; \theta)}{\partial \theta} \frac{\partial p(x; \theta)}{\partial \theta} dx_1 \dots dx_n = 0.$$

Учитывая, что

$$\int_{-\infty}^{\infty} \frac{\partial \log p(x; \theta)}{\partial \theta} \frac{\partial p(x; \theta)}{\partial \theta} dx_1 \dots dx_n = \int_{-\infty}^{\infty} \left(\frac{\partial \log p(x; \theta)}{\partial \theta} \right)^2 p(x; \theta) dx_1 \dots dx_n,$$

получим требуемое равенство:

$$E_x \left(\frac{\partial \log p(x; \theta)}{\partial \theta} \right)^2 = -E_x \left(\frac{\partial^2 \log p(x; \theta)}{\partial \theta^2} \right).$$

При нескольких неизвестных параметрах аналогичные рассуждения приведут к формуле (16).

3. В выражении для плотности вероятности $p(\mathbf{x}; \theta)$ вектор параметров θ предполагался фиксированным, и плотность вероятности являлась таким образом функцией значений \mathbf{x} , которые может принимать случайная величина \mathbf{X} . Теперь будем считать, что значение θ неизвестно, и определим функцию правдоподобия для θ следующим соотношением:

$$L_{\mathbf{x}}(\theta) = p(\mathbf{x}; \theta). \quad (17)$$

Если случайный вектор \mathbf{X} в результате наблюдений принял фиксированное значение $\mathbf{x} = \mathbf{x}_0$, то функция правдоподобия является функцией только θ . Подчеркнем, однако, что эта функция зависит от того, какие значения получены в результате наблюдений над случайной величиной \mathbf{X} .

При заданном значении $\mathbf{x} = \mathbf{x}_0$ функция $L_{\mathbf{x}_0}(\theta)$ определяет правдоподобие параметров в следующем смысле: если из пространства параметров θ выбрать различные точки θ , то те из них, для которых полученные из эксперимента значения \mathbf{x}_0 характеризуются большим значением плотности вероятности, будут более правдоподобны, нежели другие, для которых плотность вероятности мала.

Функция правдоподобия для нескольких независимых экспериментов равна произведению функций правдоподобия для каждого эксперимента:

$$L_{\mathbf{x}}(\theta) = \prod_{i=1}^n L_{\mathbf{x}_i}(\theta).$$

Это легко видеть из формул (15) и (17).

Функция правдоподобия, полученная для регулярной функции $p(\mathbf{x}; \theta)$, также является регулярной. В этом случае для нее справедливо соотношение (16):

$$E_{\mathbf{x}} \left(\frac{\partial^2 \log L_{\mathbf{x}}(\theta)}{\partial \theta_i \partial \theta_j} \right) = -E_{\mathbf{x}} \left(\frac{\partial \log L_{\mathbf{x}}(\theta)}{\partial \theta_i} \frac{\partial \log L_{\mathbf{x}}(\theta)}{\partial \theta_j} \right),$$

$$i, j = 1, 2, \dots, k.$$

4. Матрица \mathbf{J} , элементы которой

$$J_{ij} = E_{\mathbf{x}} \left(\frac{\partial \log L_{\mathbf{x}}(\theta)}{\partial \theta_i} \frac{\partial \log L_{\mathbf{x}}(\theta)}{\partial \theta_j} \right),$$

называется информационной матрицей Фишера. Диагональные члены ее - J_{ii} - называются информационным количеством Фишера и при независимых экспериментах их значения определяют количество информации о соответствующем параметре θ_i , содержащейся в данной выборке x_0 . Количество информации является мерой чувствительности выборки x к малым изменениям θ_i .

Перейдем к рассмотрению методов оценивания параметров. Различаются задачи построения точечных и интервальных оценок. Точечными оценками являются некоторые функции выборочных значений $\hat{\theta}_i = g_i(x_1, x_2, \dots, x_n)$, $i = 1, 2, \dots, k$, которые по возможности наиболее точно отражают значения искомых параметров. Интервальная оценка параметров представляет собой область, внутри которой с заданной доверительной вероятностью должны находиться истинные значения параметров.

Точечное оценивание. Существует бесконечное множество функций выборочных значений, которые можно использовать в качестве точечных оценок. Поэтому необходимо иметь возможность выбрать из них такие, которые давали бы максимальную точность, т.е. были бы оптимальными. Прежде всего отметим, что поскольку оценка является функцией выборочных значений, то она сама - случайная величина. Следовательно, свойства оценки естественно характеризовать свойствами ее распределения. Рассмотрим основные свойства оценок. Для простоты ограничимся случаем, когда оценивается один параметр θ .

1. Среднее значение оценки

$$E_0(\hat{\theta}) = \theta + \delta(\theta). \quad (18)$$

Величина $\delta(\theta)$ называется смещением.

Если $\delta(\theta) = 0$ независимо от объема выборки n , то оценки называются несмещенными. Когда $\delta(\theta) \rightarrow 0$ при $n \rightarrow \infty$, оценки $\hat{\theta}$ являются асимптотически несмещенными. Если же при $n \rightarrow \infty$ $E_0(\hat{\theta}) \rightarrow \theta$ по вероятности, то такие оценки называются остойчивыми.

2. Дисперсия оценки определяется по формуле

$$\sigma^2(\theta) = E_0[\hat{\theta} - E_0(\hat{\theta})]^2. \quad (19)$$

Для регулярных функций $p(x; \theta)$ существует нижняя граница для дисперсии оценки. Это свойство выражается неравенством Рао - Крамера:

$$\sigma^2(\theta) \geq \frac{(1 + \partial\theta/\partial\theta)^2}{J(\theta)}, \quad (20a)$$

где $J(\theta)$ – информационное количество Фишера. Оценка, для которой это условие переходит в равенство, называется эффективной в смысле неравенства Рао – Крамера. Однако равенство возможно только в классе несмещенных оценок, т.е. когда

$$\sigma^2(\theta) \geq [J(\theta)]^{-1}. \quad (20б)$$

В случае многопараметрического оценивания вместо дисперсии следует рассматривать ковариационную матрицу V , элементы которой

$$V_{ij} = E_{\theta} \{ [\hat{\theta}_i - E_{\theta}(\hat{\theta}_i)] [\hat{\theta}_j - E_{\theta}(\hat{\theta}_j)] \}.$$

Детерминант этой матрицы $|V|$ называется обобщенной дисперсией. Для обобщенной дисперсии в классе несмещенных оценок имеет место неравенство, аналогичное (20a): $|V| \geq |J^{-1}|$, а для дисперсии каждой компоненты вектора оценок справедливо

$$\sigma_i^2 \geq (J_{ii})^{-1}. \quad (20в)$$

Нижняя граница в условии (20в) достигается в классе несмещенных оценок, только если $\hat{\theta}$ – линейная функция компонент вектора x . В более общем случае несмещенная оценка $\hat{\theta}_i$ имеет дисперсию не меньшую, чем J_{ii}^{-1} , и нижняя граница может быть достигнута лишь асимптотически при $n/k \rightarrow \infty$. Тогда вектор $\hat{\theta}$ называют асимптотически эффективной оценкой параметров в смысле неравенства Рао – Крамера.

Если в выражение (19) подставить среднее значение оценки (18), то получим формулу для среднего квадрата ошибки:

$$E_{\theta} (\hat{\theta} - \theta)^2 = \sigma^2(\theta) + \delta^2(\theta). \quad (21)$$

Из двух оценок наилучшей будет та, у которой левая часть этой формулы будет меньше. Но добиться минимума среднего квадрата ошибки, уменьшая только дисперсию или только смещение, нельзя, так как уменьшение одной величины влечет за собой увеличение другой. Только в классе несмещенных оценок эффектив-

ние оценки в смысле неравенства Рао - Крамера минимизируют средний квадрат ошибки.

В настоящее время (после работ А. Вальда [9]) используются способы сравнения разных оценок, которые учитывают потери, возникающие при ошибочном оценивании параметров. Так, с каждой оценкой $\hat{\theta}$, найденной по выборочным значениям x_1, x_2, \dots, x_n , связана ошибка, например, измеряемая разностью $\hat{\theta} - \theta$ (хотя в общем случае ошибку можно измерить и по-другому). Величина ошибки от выборки к выборке меняется: в одном случае она может быть мала, а в другом - недопустимо большой. Чтобы учесть это обстоятельство для каждой оценки, назначается неотрицательная функция $\omega(\hat{\theta}; \theta)$, зависящая от $\hat{\theta}$ и θ и характеризующая ущерб при замене истинного параметра θ оценкой $\hat{\theta}$.

Функция $\omega(\hat{\theta}; \theta)$ называется функцией потерь. Выбор функции потерь произволен и зависит от конкретной задачи. Требуется, чтобы она была неотрицательной и не убывала с ростом ошибки. На практике чаще всего встречаются функции потерь вида $|\hat{\theta} - \theta|^g$. При $g = 2$ $\omega(\hat{\theta}; \theta)$ называют квадратичной функцией потерь. Если значения $\omega(\hat{\theta}; \theta)$ при увеличении ошибки оценивания остаются постоянными, то функцию потерь называют простой. Из двух оценок $\hat{\theta}_1$ и $\hat{\theta}_2$ считается наилучшей та, для которой среднее значение потерь - величина риска $E_x[\omega(\hat{\theta}; \theta)]$ - наименьшее. Осреднение здесь производится по совокупности данных x_1, x_2, \dots, x_n в предположении, что истинное значение параметра равно θ . Такая оценка удовлетворяет условию

$$E_x[\omega(\hat{\theta}; \theta)] = r(\theta) = \int \omega(\hat{\theta}; \theta) p(x; \theta) dx_1 \dots dx_n = \min.$$

Однако этот способ не лишен недостатков: во-первых, выбор функции потерь субъективен; во-вторых, не все оценки сравнимы между собой.

Другой подход к сравнению оценок параметров состоит в изучении поведения оценок в худших ситуациях. Рассматриваются функции риска для семейства функций потерь, зависящих от различных оценок $\hat{\theta}_1, \dots, \hat{\theta}_n$ параметра θ . За наилучшую оценку рекомендуется принять ту, для которой максимальная величина риска имеет минимальное значение. В этом случае оценка называется минимаксной. В таком подходе к выбору лучшей оценки есть

тоже свои трудности. Иногда минимаксная оценка может приводить к большим потерям, чем другая оценка в области пространства параметров, хотя она и дает уверенность в том, что потери в среднем не будут больше некоторого значения. Кроме того, построение таких оценок связано, как правило, с большими вычислительными трудностями.

Еще один подход к сравнению оценок основывается на трактовке параметра θ как случайной величины. Если допустить, что пространство параметров θ выборочное и параметры в нем получают значения в соответствии с плотностью вероятности $p(\theta)$ независимо от данного эксперимента, то за наилучшую оценку параметра θ можно принять ту, которая минимизирует функцию среднего риска

$$R(\theta) = \int r(\theta) p(\theta) d\theta.$$

Плотность вероятности $p(\theta)$ называется априорной, а оценка, удовлетворяющая минимуму среднего риска, называется байесовской оценкой для данной функции потерь и априорной плотности вероятности $p(\theta)$. При довольно общих условиях байесовская оценка является минимаксной для наименее благоприятного априорного распределения (например, равномерного). Затруднения, возникающие при данном подходе к сравнению оценок, касаются в основном исходной постановки задачи, так как параметр θ не всегда можно считать случайным, а при этом терлет смысл понятие распределения этой величины.

Только что описанные способы выбора наилучшей оценки пригодны в альтернативных ситуациях, ибо функция потерь зависит от оценки параметра, уже найденной по какому-то правилу, и не дает конструктивного способа определения самих оценок.

Один из общих подходов к построению точечных оценок заключается в том, чтобы ограничить класс рассматриваемых оценок и в нем найти ту, которая минимизирует средний квадрат ошибки (21). Наиболее разработана теория построения решений в классе несмещенных и асимптотически (при $n \rightarrow \infty$) несмещенных оценок, как было показано ранее, дающих эффективные (асимптотически эффективные) оценки в смысле неравенства (20б).

В статистической теории показано [16], что существуют методы оценивания параметров для регулярных функций $p(x; \theta)$, при-

водящие в общем случае к асимптотически несмещенным и эффективным оценкам в смысле неравенства Рао - Крамера, которые одновременно являются асимптотически минимаксными для широкого класса функций потерь. Такими методами являются методы максимума апостериорной вероятности, максимального правдоподобия и наименьших квадратов. К указанному семейству функций потерь относятся симметричные функции $\psi(\hat{\theta}; \theta)$, зависящие от разности $\hat{\theta} - \theta$, например функции вида $|\hat{\theta} - \theta|^k$, которые, как уже было упомянуто, чаще всего используются на практике.

Метод максимума апостериорной вероятности основывается на байесовском подходе к решению задачи. Параметр предполагается случайной величиной с известной априорной плотностью вероятностей $p(\theta)$. Пусть производятся наблюдения над случайной величиной X , которая получает значения в соответствии с условной плотностью вероятностей $p(x; \theta)$, зависящей от параметра θ . В результате эксперимента к априорным представлениям, характеризующимся плотностью вероятности $p(\theta)$, добавляется информация, извлеченная из наблюдений над величиной X . Теперь степень наших знаний о том, какие значения может принимать параметр θ , будет характеризоваться условной плотностью вероятности $p(\theta/x)$, которая носит название апостериорной плотности вероятностей параметра θ . Значение X , получаемое в результате эксперимента, обозначим через x_0 . Тогда условная плотность вероятности $p(\theta/x_0)$ при заданном x_0 , согласно теореме Байеса равна

$$p(\theta/x_0) = \frac{p(\theta)L_{x_0}(\theta)}{c}, \quad (22)$$

где $c = \int_0^1 L_{x_0}(\theta)p(\theta)d\theta, \dots, d\theta_2$; $L_{x_0}(\theta)$ - функция правдоподобия. В качестве точечной оценки в этом случае выбирается такая, которая делает апостериорную плотность вероятности максимальной, т.е. удовлетворяет условию

$$p(\theta/x_0) = p(\theta)L_{x_0}(\theta) = \max_{\theta} \quad (23)$$

Однако часто не удается определить априорную плотность вероятности для искомого параметра. Тогда оценивание параметров проводится по методу максимального правдоподобия. Функцию правдоподобия строят на основе плотности вероятности $p(x; \theta)$, в соответствии с которой в результате эксперимента случайная

величина X получила значение x_0 . Принцип максимального правдоподобия заключается в том, что компоненты вектора θ должны быть выбраны так, чтобы полученная из наблюдений выборка x_0 была бы наиболее вероятной. Это происходит тогда, когда выполняется условие

$$L_{x_0}(\theta) = \max_{\theta} . \quad (24)$$

Обычно вместо функции $L_{x_0}(\theta)$, определяемой по формуле (17), рассматривают логарифмическую функцию правдоподобия $l_{x_0}(\theta) = \log L_{x_0}(\theta)$. Поскольку обе эти функции достигают максимума в одной и той же точке, то безразлично, какую из них использовать для построения оценки. Однако в ряде случаев логарифмическая функция оказывается предпочтительнее.

Метод максимального правдоподобия может применяться как для неслучайных, так и для случайных параметров. Если параметры случайны, то условие (24) получается из условия (23) при равномерном априорном распределении. Отсюда следует, что свойства оценок максимального правдоподобия и максимальной апостериорной вероятности совпадают.

Пусть распределение $p(x; \theta)$ наблюдаемой случайной величины X неизвестно, а известны только первые два момента распределения, являющиеся функциями искомых параметров θ :

$$E_x(x_i) = \int_x x_i p(x; \theta) dx_1 \dots dx_n = r_i(\theta),$$

$$E_x[(x_i - r_i(\theta))(x_j - r_j(\theta))] = R_{ij}(\theta).$$

В случае, когда само распределение неизвестно, но известны некоторые его характеристики, максимальной неопределенности наших знаний о распределении будет соответствовать такое распределение, которое приводит к максимальному значению энтропии. Энтропия распределения вероятностей определяется следующим образом:

$$H = - \int_x p(x) \log p(x) dx_1 \dots dx_n . \quad (25)$$

Нетрудно показать, используя метод неопределенных множителей Лагранжа, что распределением, соответствующим максимальной энтропии (25) и удовлетворяющим условиям (24) и условию нормировки плотности вероятности $\int_x p(x) dx_1 \dots dx_n = 1$, является нормальное распределение

$$p(\mathbf{x}; \theta) = \frac{1}{(\sqrt{2\pi})^n |\mathbf{R}|} \exp\{-[\mathbf{x} - \mathbf{f}(\theta)]^T \mathbf{R}^{-1} [\mathbf{x} - \mathbf{f}(\theta)]\},$$

где $|\mathbf{R}|$ — определитель ковариационной матрицы \mathbf{R} , которая характеризует корреляционные связи между величинами x_i и x_j ($i, j = 1, 2, \dots$). Для нормального распределения логарифм функции правдоподобия, полученной на основании выборки $x_{01}, x_{02}, \dots, x_{0n}$, имеет вид

$$l_{x_0}(\theta) = -\log[(\sqrt{2\pi})^n |\mathbf{R}|] - [\mathbf{x} - \mathbf{f}(\theta)]^T \mathbf{R}^{-1} [\mathbf{x} - \mathbf{f}(\theta)]. \quad (26)$$

В случае, когда \mathbf{R} зависит от θ , для поиска оценки максимального правдоподобия следует искать компоненты вектора θ из условия максимизации (26). Однако часто максимизируют только второе слагаемое в правой части выражения (26). Получаемая таким способом оценка называется оценкой наименьших квадратов. Если \mathbf{R} не зависит от θ , оценка наименьших квадратов совпадает с оценкой максимального правдоподобия (в частности, для линейных задач). Вообще почти во всех практических задачах, когда истинный закон распределения случайных величин допускает аппроксимацию нормальным законом, эти оценки различаются очень мало, так что при неизвестном распределении обычно используют оценку наименьших квадратов. Если априори можно считать все наблюдения независимыми и равнозначными, то $\mathbf{R} = G^2 \mathbf{I}$ (G^2 может быть неизвестна) и оценка наименьших квадратов получается из условия минимизации суммы квадратов отклонений:

$$[\mathbf{x} - \mathbf{f}(\theta)]^T [\mathbf{x} - \mathbf{f}(\theta)] = \sum_{i=1}^n [x_i - f_i(\theta)]^2.$$

Таким образом, описанные методы оценивания параметров при определенных условиях вытекают один из другого; метод максимального правдоподобия является частным случаем байесовского подхода к решению задачи при отсутствии предварительной информации об искомых параметрах. Метод максимального правдоподобия переходит в метод наименьших квадратов, когда наблюдаемые случайные величины распределены нормально, и помеха аддитивна: $x_i - f_i(\theta) = \xi_i$, так что ее статистические характеристики не зависят от неизвестных параметров распределения. Следовательно, свойства оценок, полученных указанными методами, вообще говоря, совпадают.

Мы уже упоминали о свойствах оценок наименьших квадратов, максимального правдоподобия и оценок, соответствующих максимуму апостериорной плотности вероятности. Остановимся на них еще раз. В условиях регулярности функции правдоподобия определенные ранее оценки максимальной апостериорной плотности вероятности и максимального правдоподобия асимптотически несмещены и эффективны в смысле неравенства Рао-Крамера, а также имеют асимптотически нормальное распределение со средним значением, стремящимся к истинному значению параметра θ , и ковариационной матрицей J^{-1} :

$$p(\hat{\theta}/\theta) = \frac{1}{(\sqrt{2\pi})^n |J^{-1}|^{1/2}} \exp[-(\hat{\theta} - \theta)' J^{-1} (\hat{\theta} - \theta)],$$

где $p(\hat{\theta}/\theta)$ — плотность вероятностей оценки $\hat{\theta}$ при условии, что истинными параметрами являются компоненты вектора θ . Диагональные члены матрицы J_{ii}^{-1} есть дисперсии оценок параметров, поэтому их величинами можно характеризовать точность найденных оценок. Оценки $\hat{\theta}$ при дополнительных условиях, накладываемых на функцию правдоподобия, о которых уже говорилось, будут асимптотически минимаксными для определенного класса функций потерь. Оценки наименьших квадратов несмещены и эффективны для малой выборки экспериментальных данных, необходимо только, чтобы $n/k \gg 1$. Если x линейно зависит от параметров, то оценка наименьших квадратов является линейной оценкой и, как уже упоминалось, дисперсия при этом достигает своей нижней границы.

Задачи нелинейного оценивания по методу наименьших квадратов могут быть сведены к построению линейных оценок, если определять не сам параметр, а малую поправку к величине, принятой в качестве начального приближения (см. § 3 гл. I). Оценивание по методу наименьших квадратов играет большую роль во многих практических задачах, так как в большинстве случаев распределение наблюдаемой величины либо действительно является нормальным, либо может быть принято таковым при отсутствии информации о нем, а случайные наблюдаемые величины осложнены помехой, которую почти всегда можно считать аддитивной, ибо ошибки наблюдения, как правило, малы. (В связи с этим метод наименьших квадратов будет изложен несколько подробнее остальных в § 6 настоящей главы.)

Интервальное оценивание. Задача построения интервальной оценки заключается в нахождении границы области $\theta^* \in \Theta$, внутри которой с заданной доверительной вероятностью $1 - \epsilon$ находится истинное значение вектора параметров θ . Оптимальность доверительной границы определяется требованием, чтобы вероятность $1 - \epsilon$ была близка к единице, а объем области θ^* минимальным.

При оценивании одного параметра θ доверительная область переходит в доверительный интервал (θ_1^*, θ_2^*) , внутри которого с вероятностью $1 - \epsilon$ находится значение θ . Для простоты здесь мы рассмотрим теорию построения доверительных интервалов, которая легко распространяется на более общий случай [24].

Если искомый параметр трактовать как случайную величину, то можно пытаться найти ее распределение, а зная распределение, легко построить доверительный интервал. Такой подход, как и при точечном оценивании, основан на теореме Байеса и может быть сформулирован следующим образом: имеется совокупность случайных величин $X_1, X_2, \dots, X_n, \theta$. Величина X определяется из наблюдений, а θ не может быть измерена непосредственно. Требуется на основании наблюдений над величиной X сделать вероятностные заключения относительно θ . Очевидно, что для решения такой задачи необходимо иметь представление о распределении как X , так и θ .

Пусть задана плотность вероятности взятой отдельно величины θ , которую обозначим $p(\theta)$. Кроме того, пусть известно, что случайная величина X приобретает значения в соответствии с плотностью вероятности $p(x; \theta)$. Предположим, что в результате эксперимента вектор X приобрел значения x_0 . Тогда условная плотность вероятности θ при заданном x_0 удовлетворяет формуле Байеса (22) (все обозначения имеют тот же смысл):

$$p(\theta/x_0) = p(\theta)L_{x_0}(\theta)/c.$$

Доверительный интервал для θ можно построить следующим образом:

$$\int_{\theta_1}^{\theta_2} p(\theta/x_0) d\theta = 1 - \epsilon,$$

здесь θ_1 и θ_2 выбираются так, чтобы длина $|\theta_2 - \theta_1|$ была минимальной при заданном значении $1 - \epsilon$.

Однако часто не удается определить априорную вероятность для искомого параметра. В этом случае возможны два подхода. Один из них основывается на предположении, что $p(\theta) = \text{const}$.

Тогда границы интервала отыскиваются из условия $\int_{\theta_1}^{\theta_2} L_{x_0}(\theta) d\theta / \int_{-\infty}^{+\infty} L_{x_0}(\theta) d\theta = 1 - \epsilon$. Но в связи с тем, что такое предположение является слишком произвольным, возникают определенные трудности (см. пример в книге Д.Худсона [48]).

Другой подход к построению интервальных оценок основан на методе доверительных интервалов. Кратко суть его заключается в следующем. Определяют некоторую точечную оценку $\hat{\theta}$ для θ (например, методами, описанными ранее). Поскольку θ выражается через выборочные значения величины X (распределение вероятности которой предполагается известным при любом фиксированном θ), то можно определить и плотность вероятности оценки $p(\hat{\theta}; \theta)$. Значит, если мы зададимся некоторой доверительной вероятностью $1 - \epsilon$, то можем указать такие значения θ_1^* и θ_2^* , между которыми с вероятностью $1 - \epsilon$ находится случайная величина $\hat{\theta}$:

$$P(\theta_1^* < \hat{\theta} < \theta_2^* | \theta) = 1 - \epsilon.$$

Значения θ_1^* и θ_2^* зависят от θ : $\theta_1^* = h(\theta)$, $\theta_2^* = g(\theta)$. Определим обратные функции $\theta = h(\theta_1^*)$, $\theta = g(\theta_2^*)$. Тогда двойное неравенство $\theta_1^* < \hat{\theta} < \theta_2^*$ эквивалентно следующему:

$$g(\hat{\theta}) < \theta < h(\hat{\theta}). \quad (27)$$

В простейшем случае при $\theta_1^* = \theta - \delta$, $\theta_2^* = \theta + \delta$ это неравенство принимает вид $\hat{\theta} - \delta_2 < \theta < \hat{\theta} + \delta_1$. Соответственно вероятность неравенства (27) будет также равна $1 - \epsilon$, и таким образом определятся границы доверительного интервала для параметра θ .

Обратим внимание читателя, что при таком подходе не делается никакого предположения о том, является ли величина θ случайной или нет. В то же время из неравенства (27) как будто бы следует, что θ является случайной величиной. Необходимо, однако, иметь в виду, что случайными величинами в неравенстве (27) являются границы доверительного интервала, так что результат следует формулировать следующим образом: случайный интервал с границами θ_1^* , θ_2^* , зависящими от выборочных значений x_0 , покрывает истинное значение θ с заданной вероятностью.

§ 2. Статистическая постановка обратных геофизических задач

Покажем, что обратные задачи геофизики можно сформулировать так, чтобы свести их к задаче статистического оценивания параметров распределения. Зададим модель среды совокупностью k неизвестных и подлежащих определению параметров $m \equiv (m_1, m_2, \dots, m_k)$. Предполагается, что параметризованная таким способом модель достаточно хорошо соответствует изучаемой среде в том смысле, что вычисленные для нее значения геофизического поля $\Phi_i(m)$ в точках $i = 1, 2, \dots, n^*$, в которых производятся измерения, отличаются от соответствующих реальной среде на малое значение, которое можно рассматривать как случайную помеху. Таким образом, значения измеряемых величин y_i будут содержать ошибки, обусловленные как неадекватностью принятой модели реальной среде, так и погрешностями измерений.

Совокупность значений y_i образует конечную n -мерную случайную величину y . Ее функция распределения может быть известна, тогда она включает в себя неизвестные параметры среды m через посредство функций $\Phi_i(m)$; если же она неизвестна, то в каждой конкретной задаче исследуются ее характеристики, на основе которых можно судить о вероятности появления выборки y (см. далее § 3). Таким образом, пространство наблюдений Y можно трактовать как выборочное пространство размерностью n , каждому элементу которого соответствует некоторая вероятность.

Параметры модели при этом можно рассматривать как параметры распределения случайной величины y , по выборочным значениям которой, получаемым в результате наблюдений, требуется получить их оценку.

Кроме того, могут быть известны некоторые априорные сведения о параметрах изучаемой среды: одни параметры заданы точно, например глубина залегания осадочного чехла, если наблюдения проводятся вблизи глубокой скважины; для других лишь определены возможные пределы их изменения. Бывает так, что в разведкуемом районе уже проводились какие-то геолого-геофизические

* Под точками мы понимаем значения текущей координаты, в зависимости от которой определяется характеристика геофизического поля. Это могут быть координаты точек на земной поверхности, в которых производятся наблюдения, значения периода волны и т.д.

исследования, на основании которых известно, что значения некоторых из искомых параметров концентрируются около определенной величины. Тогда можно трактовать неизвестные параметры как случайные величины и характеризовать степень неопределенности наших знаний о том, какие значения может принимать параметр m_i , априорной плотностью вероятности $p_0(m_i)$.

Итак, в статистической постановке обратная задача геофизики может быть сформулирована следующим образом: задан случайный вектор $\mathbf{y} = (y_1, y_2, \dots, y_n)$ из выборочного пространства Y размерностью n с плотностью вероятности $p(\mathbf{y}; \mathbf{m})$, зависящей от параметров \mathbf{m} через посредство известной функции $\Phi(\mathbf{m})$. Могут быть заданы в той или иной форме априорные сведения об изучаемой среде. Требуется по известному вектору \mathbf{y} оценить параметры \mathbf{m} и определить погрешности найденных оценок.

Пример 1. Определение скоростного разреза по географу отраженных сейсмических волн. Пусть задана модель с горизонтальной отражающей границей и однородной покрывающей толщей. Времена пробега отраженной волны t_i в точках x_i по профилю при взрыве в точке $x = 0$ вычисляются по формуле $t_i = \Phi_i(h, x) = v^{-1} \sqrt{x_i^2 + 4h^2}$, где v — скорость волны в среде; h — глубина залегания отражающей границы.

Из наблюдений получают набор значений времен пробега $t_i^{\text{эксн}}$ в точках x_i , которые отличаются от вычисленных значений в тех же точках на случайную ошибку ϵ_i . Если можно считать, что эта ошибка распределена нормально с нулевым средним и дисперсией σ^2 , тогда случайная величина $t_i^{\text{эксн}}$ также распределена нормально со средним значением $\Phi_i(h, v)$ и дисперсией σ^2 :

$$p(t_i^{\text{эксн}}) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left[-\frac{1}{2\sigma^2} (t_i^{\text{эксн}} - \Phi_i(h, v))^2 \right].$$

Полная совокупность наблюдений $\mathbf{y} = (t_1^{\text{эксн}}, t_2^{\text{эксн}}, \dots, t_n^{\text{эксн}})$ будет иметь многомерное нормальное распределение

$$p(\mathbf{y}; \mathbf{m}) = \prod_{i=1}^n p(t_i^{\text{эксн}}) = \frac{1}{(\sqrt{2\pi\sigma})^n} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (t_i^{\text{эксн}} - \Phi_i(\mathbf{m}))^2 \right],$$

где $\mathbf{m} = (h, v)$.

Задача сводится к нахождению неизвестных параметров распределения v , h , σ . Первые два параметра являются характеристиками скоростного разреза, а последний — характеристикой помехи.

Пример 2. Определение удельного сопротивления в Земле по данным вертикального электрического зондирования. Модель Земли представим в виде горизонтально-слоистой среды. Каждый i -й слой ($i = 1, 2, \dots, N$) характеризуется мощностью (толщиной) h_i и сопротивлением ρ_i (для простоты среду будем считать изотропной). Тогда значения кажущегося сопротивления $\rho^{теор}(r_j, m)$ на r_j разnose можно вычислить по формуле [27]

$$\rho^{теор}(r_j, m) = \rho_1 \left[1 + r_j^2 \int_0^\infty B(m, \tau) J_1(\tau r_j) d\tau \right],$$

где $J_1(\tau r)$ - функция Бесселя первого рода, первого порядка; $B(m, \tau)$ - некоторая функция, зависящая от параметров среды $m = (h_i, \rho_i)$ и переменной интегрирования τ .

Пусть наблюдения $\rho_j^{ансп}$ будут отличаться от $\rho^{теор}(r_j, m)$ на величину случайной ошибки ξ_j , имеющей нормальное распределение с нулевым средним значением и дисперсией σ_j^2 . Предположим, что значения σ_j^2 не определены из эксперимента, но известно, что $\sigma_j^2 = \sigma_0^2 (\rho^{теор}(r_j, m))^2$. Тогда случайная величина $\rho_j^{ансп}$ также распределена нормально с дисперсией σ_j^2 , но со средним $\rho^{теор}(r_j, m)$.

Данная задача отличается от предыдущей тем, что здесь дисперсия случайной ошибки является функцией параметров среды. Многомерное распределение случайного вектора $\mathbf{y} = (\rho_1^{ансп}, \rho_2^{ансп}, \dots, \rho_n^{ансп})$ будет иметь вид

$$p(\mathbf{y}, m) = \frac{1}{(\sqrt{2\pi} \sigma_0)^n \prod_{i=1}^n \rho^{теор}(r_i, m)} \exp \left[-\frac{1}{2\sigma_0^2} \sum_{i=1}^n \left(\frac{\rho^{ансп} - \rho^{теор}(r_i, m)}{\rho^{теор}(r_i, m)} \right)^2 \right]. \quad (28)$$

Если число слоев равно N , то распределение $p(\mathbf{y}, m)$ имеет $2N - 1$ неизвестных параметров, из которых $2N - 1$ - параметры модели и один параметр σ_0^2 - параметр случайной помехи.

Пример 3. Определение продольного удельного сопротивления в Земле по данным магнитотеллурического зондирования [37]. Пусть в некоторой точке дневной поверхности проводится магнитотеллурическое зондирование (МТЗ). Исходными данными для решения обратной задачи МТЗ являются значения вещественной $x(r_i, m)$ и мнимой $y(r_i, m)$ частот комплексного входного импедан-

са $Z(T_i, \mathbf{m})$ ($i = 1, 2, \dots, n$; n - объем выборки; T_i - период электромагнитной волны).

Модель Земли, как и раньше, представим в виде горизонтально слоистой среды с известным числом слоев N . Тогда значения $x_0(T_i, \mathbf{m})$ и $y_0(T_i, \mathbf{m})$ для периода T_i можно получить из формулы для комплексного импеданса $Z_{0N}(5)$:

$$Z_{0N-s}(T_i, \mathbf{m}) = Z_{01}(T_i) \operatorname{cth} \left\{ h_{s+1} h_{s+1} + \operatorname{arclth} \sqrt{\rho_{s+1} / \rho_s} R_{N-(s+1)} \right\}, \quad (29)$$

где составляющие вектора $\mathbf{m} = (h_1, h_2, \dots, h_{N-1}, \rho_1, \dots, \rho_N)$. Все обозначения в формуле (29) имеют тот же смысл, что и в формуле (5). Будем полагать, что измеренные значения x_i и y_i отличаются от вычисленных $x_0(T_i, \mathbf{m})$, $y_0(T_i, \mathbf{m})$ на величины случайных ошибок ξ_i и η_i соответственно, которые независимы и распределены нормально с нулевым средним и одинаковой дисперсией σ^2 . Тогда совместную плотность распределения случайных величин x_i и y_i можно записать для данного i в виде

$$p(x_i, y_i; \mathbf{m}) = \frac{1}{2\pi\sigma^2} \exp \left[-\frac{[x_i - x_0(T_i, \mathbf{m})]^2 + [y_i - y_0(T_i, \mathbf{m})]^2}{2\sigma^2} \right],$$

а общую плотность распределения двух векторов $\mathbf{x} = (x_1, x_2, \dots, x_n)$ и $\mathbf{y} = (y_1, y_2, \dots, y_n)$ -

$$p(\mathbf{x}, \mathbf{y}, \mathbf{m}) = \frac{1}{(2\pi\sigma^2)^n} \exp \left\{ -\sum_{i=1}^n \frac{[x_i - x_0(T_i, \mathbf{m})]^2 + [y_i - y_0(T_i, \mathbf{m})]^2}{2\sigma^2} \right\} \quad (30)$$

Предположим теперь, что предварительно, например по данным вертикального электрического зондирования (ВЭЗ), были получены оценки этих же параметров $\hat{\mathbf{m}} = \{h_i, \rho_i\}$. Как уже говорилось в § 1, оценки параметров, полученные по данным случайной выборки, являются случайными и соответственно сами имеют определенное распределение. Это распределение оценок $\hat{\mathbf{m}}$ можно рассматривать как априорное распределение искомых параметров при решении задачи МЭЗ. Пусть плотность вероятности этих оценок $p_0(\hat{\mathbf{m}})$. Тогда можно построить плотность апостериорной вероятности значений \mathbf{m} , учитывающую информацию о параметрах, содержащуюся в данных МЭЗ:

$$p(\mathbf{m} / \mathbf{x}, \mathbf{y}) = p_0(\mathbf{m}) p(\mathbf{x}, \mathbf{y}; \mathbf{m}).$$

В этом случае оценка параметров \mathbf{m} может быть получена путем максимизации апостериорной плотности вероятности этих параметров.

Пример 4. Определение параметров геоэлектрического разреза в случае анизотропной горизонтально слоистой модели [30, 61]. Рассмотрим геоэлектрический разрез, состоящий из N горизонтальных анизотропных слоев, каждый из которых характеризуется продольным сопротивлением ρ_l^i , поперечным сопротивлением ρ_t^i , коэффициентом анизотропии $\lambda_i = \sqrt{\rho_l^i / \rho_t^i}$ ($i=1, 2, \dots, N$) и мощностью h_i ($i=1, 2, \dots, N-1$). Поскольку анизотропия существенно влияет на постоянное электрическое поле, то по исходным данным ВЭЗ можно оценить среднеквадратичное сопротивление анизотропного пласта $\rho^i = \sqrt{\rho_l^i \rho_t^i}$ и завышенную мощность $h_i^* = \lambda_i h_i$, в то же время методами индукционных зондирований определяется продольная проводимость слоя ρ_l^i и его истинная мощность h_i . Поэтому сведения о параметрах анизотропии можно получить, комплексирова данные зондирований на постоянном и переменном токе.

Пусть в изучаемой точке земной поверхности проводятся совместно ВЭЗ и МТЗ. В результате наблюдений образуется выборка данных ВЭЗ $y_1^{ВЭЗ}, y_2^{ВЭЗ}, \dots, y_{n_1}^{ВЭЗ}$ для разносов r_1, r_2, \dots, r_{n_1} (n_1 - объем выборки ВЭЗ) и выборка данных МТЗ - значения $x_1^{МТЗ}, y_1^{МТЗ}, x_2^{МТЗ}, \dots, x_{n_2}^{МТЗ}, y_{n_2}^{МТЗ}$ для периодов T_1, T_2, \dots, T_{n_2} (n_2 - объем выборки МТЗ). Поскольку экоперименты независимы, то с учетом (15) общая плотность вероятности векторов $y_1(y_1^{ВЭЗ}, \dots, y_{n_1}^{ВЭЗ})$ и $y_2(x_1^{МТЗ}, y_1^{МТЗ}, \dots, x_{n_2}^{МТЗ}, y_{n_2}^{МТЗ})$ будет

$$p(y_1, y_2; m) = p(y_1; m) p(y_2; m) = \left[(\sqrt{2\pi})^n G_{ВЭЗ}^{n_1} G_{МТЗ}^{2n_2} \times \right. \\ \left. \times \prod_{p=1}^{n_1} \Phi(r_p; m) \right]^{-1} \exp \left\{ -\frac{1}{2G_{ВЭЗ}^2} \sum_{p=1}^{n_1} \left(\frac{y_p^{ВЭЗ} - \Phi(r_p, m)}{\Phi(r_p, m)} \right)^2 - \right. \\ \left. - \frac{1}{2G_{МТЗ}^2} \sum_{q=1}^{n_2} \left[[x_q^{МТЗ} - x_0^{МТЗ}(T_q, m)]^2 + [y_q^{МТЗ} - y_0^{МТЗ}(T_q, m)]^2 \right] \right\},$$

где $n = n_1 + 2n_2$, $\Phi(r_p, m) = \rho_n^{top}(r_p, m)$; остальные обозначения имеют тот же смысл, что и в формулах (28), (30). Независимыми параметрами плотности вероятности $p(y_1, y_2; m)$ являются компоненты вектора $m = \{h_1, h_2, \dots, h_{N-1}, \rho_1^i, \dots, \rho_{N-1}^i, \rho_1^t, \dots, \rho_{N-1}^t, G_{ВЭЗ}^2, G_{МТЗ}^2\}$, из которых $G_{ВЭЗ}^2$ и $G_{МТЗ}^2$ - параметры распределения ошибок случайных векторов y_1 и y_2 , остальные - параметры анизотропного геоэлектрического разреза.

§ 3. Функции распределения данных геофизических наблюдений

Как было сказано в предыдущем параграфе, для решения обратной задачи геофизики можно использовать выводы теории статистического оценивания, если пространство наблюдений трактовать как выборочное пространство, каждой точке которого приписывается определенная вероятность, зависящая от параметров модели среды. Иначе говоря, каждая точка в пространстве наблюдений рассматривается как случайный вектор, характеризующийся определенной плотностью вероятности. Очевидно, что оценки параметров модели, получаемые на основе того или иного статистического критерия, будут зависеть от вида функции распределения исходных данных. Для получения решения, наиболее точно описывающего реальную среду, мы должны выбрать функцию распределения так, чтобы она по возможности лучшим образом отражала действительное распределение наблюдаемой величины. А для этого следует прежде всего понять, в чем состоит случайность вектора наблюдений y и какие факторы ее обуславливают.

Вследствие того, что мы всегда вынуждены идеализировать модель среды, нельзя добиться точного совпадения наблюдения y с полем, соответствующим модели, т.е. с $\Phi(m)$. Для любой модели m из пространства M всегда будет иметь место некоторое отклонение $e = y - \Phi(m)$, или невязка. Статистические методы интерпретации основаны на предположении, что невязку можно рассматривать как случайную помету (ошибку наблюдений).

Очевидно, что невязка будет включать в себя ошибки измерений и случайные ошибки, зависящие от условий эксперимента. И те и другие могут быть существенно уменьшены путем многократного повторения наблюдений. Но невязка при этом никаким образом не может быть сведена к нулю из-за того, что реальная среда не соответствует принятой для нее модели. Это несоответствие модели реальной среде и приводит к тому, что в данных наблюдений всегда будет присутствовать неучтенная компонента. По своей природе эта компонента неслучайна, так как обусловлена особенностями среды. Поэтому если эксперимент будет повторяться в идентичных условиях (на одном и том же профиле, для одних и тех же источников и т.д.), то будут получаться одни и те же данные, и, следовательно, нельзя будет исключить эту компонен-

ту поля. Однако если распределение невязок в последовательных точках, в которых производится наблюдения, не имеет систематического характера, то невязки можно рассматривать как случайные, т.е. обусловленные такими элементами среды (или источника поля), которые нельзя описать в рамках модели и которые мы вынуждены принять случайными. Если же систематическое отклонение выявляется, то его обычно можно исключить, введя соответствующие осложнения в модель.

Очевидно, что вместо распределения вектора y можно рассматривать распределение невязки $e = y - \Phi(m)$. В общем случае параметры распределения невязки могут зависеть от параметров модели m . Задача, таким образом, сводится к определению в каждом конкретном случае вида распределения e . Как и в других задачах, где приходится иметь дело со случайными величинами, для определения функции распределения здесь могут быть использованы два подхода: построение эмпирической кривой распределения и аналитический вывод функции распределения, если известны основные факторы, определяющие ошибку, и могут быть приняты более или менее правдоподобные гипотезы о рассматриваемом явлении.

Во многих случаях оказывается, что для e допустимо принимать нормальное распределение с нулевым средним значением. Учет того, что истинное распределение отличается от нормального, важен, как мы покажем далее, в основном только тогда, когда ошибки велики, т.е. велика дисперсия наблюдений. При малых ошибках многие распределения в предельном случае переходят в нормальное.

Рассмотрим некоторые примеры.

Пример 1. Распределение амплитуд продольной волны от землетрясений. На характер распределения амплитуд волн от землетрясений, наблюдаемых на земной поверхности, влияют два основных фактора: неравномерность излучения волн очагом и наличие локальных неоднородностей в распределении скорости и плотности в Земле. Отвлечемся от второго фактора и рассмотрим влияние только неравномерности излучения на разброс данных об амплитудах. Будем считать, что очаг землетрясения можно моделировать двойным диполем, так что функция направленности излучения имеет вид

$$B(\vartheta, \varphi) = \sin^2 \vartheta |\sin 2\varphi|,$$

где ϑ , φ - угловые координаты в системе x, y, z , связанной с ориентацией плоскости разрыва и направлением подвиги в очаге. Естественно предположить, что направления осей в источнике (т.е. ориентация координатной системы x, y, z) равновероятны. Это соответствует выборкам данных от землетрясений в разных районах, полученных на разных станциях. В работе [11] на основе этих предположений выведено выражение для функции распределения логарифма приведенной амплитуды продольной волны, исправленной за магнитуду землетрясения M):

$$A^* = \log A_{\text{навл}} - M_n.$$

Магнитуда M_n определяется путем осреднения данных об амплитудах, полученных на n станциях, и, таким образом, она сама является случайной величиной. На рис. 7, а приведены кривые плотности вероятности для разных значений n , а на рис. 7, б - гистограмма ошибок A^* (кривая 1), построенная по данным дальневосточных землетрясений в интервале эпицентральных расстояний 600-1200 км, и теоретическая кривая плотности вероятности для $n = 6$ (кривая 2). Сравнительно хорошее согласие эмпирического распределения и теоретической кривой говорят о том, что исходное предположение о причине разброса данных об амплитудах является оправданным.

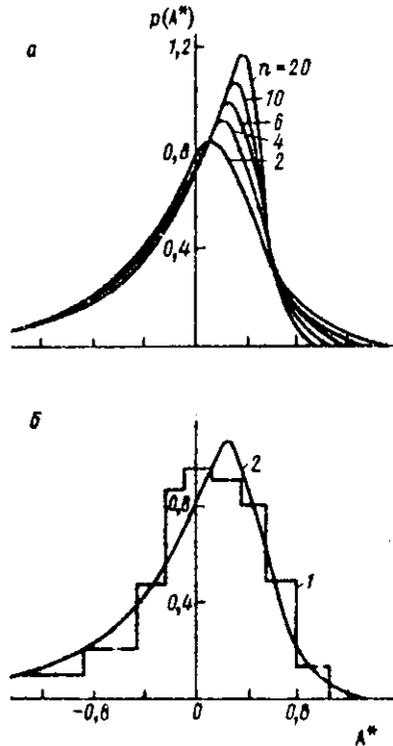


Рис. 7.

Отметим, что дисперсия амплитудных данных в этом примере чрезвычайно большая. Чтобы использовать амплитудную кривую для нахождения распределения скорости и добротности в Эсмье, необходимо иметь данные об амплитудах с точностью не ниже 0,1, поэтому в данном случае очень важно иметь максимально точное представление о характере распределения наблюдений.

Пример 2. Распределение времен пробега волн от землетрясений. Г. Джеффрис [59] при построении таблиц времен пробега показал, что распределение наблюдаемых времен вступления волн отличается от нормального относительно более высокой вероятностью больших ошибок. Для описания такого распределения он принял закон в виде следующей суммы:

$$p(t) = \frac{1-q}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(t-t_0)^2}{2\sigma^2}\right] + qg(t), \quad (31)$$

где $g(t)$ — такая функция, о которой известно только, что она плавно изменяется в области примерно нескольких σ в обе стороны от t_0 и $\int_{-\infty}^{+\infty} g(t) dt = 1$.

Параметры этого распределения могут быть оценены по эмпирической кривой распределения.

Если в первом приближении принять $g = \text{const}$ внутри некоторого конечного, хотя и достаточно широкого интервала, а вне этого интервала $g = 0$, то при больших значениях невязок $p(t) \approx qg$, так что значение qg оценивается как средний уровень эмпирического распределения при больших отклонениях от среднего значения (обозначим его через p_{∞}). А при малых отклонениях, т.е. в окрестности максимума распределения,

$$p_0 \approx qg + \frac{1-q}{\sqrt{2\pi}\sigma} = p_{\infty} + \frac{1-q}{\sqrt{2\pi}\sigma}.$$

Величину σ можно оценить по характеру распределения невязок в окрестности максимума. Если число невязок в интервале $(t_0 - \alpha, t_0 + \alpha)$ равно N_{α} , а в интервале $(t_0 - \beta, t_0 + \beta)$ равно N_{β} , то при малых α и β

$$\frac{N_{\alpha}}{N_{\beta}} \approx \frac{F(\alpha/\sigma) - F(-\alpha/\sigma)}{F(\beta/\sigma) - F(-\beta/\sigma)},$$

здесь $F(x)$ — интеграл вероятности. Из этого соотношения с помощью таблиц нормального распределения оценивается σ .

Характер распределения времен пробега волны $t = t_{\text{набл}} - t_{0\text{ч}}$ определяется природой ошибки $\delta t = t - t_0$. Дело в том, что наблюдаемая ошибка δt складывается из целого ряда различных ошибок: вследствие неточного определения времени в очаге ($t_{0\text{ч}}$); в результате ошибки в определении эпицентрального расстояния; вследствие влияния локальных неоднородностей среды на пути очаг - станция; и, наконец, из-за ошибки интерпретации, обусловленной неверным выделением фазы на сейсмограмме или неверным отсчетом времени. Интерпретационная ошибка может быть довольно большой по сравнению со всеми остальными, но она присутствует не всегда, а с некоторой вероятностью $q < 1$. Это приводит к тому, что в выражении для плотности вероятности ошибки появляется второе слагаемое. Но если из выборки данных о времени пробега исключить такие значения, которые характеризуются большим отклонением от среднего, то влияние интерпретационной ошибки на характер распределения наблюдений можно практически свести к нулю, так что распределение оставшихся данных будет приблизительно нормальным. Такой способ отбрасывания больших ошибок часто применяют для того, чтобы можно было проводить интерпретацию в предположении о нормальности распределения.

Пример 3. Распределение фазы и амплитуды импеданса.

При МТЗ регистрируются естественные электрическое и магнитное поля Земли. Исходными данными для решения обратной задачи являются значения входного импеданса Z , который для горизонтально однородной среды равен E_x/H_y . Поскольку вещественная и мнимая части импеданса x и y определяются из первичных наблюдений методом наименьших квадратов, то можно считать, что величины x и y распределены нормально со средними значениями x_0 и y_0 и одной и той же дисперсией σ^2 . При интерпретации же используются модуль импеданса $R = \sqrt{x^2 + y^2}$ (амплитуда), или фаза $\varphi = \arctg(y/x)$, или совместно R и φ . Рассмотрим, каково будет распределение амплитуды и фазы [31].

Совместная плотность вероятности величин x и y определяется выражением

$$p(x, y) = \frac{1}{2\pi\sigma^2} \exp \left\{ -\frac{(x-x_0)^2 + (y-y_0)^2}{2\sigma^2} \right\},$$

а совместная плотность вероятности R и φ - выражением

$$p(R, \varphi) = p(x(R, \varphi), y(R, \varphi)) \left| \frac{\partial(x, y)}{\partial(R, \varphi)} \right|, \quad (32)$$

где $\left| \frac{\partial(x, y)}{\partial(R, \varphi)} \right|$ - якобиан преобразования от x, y к R, φ , равный

$$\begin{vmatrix} \cos \varphi & -R \sin \varphi \\ \sin \varphi & R \cos \varphi \end{vmatrix} = R.$$

Выражение (32) преобразуется к виду

$$p(R, \varphi) = \frac{R}{2\pi G^2} \exp\left(-\frac{R^2 + R_0^2}{2G^2}\right) \exp\left(\frac{RR_0 \cos(\varphi - \varphi_0)}{G^2}\right), \quad (33)$$

где $R_0 = \sqrt{x_0^2 + y_0^2}$, $\varphi_0 = \arctg(x_0/y_0)$.

Распределение амплитуды может быть получено путем интегрирования (36) по интервалу изменения φ , т.е. от 0 до 2π :

$$\begin{aligned} p(R) &= \frac{R}{2\pi G^2} \exp\left(-\frac{R^2 + R_0^2}{2G^2}\right) \int_0^{2\pi} \exp\left(\frac{RR_0 \cos(\varphi - \varphi_0)}{G^2}\right) d\varphi = \\ &= \frac{R}{G^2} \exp\left(-\frac{R^2 + R_0^2}{2G^2}\right) I_0\left(\frac{RR_0}{G^2}\right), \quad R > 0, \end{aligned} \quad (34)$$

где $I_0(x)$ - модифицированная функция Бесселя. А распределение фазы получается в результате интегрирования (33) по области изменения R , т.е. от 0 до ∞ :

$$\begin{aligned} p(\varphi) &= \frac{1}{2\pi G^2} \exp\left(-\frac{R_0^2}{2G^2}\right) \int_0^{\infty} \exp\left[-\frac{R^2 - 2RR_0 \cos(\varphi - \varphi_0)}{2G^2}\right] dR = \\ &= \frac{\exp(-1/2G_0^2)}{2\pi} \left\{ 1 + \sqrt{\frac{\pi}{2}} \exp\left(\frac{u^2}{2G_0^2}\right) \left(\frac{u}{G_0}\right) \left[1 + F\left(\frac{u}{G_0}\right) \right] \right\}, \end{aligned} \quad (35)$$

где $G_0 = G/R_0$, $u = \cos(\varphi - \varphi_0)$, $F(z) = \sqrt{2/\pi} \int_0^z \exp(-t^2/2) dt$ - интеграл вероятностей.

Распределение амплитуды (34) является обобщенным рэлеевским распределением. В случае $R_0/G \gg 1$, если R не слишком мало по сравнению с G , функцию Бесселя можно заменить ее асимптотическим представлением

$$f_0\left(\frac{RR_0}{\sigma^2}\right) \approx \frac{\sigma}{\sqrt{2\pi RR_0}} \exp \frac{RR_0}{\sigma^2}.$$

и тогда

$$p(R) = \frac{1}{\sqrt{2\pi\sigma}} \sqrt{\frac{R}{R_0}} \exp\left(-\frac{(R-R_0)^2}{2\sigma^2}\right).$$

Это распределение отличается от нормального наличием множителя $\sqrt{R/R_0}$. Но такое различие оказывается существенным только в области, где значения R маловероятны, т.е. при значениях R , достаточно далеких от R_0 . В окрестности же максимума плотности вероятности, т.е. при значениях R , близких к R_0 , это распределение можно считать нормальным.

Распределение фаз при малых σ_0 также допустимо считать нормальным. Действительно, при $\sigma_0 \ll 1$ первым членом в фигурных скобках можно пренебречь по сравнению со вторым, а второй член приближенно равен $(u/\sigma_0) \exp[u^2/(2\sigma_0^2)]$. Таким образом, распределение фаз при малых σ_0 принимает вид

$$p(\varphi) \approx \frac{1}{\sqrt{2\pi\sigma_0}} u \exp\left(-\frac{1-u^2}{2\sigma_0^2}\right). \quad (36)$$

В окрестности максимума, т.е. при φ , близких к φ_0 , можно принять, что $1-u^2 \approx (\varphi-\varphi_0)^2$, $u \approx 1$, и тогда распределение (36) переходит в нормальное:

$$p(\varphi) \approx \frac{1}{\sqrt{2\pi\sigma_0}} \exp\left(-\frac{(\varphi-\varphi_0)^2}{2\sigma_0^2}\right).$$

В случае, когда импеданс определяется с большой относительной погрешностью (это соответствует случаю больших периодов глобальной кривой зондирования, при которых значения R становятся очень малыми [22]), в выражении для плотности вероятности фазы (35) можно пренебречь вторым членом в фигурной скобке по сравнению с первым, а экспоненциальный множитель принять равным единице. При этом распределение фазы оказывается равномерным в интервале $0 < \varphi < 2\pi$: $p(\varphi) \approx (2\pi)^{-1}$.

Приведенные примеры показывают, что при малых ошибках распределение наблюдений в большинстве случаев можно считать нормальным. В некоторых случаях оно может быть искусственно

сделано нормальным, если из материала наблюдений исключить данные, характеризующиеся большими отклонениями от среднего. Но поскольку для решения обратных задач геофизики обычно используются достаточно качественные наблюдения, то их распределение можно практически всегда считать нормальным. Если же ошибки наблюдений принципиально невозможно уменьшить (как, например, при использовании данных об амплитудах сейсмических волн), то в этих случаях необходимо тщательное изучение распределения таких данных.

Сделаем еще некоторые замечания по поводу характера распределения той компоненты невязки, которая обусловлена неадекватностью модели среды. Как уже говорилось, эту компоненту невозможно учесть, и при выборе модели среды следует добиваться только того, чтобы ее можно было действительно считать случайной, т.е. чтобы она не имела систематического характера. Естественное требование к распределению этой компоненты состоит в том, что ее среднее значение должно быть равно нулю. Обычно просто принимают, что эта компонента распределена нормально с нулевым средним.

Таким образом, в целом распределение невязки можно считать нормальным, учитывая, что в большинстве случаев и случайная ошибка, входящая в невязку, оказывается распределенной по нормальному закону.

§ 4. Оценивание параметров методом максимального правдоподобия

Пусть вид распределения величин u_i , получаемых из наблюдений, известен, параметры этого распределения зависят от искоемых параметров среды m , а сведения об априорном распределении параметров отсутствуют. В этом случае, как уже говорилось в § 1, при неограниченном увеличении объема выборки данных наблюдений оценки максимального правдоподобия несмещены, эффективны (в классе асимптотически несмещенных оценок минимизируют средний квадрат ошибки) и имеют нормальное распределение со средним значением, равным истинному значению параметров, и ковариационной матрицей, обратной информацион-

62

ной матрице Фишера. Отметим, что для того чтобы оценки обладали указанными свойствами, необходимо увеличение количества информации по Фишеру с ростом объема выборки.

Определение функции правдоподобия было дано в § 1 (см. формулу (17)). Оценкой максимального правдоподобия является такая оценка, которая максимизирует функцию $L_y(m)$, или (что то же самое) $l_y(m) = \log L_y(m)$, называемую логарифмической функцией правдоподобия. Отыскание максимума функции $l_y(m)$ сводится к решению системы уравнений

$$\frac{\partial l_y(m)}{\partial m_q} = 0, \quad q = 1, 2, \dots, k, \quad (37)$$

которые называются уравнениями правдоподобия.

Эти уравнения в общем являются трансцендентными, и для их решения обычно используется метод последовательных приближений (метод Ньютона - Рафсона). При этом оказывается возможным на основании свойства (16) упрощать линейные уравнения, которые решаются на каждом шаге итераций.

Как правило, всегда есть возможность выбрать некоторое начальное приближение $m^{(0)}$, в окрестности которого функция $l_y(m)$, максимум которой отыскивается, была бы регулярна и унимодальна (имела бы единственный максимум). Разложим выражение (37) в ряд в окрестности $m^{(0)}$ и ограничимся первыми членами ряда:

$$\frac{\partial l_y(m)}{\partial m_q} = \left(\frac{\partial l_y(m)}{\partial m_q} \right)_{m^{(0)}} + \sum_{p=1}^k \frac{\partial^2 l_y(m)}{\partial m_q \partial m_p} \Delta m_p,$$

где $\Delta m_p = m_p - m_p^{(0)}$. При этом система (37) сведется к системе линейных уравнений относительно поправок Δm_p :

$$A \Delta m = b, \quad (38)$$

$$\text{где } b_q = \left(\frac{\partial l_y(m)}{\partial m_q} \right)_{m^{(0)}}, \quad A_{qp} = - \left(\frac{\partial^2 l_y(m)}{\partial m_q \partial m_p} \right)_{m^{(0)}}. \quad (39)$$

Используя полученную оценку $m^{(1)} = m^{(0)} + \Delta m$ в качестве нового начального приближения, мы можем определять поправку к $m^{(0)}$, решая снова систему уравнений (38), в которой b и A вычислены уже в точке $m = m^{(1)}$, и т.д. до сходимости процесса.

Решение системы (38) сопряжено с определенными вычислительными трудностями, связанными с тем, что на каждом шаге

приближений мы должны вычислять вторую производную функции правдоподобия по параметрам. А поскольку параметры входят в выражение для плотности распределения $p(y; m)$ через посредство оператора решения прямой задачи $\Phi(m)$, который в задачах геофизики часто достаточно сложен и иногда даже не допускает аналитического представления, вычисление вторых производных $\partial^2 L_y(m) / \partial m_p \partial m_q$ сводится к вычислению вторых производных $\partial^2 \Phi_i(m) / \partial m_p \partial m_q$. Для того чтобы обойти эту трудность, поступают следующим образом.

Поскольку в результате решения системы (38) находятся не окончательная оценка, максимизирующая функцию правдоподобия, а лишь некоторое приближение, то вполне допустимо несколько изменить матрицу системы (38) так, чтобы процесс последовательных приближений остался сходящимся к решению (37), хотя приближения на каждом шаге и оказались бы отличны от тех, которые получаются в результате последовательных решений системы (38). Важно лишь, чтобы левые части этой системы сохраняли выражение (39) и в процессе последовательных приближений стремились бы к нулю.

Так как исходные наблюдения, используемые для построения функции правдоподобия, являются случайной выборкой из генеральной совокупности, имеющей распределение $p(y; m)$, то величину $\partial^2 L_y(m) / \partial m_p \partial m_q$, являющуюся выборочным значением $\partial^2 \log p(y; m) / \partial m_p \partial m_q$, можно приближенно заменить ее математическим ожиданием. А согласно формуле (16)

$$-E_y \left(\frac{\partial^2 \log p(y; m)}{\partial m_p \partial m_q} \right) = E_y \left(\frac{\partial \log p(y; m)}{\partial m_p} \frac{\partial \log p(y; m)}{\partial m_q} \right) = E_y \left(\frac{\partial L_y(m)}{\partial m_p} \frac{\partial L_y(m)}{\partial m_q} \right). \quad (40)$$

Поэтому в качестве элементов матрицы системы (38) можно использовать выражения, определяемые правой частью выражения (40).

Покажем, насколько такая замена упрощает систему (38).

П р и м е р. Определение параметров геоэлектрического разреза по данным МТЗ. Исходными данными в общем случае являются значения амплитуды R_i и фазы φ_i комплексного импеданса Z_i , соответствующие периодам электромагнитной волны T_i ($i = 1, 2, \dots, n$). Параметры разреза - мощности слоев h_j ($j = 1, 2, \dots, N - 1$) и удельные сопротивления ρ_j ($j = 1, 2, \dots, N$) - об-

разуют $(2N - 1)$ -мерный вектор \mathbf{m} . В § 3 было приведено выражение для совместной плотности амплитуды и фазы импеданса (см. формулу (33)), воспользуемся им. Считая, что наблюдения на разных периодах являются независимыми, мы можем записать функцию правдоподобия в виде следующего произведения:

$$L_{R,\varphi}(\mathbf{m}) = \prod_{i=1}^n p(R_i, \varphi_i; \mathbf{m}),$$

тогда логарифмическая функция правдоподобия будет иметь вид

$$l_{R,\varphi}(\mathbf{m}) = \sum_{i=1}^n \log \frac{R_i}{2\pi\sigma^2} - \frac{1}{2\sigma^2} \sum_{i=1}^n \left\{ (R_i - R_{0i})^2 + 2R_i R_{0i} [1 - \cos(\varphi_i - \varphi_{0i})] \right\},$$

где R_i, φ_i - значения амплитуды и фазы, полученные из наблюдений; R_{0i}, φ_{0i} - значения амплитуды и фазы, вычисленные для модели с параметрами $\mathbf{m} = (h_1, h_2, \dots, h_{N-1}, \varphi_1, \varphi_2, \dots, \varphi_N)$.

Правые части уравнений (38) будут иметь вид

$$\frac{\partial l_{R,\varphi}(\mathbf{m})}{\partial m_q} = -\frac{1}{\sigma^2} \sum_{i=1}^n \left\{ [R_{0i} - R_i \cos(\varphi_i - \varphi_{0i})] \frac{\partial R_{0i}}{\partial m_q} - R_i R_{0i} \sin(\varphi_i - \varphi_{0i}) \frac{\partial \varphi_{0i}}{\partial m_q} \right\}. \quad (41)$$

Коэффициенты системы A_{pq} определим в соответствии с формулой (40). Учитывая, что наблюдения независимы, получим

$$A_{pq} = E_{R,\varphi} \left(\frac{\partial l}{\partial m_p} \frac{\partial l}{\partial m_q} \right) = E_{R,\varphi} \left\{ \sum_{i=1}^n (R_{0i} - R_i \cos(\varphi_i - \varphi_{0i}))^2 \frac{\partial R_{0i}}{\partial m_q} \frac{\partial R_{0i}}{\partial m_p} - R_i R_{0i} \sin(\varphi_i - \varphi_{0i}) [R_{0i} - R_i \cos(\varphi_i - \varphi_{0i})] \left(\frac{\partial R_{0i}}{\partial m_q} \frac{\partial \varphi_{0i}}{\partial m_p} + \frac{\partial R_{0i}}{\partial m_p} \frac{\partial \varphi_{0i}}{\partial m_q} \right) + R_i^2 R_{0i}^2 \sin^2(\varphi_i - \varphi_{0i}) \frac{\partial \varphi_i}{\partial m_q} \frac{\partial \varphi_{0i}}{\partial m_p} \right\}. \quad (42)$$

При вычислении математического ожидания учтем следующие соотношения:

$$\begin{aligned} E_{\varphi}(\sin(\varphi_i - \varphi_{0i})) &= 0, \\ E_{\varphi}(\sin(\varphi_i - \varphi_{0i}) \cos(\varphi_i - \varphi_{0i})) &= 0, \\ E_{R,\varphi}(R_{0i} R_i \cos(\varphi_i - \varphi_{0i})) &= E_{x,y}(x_i x_{0i} + y_i y_{0i}) = x_{0i}^2 + y_{0i}^2 = R_{0i}^2, \\ E_{R,\varphi}(R_i^2 \sin^2(\varphi_i - \varphi_{0i})) &= \\ &= E_{x,y}(x_i^2 \sin^2 \varphi_{0i} - 2x_i y_i \sin \varphi_{0i} \cos \varphi_{0i} + y_i^2 \cos^2 \varphi_{0i}) = \sigma^2, \\ E_{R,\varphi}(R_i^2 \cos^2(\varphi_i - \varphi_{0i})) &= \\ &= E_{x,y}(x_i^2 \cos^2 \varphi_{0i} + 2x_i y_i \sin \varphi_{0i} \cos \varphi_{0i} + y_i^2 \sin^2 \varphi_{0i}) = R_{0i}^2 + \sigma^2. \end{aligned} \quad (43)$$

Подставляя эти соотношения в правую часть формулы (42) и выполняя элементарные преобразования, получим

$$A_{pq} = \frac{1}{G^2} \sum_{i=1}^n \left(\frac{\partial R_{0i}}{\partial m_q} \frac{\partial R_{0i}}{\partial m_p} + R_{0i}^2 \frac{\partial \varphi_{0i}}{\partial m_q} \frac{\partial \varphi_{0i}}{\partial m_p} \right). \quad (44)$$

Выражения (41) и (44) можно упростить, если $\frac{\partial R_0}{\partial m_q}$ и $\frac{\partial \varphi_0}{\partial m_q}$ выразить через производные импеданса $\frac{\partial Z_0}{\partial m_q}$. Принимая во внимание, что $R_0 = \sqrt{Z_0 Z_0^*}$, а $\varphi_0 = -i(\ln Z_0 - \ln Z_0^*)/2$ (звездочкой обозначены комплексно сопряженные величины), найдем

$$\begin{aligned} \frac{\partial R_0}{\partial m_q} &= |Z'_q| \cos(\psi_q - \varphi_0), \\ \frac{\partial \varphi_0}{\partial m_q} &= \frac{|Z'_q|}{R_0} \sin(\psi_q - \varphi_0), \end{aligned} \quad (45)$$

где $Z'_q = \frac{\partial Z_0}{\partial m_q}$; $|Z'_q|$ и ψ_q - модуль и фаза производной импеданса по параметру m_q . Подставляя формулы (45) в выражения (41) и (44), получим

$$\begin{aligned} \delta_q &= -\frac{1}{G^2} \sum_{i=1}^n |Z'_{qi}| [R_{0i} \cos(\psi_{qi} - \varphi_{0i}) - R_i \cos(\varphi_i - \psi_{qi})], \\ A_{pq} &= \frac{1}{G^2} \sum_{i=1}^n |Z'_{pi}| |Z'_{qi}| \cos(\psi_{qi} - \psi_{pi}). \end{aligned}$$

Формулы для вычисления $\partial Z_0 / \partial m_q$ можно найти в работе [37].

В некоторых случаях уравнения правдоподобия (37) можно без особого ущерба для решения заменить более простыми приближенными уравнениями, учитывая особенности функции распределения. В качестве примера рассмотрим задачу о построении сейсмического голографа путем усреднения наблюдаемых времен пробега t_i в предположении, что они распределены в соответствии с законом Джеффриса (31). Такая задача не является в полном смысле обратной, так как здесь еще не ставится вопрос об оценке параметров среды, однако полученный в результате статистической обработки голограф можно использовать для определения скоростного разреза методом прямого обращения $l(\Delta) \rightarrow v(r)$. По-

этому для решения обратной задачи требуется построение годографа $l(\Delta)$ с максимально возможной точностью.

Для простоты будем считать, что приближенный годограф $l_0(\Delta)$ известен и требуется определить поправку к нему α , постоянную в некотором достаточно узком интервале эпицентральных расстояний, на основании данных о временах пробега в этом интервале. Функция правдоподобия в этом случае имеет вид

$$l(\alpha) = \sum_{i=1}^n \log \left\{ \frac{1-q}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(l_i - l_0(\Delta_i) - \alpha)^2}{2\sigma^2} \right] + qg \right\}.$$

Обозначим $l_i - l_0(\Delta_i) = \delta l_i$. Будем считать, что величины qg , σ^2 и $(1-q)(\sqrt{2\pi}\sigma)^{-1}$ определены из распределения $p(l)$, как было описано в § 3. Уравнение, определяющее максимум функции $l(\alpha)$, будет соответственно следующим:

$$\sum_{i=1}^n (\delta l_i - \alpha) \left\{ \sigma^2 \left[1 + \frac{qg\sqrt{2\pi}\sigma}{1-q} \exp \left(\frac{(\delta l_i - \alpha)^2}{2\sigma^2} \right) \right] \right\}^{-1}. \quad (46)$$

Обозначим $\mu = \frac{qg\sqrt{2\pi}\sigma}{1-q}$, $\frac{1}{w_i} = 1 + \mu \exp \left(\frac{(\delta l_i - \alpha)^2}{2\sigma^2} \right)$. (47)

Тогда уравнение (46) можно записать в виде

$$\sum_i (\delta l_i - \alpha) w_i = 0,$$

и оценкой α будет средневзвешенное значение $\sum w_i \delta l_i / \sum w_i$.

Обратим внимание читателя, что здесь весовые коэффициенты w_i сами зависят от неизвестного параметра α . Но в связи с тем, что эта зависимость не очень сильная, можно определять w_i приближенно, оценивая α в предположении нормальности распределения ($\hat{\alpha} = \sum \delta l_i / n$), и считать коэффициенты уже заранее известными. Такое допущение эквивалентно тому, что параметр α оценивается на условия минимизации взвешенной суммы квадратов отклонений $\sum_{i=1}^n (\delta l_i - \alpha)^2 w_i$, в которой весовые коэффициенты определены заранее и не зависят от оцениваемого параметра.

Зависимость коэффициента w от отклонения $x = \delta t - \alpha$ изображена на рис. 8 (жирная линия). Видно, что таким наблюдениям, которые мало отклоняются от среднего значения ($|x| \leq 2\sigma$), придается большой вес, а наблюдениям, характеризующимся значительными отклонениями ($|x| > 2\sigma$) - малый.

Часто для упрощения весовую функцию аппроксимируют П-образной функцией (тонкая линия): наблюдения с существенными отклонениями просто исключаются из рассмотрения, а остальные берутся с равными весами. Этот метод усечения данных используется тогда, когда известно, что наблюдения содержат большее количество значительных отклонений, чем должно было бы быть в предположении о нормальности распределения.

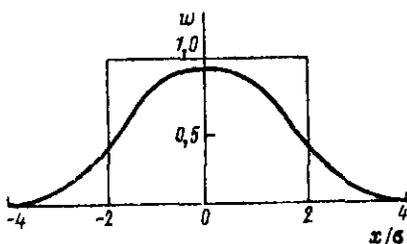


Рис. 8.

Рассмотрим теперь наиболее часто встречающийся на практике случай, когда наблюдения y распределены по нормальному закону со средним значением $\Phi(m)$ и ковариационной матрицей R_y , элементы которой не зависят от искомого параметра m . Как уже указывалось в § 3, большинство реальных распределений сводится к нормальному, если дисперсии наблюдений достаточно малы.

Функция правдоподобия в этом случае имеет вид

$$L_y(m) = -\frac{1}{2} \left\{ \log |(2\pi)^T R_y| + [y - \Phi(m)]^T R_y^{-1} [y - \Phi(m)] \right\}. \quad (48)$$

Вектор правых частей системы (38) будет иметь компоненты

$$\theta_q = \frac{\partial L}{\partial m_q} = \left(\frac{\partial \Phi}{\partial m_q} \right)^T R_y^{-1} [y - \Phi(m)] = [y - \Phi(m)]^T R_y^{-1} \frac{\partial \Phi(m)}{\partial m_q},$$

а элементы матрицы A_{pq} определяются в соответствии с формулой (40) следующим образом:

$$\begin{aligned} A_{pq} &= E_y \left(\frac{\partial L}{\partial m_p} \frac{\partial L}{\partial m_q} \right) = \\ &= E_y \left\{ \left(\frac{\partial \Phi(m)}{\partial m_p} \right)^T R_y^{-1} [y - \Phi(m)] [y - \Phi(m)]^T R_y^{-1} \left(\frac{\partial \Phi(m)}{\partial m_q} \right) \right\} = \end{aligned}$$

$$= \left(\frac{\partial \Phi(\mathbf{m})}{\partial m_p} \right)^T \mathbf{R}_y^{-1} \mathbf{E}_y \{ [\mathbf{y} - \Phi(\mathbf{m})][\mathbf{y} - \Phi(\mathbf{m})]^T \} \mathbf{R}_y^{-1} \left(\frac{\partial \Phi(\mathbf{m})}{\partial m_q} \right).$$

Учитывая, что $\mathbf{E}_y \{ [\mathbf{y} - \Phi(\mathbf{m})][\mathbf{y} - \Phi(\mathbf{m})]^T \} = \mathbf{R}_y$, получим для A_{pq} следующее выражение:

$$A_{pq} = \left(\frac{\partial \Phi(\mathbf{m})}{\partial m_p} \right)^T \mathbf{R}_y^{-1} \left(\frac{\partial \Phi(\mathbf{m})}{\partial m_q} \right).$$

Таким образом, поправки $\Delta \mathbf{m}^{(r)}$ на каждом r -м шаге итеративного процесса можно определять по формуле

$$\Delta \mathbf{m}^{(r)} = (\mathbf{B}^T \mathbf{R}_y^{-1} \mathbf{B})^{-1} \mathbf{B}^T \mathbf{R}_y^{-1} [\mathbf{y} - \Phi(\mathbf{m}^{(r-1)})],$$

где $B_{iq} = \left(\frac{\partial \Phi_i(\mathbf{m})}{\partial m_q} \right)_{\mathbf{m}^{(r-1)}}$.

В следующем параграфе при рассмотрении решения обратной задачи методом наименьших квадратов, к которому сводится метод максимального правдоподобия в случае нормального распределения исходных данных, будет показано, что такое же выражение для поправки получается, если разлагать в ряд не $\frac{\partial l_y(\mathbf{m})}{\partial m_q}$, а функцию $\Phi(\mathbf{m})$, т.е. линеаризовать исходную задачу.

Часто оказывается, что логарифмическая функция правдоподобия содержит слагаемые или множители, не зависящие от искомым параметров \mathbf{m} . В выражении (48) таким является первое слагаемое. Очевидно, что при нахождении максимальных значений функции $L_y(\mathbf{m})$ эти члены не играют роли и могут быть отброшены. Функция $\lambda(\mathbf{m})$, полученная путем отбрасывания таких членов, называется целевой функцией, или функцией отклика. В некоторых работах [14, 37] ее называют функцией достаточного признака. Обычно $\lambda(\mathbf{m})$ берут с обратным знаком и определяют оценку $\hat{\mathbf{m}}$ из условия минимизации целевой функции.

Как увидим далее, все статистические методы сводятся к нахождению экстремума той или иной функции отклика, или (что то же самое) к решению системы уравнений $\frac{\partial \lambda(\mathbf{m})}{\partial m_q} = 0$. Обычно эта система решается описанным ранее способом последовательных приближений, но в некоторых случаях такой подход к решению экстремальной задачи встречает определенные трудности. Одна из

них связана с необходимостью вычисления производных $\frac{\partial \Phi(m)}{\partial m_q}$. Если функция $\Phi(m)$ не допускает аналитического представления, то производные $\frac{\partial \Phi(m)}{\partial m_q}$ приходится определять путем численного дифференцирования, что при большом числе параметров сильно увеличивает объем вычислений. Вторая трудность заключается в том, что функция отклика может иметь не один, а несколько минимумов. И если априорных сведений, используемых для выбора начального приближения, окажется недостаточно, то полученное решение может соответствовать не абсолютному минимуму, а одному из относительных. В этом случае приходится искать решение из разных начальных точек, выбирая их из пространства параметров случайно или по некоторому закону, как это будет показано в § 7. Кроме того, рельеф функции отклика может быть настолько сложным, что метод Ньютона – Рафсона даже при достаточно хорошем выборе начального приближения будет расходиться. Тогда следует использовать другие методы поиска экстремума функции, которые также будут описаны в § 7.

§ 5. Оценивание параметров методом наименьших квадратов

При использовании метода максимального правдоподобия требуется знание математической формы распределения случайных исходных данных, так как задача оценивания параметров модели сводится к оцениванию параметров распределения. Однако часто тип распределения неизвестен, и чтобы его определить, требуется всякий раз проводить большой объем исследований, связанных с анализом случайных ошибок наблюдений.

Наиболее широкое распространение в практических задачах получил метод наименьших квадратов, который по своему подходу отличается от предыдущего простотой. Для его использования не требуется знать распределения случайных ошибок, достаточно иметь сведения о его первых двух моментах. Вместе с тем для случая линейной зависимости наблюдений от параметров метод наименьших квадратов в классе несмещенных линейных оценок минимизирует средний квадрат ошибок ($2\sigma^2$). С теорией метода

наименьших квадратов читатель может ознакомиться в курсах по математической статистике, например [8, 21, 33].

Рассмотрим вначале метод наименьших квадратов для линейных задач, т.е. таких, для которых $\Phi(\mathbf{m}) = \mathbf{A}\mathbf{m}$. Здесь \mathbf{A} - матрица $n \times k$ с известными коэффициентами, $\mathbf{m} \in \mathcal{M}^k$, $\mathbf{y} \in \mathcal{Y}^n$, $n > k$.

Будем считать, что невязки $\mathbf{e}_i = \mathbf{y}_i - \Phi_i(\mathbf{m})$ независимы, имеют нулевое среднее значение и одну и ту же дисперсию σ^2 , не зависящую от \mathbf{m} , т.е.

$$E(\mathbf{e}\mathbf{e}^T) = \sigma^2 \mathbf{I}. \quad (49)$$

Если оказывается, что невязки коррелированы или характеризуются разными дисперсиями, т.е. $E_{\mathbf{y}}(\mathbf{e}\mathbf{e}^T) = \mathbf{R}_{\mathbf{e}}$ ($\mathbf{R}_{\mathbf{e}}$ - некоторая ковариационная матрица, отличная от (49)), то преобразованием

$$\mathbf{y}' = \mathbf{R}_{\mathbf{e}}^{-1/2} \mathbf{y} \quad (50)$$

можно перейти к новым исходным данным \mathbf{y}' , для которых ковариационная матрица равна \mathbf{I} .

Метод наименьших квадратов заключается в том, что в качестве оценок параметров \mathbf{m} принимаются такие, которые минимизируют сумму квадратов невязок:

$$S = \sum_{i=1}^n \mathbf{e}_i^2 = \mathbf{e}^T \mathbf{e} = (\mathbf{y} - \mathbf{A}\mathbf{m})^T (\mathbf{y} - \mathbf{A}\mathbf{m}). \quad (51)$$

Оценки $\hat{\mathbf{m}}_{\text{н.к.}}$, удовлетворяющие условию (51), называются оценками наименьших квадратов. Минимизация S эквивалентна реше-

нию системы уравнений $\frac{\partial S}{\partial m_i} = 0$ ($i = 1, 2, \dots, k$), или

$$\mathbf{A}^T \mathbf{A} \mathbf{m} = \mathbf{A}^T \mathbf{y}. \quad (52)$$

Уравнения (52) называются нормальными уравнениями. Предполагая, что матрица $\mathbf{A}^T \mathbf{A}$ невырождена, мы можем записать решение системы нормальных уравнений в виде

$$\hat{\mathbf{m}}_{\text{н.к.}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}. \quad (53)$$

Случай вырожденной матрицы будет рассмотрен в гл. 7. Если ковариационная матрица ошибок наблюдений отличается от выражения (49), то, учитывая преобразование (50), легко показать, что оценка наименьших квадратов для этого случая будет иметь вид

$$\hat{\mathbf{m}}_{\text{н.к.}} = (\mathbf{A}^T \mathbf{R}_{\mathbf{e}}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{R}_{\mathbf{e}}^{-1} \mathbf{y}.$$

Оценки наименьших квадратов обладают следующими важными свойствами, не зависящими от распределения ошибок исходных наблюдений:

1. Оценки наименьших квадратов несмещены, т.е.

$$E(\hat{m}) = m. \quad (54)$$

Действительно, принимая во внимание, что $E(\varepsilon) = 0$, и подставляя в левую часть формулы (54) выражение для оценки (56), получим

$$E(\hat{m}) = E[(A^T A)^{-1} A^T (A m + \varepsilon)] = m + E(\varepsilon) = m.$$

2. Рассмотрим ковариационную матрицу оценок наименьших квадратов $R_{\hat{m}} = E[(\hat{m} - m)(\hat{m} - m)^T]$. Подставим формулу (53) в выражение (54) и учтем, что $E(\varepsilon) = 0$, тогда

$$\begin{aligned} R_{\hat{m}} &= E\{[(A^T A)^{-1} A^T (A m + \varepsilon) - m][(A^T A)^{-1} A^T (A m + \varepsilon) - m]^T\} = \\ &= (A^T A)^{-1} A^T E(\varepsilon \varepsilon^T) A (A^T A)^{-1}. \end{aligned}$$

Согласно формуле (49) $E(\varepsilon \varepsilon^T) = \sigma^2 I$, и таким образом

$$R_{\hat{m}} = \sigma^2 (A^T A)^{-1}. \quad (55)$$

Диагональные члены $R_{\hat{m}}$ являются дисперсиями оценок наименьших квадратов.

Второе свойство оценок состоит в том, что среди всех несмещенных оценок, являющихся линейными функциями исходных данных, оценки наименьших квадратов обладают минимальными дисперсиями. Это утверждение известно в литературе как теорема Гаусса - Маркова. Мы здесь не будем останавливаться на ее доказательстве, отсылая читателя к любому курсу по статистике [10, 23, 48].

3. Если дисперсия наблюдений σ^2 неизвестна, то ее можно оценить по формуле

$$\hat{\sigma}^2 = \frac{1}{n-k} (y - A\hat{m})^T (y - A\hat{m}).$$

Разность $y - A\hat{m}$ называют остаточным вектором, а $\hat{\sigma}^2$ - остаточной дисперсией. Эта оценка является несмещенной. Подставляя ее вместо σ^2 в формулу (55), можно получить несмещенную оценку для ковариационной матрицы оценок наименьших квадратов.

Описанные свойства оценок наименьших квадратов справедливы при любых, в том числе и малых, выборках исходных данных

(достаточно только выполнения условия $n > k$) и не зависят от закона распределения ошибок наблюдений. Однако, не зная закона распределения ошибок, нельзя ничего сказать о виде распределения оценок наименьших квадратов. Можно только утверждать, что если ошибки ε распределены нормально, то и оценки наименьших квадратов имеют нормальное распределение. Это следует из того, что оценки являются линейными функциями наблюдений, а любая линейная функция нормально распределенных величин распределена нормально.

Легко видеть, что в этом случае оценки наименьших квадратов совпадают с оценками максимального правдоподобия, так как

$$\log L_y(m) = -\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{A}m)^T(\mathbf{y} - \mathbf{A}m) - n \log \sigma - \frac{n}{2} \log(2\pi),$$

а поскольку σ не зависит от m , то нахождение максимума $\log L_y(m)$ эквивалентно минимизации $(\mathbf{y} - \mathbf{A}m)^T(\mathbf{y} - \mathbf{A}m)$. Заметим, что в этом случае кроме второго свойства, справедливого только в классе линейных и несмещенных оценок, оценки наименьших квадратов обладают еще и свойством, характерным для оценок максимального правдоподобия, т.е. они являются асимптотически эффективными в значительно более широком классе оценок.

В геофизике редко приходится иметь дело с линейными задачами. Как уже говорилось в § 3 гл. I, линейная зависимость наблюдений от параметров модели имеет место только в некоторых задачах гравиметрии и магнитометрии. Но к описанной процедуре линейного оценивания по методу наименьших квадратов сводятся и задачи оценивания в нелинейных моделях путем их линеаризации.

Рассмотрим теперь применение метода наименьших квадратов в нелинейных задачах и свойства получаемых оценок.

Пусть наблюдения являются нелинейными функциями параметров модели $\Phi(m)$. Так же, как и в случае линейной задачи, определим невязку $\varepsilon = \mathbf{y} - \Phi(m)$ и будем считать, что ее среднее значение равно нулю, а матрица ковариаций имеет вид (49).

Оценка наименьших квадратов определяется, как и в линейном случае, из условия минимизации суммы квадратов невязок:

$$S(m) = [\mathbf{y} - \Phi(m)]^T [\mathbf{y} - \Phi(m)]. \quad (56)$$

Если ϵ имеет нормальное распределение, то получаемая таким способом оценка совпадает с оценкой максимального правдоподобия. В § 3 отмечалось, что в большинстве задач геофизики предположение о нормальности распределения ϵ может считаться оправданным. Это является основанием для широкого использования метода наименьших квадратов при решении обратных задач геофизики. Оценки наименьших квадратов при этом обладают свойствами оценок максимального правдоподобия: состоятельностью и асимптотической эффективностью.

Если распределение ϵ не является нормальным, то оценки наименьших квадратов не обладают оптимальными свойствами. Несмотря на это, они часто используются благодаря простоте их построения.

В ряде случаев, когда распределение заведомо отличается от нормального, удается некоторым преобразованием исходных данных добиться нормальности распределения. Одним из примеров этого может служить усечение выборки путем отбрасывания данных, характеризующихся большими невязками (если, конечно, таких данных не слишком много). Как уже упоминалось в § 3, во многих случаях оказывается, что оставшиеся данные можно считать распределенными по нормальному закону. Другой пример - взвешивание весовых множителей к исходным невязкам, как было показано в § 4 для случая, когда невязка распределена по закону Дюффресса. Если вместо исходных невязок $y - \Phi(m)$ рассматривать взвешенные невязки $e'_i = (y_i - \Phi_i(m)) \sqrt{\omega_i}$, то их распределение при фиксированных весовых коэффициентах ω_i оказывается приблизительно нормальным с нулевым средним и равными дисперсиями, и так можно оценивать параметры на условия минимума сумм квадратов взвешенных невязок. Правда, в этом случае ω_i сама зависят от искомого параметра (см. формула (47)). Но оказывается, что полученная оценка слабо зависит от значений ω_i . Это позволяет оценивать веса приблизительно, используя для их вычисления оценки, полученные без учета весовых множителей.

Использование оценок наименьших квадратов оправдано и тогда, когда о распределении невязки неизвестно ничего, кроме ее среднего значения и каких-то корреляционных свойств. В этом случае, как уже говорилось в § 1, максимально антропийным за-

коном распределения является нормальный закон. И тогда оценкой, соответствующей нашим представлениям о свойствах исходного материала, будет оценка именно по методу наименьших квадратов.

Рассмотрим теперь процедуру построения оценок по методу наименьших квадратов в нелинейных задачах.

Наиболее широко распространенный подход к нахождению оценок, минимизирующих выражение (55), состоит в линеаризации исходного оператора $\Phi(m)$:

$$\Phi(m) = P(m^{(0)}) + A(m^{(0)}) \Delta m,$$

где

$$A_{ik}(m^{(0)}) = \left(\frac{\partial \Phi_i}{\partial m_k} \right)_{m^{(0)}}.$$

Рассматривая вместо y разность между исходными наблюдениями и соответствующими начальному приближению $m^{(0)}$, $\Delta y = y - \Phi(m^{(0)})$, получим, что Δy выражается линейно через поправку Δm :

$$\Delta y = A \Delta m + z.$$

Согласно предыдущему оценка $\Delta \hat{m}$ имеет вид

$$\Delta \hat{m} = (A^T A)^{-1} A^T \Delta y.$$

Нахождение оценки \hat{m} , минимизирующей выражение (56), производится методом последовательных приближений точно так же, как описано в предыдущем параграфе. Определение поправок $\Delta m^{(n)}$ производится до тех пор, пока значения $S(\hat{m}^{(n)})$ не стабилизируются, а поправки $\Delta m^{(n)}$ не станут достаточно малы.

Вычислительные трудности, связанные с реализацией этого метода, те же, что и при решении уравнений правдоподобия. Во-первых, начальное приближение $m^{(0)}$ должно быть выбрано настолько близким к решению, чтобы процесс последовательных приближений сходился. Во-вторых, на каждом шаге итераций необходимо вычислять производные $\frac{\partial \Phi(m)}{\partial m_j}$, что требует большого объема расчетов.

Другой подход к решению задачи состоит в отыскании экстремума $S(m)$ методами, не требующими вычисления производных, например такими, которые описаны в § 7.

Рассмотрим в качестве примера определение параметров скоростного разреза по дисперсионной кривой фазовой скорости волны

Лява. В качестве исходных данных взяты 30 значений фазовой скорости в интервале периодов $9 < T < 52$ с с шагом по частоте $\Delta\omega = 0,02$ рад/с, рассчитанные для модели, состоящей из двух однородных слоев ($i = 1, 2$) на полупространстве ($i = 3$) с параметрами, приведенными ниже:

i	1	2	3
θ_i , км/с	3,0	3,7	4,6
ρ_i , г/см ³	2,6	2,9	3,3
h_i , км	10	20	-

Поиск решения производился в классе однослойных моделей (однородный слой на полупространстве). Таким образом, невязка была обусловлена неадекватностью модели среды. В качестве начального приближения была взята модель со следующими параметрами: $\theta_1 = 3,4$ км/с, $\theta_2 = 4,4$ км/с, $\rho_1 = 2,8$ г/см³, $\rho_2 = 3,3$ г/см³, $h = 40$ км. Поскольку дисперсия слабо зависит от распределения плотности, то значения плотностей были фиксированы и определению подлежали только три параметра: θ_1 , θ_2 и h .

Решение, полученное в результате пяти итераций, оказалось следующим: $h = 22,6$ км, $\theta_1 = 3,23$ км/с, $\theta_2 = 4,57$ км/с. Скорость в полупространстве получилась близкой к истинной, скорость в слое — близкой к средней скорости в верхних двух слоях, а мощность слоя сильно отличается от суммарной мощности двух слоев (22,6 км вместо 30 км). При этом оказалось, что невязки, соответствующие окончательному решению, имеют систематический характер: они положительны на краях рассматриваемого интервала частот и отрицательны в его середине. Такое распределение невязок не соответствует исходному предположению о некоррелированности ошибок в исходных данных. Это несоответствие и привело к сильному смещению оценки одного из искомых параметров (мощности слоя). Таким образом, после построения решения всегда следует проверять остаточный вектор на соответствие исходным предположениям о характере распределения невязки.

§ 6. Оценка погрешности решения

Ограничимся рассмотрением погрешности только оценки максимального правдоподобия. Если оценивание производится по методу наименьших квадратов, то, как отмечалось в предыдущем параграфе, такую оценку можно рассматривать как оценку максимального правдоподобия в случае нормального распределения исходных данных. Таким образом, все дальнейшие выводы могут быть перенесены и на случай оценки наименьших квадратов.

Оценка максимального правдоподобия \hat{m} представляет собой такое значение параметра m , которое максимизирует вероятность получить заданную выборку наблюдений. Эта оценка не совпадает с истинным значением параметра m_0 , а находится в некоторой его окрестности. Поскольку в результате решения обратной задачи требуется на основе выборки наблюдений y дать заключение об истинном значении параметра m_0 , необходимо уметь строить такую окрестность оценки \hat{m} , в которой с заданной доверительной вероятностью может находиться истинное значение искомого параметра m_0 . Размеры этой окрестности будут характеризовать погрешность найденного решения.

Выводы о размерах и форме доверительной области можно сделать, исходя из поведения функции правдоподобия $L_y(m)$ в окрестности ее максимума. В $(k+1)$ -мерном пространстве $(m_1, m_2, \dots, m_k, L_y)$ эта функция изображается некоторой гиперповерхностью, выпуклой в окрестности максимума (рис. 9). Если провести сечение такой поверхности плоскостью $l = \tilde{l}$, то проекция этого сечения в пространстве M^k образует некоторую область (заштрихована), граница которой состоит из точек, обладающих одинаковым правдоподобием $l = \tilde{l}$, а внутри нее $l > \tilde{l}$. Такую область естественно принять в качестве доверительной области для истинного значения параметра. Но при таком подходе к нахождению доверительной области встает следующий вопрос: каков будет контур этой области? как определить величину \tilde{l} , соответствующую заданной доверительной вероятности?

Вообще говоря, если выборка наблюдений не слишком велика (точнее, если в выборке содержится недостаточное количество информации о параметрах), то граница области в пространстве параметров, определяемая уравнением $L_y(m) = \tilde{l}$, оказывается до-

статочной сложности. Поиск такой области может производиться лишь численно, путем вычисления значений $L_y(m)$ в окрестности \hat{m} . Такой способ будет описан в § 7.

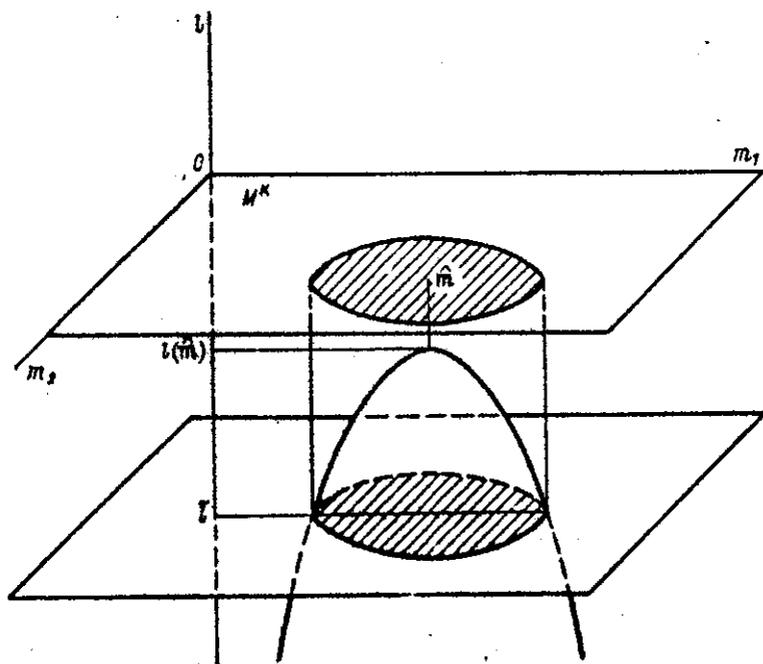


Рис. 9.

Для больших выборок ($n \gg k$) при условии, что общее количество информации о параметрах, определяемое детерминантом информационной матрицы Фишера J , достаточно велико, поверхность, характеризуемую функцией $L_y(m)$ в окрестности ее максимума, можно аппроксимировать k -мерным параболоидом:

$$L_y(m) \approx L_y(\hat{m}) + \frac{1}{2} (m - \hat{m})^T A (m - \hat{m}), \quad (57)$$

где элементы матрицы A $A_{ij} = -\left(\frac{\partial^2 L_y(m)}{\partial m_i \partial m_j}\right)_{\hat{m}}$. Поскольку $n \rightarrow \infty$,

оценка \hat{m} , соответствующая максимуму функции правдоподобия, ско-

дятся по вероятности к истинному значению параметра m_0 (см. § I), так что мы не внесем существенной ошибки, если будем вычислять элементы матрицы A не в точке \hat{m} , а при истинном значении параметра m_0 . Кроме того, благодаря тому, что объем выборки велик, можно вычислять A_{ij} не при конкретных значениях y_i , соответствующих этой выборке, а заменить их математическими ожиданиями [10]:

$$A_{ij} = E_y \left(\frac{\partial^2 \log p(y; m)}{\partial m_i \partial m_j} \right)_{m_0}$$

или, согласно формуле (16),

$$A_{ij} = -E_y \left(\frac{\partial \log p(y; m)}{\partial m_i} \frac{\partial \log p(y; m)}{\partial m_j} \right)_{m_0} = -J_{ij}.$$

Таким образом, формулу (57) можно записать в виде

$$L_y(m) \approx L_y(\hat{m}) - \frac{1}{2} (m - \hat{m})^T J (m - \hat{m}).$$

Отсюда следует, что область, характеризующаяся заданным правдоподобием L , будет представлять k -мерный эллипсоид, описываемый выражением

$$(m - \hat{m})^T J (m - \hat{m}) \leq \beta = 2(L_y(\hat{m}) - L). \quad (58)$$

Покажем теперь, что эта область является одновременно и доверительной областью для истинного значения параметра m_0 . С этой целью воспользуемся идеей метода построения доверительных интервалов, изложенной в § I. Вспомним, что оценка максимального правдоподобия \hat{m} асимптотически имеет нормальное распределение со средним значением, равным истинному значению параметра m_0 , и ковариационной матрицей $R = J_0^{-1}$. Для k -мерного нормального распределения всегда можно определить такое C_α , что оценка \hat{m} с заданной вероятностью $(1 - \alpha) \cdot 100\%$ будет внутри области, удовлетворяющей неравенству

$$(\hat{m} - m_0)^T R^{-1} (\hat{m} - m_0) < C_\alpha. \quad (59)$$

Если дисперсии наблюдений σ_y^2 известны (не подлежат оценкам одновременно с параметрами), то C_α выбирается по таблицам χ^2 -распределения с k степенями свободы, в противном случае — по таблицам $F_{k, n-k}$ -распределения Фишера с k , $n - k$ степенями свободы.

Сравнивая выражения (58) и (59), нетрудно видеть, что если в формуле (58) в качестве β взять $C_{\alpha, \beta}$ то область, описываемая этой формулой, будет доверительной областью для истинного значения параметра m_0 , т.е. с доверительной вероятностью $(1 - \alpha) \cdot 100\%$ истинное значение параметра m_0 будет находиться внутри этой области.

В частном случае, когда искомым параметр является скалярной величиной ($k = 1$), доверительный эллипсоид превращается в доверительный интервал. Дисперсия оценки в этом случае рассчитывается по формуле $\hat{C}^2 = \hat{C}_y^2 \left[\sum_{i=1}^n \left(\frac{\partial \Phi_i}{\partial m} \right)_{\hat{m}}^2 \right]^{-1}$ (\hat{C}_y^2 оценивается одновременно с m). Используя это значение \hat{C}^2 , можно построить доверительный интервал для параметра m следующим образом:

$$\hat{m} - \tau_{\alpha, n-1} \hat{C} < m < \hat{m} + \tau_{\alpha, n-1} \hat{C}, \quad (60)$$

где $\tau_{\alpha, n-1}$ — коэффициент Стьюдента для доверительной вероятности $(1 - \alpha) \cdot 100\%$. Если дисперсия C_y^2 известна из эксперимента, то доверительный интервал строится при помощи таблиц для нормального распределения.

Рассмотрим геометрическую интерпретацию доверительной области для двух параметров. Функция $l(m)$ в этом случае имеет вид

$$l(m_1, m_2) = l(\hat{m}_1, \hat{m}_2) - \frac{(m_1 - \hat{m}_1)^2}{2C_1^2(1-r^2)} - \frac{(m_2 - \hat{m}_2)^2}{2C_2^2(1-r^2)} + \frac{r(m_1 - \hat{m}_1)(m_2 - \hat{m}_2)}{C_1 C_2(1-r^2)}.$$

Здесь C_i^2 — дисперсии оценок \hat{m}_i , равные диагональным членам ковариационной матрицы R ; r — коэффициент корреляции оценок параметров \hat{m}_1 и \hat{m}_2 , определяемый внедиагональным элементом матрицы R , $r = R_{12} / (C_1 C_2)$.

Граница доверительной области в пространстве двух параметров представляет эллипс с центром в точке \hat{m}_1, \hat{m}_2 . На рис. 10 изображены примеры таких областей (жирные линии), соответствующих одной и той же доверительной вероятности (95%) при разной дисперсии исходных данных. С увеличением дисперсии доверительная область расширяется и может оказаться такой, что будет включать в себя отрицательные значения параметра. На рисунке

заштрихована часть области, содержащая отрицательные значения параметра m_1 .

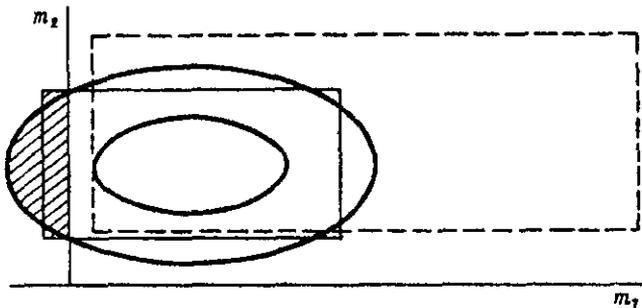


Рис. 10.

Если параметры – реальные физические величины, которые могут принимать только положительные значения, то такая доверительная область уже не является допустимой. С подобной ситуацией нередко приходится сталкиваться при решении обратных задач в случаях, когда дисперсии исходных данных велики, или при наличии корреляционных связей между параметрами. Как читатель увидит далее, дисперсии отдельных параметров в таких случаях могут становиться очень большими (>100%). Чтобы доверительная область не выходила за пределы допустимых значений параметров, можно сделать следующее.

Будем вместо параметра m рассматривать параметр ξ , связанный с m соотношением $\xi = \log m$. При этом область изменения параметра ξ уже включает в себя все значения от $-\infty$ до $+\infty$, так что ограничений на размер доверительной области для ξ не накладывается. Доверительный интервал для параметра ξ определяется, как и ранее: $\hat{\xi} - \hat{\gamma} \tau_{\alpha, n-1} < \xi < \hat{\xi} + \hat{\gamma} \tau_{\alpha, n-1}$, где $\hat{\gamma} = \hat{\sigma}/m$, а получаемый из него преобразованием доверительный интервал для m

$$\exp(\log \hat{m} - \hat{\gamma} \tau_{\alpha, n-1}) < m < \exp(\log \hat{m} + \hat{\gamma} \tau_{\alpha, n-1}) \quad (6f)$$

попадает уже только в область положительных значений параметра m .

81

На практике обычно для оценки погрешности пользуются доверительными интервалами, так что доверительная область определяется как совокупность доверительных интервалов (60) для всех параметров. Такая область представляет в пространстве n n -мерный параллелепипед. На рис.10, иллюстрирующем двумерный случай, эти области изображены прямоугольниками. Тонкой сплошной линией изображен контур области для исходных параметров m_1, m_2 (так же, как и эллипс, она включает отрицательные значения параметра m_1), а пунктиром околнурена область, построенная в соответствии с формулой (61). Последняя уже не содержит отрицательных значений параметра, но при этом она сильно расширяется в сторону положительных значений.

Обратим внимание еще на то, что в случае, когда параметры некоррелированы, прямоугольная и эллипсоидальная области различаются не слишком сильно, что позволяет вместо истинной доверительной области принимать такую, которая характеризуется совокупностью доверительных интервалов для отдельных параметров. Иная ситуация возникает в случае коррелированных параметров. Если существует корреляция между параметрами ($r \neq 0$), главные оси эллипса, представляющего доверительную область, оказываются повернутыми относительно координатных осей на некоторый угол, и эллипс становится более вытянутым. При $r \rightarrow 1$ он вырождается в прямую линию. Это иллюстрируется рис.11, где изображены контуры доверительных областей для разных коэффициентов корреляции (r) при условии, что общее количество информации о параметрах $\{J\}$, содержащееся в выборке, сохраняется неизменным. Величина дисперсии параметров возрастает с увеличением коэффициента корреляции: $\sigma_i^2 = \sigma_{0i}^2 / \sqrt{1-r^2}$ (σ_{0i} - дисперсия i -го параметра, соответствующая отсутствию корреляции между m_1 и m_2). При $r \rightarrow 1$ дисперсия бесконечно возрастает и параметры невозможно определять раздельно.

При наличии корреляционных связей между параметрами представлять приближенно доверительную область с помощью доверительных интервалов уже неправомерно. Это хорошо видно из рис. 11, где контуры таких областей изображены пунктиром.

Итак, при больших выборках и при условии, что количество информации о параметрах в выборке достаточно велико, до-

рительная область для параметров представляет собой k -мерный эллипсоид, конфигурация которого определяется ковариационной матрицей $R=J^{-1}$. Следовательно, ковариационная матрица полностью характеризует погрешность найденной оценки.

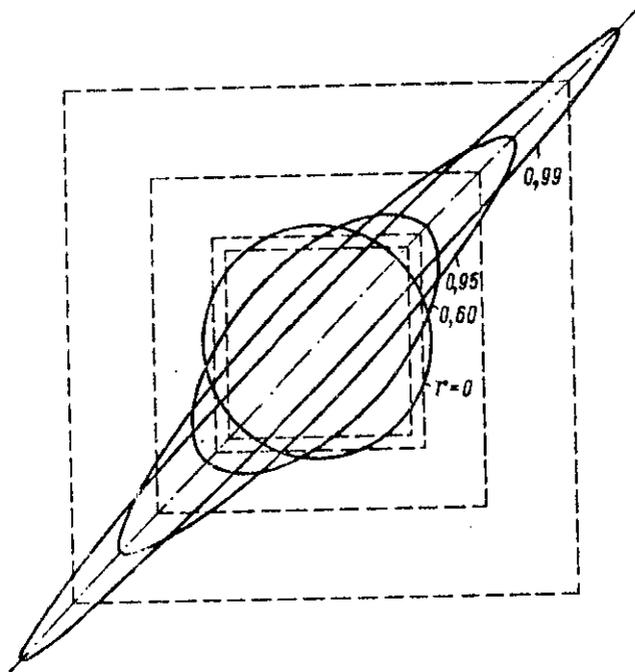


Рис. 11.

В случае малой выборки исходных данных, как уже упоминалось, распределение оценки параметров в общем не является нормальным, а рельеф функции правдоподобия нельзя аппроксимировать поверхностью параболоида, так что описанная процедура построения доверительных областей уже не может быть использована. Если в этом случае пытаться строить доверительные области, исходя только из оценки ковариационной матрицы, то можно допустить значительную ошибку. Такую ситуацию иллюстрирует рис. 12, где сплошными линиями изображены контуры равного правдоподобия,

а пунктиром — доверительные эллипсы, построенные на основе оценки ковариационной матрицы, соответствующие этим значениям правдоподобия. Даже для довольно низкой доверительной вероятности эти области заметно различаются. На рисунке точкой m отмечено истинное значение параметра. Оно соответствует достаточно высокому значению правдоподобия, но при этом не попадает внутрь соответствующего эллипса. В подобных ситуациях следует определять в пространстве параметров область, в которой значение функции правдоподобия оказывается выше некоторого порогового уровня. Методы поиска таких областей будут описаны в следующем параграфе.

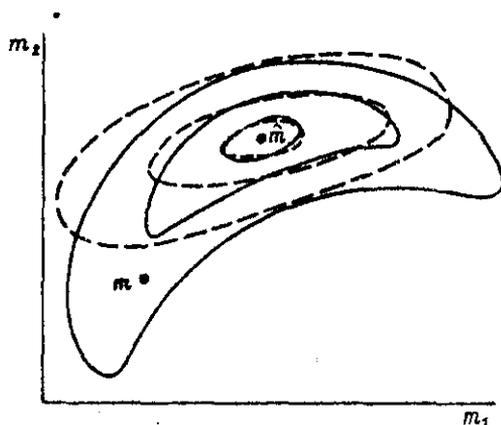


Рис. 12.

П р и м е р. Рассмотрим результаты интерпретации кривой кажущегося сопротивления МТЗ $\varphi_k(\tau)$, рассчитанные для модели, параметры которой — мощности слов h_1, h_2 и удельные сопротивления $\varphi_1, \varphi_2, \varphi_3$ — приведены в первой строке табл. I. Чтобы приблизить исходные данные к экспериментальным, рассчитанные для этой модели значения φ_k были осложнены случайными ошибками, распределенными нормально. Решение строилось для трех выборок данных, приведенных на рис. 13. Первые две выборки (I и II) содержали значений φ_k в интервале периодов $\tau \in (0,5; 10^8)$ с, объем первой выборки $n = 50$, второй — $n = 22$.

Таблица I

Примечание	Удельное сопротивление, Ом.м		Мощность, кМ		Относительная погрешность, %					
	ρ_1	ρ_2	ρ_3	h_1	h_2	γ_{ρ_1}	γ_{ρ_2}	γ_{ρ_3}	γ_{h_1}	γ_{h_2}
Истинные значения...	I	0,0256	I	I	2	-	-	-	-	-
Начальное приближе- ние	I,2	0,042	2,5	I,125	I,7	-	-	-	-	-
Решение по выбор- ке: I	I,009	0,0257	0,991	I,014	2,038	I,5	6	8	2	7
II	I,03	0,0265	0,981	0,985	2,07	6,8	2I	12	2	22
III	0,913	0,0361	0,967	0,7	2,85	200	150	25	200	200
Закрепление парамет- ров ρ_1, h_1 истинными значениями (выборка III)	I	0,027	0,94	I	2,1	-	3,5	9,15	-	33
Закрепление парамет- ров ρ_1, h_1 значениями, соответствующими началь- ному приближению (выбор- ка III)	I,2	0,021	0,932	I,125	I,65	-	3,3	11,5	-	37

6.

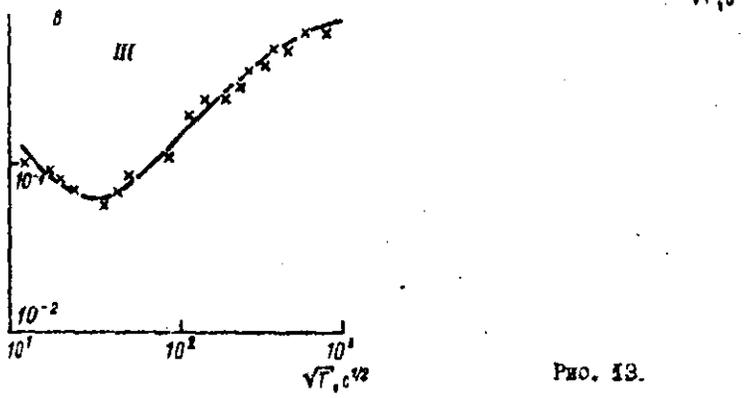
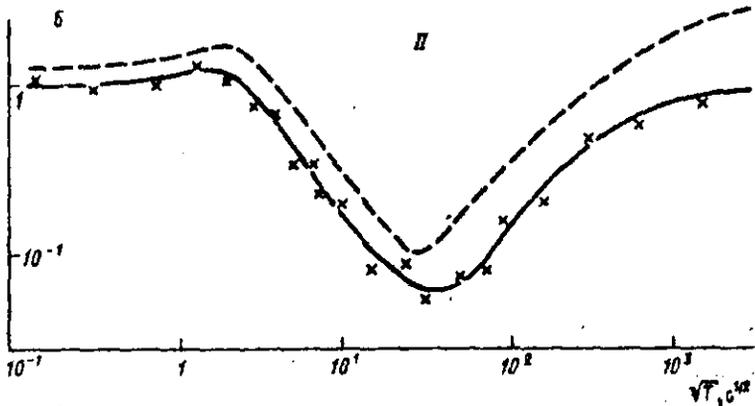
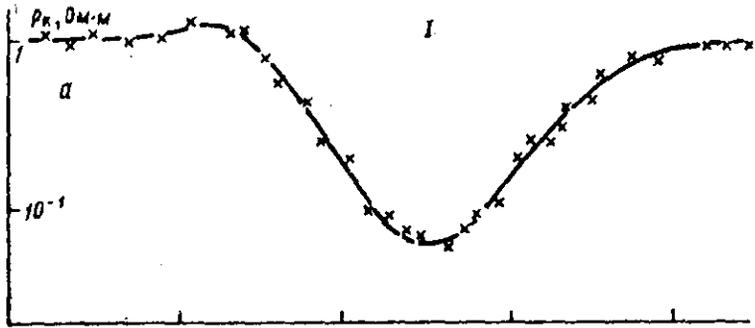


Рис. 13.

Третья выборка ($n = 22$) содержала значения φ_n в более узком интервале периодов: $T \in (10^0, 10^6)$ с. Решение строилось последовательными приближениями по методу Ньютона - Рафсона (см. § 4). Параметры, отвечающие начальному приближению, приведены во второй строке табл.1, а кривая кажущегося сопротивления, соответствующая начальному приближению, изображена пунктиром на рис.13,б. В качестве оценки погрешности решения принималась величина стандартного отклонения, нормированная на величину соответствующего параметра $\gamma = (\sigma_{m_i}/m_i) \cdot 100\%$.

Результаты интерпретации сведены в табл.1. Найденные оценки параметров наиболее близки к истинным значениям для выборки I. Относительные погрешности в этом случае не превышают 8%. Для выборки II точность интерпретации уже существенно снизилась вследствие уменьшения объема выборки. Для выборки III погрешности настолько велики, что решение теряет смысл. Это объясняется тем, что в выбранном интервале периодов данные не содержат достаточно информации об искомым параметрах. Решение делается неустойчивым, и при использовании стандартного метода Ньютона - Рафсона процесс последовательных приближений расходится. Для построения решения приходится использовать в этом случае модификацию Левенберга - Маркуардта, которая также будет описана в § 7.

Коэффициенты корреляции оценок параметров приведены в табл.2. В случае выборки I параметры оказываются практически

Т а б л и ц а 2

Параметры						Выборка
φ_1	φ_2	φ_3	h_1	h_2		
I	0,0826	0,0039	0,148	0,0846	φ_1	I II III
	0,233	0,0048	0,298	0,0983		
	0,987	0,381	0,9997	0,9977		
	I	0,0645	0,647	0,931	φ_2	I II III
		0,0818	0,727	0,974		
		0,418	0,990	0,9995		
		I	0,0362	0,158	φ_3	I II III
			0,0388	0,234		
			0,386	0,405		
			I	0,606	h_1	I II III
				0,696		
				0,9989		
				I	h_2	I-III

некоррелированными, для выборки II корреляция заметно увеличивается, что приводит к повышению относительной погрешности. Для выборки III коэффициенты корреляции параметров (d_1, ρ_1) ; (d_1, d_2) ; (ρ_2, ρ_3) становятся практически равными единице, вследствие чего соответствующие параметры не могут быть совместно разрешены.

§ 7. Некоторые методы поиска экстремума

При использовании оптимизационного подхода к решению обратной задачи возникают трудности, связанные с построением решения. Поскольку "решением" является точка, соответствующая экстремуму, либо множество точек, отвечающих значениям целевой функции ниже (или выше) некоторого порогового значения, то задача оценивания параметров сводится к поиску минимума (максимума) функции в многомерном пространстве. Хотя такие задачи относятся к области вычислительной математики, мы на некоторых из них остановимся довольно подробно, так как они собственно и определяют алгоритмы решения обратной задачи в геофизике.

Функция $\lambda(m)$, экстремум которой отыскивается, определяет, как уже упоминалось в § 6, гиперповерхность в $(k+1)$ -мерном пространстве (k — число параметров). Проекция сечения поверхности гиперплоскостью $\lambda = \text{const}$ образует в пространстве параметров контуры равных значений $\lambda(m)$. Мы их будем называть линиями уровня. Поскольку явный вид функции $\lambda(m)$ зависит от функции правдоподобия, а она строится на основании имеющихся исходных данных, то фигура поверхности определяется результатами эксперимента, а также выбранной моделью $\Phi(m)$.

При большой выборке экспериментальных данных ($n \gg k$), в которой содержится достаточно информации относительно всех исходных параметров, поверхность в окрестности максимума будет представлять собой гиперпараболоид (см. (E7)). Пример такой фигуры иллюстрируется рис. 14 [47].

Если о каких-либо параметрах в исходных данных имеется так мало информации, что изменение их слабо сказывается на $\lambda(m)$, то на поверхности целевой функции возникают хребты (овраги) (рис. 15, а). Если параметры завышены, то рельеф функций

$\lambda(m)$ существенно осложняется: главные оси фигур поворачиваются относительно координатных осей, и при сильной корреляционной связи на поверхности образуются один или несколько гребней (овраги с котловинами). Возникают ложные экстремумы (рис. 15, б [43]).

В силу нелинейности модели $\Phi(m)$ гребни и хребты имеют искривленную форму. Линии уровня таких поверхностей можно, например, увидеть на рис. 13 или на рис. 16. Искривленные хребты, гребни в геофизических задачах встречаются очень часто.

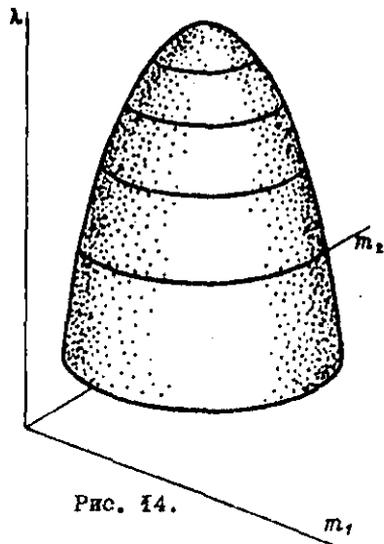


Рис. 14.

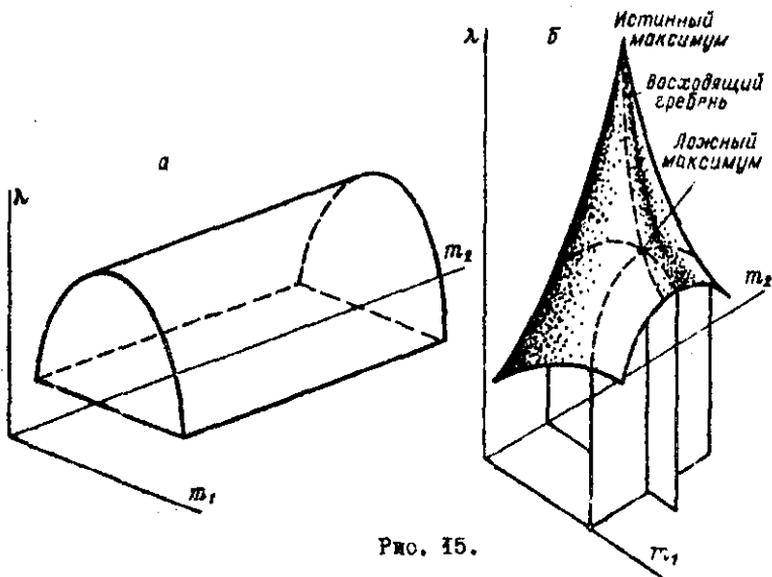


Рис. 15.

В настоящее время существует много разнообразных методов поиска экстремума функции многих переменных. Они широко представлены в литературе [3, 20, 43, 47]. Мы остановимся лишь на тех из них, которые чаще используются в геофизических задачах.

Для определенности будем решать задачу $\lambda(m) = \min$, поскольку поиск максимальных значений функции можно всегда превратить в задачу минимизации, сменив знак у функции (см. § 4). Почти все существующие методы минимизации нелинейной функции используют последовательную линейную минимизацию. Начиная с заданной начальной точки $m^{(0)}$ переходят в новую точку $m^{(1)}$, двигаясь в параметрическом пространстве по некоторому направлению p_0 с шагом h_0 . Затем точка $m^{(1)}$ принимается за начальную, и движение к следующей точке $m^{(2)}$ производится по направлению p_1 с шагом h_1 и т.д. Методы различаются между собой выбором направления p_r и величины шага h_r .

В § 4 при построении решения использовался метод Ньютона - Рафсона. Напомним, что в этом методе целевая функция $\lambda(m)$ на каждом шаге итерации аппроксимируется параболической функцией $\tilde{\lambda}(m)$, полученной путем разложения $\lambda(m)$ в ряд Тейлора. Решением системы уравнений

$$\left. \frac{\partial \tilde{\lambda}(m)}{\partial m_i} \right|_{m^{(r)}} = 0, \quad i=1, 2, \dots, k, \quad (62)$$

является поправка $\Delta m^{(r)} = A^{-1} \theta$ (A - матрица средних значений вторых производных от $\lambda(m)$ по m , вычисленных в точке

$m^{(r)}$, а $\theta_i = \left. \frac{\partial \tilde{\lambda}(m)}{\partial m_i} \right|_{m^{(r)}}$, которая определяет направление дальнейшего шага, и тогда $m^{(r+1)} = m^{(r)} + A^{-1} \theta$.

Метод Ньютона - Рафсона применим тогда, когда матрица A невырождена и имеет обратную. Он хорошо сходится, если начальная точка $m^{(0)}$ достаточно близка к решению и первые производные функции $\lambda(m)$ по m_i не меняют знака. Если рельеф целевой функции овражный, то матрица A будет плохо обусловленной, и на r -м шаге итераций поправки $\Delta m^{(r)}$ могут оказаться очень большими в направлении, обратном направлению "оврага", так что дальнейший процесс последовательных приближений либо станет неэффективным, либо разойдется совсем.

В таких случаях К. Девенберг предложил применить демпфирование поправок [63]. Сущность демпфирования сводится к то-

му, что одновременно с целевой функцией минимизируется сумма квадратов поправок, т.е. вводится в рассмотрение функция

$$\bar{\lambda}(m) = \tilde{\lambda}(m) + \alpha \sum_q \omega_q \Delta m_q^2, \quad q=1, 2, \dots, k$$

(ω_q - весовые множители, определяющие вклад каждого слагаемого, $\alpha > 0$).

Вместо системы уравнений (62) решается система $\left. \frac{\partial \bar{\lambda}}{\partial m_i} \right|_{m^{(r)}} = 0$, которая эквивалентна системе линейных уравнений

$$(A + K) \Delta m = 0, \quad (63)$$

где K - диагональная матрица, имеющая элементы $K_{qq} = \alpha \omega_q$, $q = 1, 2, \dots, k$.

Таким образом, направление поиска будет определяться вектором

$$\Delta m = (A + K)^{-1} 0.$$

Демпфирование поправок при удачном подборе K_{qq} регулирует сходимость процесса, не позволяя слишком менять направление при последовательном продвижении. При $K_{qq} \rightarrow 0$ $\Delta m \rightarrow A^{-1} 0$, т.е. имеет место метод Кэтона - Рафсона, где матрица A может быть плохо обусловленной. Если же $K_{qq} \rightarrow \infty$, то новая матрица будет хорошо обусловлена благодаря присутствию хорошо обусловленной матрицы K , однако процесс последовательного приближения сильно замедляется. Поэтому малые или слишком большие значения K_{qq} непригодны.

Д.В. Маркуартом был предложен довольно простой способ отскакивания величин K_{qq} [64]. Прежде всего система уравнений (63) заменяется равносильной: $(A^* + K^*) \Delta m^* = 0^*$, где элементы матрицы $A_{pq}^* = A_{pq} / \sqrt{A_{pp} A_{qq}}$, $0_p^* = 0_p / \sqrt{A_{pp}}$, $\Delta m_p^* = \Delta m_p \sqrt{A_{pp}}$, а $K^* = \chi I$, I - единичная матрица; χ - скалярная величина. Диагональные члены матрицы A^* равны единице. Величина χ принимается равной нулю до тех пор, пока параболическая аппроксимация дает хорошее приближение. Как только сходимость нарушается, χ получает значение, равное, например, 10^{-2} . Обозначим это через $\chi^{(r-1)}$. Далее вычисляются поправки $\Delta m^{(r)}$ при $\chi^{(r-1)}/\nu$ и $\Delta m^{(r)}$ при $\chi^{(r-1)}$ (ν - заданное положительное число). По формуле $m^{(r+1)} = m^{(r)} + \Delta m^{(r)}$ находят две точки $m^{(r+1)}$ ($\chi^{(r-1)}/\nu$) и $\bar{m}^{(r+1)}$ ($\chi^{(r-1)}$). Из них выбирается та, для которой выполняется условие

$$\lambda(m^{(r+1)}) < \lambda(\bar{m}^{(r)}), \quad (64)$$

и на следующем шаге используется $\chi^{(r)} = \chi^{(r-1)}/\lambda$ или $\chi^{(r)} = \chi^{(r-1)}$ соответственно. Если обе точки удовлетворяют условию (64), то $\chi^{(r)} = \chi^{(r-1)}/\lambda$. Пока величина χ мала, процесс последовательных приближений сходится довольно быстро.

В том случае, когда ни одна из точек не удовлетворяет условию (64), $\chi^{(r-1)}$ увеличивают, последовательно умножая на λ (обычно принимают $\lambda = 10$), пока не будет выполнено условие (64). После этого принимают $\chi^{(r)} = \chi^{(r-1)}\lambda^\omega$ (ω - целое положительное число). Такая ситуация встречается тогда, когда между параметрами существует сильная связь, что и вызывает неразумно большие значения χ . Как уже указывалось, сходимость процесса настолько замедляется, что продолжать его не имеет смысла. В таких ситуациях разумней использовать методы, позволяющие пройти вдоль оврага к котловине.

Одним из таких довольно эффективных методов является метод оврагов [12]. Из заданной начальной точки $m^{(0)}$ каким-либо методом минимизации (например, Ньютона - Рафсона) находится точка s_0 . Поскольку рельеф функции $\lambda(m)$ овражный, то точка s_0 окажется на дне оврага (рис.16). Затем выбирается другая точка $\bar{m}^{(0)}$ не слишком далеко от первой, и из нее таким же

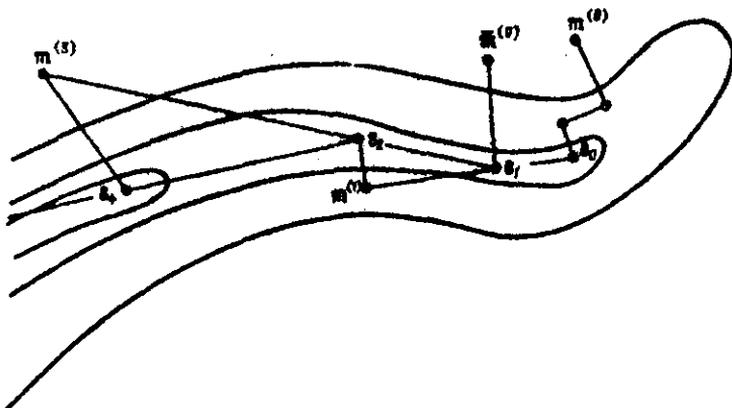


Рис. 16.

способом производится спуск на дно оврага в точку s_1 . Следующий шаг к точке $m^{(1)}$ делается вдоль направления $(s_1 - s_0)$ в сторону убывания функции $\lambda(m)$:

$$m^{(1)} = m^{(0)} \pm h(s_1 - s_0). \quad (65)$$

Здесь берется знак плюс, если $\lambda(s_1) < \lambda(s_0)$, и минус - в противоположном случае, так что движение направлено в сторону понижения дна оврага. Величина h подбирается с учетом особенностей целевой функции. Из точки $m^{(1)}$ вновь производится спуск на дно оврага в точку s_2 . Дальнейший шаг делается в направлении $s_2 - s_1$ к точке $m^{(2)} = m^{(1)} \pm h(s_2 - s_1)$ и т.д. Вычисления вдоль оврага продолжают в окрестности минимума $\lambda(m)$ до тех пор, пока $\lambda(s^{(r+1)}) < \lambda(s^{(r)})$. Начиная с $\lambda(s^{(r)})$ положение точки минимума лучше уточнять другими методами.

Если нахождение производных сопряжено с вычислительными трудностями (например, если функция, описывающая модель, задана в неявном виде - см. пример I в гл. I), обычно применяются методы минимизации, при использовании которых достаточно знать значения только самой целевой функции. Простейшим из них является метод прямого поиска, предложенный Р. Хуком и Т.А. Джимсом [57]. Его часто в литературе называют методом конфигураций [43]. Идея его состоит в следующем. Задается некоторая начальная точка $m^{(0)}$. К каждому из параметров $m_q^{(0)}$ по очереди делаются добавки

$$\Delta m_q^{(0)} = h_q \quad \text{или} \quad \Delta m_q^{(0)} = -h_q^{(0)}$$

($q = 1, \dots, k$). В $2k+1$ точках (центральной $m^{(0)}$ и $2k$ - раскваченных $m^{(0)} \pm h_i \delta_i$, где δ_i - k -мерный вектор, имеющий все компоненты равными нулю, кроме i -й, которая равна единице) вычисляется целевая функция $\lambda(m)$. Из этих точек выбирается та, в которой значение $\lambda(m)$ минимально, и эта точка принимается за центральную и т.д., до тех пор пока минимум целевой функции не будет достигнут в центральной точке.

Таким образом, на каждом шаге итераций задается направление, ведущее к минимуму функции $\lambda(m)$. Поскольку такая процедура требует большого объема вычислений, вначале обычно производят рекогносцировку функции отклика с большим шагом и, определив приблизительно положение минимума, уточняют его с более мелким шагом, уменьшая h_i каждый раз, пока не попадут в достаточно узкую окрестность точек.

Метод прямого поиска позволяет двигаться по обврагу в направлении котловин. Но так как пробные шаги осуществляются только параллельно координатным осям и неясно, что делается в других направлениях, то можно саму котловину (вершину гребни) пропустить из-за неудачного шага. В этом аспекте метод обврагов имеет существенное преимущество перед методом прямого поиска.

Описанные методы минимизации функции многих переменных пригодны для поиска локального минимума. Если поверхность $\lambda(\mathbf{m})$ имеет несколько минимумов и начальная точка $\mathbf{m}^{(0)}$ находится не вблизи точки абсолютного минимума, то эти методы не исключают возможности "застревания" в локальном минимуме, не являющемся абсолютным экстремумом. Для того чтобы обеспечить хорошее первое приближение, часто приходится исследовать параметрическое пространство: значения функции $\lambda(\mathbf{m})$ вычисляются в узлах сетки, покрывающей некоторую область, в которой априори предполагается возможное расположение минимальных значений $\lambda(\mathbf{m})$; за $\mathbf{m}^{(0)}$ берется точка, в которой $\lambda(\mathbf{m})$ имеет наименьшее значение, если только поведение $\lambda(\mathbf{m})$ в соседних точках указывает на непрерывное стремление целевой функции к единственному минимуму. В противном случае такое же исследование повторяется в окрестности $\mathbf{m}^{(0)}$ с большей детальностью.

Простейший способ построения сетки - разместить ее узлы вдоль координатных осей через равные интервалы. Но при большом числе параметров число узлов сетки оказывается настолько велико, что этот способ становится практически неосуществим. В таком случае удобным способом построения сетки является метод Монте-Карло [1], который заключается в том, что в качестве узлов берутся случайные точки в заданной области пространства параметров. Преимущество такого метода в том, что на число проверяемых точек не накладывается никаких ограничений, так что он становится приемлем для практической реализации. При этом точки (узлы сетки) будут равномерно заполнять исследуемую область.

На начальном этапе развития методов решения обратных задач метод Монте-Карло использовался не только для предварительного исследования области, в которой предполагается существование решения, но и непосредственно для нахождения допустимых решений [1, 49, 69]. При этом задается пороговое зна-

чение c для $\lambda(m)$, которое определяет область допустимых решений: точка m принимается в качестве одного из допустимых решений, если $\lambda(m) < c$.

На рис.17 изображены результаты определения моделей распределения плотности в мантии Земли, полученные Ф.Прессом [69] по данным 97 значений периодов собственных колебаний с учетом ограничений, накладываемых значениями массы и момента инерции

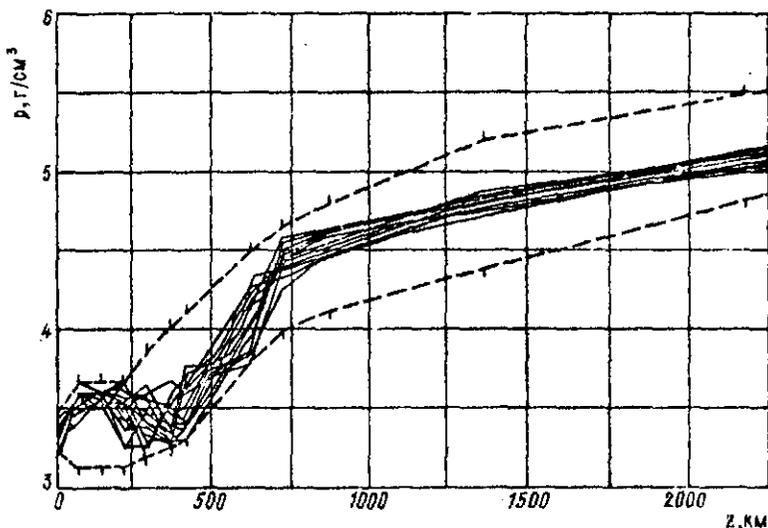


Рис.17.

Земли. Пунктиром показаны пределы, внутри которых выбирались плотностные разрезы методом Монте-Карло. Из 881 разреза было отобрано 27, изображенных на этом рисунке. Распределения плотности, выбранные в качестве допустимых решений, укладываются в значительно более узкую полосу, чем исходная, и позволяют сделать определенные заключения об особенностях распределения плотности, в частности о наличии зоны повышенной плотности в верхней мантии.

Более надежным методом, позволяющим оконтурить область, точки которой относятся к классу допустимых решений, является метод "ежа" [28]. Суть его заключается в следующем. Каким-ли-

бо методом (например, прямого поиска или Монте-Карло) перебираются узлы координатной сетки в пространстве параметров и отыскивается такой, в котором

$$\lambda(\mathbf{m}) < c,$$

где c — некоторое пороговое значение. Как только такая точка найдена, функция $\lambda(\mathbf{m})$ вычисляется в соседних с ней узлах выбранной сетки значений параметров. Под соседними узлами понимают точки, которые отличаются от исходной не более, чем на один шаг по каждой координате. Но так как при большом числе параметров таких узлов оказывается слишком много ($3^k - 1$), то обычно исследуют не все узлы, а только такие, у которых не более чем q координат ($q \ll k$), отличаются на один шаг от координаты центральной точки. Те узлы, в которых выполняется указанное условие, относятся к области решения, и для них производится та же процедура перебора соседних узлов. Этот процесс продолжается до тех пор, пока можно непрерывно переходить от одного узла к другому. Затем одним из обычных методов отыскивается новая часть области минимума. При этом в каждом узле вычисление $\lambda(\mathbf{m})$ производится один раз.

Затем следует проверить, являются ли найденные части области изолированными. С этой целью в двух частях области берется по узлу с наименьшим значением целевой функции. Пусть это будут узлы с номерами i и j . Им соответствуют параметры m_i и m_j . Для точек $m_\alpha = m_i + \alpha(m_j - m_i)$ вычисляют $\lambda(m_\alpha)$, где α меняется от 0 до 1 с небольшим шагом. Если все $\lambda(m_\alpha)$ удовлетворяют условию $\lambda(\mathbf{m}) < c$, то найденная область минимума односвязная.

Таким образом, с помощью метода "жа" удается оконтурировать всю область минимума многомерной целевой функции, внутри которой значения этой функции меньше заданного порогового условия.

Для иллюстрации сказанного рассмотрим рис. 18. На нем дана координатная сетка для двух параметров; пересечения — узлы сетки; жирные линии изображают грубую сетку, тонкие (вместе с жирными) — более мелкую (шаг по одной координате вдвое мельче). Область минимума не односвязна. Она состоит из двух изолированных частей: A и B . Пусть методом Монте-Карло най-

дена точка 1; ближайший к ней узел (точка 2) принадлежит области А. Далее рассматривается функция $\lambda(m)$ в узлах, ближайших к точке 2. При грубой сетке и $q=1$ это узлы 3-6. Из них узел 3 попадает в искомую область. Поэтому следующий шаг -

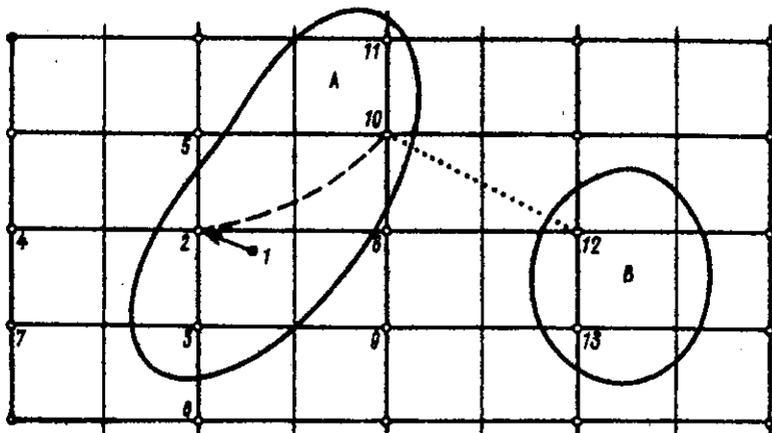


Рис. 18.

проверка узлов 7-9 и т.д. В результате последовательного перебора координатной сетки получены три части искомой области минимума с узлами 2,3; 10,11; 12,13. Детальный просчет целевой функции вдоль пунктирной линии показывает, что две части области А соединяются, в то время как значения $\lambda(m)$ вдоль точечной линии указывает, что часть В является изолированной.

§ 8. Выбор параметров

Рассмотренные в этой главе способы решения обратной задачи исходят из того, что исследуемая среда каким-то образом параметризована, т.е. заменена моделью, описываемой конечным числом параметров $m = (m_1, m_2, \dots, m_k)$. До сих пор мы не касались вопроса, как выбрать параметры, хотя и отмечали нежелательные эффекты, связанные с неудачной параметризацией. Если параметров взято слишком мало, то модель будет неадекватна

исходным данным, а это приведет к тому, что невязка $\mathbf{a} = \mathbf{y} - \Phi(\mathbf{m})$, будет иметь систематическую компоненту, в то время как при интерпретации мы ее рассматриваем как случайную. Если параметров слишком много, то между их оценками возникают сильные корреляционные связи, и при этом величина дисперсии оценок может настолько возрасти, что решение станет бессмысленным.

Сильная корреляционная зависимость между оценками параметров возникает и в случае, если между параметрами существует связь, которая при определенных условиях может рассматриваться как функциональная. Например, входной импеданс естественного электромагнитного поля Z зависит, вообще говоря, от мощностей h_i и удельных сопротивлений ρ_i слоев (см. пример 4 в § 2 гл. I). Однако если длина волны в слое во много раз (по крайней мере больше, чем в 10 раз) превосходит мощность некоторого слоя, то Z становится зависящим только от продольной проводимости этого слоя: $S_i = h_i / \rho_i$, так что параметры h_i и ρ_i не могут быть определены раздельно. И если используемые для решения обратной задачи данные МТЗ получены на таких периодах, для которых отношение длины волны к мощности какого-то из слоев велико, то коэффициент корреляции между оценками соответствующих этому слою параметров приблизится к единице, а их дисперсии станут очень большими. Из-за того, что такая выборка данных не несет информации отдельно о параметрах ρ_i и h_i , общее количество информации о них в данной выборке, определяемое детерминантом информационной матрицы Фишера, становится близким к нулю.

Дисперсии оценок параметров становятся большими не только в случае сильных корреляционных связей между ними, но и когда используемые характеристиками поля слабо зависят от этих параметров. В этом случае количество информации по Фишеру, содержащейся в данной выборке о таких параметрах, также оказывается малым.

В обоих случаях параметризация является неудачной и требует корректировки. Такую корректировку можно проводить по-разному.

Можно не изменять модель, но некоторым параметрам придавать фиксированные значения. Это должны быть либо такие параметры, изменения которых слабо влияют на рассматриваемые ха-

характеристики поля, либо такие, оценки которых обнаруживают сильную корреляцию с оценками других параметров. При фиксации параметров следует, конечно, использовать априорную информацию о среде. Например, если эксперимент проводится вблизи скважин, то можно фиксировать глубину осадочного чехла; если для данного района имеются данные акустического каротажа, то можно принимать известными скорости сейсмических волн. Но поскольку точные значения закрепляемых параметров нам неизвестны и мы вынуждены принимать для них некоторые приближенные значения, это будет приводить к тому, что оставшиеся параметры будут оцениваться с ошибкой [15]. Рассмотрим, к чему сводятся эти ошибки.

Пусть мы фиксируем $k-j$ компонент вектора m ($m_{j+1}, m_{j+2}, \dots, m_k$), которые можно рассматривать как $(k-j)$ -мерный вектор $\eta = (\eta_1, \eta_2, \dots, \eta_{k-j})$. Компоненты вектора m , которые подлежат оцениванию (m_1, m_2, \dots, m_j) обозначим через $\xi = (\xi_1, \xi_2, \dots, \xi_j)$, так что $m = \begin{pmatrix} \xi \\ \eta \end{pmatrix}$. Для простоты будем считать, что случайные ошибки исходных данных независимы и распределены нормально с одинаковой дисперсией σ_y^2 . Среднее значение $E(y) = \Phi(m)$ обозначим через y_0 . Оператор $\Phi(m)$ будем считать линейным: $\Phi(m) = Am$ (если это не так, то линеаризуем задачу и сведем ее к линейному оцениванию поправок к искомым параметрам относительно некоторого начального приближения).

Таким образом, вектор m оценивается из условия минимизации

$$\lambda(m) = (y - Am)^T (y - Am).$$

Представим матрицу A в виде [13]

$$A = \left(\begin{array}{ccc|ccc} C_{11} & \dots & C_{1j} & D_{11} & \dots & D_{1,k-j} \\ \vdots & & \vdots & \vdots & & \vdots \\ C_{n1} & \dots & C_{nj} & D_{n1} & \dots & D_{n,k-j} \end{array} \right),$$

так что $C\xi + D\eta = A \begin{pmatrix} \xi \\ \eta \end{pmatrix} = Am$. Тогда, учитывая, что вектор η мы фиксируем ($\eta = \bar{\eta}$), условие, из которого определяется оценка вектора ξ , можно записать в виде

$$(u - C\xi)^T (u - C\xi) = \min,$$

здесь $u = y - D\bar{\eta}$. Таким образом,

$$\hat{\xi} = (C^T C)^{-1} C^T u. \quad (66)$$

Пусть ξ_0, η_0 — значения параметров, отвечающие среднему значению вектора наблюдений, т.е. $C\xi_0 + D\eta_0 = y_0$. Тогда

$$\xi_0 = (C^T C)^{-1} C^T (y_0 - D\eta_0). \quad (67)$$

Определим теперь смещение оценки вектора ξ , получаемой по формуле (66) с учетом того, что вектор η принимает некоторые фиксированные значения $\bar{\eta}$, отличные от η_0 . Величина смещения определяется следующим образом (см. в § I формулу (18)):

$$b(\xi) = E(\hat{\xi}) - \xi_0.$$

Подставляя сюда $\hat{\xi}$ и ξ_0 из выражений (66) и (67), получим

$$b(\xi) = (C^T C)^{-1} C^T E(u) - (C^T C)^{-1} C^T (y_0 - D\eta_0).$$

С учетом того, что $E(u) = y_0 - D\bar{\eta}$, смещение оценки незакрепленных параметров $\hat{\xi}$ будет иметь вид

$$b(\xi) = (C^T C)^{-1} C^T D(\eta_0 - \bar{\eta}). \quad (68)$$

Из этой формулы сразу видно, что если мы приняли значения параметров $\bar{\eta}$ равными их точным значениям η_0 , то оценка остальных параметров оказывается несмещенной, а дисперсия этих параметров определится из ковариационной матрицы

$$R_{\xi} = C_y^2 (C^T C)^{-1}. \quad (69)$$

Если же параметры η не фиксировать, а оценивать их совместно с параметрами ξ , то дисперсия оценки $\hat{\xi}$ определяется из ковариационной матрицы

$$R_{\xi} = C_y^2 (A^T A)^{-1}. \quad (70)$$

При сравнении выражений (69) и (70) видно, что, выбирая параметры, значения которых предполагается фиксировать, следует исходить из того, чтобы оставшиеся параметры были некоррелированы, тогда дисперсии оценок параметров ξ , определяемые диагональными элементами матрицы $(C^T C)^{-1}$, уменьшатся по сравнению с соответствующими диагональными элементами матрицы $(A^T A)^{-1}$. Но улучшение точности оценок параметров ξ достигается, лишь если принять $\bar{\eta} = \eta_0$. Если $\bar{\eta} \neq \eta_0$, то средний квадрат ошибки параметров ξ возрастает из-за наличия смещения. Он определяется диагональными элементами матрицы

$$E[(\xi - \xi_0)(\xi - \xi_0)'] = C_y^2 (C' C)^{-1} + (C' C)^{-1} C' D (\bar{\eta} - \eta_0) (\bar{\eta} - \eta_0)' D' C (C' C)^{-1}. \quad (71)$$

Первый член здесь определяет дисперсию, а второй — квадрат смещения оценки.

Таким образом, при закреплении параметров надо добиваться еще и того, чтобы смещение $\delta(\xi)$ не оказалось бы слишком большим.

Анализ формулы для смещения оценки (68) показывает, что $\delta(\xi)$ можно сделать малым при выполнении следующих условий. Во-первых, если исходные параметры η были коррелированными, а в результате закрепления части параметров корреляция между оставшимися сильно снизилась, то элементы матрицы $(C' C)^{-1}$ будут, как уже говорилось, малы по сравнению с элементами матрицы $(A' A)^{-1}$, так что квадрат смещения может быть мал по сравнению с дисперсией исходных параметров (если, конечно, $\bar{\eta} - \eta_0$ не слишком велико). Во-вторых, если исходные параметры и не были сильно коррелированы, но дисперсии некоторых из них были велики из-за того, что исходная выборка содержала о них мало информации, т.е. используемые характеристики поля мало изменялись при вариациях этих параметров, то при закреплении именно таких параметров элементы матрицы D оказываются малыми, и смещение оставшихся параметров не будет велико, даже если при фиксации значений η допущена ошибка.

Таким образом, закреплять следует такие параметры, которые слабо влияют на рассматриваемые характеристики поля, а также такие, закрепление которых сильно уменьшает корреляцию оставшихся параметров.

Другой способ корректировки параметризации — это изменение модели путем "укрупнения" слоев, что также приводит к уменьшению числа оцениваемых параметров. Укрупнять слои следует до тех пор, пока внедиагональные члены корреляционной матрицы r_{ij} ($r_{ij} = R_{ij} / \sqrt{R_{ii} R_{jj}}$) не станут заметно отличаться от единицы, что будет указывать на отсутствие корреляционных связей между выбранными параметрами.

Как при одном, так и при другом способе изменения параметризации можно исходить из анализа информационных свойств используемых наблюдений. Такой подход основан на исследовании информативных областей [29].

Рассмотрим диагональные члены информационной матрицы Фишера:

$$J_{pp} = E_{\mathbf{y}} \left(\frac{\partial \ln l_{\mathbf{y}}(\mathbf{m})}{\partial m_p} \right)_{\mathbf{m}=\mathbf{m}_0}^2, \quad p = 1, 2, \dots, k, \quad (72)$$

где $l_{\mathbf{y}}(\mathbf{m})$ — логарифмическая функция правдоподобия для заданной выборки \mathbf{y} ; \mathbf{m}_0 — вектор истинных значений параметров. Величина J_{pp} характеризует количество информации, содержащейся в выборке y_1, \dots, y_n относительно параметра m_p . Как упоминалось в § 1 J_{pp} является мерой чувствительности функции правдоподобия к малым изменениям m_p . Когда в исходных данных отсутствует информация о каком-либо из параметров m_p ($p = 1, 2, \dots, k$) $l_{\mathbf{y}}(\mathbf{m})$ не реагирует на вариации m_p и $J_{pp} = 0$. Это свойство функции правдоподобия можно использовать для выбора закрепляемых параметров.

В частном случае некоррелированных случайных величин y_1, \dots, y_n , распределенных нормально с постоянной дисперсией σ_y^2 ,

$$J_{pp} = \frac{1}{\sigma_y^2} \sum_{i=1}^n \left(\frac{\partial \Phi_i(\mathbf{m})}{\partial m_p} \right)^2. \quad (72a)$$

Каждое слагаемое в формуле (72a) $\frac{1}{\sigma_y^2} \left(\frac{\partial \Phi_i(\mathbf{m})}{\partial m_p} \right)^2$ представляет собой количество информации, которое содержится в i -м наблюдении y_i о параметре m_p . Чтобы устранить влияние размерности параметров, нормируем величину $\frac{1}{\sigma_y^2} \left(\frac{\partial \Phi_i(\mathbf{m})}{\partial m_p} \right)^2$ на m_p^2 , и будем рассматривать

$$J_p^{(i)} = \frac{1}{\sigma_y^2} \left(\frac{\partial \Phi_i(\mathbf{m})}{\partial m_p} m_p \right)^2, \quad i = 1, 2, \dots, n,$$

как значения функции $J_p(x)$ в точках x_i , в которых проводятся наблюдения (см. § 2). Правая часть этого выражения должна вычисляться согласно формуле (72) при истинных значениях параметров $\mathbf{m}_0 = (m_{01}, \dots, m_{0k})$. Но поскольку истинные значения нам неизвестны, а для исследования распределения информации по x достаточно приближенно знать поведение функций $J_p(x)$, то при вычислениях этих функций вместо истинных используются априорные значения параметров.

Построенная таким образом функция $J_p(x)$ отражает распределение информации о параметре m_p по x . Интервал (x_{p1}, x_{p2}) , внутри которого $J_p(x)$ принимает значения, близкие к

максимальному, будем называть информативной областью параметра m_p . Внутри информативной области проявляется наибольшая зависимость функции правдоподобия от данного m_p . Сравнение максимальных значений $J_p(x)$ для разных параметров позволяет оценить, в какой степени наблюдаемые характеристики поля зависят от того или иного параметра. Кроме того, из сравнения поведения кривых $J_p(x)$ для разных p можно судить о том, насколько сильны корреляционные связи между оценками выбранных параметров. Если кривые $J_p(x)$ и $J_q(x)$ в некотором интервале x приблизительно совпадают, то в случае, когда наблюдения соответствуют значениям x только из этого интервала, оценки параметров m_p и m_q будут сильно коррелированы.

Заданной выборке наблюдений соответствуют определенные точки x_i ($i = 1, 2, \dots, n$) на оси x . Для каждой заданной выборки можно определить, попадают ли точки x_i внутрь информативных областей для выбранных параметров, и на основании этого можно судить о том, какую информацию содержит данная выборка о том или ином параметре. Если значения x_i оказываются вне информативной области для некоторого параметра m_p , то его значение следует фиксировать.

Если в интервале наблюдений кривые $J_p(x)$, отражающие распределение информации по x , совпадают для двух или более параметров, то следует фиксировать все такие параметры, кроме одного, иначе из-за сильной корреляции дисперсии их окажутся очень большими.

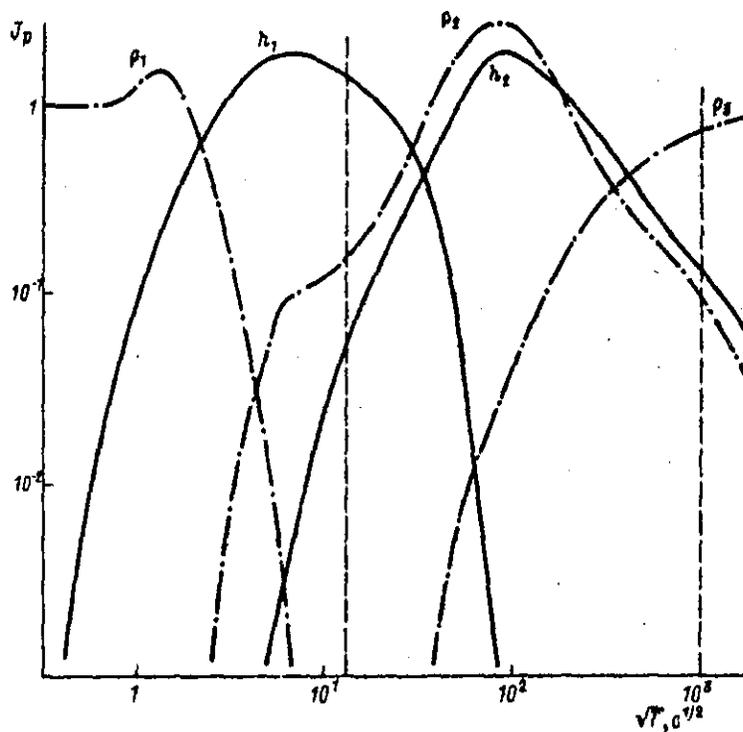
Подобный анализ может быть полезен и при выборе укрупненных параметров: замене двух или нескольких соседних слоев одним. Это целесообразно делать в случаях, когда оценки параметров соседних слоев обнаруживают сильную корреляционную связь.

П р и м е р. Проиллюстрируем изложенное на примере информационного анализа выборки III из примера I § 6.

На рис. 19 изображены информативные области для параметров $h_1, h_2, \varphi_1, \varphi_2, \varphi_3$ в интервале периодов $T \in (10^{-2}, 10^3)$ с. Вертикальными линиями отмечен диапазон периодов, соответствующих выборке III. Видно, что наблюдения не попадают в информативную область параметра φ_1 и лишь частично попадают в информативные области параметров h_1 и φ_3 . Обратим внимание читателя на то, что из-за отсутствия информации о параметрах верхнего слоя

повысилась корреляция между оценками параметров соседних слоев (см. табл. 2). Из рассмотрения информативных областей следует, что в данном случае наиболее разумно фиксировать параметры h_1 и ρ_1 . В табл. I были приведены результаты оценки параметров по выборке III в случае, когда оценивались все параметры совместно и когда h_1 и ρ_1 принимали фиксированные значения. В последнем случае значения h_1 и ρ_1 считались равными точным значениям и соответствующими начальному приближению этих параметров.

Закрепление параметров h_1 и ρ_1 приводит к существенно уменьшению погрешности оценки параметров второго слоя h_2 , ρ_2 . Замена точных значений h_2 и ρ_2 приближенными практически не сказывается на оценках.



Большая погрешность оценки ρ_3 объясняется тем, что в исходных данных недостаточно информации о ρ_3 . Для увеличения точности оценки этого параметра необходимо расширить интервал периодов по крайней мере до 10^8 с.

§ 9. Оценка параметров при наличии априорной информации

При решении обратных задач большое значение имеет учет априорных сведений о среде. Построение модели среды всегда основывается на каких-то представлениях о характере изучаемой структуры, представлениях, существующих до проведения эксперимента. При выборе начального приближения для параметров, как уже упоминалось ранее, используются априорные сведения о возможных значениях этих параметров. Но в оптимизационной схеме решения обратных задач априорную информацию о модели можно учитывать и в более сложной форме. Одна из возможностей — учет связей между параметрами или ограничений на параметры типа неравенств. Такие ограничения могут быть либо следствием физической природы изучаемой модели, либо результатом других исследований. Другая возможность — использование результатов оценки параметров модели по данным предшествующих наблюдений путем задания априорной вероятности параметров. В этом случае параметры можно оценивать по методу максимума апостериорной вероятности. Слабым местом такого подхода к оценке параметров, как уже упоминалось в § I, является то, что параметры в общем не являются случайными величинами, и для них поэтому теряет смысл понятие вероятности. Но если информация о параметрах есть результат некоторых проведенных ранее независимых наблюдений, то использование в качестве априорного распределения параметра распределения этой оценки (а оценка, напомним, является случайной величиной!) эквивалентно тому, что в качестве исходных данных для решения обратной задачи используются совместно результаты двух экспериментов — настоящего (I) и предшествующего (II). Покажем это.

Пусть выборка данных, используемых для решения обратной задачи, — Y_1 , а выборка, на основании которой построено апри-

орное распределение параметров, $-y_{II}$. Функция правдоподобия, учитывающая данные двух выборок, представляется суммой

$$L_{y_I, y_{II}}(m) = \log p(y_I; m) + \log p(y_{II}; m) = L_{y_I}(m) + L_{y_{II}}(m). \quad (73)$$

Предположим, что оценки максимального правдоподобия, полученные отдельно по данным выборок I и II, различаются не слишком сильно. Разложим $L_{y_{II}}(m)$ в ряд Тейлора в окрестности максимума \hat{m}_a и ограничимся квадратичным членом:

$$L_{y_{II}}(m) \approx L_{y_{II}}(\hat{m}_a) + \frac{1}{2}(m - \hat{m}_a)^T J(m - \hat{m}_a). \quad (74)$$

Здесь J - матрица, имеющая элементы $(\partial^2 L_{y_{II}}(m) / \partial m_i \partial m_j)_{\hat{m}_a}$. Как показано в § 6, приближенно можно считать, что эта матрица с точностью до знака совпадает с информационной матрицей Фишера, так что $J = R_{\hat{m}_a}^{-1}$, где $R_{\hat{m}_a}$ - ковариационная матрица оценки \hat{m}_a .

Поскольку $L_{y_{II}}(\hat{m}_a)$ не зависит от m , то максимизация (73) эквивалентна максимизации функции

$$L_{m_a, y_I}(m) = \log p(y_I; m) - \frac{1}{2}(m - \hat{m}_a)^T R_{\hat{m}_a}^{-1}(m - \hat{m}_a).$$

Если теперь формально трактовать параметры m как случайные величины, имеющие априорное нормальное распределение со средним значением \hat{m}_a и ковариационной матрицей $R_{\hat{m}_a}$, то $\tilde{L}(m)$ с точностью до постоянного слагаемого есть логарифм апостериорной вероятности параметра m , и, таким образом, значение \hat{m} , максимизирующее $\tilde{L}(m)$, есть оценка максимума апостериорной вероятности. Часто $L_{m_a, y_I}(m)$ называют обобщенной [37] или апостериорной функцией правдоподобия.

Рассмотрим, к чему сведется оценка максимума апостериорной вероятности в случае, когда наблюдения y_j независимы и распределены нормально с равными дисперсиями $\sigma_{y_j}^2$, а априорная ковариационная матрица оценок параметров $R_{\hat{m}_a} = \sigma_{\hat{m}_a}^2 I$. В этом случае оценка получается из условия минимизации следующей целевой функции:

$$\lambda(m) = \frac{1}{\sigma_{y_I}^2}(y - \Phi(m))^T (y - \Phi(m)) + \frac{1}{2\sigma_{\hat{m}_a}^2}(m - \hat{m}_a)^T (m - \hat{m}_a).$$

Нахождение минимума $\lambda(m)$ эквивалентно решению системы нелинейных уравнений

$$\frac{\partial \lambda}{\partial m_g} = 0, \quad g = 1, 2, \dots, k.$$

Используя метод Ньютона - Рафсона (см. § 4), мы можем свести решение системы нелинейных уравнений к системе линейных урав-

нений относительно поправки $\Delta m^{(s)} = m - m^{(s-1)}$ ($m^{(s-1)}$ - приближенное искомого вектора параметров на $(s-1)$ -м шаге итеративного процесса):

$$A \Delta m^{(s)} = B, \quad (75)$$

$$\text{где } A_{ij} = E \left(\frac{\partial \lambda(m)}{\partial m_i} \frac{\partial \lambda(m)}{\partial m_j} \right)_{m=m^{(s-1)}} = \frac{1}{\sigma_y^2} \frac{\partial \Phi(m)}{\partial m_i} \frac{\partial \Phi(m)}{\partial m_j} + \frac{1}{\sigma_{m_0}^2} \delta_{ij},$$

$$B_j = \frac{1}{\sigma_y^2} \left(\frac{\partial \Phi(m)}{\partial m_j} \right)_{m=m^{(s-1)}} (y - \Phi(m^{(s-1)})) + \frac{1}{\sigma_{m_0}^2} \mu_j^{(s)},$$

$$\mu_j^{(s)} = \hat{m} - m^{(s-1)}, \quad \delta_{ij} = \begin{cases} 1 & \text{при } i=j, \\ 0 & \text{при } i \neq j. \end{cases}$$

Легко видеть, что при отсутствии априорной информации о распределении параметра m ($\sigma_{m_0}^2 \rightarrow \infty$) система (75) переходит в систему нормальных уравнений метода наименьших квадратов. Как уже упоминалось в предыдущих параграфах, если между оценками параметров существует сильная корреляционная связь, то матрица системы нормальных уравнений оказывается плохо обусловленной. Если же учитывается априорное распределение параметров в приведенной выше форме, обусловленность этой матрицы повышается в результате прибавления к ее диагональным элементам положительных слагаемых.

Таким образом, учет априорного распределения приводит к повышению устойчивости решения. Более подробно вопросы повышения устойчивости решения будут рассмотрены в гл. III.

Отметим недостаток метода учета информации о параметрах, заданной в виде априорного распределения (точнее, путем задания априорного среднего значения и ковариационной матрицы). Из оказанного ранее следует, что представление в таком виде информации о предыдущих исследованиях правомерно лишь в случае, когда априорная оценка близка к той, которая получается из данного эксперимента, а априорные дисперсии параметров не слишком велики, - иначе приближение (74) перестанет быть справедливым. Но чаще всего априорные данные имеют малую точность, и в этом случае распределение оценки (полученной по предыдущим исследованиям), которое мы принимаем в качестве априорного распределения параметров, может сильно отличаться от нормального (см. § 6). Поэтому принятие нормального распределения в качестве априорного может привести к сильным искажениям.

Тем более, что обычно используют не полностью ковариационную матрицу оценок $R_{\hat{a}_0}$, а лишь сведения о дисперсиях параметров. Все это приводит к тому, что использование априорной информации в такой форме содержит значительный произвол и может приводить к неверным выводам.

Более корректным с точки зрения общих представлений об изучаемой среде является задание априорной информации путем указания области допустимых значений параметров в пространстве M . Даже при отсутствии предварительных геофизических исследований можно указать реальные пределы возможных изменений того или иного параметра. Например, заведомо можно считать, что плотность вещества земной коры может принимать значения от 1 до 5 г/см³; если в качестве параметров взяты мощности слоев, то они обязаны быть неотрицательными и т.д. В действительности, благодаря знаниям, полученным из предыдущих исследований, эти пределы могут быть значительно сужены.

Простейшая форма ограничений, определяющих область допустимых значений для параметров в пространстве M , — это задание пределов изменения для каждого из параметров:

$$m_i^{\min} \leq m_i \leq m_i^{\max}. \quad (76)$$

В более общем виде область допустимых значений может быть задана неравенствами

$$g_j(m) \leq 0, \quad j=1, 2, \dots, s, \quad (77)$$

где $g_j(m)$ — некоторая, в общем нелинейная функция параметров m .

Задача, таким образом, сводится к нахождению минимума (или максимума) целевой функции $\lambda(m)$ при ограничениях (76) или (77). Заметим, что если $\lambda(m)$ — целевая функция, построенная на основе функции правдоподобия, то задачу нахождения экстремума этой функции при ограничениях (76) можно трактовать как определение оценки максимальной апостериорной вероятности, если считать, что параметры являются случайными величинами, априорно подчиняющимися равномерному распределению (76). В этом случае, однако, нельзя для нахождения оценки использовать методы поиска экстремума, описанные в § 7, поскольку апостериорная вероятность не будет уже непрерывной функцией. Максимум апостериорной функции правдоподобия в данном случае будет либо

совпадать с экстремумом $\lambda(m)$ (если он оказывается внутри многомерного параллелепипеда, определяемого неравенствами (76)), либо окажется на границе такой области. На рис. 20 изображены обе эти возможности для случая, когда модель описывается одним параметром (пространство M - одномерное).

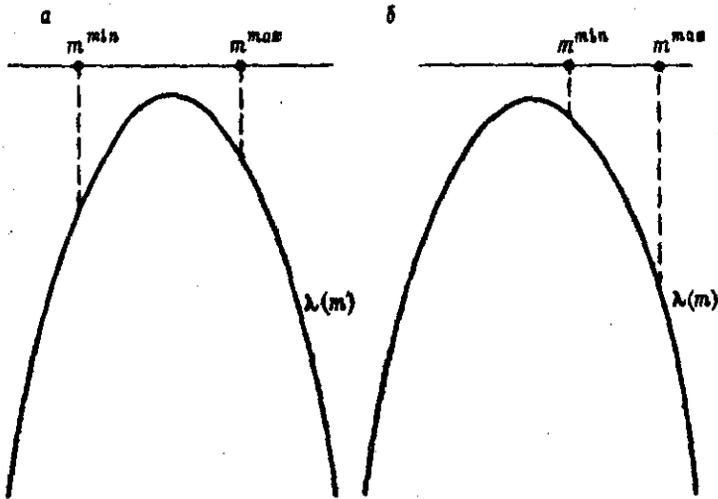


Рис. 20.

Сказанное относится полностью и к случаю, когда область допустимых значений параметров определена неравенствами (77): для поиска экстремума целевой функции не могут быть использованы описанные в § 7 методы. Для решения экстремальных задач при наличии ограничений типа (76) или (77) применяются методы математического программирования.

Мы начнем с рассмотрения более общего случая, когда ограничения заданы неравенствами типа (77), поскольку линейные неравенства (76) могут рассматриваться как частный случай таких ограничений. Наиболее распространенным способом решения этих задач является метод штрафных функций [20]. Его идея заключается в следующем.

Для конкретности будем считать, что требуется определить минимум целевой функции $\lambda(m)$ при ограничениях (77). Вводится

последовательность функций $\{P_r(m), r=1,2,\dots\}$, неотрицательных и непрерывных в пространстве M , обладающих следующим свойством: при $r \rightarrow \infty$ $P_r(m) \rightarrow 0$, если выполнены условия (77), и $P_r(m) \rightarrow \infty$, если хотя бы одно из этих условий нарушено. Функцию $P_r(m)$ называют штрафом или штрафной функцией. Пример штрафной функции

$$P_r(m) = r \sum_{i=1}^k \{g_i(m)[1 + \operatorname{sgn} g_i(m)]\}^2.$$

Исходная экстремальная задача на поиск минимума $\lambda(m)$ при наличии ограничений (77) сводится теперь к поиску безусловного минимума функции:

$$Q_r(m) = \lambda(m) + P_r(m). \quad (78)$$

Функция $P_r(m)$ играет роль своеобразного штрафа за нарушение условий (77): чем больше r , тем быстрее возрастает функция $P_r(m)$ за пределами области, внутри которой должно находиться решение. При $r \rightarrow \infty$ решение задачи минимизации (78) сходится к решению исходной задачи.

Недостатком этого метода является медленная сходимость, вызванная тем, что при больших значениях r рельеф функции $Q_r(m)$ становится очень сложным: вблизи границы области появляются крутые откосы и овраги. Для ускорения сходимости используют следующий прием. Вначале берут небольшое значение r и находят минимум $Q_r(m)$. Затем увеличивают r , а в качестве начального приближения для поиска используют решение, полученное при предыдущем значении r . Такую процедуру повторяют до тех пор, пока штраф не станет достаточно малым.

Хотя ограничения типа линейных неравенств (76) могут рассматриваться как частный случай ограничений общего вида (77) и, следовательно, при решении задач с такими ограничениями может также применяться метод штрафных функций, оказывается более целесообразным в этом случае использовать методы линейного программирования. Метод решения обратной задачи для случая, когда распределение ошибок наблюдений нормальное, а ограничения на параметры имеют вид (76), описан в [56]. Мы будем следовать здесь этому изложению.

Примем для простоты, что наблюдения независимы и имеют одинаковую дисперсию. Тогда оценку m следует производить пу-

тем минимизации суммы квадратов отклонений

$$[y - \Phi(m)]^T [y - \Phi(m)], \quad (79)$$

где $\Phi(m)$ - оператор решения прямой задачи. Минимизацию (79) следует производить при условиях (76).

Линеаризуем задачу с тем, чтобы свести ее к линейному оцениванию по методу наименьших квадратов. Выберем некоторое начальное приближение $m^{(0)}$ (которое, естественно, должно удовлетворять условиям (76)) и вместо m будем искать поправку к этому начальному приближению: $\Delta m = m - m^{(0)}$. Тогда приближенно можно записать

$$\Phi(m) = \Phi(m^{(0)}) + A \Delta m,$$

здесь матрица A имеет элементы $A_{ij} = \frac{\partial \Phi_i(m^{(0)})}{\partial m_j}$. Обозначим $u = y - \Phi(m^{(0)})$ и получим, что вместо функции отклика (79) следует минимизировать

$$(A \Delta m - u)^T (A \Delta m - u) \quad (80)$$

при условиях*

$$\mu_1 \leq \Delta m \leq \mu_2, \quad (81)$$

где $\mu_1 = m^{\min} - m^{(0)}$, $\mu_2 = m^{\max} - m^{(0)}$. Систему неравенств (81) можно записать в виде

$$B \Delta m \leq \gamma, \quad (82)$$

где матрица $B = \begin{pmatrix} I_k \\ -I_k \end{pmatrix}$, а $\gamma = \begin{pmatrix} \mu_2 \\ -\mu_1 \end{pmatrix}$.

Минимизация (80) эквивалентна нахождению минимума целевой функции

$$\lambda(\Delta m) = \Delta m^T A^T A \Delta m - 2u^T A \Delta m. \quad (83)$$

Для того чтобы свести задачу минимизации (83) при условиях (82) к задаче линейного программирования, введем еще одну дополнительную переменную z и вместо минимума нелинейной функции (83) будем искать минимум z при условии

$$\Delta m^T A^T A \Delta m - 2u^T A \Delta m - z \leq 0. \quad (84)$$

Обозначим для удобства $A^T A = C$, $2A^T u = -P$, тогда условие (84) можно записать в виде

$$\Delta m^T C \Delta m + P^T \Delta m - z \leq 0. \quad (85)$$

* Под векторным неравенством мы будем понимать систему неравенств для всех компонент вектора.

Система неравенств (85) не является, однако, системой линейных неравенств. Для линеаризации этой системы предлагается заменить квадратичную (функцию (83) кусочно-линейной. Для этого выберем в области $\mu_1 \leq \Delta m \leq \mu_2$, K точек Δm_g и будем считать, что приближенно область, определяемая неравенством (85), совпадает с областью, определяемой K линейными неравенствами:

$$(2C\Delta m_g + p)^T \Delta m - x \leq \Delta m_g^T C \Delta m_g, \quad g = 1, 2, \dots, K. \quad (86)$$

Геометрически это означает, что границу области (85), являющуюся гиперпараболомидом в $(K+1)$ -мерном пространстве переменных $\Delta m_1, \Delta m_2, \dots, \Delta m_K, x$ мы заменим набором K гиперплоскостей, касательных к этому гиперпараболомиду в точках Δm_g .

Теперь задача оказывается сведена к следующей задаче линейного программирования:

$$C^T x = \text{max}, \quad Dx \leq b, \quad (87)$$

где $x = \begin{pmatrix} \Delta m \\ x \end{pmatrix}$, $C^T = (0, 0, \dots, 0, -1)$, а D и b могут быть легко построены, если объединить условия (82) и (86).

При решении задачи (87) используется теория двойственности, являющаяся центральной в линейном программировании. Если исходная задача состоит в максимизации, то двойственная к ней — в минимизации, и решение одной прямо приводит к решению другой. Мы здесь не будем подробно пояснять эту теорию, отсылая читателя к специальным руководствам [4, 39].

К задаче (87) двойственной является следующая:

$$\theta^T p = \min, \quad D^T p = c, \quad p \geq 0. \quad (88)$$

По решению задачи (88) p^* легко может быть построено решение задачи (87) с учетом следующего свойства решений взаимодвойственных задач: если $p_i^* > 0$, то

$$\sum_j a_{ij} x_j = b_i. \quad (89)$$

так что нахождение решения (87) сводится просто к решению системы линейных уравнений (89) для соответствующих индексов i . А решение задачи (88) осуществляется с помощью симплекс-метода. Мы здесь не будем останавливаться на его деталях: описание симплекс-метода можно найти в любом курсе по линейному программированию, например в работах [4, 39].

Итак, если ввести новые переменные $u = \{u_1, u_2, \dots, u_{2K}\}$ и w_s ($s = 1, 2, \dots, K$), то задача, двойственная к исходной, может быть сформулирована следующим образом:

$$\begin{aligned} \gamma^T u + \sum_{s=1}^K (\Delta m_s^T C \Delta m_s) \omega_s &= \min, \\ \beta^T u + \sum_{s=1}^K (p + 2C \Delta m_s) \omega_s &= 0, \\ \sum_{s=1}^K \omega_s &= 1, \quad u > 0, \quad \omega_s \geq 0. \end{aligned}$$

Построенное таким способом решение для Δm получается, конечно, приближенным. Но его можно использовать для выбора следующего начального приближения $m^{(1)} = m^{(0)} + \Delta m$ и повторить процесс определения поправок $\Delta m^{(r)}$ до стабилизации решения.

§ 10. Равномерное распределение ошибок наблюдений

До сих пор мы считали, что функция распределения наблюдений является регулярной (см. § 1). Это позволяло строить функцию правдоподобия и путем ее максимизации находить оценки параметров. При этом как в случае отсутствия априорных ограничений, так и при их наличии обратная задача сводилась к нахождению экстремума некоторой функции отклика в пространстве \mathcal{M} . Теперь рассмотрим случай, когда распределение наблюдений оказывается равномерным:

$$\Phi(m) - \epsilon \leq y \leq \Phi(m) + \epsilon. \quad (90)$$

Как уже упоминалось в § 1, такая функция распределения не удовлетворяет условию регулярности, так что в этом случае становится невозможным использовать выводы статистической теории оценивания для нахождения оценок параметров. В то же время во многих геофизических задачах можно считать, что наблюдения удовлетворяют условию (90), т.е. подчиняются равномерному распределению.

Двойное неравенство (90) можно записать в виде

$$y - \epsilon \leq \Phi(m) \leq y + \epsilon \quad (91)$$

и искать такие параметры m , которые удовлетворяют этому условию. Очевидно, что в этом случае "решением" будет уже не точка, а некоторая область в пространстве \mathcal{M} , причем все точки этой области будут равноправны: из них нельзя выделить в большей или в меньшей степени правдоподобные. Эта область может быть сужена в результате наложения априорных ограничений на па-

параметры. В гл. IV будет рассмотрен случай, когда из всех возможных решений, удовлетворяющих неравенству (91), выбирается такое, которое удовлетворяет минимуму нормы в метрике L_2 . Здесь же мы рассмотрим решение обратной задачи, удовлетворяющее неравенству (91) при условии, что параметры подчинены ограничениям типа линейных неравенств (76). Для простоты будем считать оператор прямой задачи линейным, так что обратная задача может быть сформулирована как определение области в пространстве M , точки которой удовлетворяют следующей системе линейных неравенств:

$$\begin{aligned} y - \varepsilon &\leq Am \leq y + \varepsilon, \\ m^{\min} &\leq m \leq m^{\max}. \end{aligned} \quad (92)$$

Способ построения такой области предложен в работе [53] для решения обратной задачи гравиметрии, в которой оператор прямой задачи действительно является линейным.

Пусть размерность пространства параметров модели k , а размерность пространства наблюдений n . Введем новый неизвестный $(k+n)$ -мерный вектор x :

$$\begin{aligned} x_i &= m_i - m_i^{\min}, \quad i=1, 2, \dots, k, \\ x_{k+i} &= y_i + \varepsilon_i - (Am)_i, \quad i=1, 2, \dots, n. \end{aligned}$$

Тогда задача (92) может быть записана в форме

$$Gx = b, \quad 0 \leq x \leq X, \quad (93)$$

где матрица G имеет размер $n \times (k+n)$ и строится добавлением справа к матрице A единичной матрицы I_n , т.е. $G = \|A, I_n\|$,

$$b = y + \varepsilon - Am^{\min},$$

а вектор X определяется следующим образом:

$$\begin{aligned} X_i &= m_i^{\max} - m_i^{\min}, \quad i=1, 2, \dots, k, \\ X_{i+k} &= 2\varepsilon_i, \quad i=1, 2, \dots, n. \end{aligned}$$

Геометрическая интерпретация решения задачи (93) дана на рис.21 для $k=2$, $n=1$. Решением является часть плоскости $Gx = b$, вырезаемая параллелепипедом $0 \leq x \leq X$. Точки этой области, являющиеся в общем случае внутренними точками некоторого выпуклого многогранника, определяются соотношением

$$x = \sum_{s=1}^k \nu_s x^{(s)}, \quad (94)$$

где $x^{(s)}$ - точки, соответствующие вершинам многогранника; ν_s - положительные числа, такие, что $\sum_s \nu_s = 1$.

Таким образом, задача сводится к нахождению частных решений $x^{(j)}$ называемых допустимыми базисными решениями (ЭБ). Определение этих частных решений производится следующим образом.

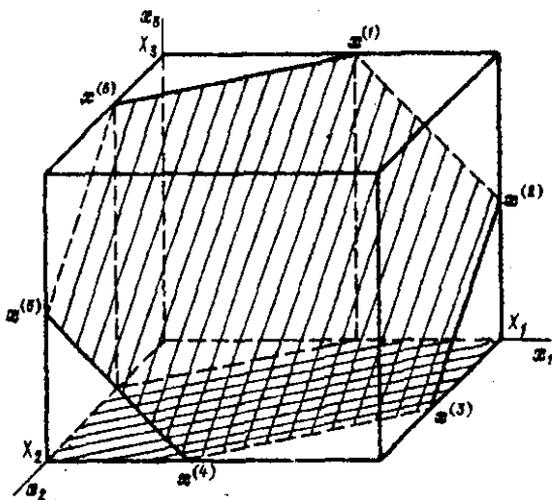


Рис. 21.

Из матрицы G извлекается n столбцов G_j с номерами $\{j_1, j_2, \dots, j_n\}$, из которых строится квадратная матрица G_j (считаем, что ее ранг равен n). Набор чисел $\{j_1, j_2, \dots, j_n\}$ обозначим через J . Из оставшихся столбцов G_j выбирается $q \leq k$ столбцов с номерами $\{j_{p_1}, j_{p_2}, \dots, j_{p_q}\}$, совокупность которых обозначим через J_q . Для выбранных множеств J и J_q решается линейная система

$$G_j x = b - \sum_{j \in J_q} \lambda_j G_j.$$

Базисное решение, соответствующее заданным J и J_q , определяется следующим образом:

$$\begin{aligned} x_{j_i} &= z_i, \quad j_i \in J, \\ x_j &= \lambda_j, \quad j \in J_q, \\ x_j &= 0, \quad j \notin J \cup J_q. \end{aligned}$$

Допустимым базисным решением будет такое, которое удовлетворяет условию $0 \leq x \leq X$. Поскольку подмножества J и J_0 могут быть извлечены из исходного набора индексов $2^k C_{n+k}^n$ различными способами, то число допустимых базисных решений $K \leq 2^k C_{n+k}^n$.

Теперь вернемся к исходной задаче, которая состояла в том, чтобы определить область в пространстве M^k , удовлетворяющую (92). Запишем выражение (94) для k первых компонент вектора x :

$$x_i = m_i - m_i^{\min} = \sum_{s=1}^k \nu_s (m_i^{(s)} - m_i^{\min}).$$

Учитывая, что $\sum \nu_s = 1$, получим

$$m_i = \sum_{s=1}^k \nu_s m_i^{(s)}, \quad i=1, 2, \dots, k. \quad (95)$$

В пространстве M^k область, определяемая соотношением (95), представляет собой в общем случае выпуклый многогранник. Для иллюстрации этого обратимся снова к рис. 21. Поскольку в данном случае размерность пространства M принята равной двум, так что $x_1 = m_1 - m_1^{\min}$, $x_2 = m_2 - m_2^{\min}$, то пространство M будет совпадать с плоскостью (x_1, x_2) , причем началу координат $x_1 = x_2 = 0$ будет соответствовать в пространстве M точка (m_1^{\min}, m_2^{\min}) . Область решений будет в этом случае заштрихованным участком плоскости (x_1, x_2) .

Сделаем теперь некоторые замечания, касающиеся того, как подходить к решению задачи (91) при линейных ограничениях на параметры, если оператор прямой задачи оказывается нелинейным. Очевидно, что в этом случае путем линейризации задачи и использования метода последовательных приближений невозможно приблизиться к точному решению, так как область в пространстве M , удовлетворяющая (91), в общем случае будет ограничена криволинейной гиперповерхностью. Поэтому можно только пытаться аппроксимировать эту область многогранником, что осуществляется, например, следующим образом. Выберем некоторое начальное приближение для параметров $m^{(0)}$ и определим область для поправки к нему описанным ранее методом. Эта область будет представлять некоторый выпуклый многогранник. Определим центр тяжести этого многогранника $\Delta m^{(1)}$ и построим следующее приближение: $m^{(1)} = m^{(0)} + \Delta m^{(1)}$. Такую процедуру будем повторять до тех пор, пока $\Delta m^{(r)}$ не станет близкой к нулю.

§ I. Понятие регуляризирующего оператора

В § 4 гл. I было определено понятие некорректно поставленной задачи и указано, что большинство обратных задач геофизики относится к классу некорректных. Некорректность обратных задач приводит к тому, что точное решение оказывается искать бессмысленно, и уже в самой постановке задачи ставится требование поиска приближенного решения, удовлетворяющего тем или иным условиям.

Пусть оператор решения прямой задачи

$$\Phi(m) = y, \quad m \in M, \quad y \in Y. \quad (96)$$

Мы будем считать, что решение обратной задачи — нахождение m по заданному y — при точных входных данных единственно. Но при наличии даже достаточно малых возмущений правой части (96) может оказаться, что решения $m \in M$, удовлетворяющего точно (96), не существует или оно находится очень далеко от m_0 , соответствующего невозмущенной правой части y_0 . В таком случае, как было указано в предыдущей главе, для построения приближенного решения (96) используются методы оптимизации, основанные на учете статистических свойств помехи. Но и эти оценки могут быть неустойчивы по отношению к малым вариациям y , и при различных реализациях помехи $e = y - y_0$ они могут сильно различаться. Эта неустойчивость выражается в том, что дисперсия оценок оказывается очень большой, причем иногда настолько, что сами оценки теряют смысл. В таких случаях основное требование, предъявляемое к приближенному решению, заключается в том, что решение должно быть устойчиво по отношению к вариациям исходных данных в пределах их погрешности. При этом выгоднее допустить некоторое смещение оценок, но добиться уменьшения их дисперсии.

Для построения приближенных решений некорректных задач широкого класса А.Н.Тихоновым был разработан метод [18], осно-

ванный на принципе регуляризации. Этот метод успешно применяется и для решения обратных задач геофизики [24, 34, 35, 36, 38, 41].

Согласно принципу регуляризации приближенное решение должно сколь угодно точно аппроксимировать истинное решение m_0 при приближении y к y_0 . Иначе говоря, пусть y отличается от y_0 в метрике пространства Y на величину δ . Приближенное решение, удовлетворяющее принципу регуляризации, должно быть выбрано так, чтобы при $\delta \rightarrow 0$ m стремилось бы к m_0 в метрике пространства M . Оператор $R_\delta(y)$ построения решения m называется регуляризирующим, если он обладает следующими свойствами: 1) определен при всех достаточно малых δ ; 2) для любого $\epsilon > 0$ найдется $\delta(\epsilon)$ такое, что если $\|y - y_0\| < \delta$, то $\|m - m_0\| < \epsilon$, где $m = R_\delta(y)$. Решение m , полученное с помощью регуляризирующего оператора, называется регуляризованным решением (96). Оператор $R_\delta(y)$ включает в себя некоторый параметр $\alpha = \alpha(\delta)$, который называется параметром регуляризации.

Таким образом, задача нахождения приближенного решения сводится, во-первых, к построению регуляризирующего оператора $R_\delta(y; \alpha)$ и во-вторых, к выбору параметра регуляризации α , согласованного с точностью исходных данных δ .

Способ построения регуляризирующих операторов для уравнения (96) основан на априорных представлениях о характере решения. Такой дополнительной информацией о решении может быть, например, требование гладкости решения (если $m = m(r)$) или близости решения к некоторой заданной точке в пространстве M и т.д. На основании этих представлений строится так называемый стабилизирующий функционал $\Omega(m)$, и регуляризованное решение (96) определяется как одно из возможных приближенных решений, удовлетворяющих условию

$$\|\Phi(m) - y\| \leq \delta \quad (97)$$

и минимизирующее функционал $\Omega(m)$.

Очевидно, что если минимум $\Omega(m)$ достигается внутри области, определяемой (97), то его и следует принимать в качестве решения. В противном случае, при условии квазимонотонности функционала $\Omega(m)$, наименьшее значение функционала достигается на границе области (97), т.е. при значениях m , дающем минимум функционалу $\Omega(m)$ при условии $\|\Phi(m) - y\| = \delta$. Это-

задача на нахождение условного экстремума, и она может быть решена методом неспределенных множителей Лагранжа. Сведем ее к отысканию минимума функционала:

$$\mathcal{M}^{\alpha}(m, y) = \|\Phi(m) - y\|^2 + \alpha \Omega(m). \quad (98)$$

При этом параметр α должен быть определен из условия $\|\Phi(m) - y\| = \delta$. Функционал $\mathcal{M}^{\alpha}(m, y)$ называется сглаживающим функционалом.

Все сказанное справедливо для произвольного выбора метрики как в пространстве Y , как и в пространстве M . В задачах геофизики практически всегда метрику пространства Y можно принимать квадратичной. Это следует из того, что $\Phi(m) - y$ можно рассматривать как случайную помеху, распределение которой практически во всех случаях можно свести к нормальному (см. гл. II). А выборочное пространство с нормальным распределением всегда можно некоторой трансформацией привести к пространству с квадратичной метрикой. Действительно, если ковариационная матрица помехи R , то линейное преобразование $y' = R^{-1/2}y$ переводит пространство Y в Y' , с нормальным распределением, характеризующимся ковариационной матрицей I , что эквивалентно пространству с квадратичной метрикой.

Вид оператора $\Phi(m)$ в задачах геофизики, как уже говорилось в гл. I, может быть самым разнообразным. Если этот оператор нелинеен (или вообще не представим в аналитическом виде), то нахождение регуляризованного решения сводится к вычислительной задаче поиска экстремума в пространстве M . При этом благодаря введению стабилизирующего функционала экстремум $\mathcal{M}^{\alpha}(m, y)$ является локализованным, и для его поиска можно применять итеративный метод Ньютона - Рафсона, или (что то же самое) линеаризовать оператор $\Phi(m)$ и искать поправки $\Delta m^{(i)}$ к последовательным приближениям $m^{(i)}$. В зависимости от того, какова размерность пространств M и Y (конечная или бесконечная) линейный оператор, определяющий поправку $\Delta y^{(i)} = y - \Phi(m^{(i)})$ через поправку $\Delta m^{(i)} = m - m^{(i)}$, может иметь вид либо

$$A \Delta m = \Delta y,$$

если $\Delta m \in M^k$, $\Delta y \in Y^n$, при этом A есть матрица $n \times k$, либо

$$\Delta y(x) = \int_a^b G(x, r) \Delta m(r) dr,$$

если $\Delta m(r)$ и $\Delta y(x)$ принадлежат функциональным пространствам; соответственно $G(x, r)$ является ядром интегрального преобразования.

Возможен и более общий случай, включающий оба упомянутых способа задания модели, а именно когда модель соответствует слоисто-неоднородной среде с непрерывным распределением искомого параметра $m(r)$ внутри каждого слоя $r_s < r < r_{s+1}$ ($s = 1, 2, \dots$) и скачками $\mu^s = m(r_s^+) - m(r_s^-)$ на границах слоев. В этом случае

$$\Delta y(x) = \sum_s \int_{r_s}^{r_{s+1}} G_s(x, r) \Delta m(r) dr + \sum_s \Delta \mu^s a_s(x). \quad (99)$$

Мы рассмотрим в последующих параграфах только первые два случая, поскольку задача (99) может быть сведена к ним.

§ 2. Построение регуляризованного решения линейной системы

Рассмотрим линейризованную задачу

$$A m = y, \quad (100)$$

понимая под m поправку $\Delta m^{(i)}$, а под y соответственно поправку к исходным наблюдениям относительно рассчитанной для модели, соответствующей $(i-1)$ -му приближению. Согласно принятому в § I предположению, будем считать, что решение (100) при точных входных данных единственно. Это означает, что ранг матрицы A должен быть равен числу искомого параметров k , т.е. размерности пространства M . При этом число уравнений может быть больше числа неизвестных ($n \geq k$). Если правые части выражения (100) содержат погрешности, то, как было показано в § 5 гл. II, приближенное решение может быть построено методом наименьших квадратов. Полученные таким способом оценки являются несмещенными и имеют наименьшую дисперсию в классе линейных несмещенных оценок. Но в случае, если матрица системы нормальных уравнений $A^T A$ оказывается плохо обусловленной, то решение по методу наименьших квадратов неустойчиво и его дисперсия может стать около угодно большой. При этом имеет смысл оторвать приближенное решение в классе смещенных оценок, но характеризующихся меньшей дисперсией. Именно таким свойством обладает решение, получаемое по методу регуляризации.

Стабилизирующий функционал $B_2(m)$ в этом случае выбирается в виде $\|m - m_0\|^2$, где m_0 — некоторое априорное значение, которое может и не быть очень близким к истинному решению. Принимая метрики пространств M и Y квадратичными, получим, что регуляризованное решение минимизирует следующий функционал:

$$(Am - y)^T (Am - y) + \alpha (m - m_0)^T (m - m_0)$$

или (что то же самое) отвечает решению системы

$$(A^T A + \alpha I) m = A^T y + \alpha m_0.$$

Это решение имеет вид $m = (A^T A + \alpha I)^{-1} (A^T y + \alpha m_0)$.

При $\alpha = 0$ регуляризованное решение переходит в решение по методу наименьших квадратов. Его можно рассматривать как оценку искомого параметра.

Покажем, чему равно смещение регуляризованной оценки. Пусть $F(y) = y_0$. Поскольку решение по методу наименьших квадратов является, как известно, несмещенным, то смещение регуляризованного решения можно определять относительно него, т.е. относительно $m_0 = (A^T A)^{-1} A^T y_0$. Таким образом,

$$\begin{aligned} \delta_\alpha &= F(m) - m_0 = \\ &= (A^T A + \alpha I)^{-1} (A^T y_0 + \alpha m_0) - (A^T A)^{-1} A^T y_0 = \\ &= (I + \alpha (A^T A)^{-1})^{-1} (A^T A)^{-1} (A^T y_0 + \alpha m_0) - (A^T A)^{-1} A^T y_0 = \\ &= (I + \alpha (A^T A)^{-1})^{-1} [\alpha (A^T A)^{-1} m_0 - \alpha (A^T A)^{-1} m_0] = \\ &= \alpha (A^T A + \alpha I)^{-1} (m_0 - m_0). \end{aligned} \quad (101)$$

Отсюда видно, что смещение регуляризованной оценки зависит от величины параметра регуляризации α (при $\alpha = 0$ смещение становится равным нулю) и от того, насколько априорное значение m_0 отличается от истинного решения m_0 .

Теперь рассмотрим дисперсию регуляризованного решения. Прежде всего заметим, что поскольку мы приняли метрику пространства Y квадратичной, мы тем самым подразумеваем, что случайные погрешности компонент вектора y независимы и распределены нормально с дисперсией $\sigma^2 = \delta^2/n$. Для такого случая, как уже было показано в § 5 гл. II, ковариационная матрица оценок наименьших квадратов имеет вид

$$R_{n,k} = \sigma^2 (A^T A)^{-1}.$$

Для регуляризованного решения имеем с учетом выражения (101)

$$m - E(m) = (A^T A + \alpha I)^{-1} A^T (y - y_0).$$

Подставляя эту формулу в выражение для ковариационной матрицы

$$R = E \{ (m - E(m))(m - E(m))^T \}$$

и учитывая, что $E \{ (y - y_0)(y - y_0)^T \} = \sigma^2 I$,

$$\begin{aligned} \text{получаем } R_{\text{рег}} &= \sigma^2 (A^T A + \alpha I)^{-1} A^T A (A^T A + \alpha I)^{-1} = \\ &= \sigma^2 (I + \alpha (A^T A)^{-1})^{-1} (A^T A)^{-1}. \end{aligned} \quad (102)$$

Если матрица $A^T A$ является плохо обусловленной, то элементы $(A^T A)^{-1}$ очень велики, и дисперсия оценки, получаемой по методу наименьших квадратов, оказывается очень большой. Элементы же ковариационной матрицы регуляризованной оценки могут быть значительно меньше, как видно из выражения (102).

В качестве примера приведем матрицу

$$A^T A = \begin{pmatrix} 1,43 & 1 \\ 1 & 0,7 \end{pmatrix}.$$

Эта матрица, как легко убедиться, является плохо обусловленной. Обратная ей матрица

$$(A^T A)^{-1} = \begin{pmatrix} 700 & -1000 \\ -1000 & 1430 \end{pmatrix}.$$

Если принять величину параметра регуляризации α равной $0,001$, то матрица $R_{\text{рег}}$, определяемая по формуле (102), будет (при $\sigma^2 = 1$) следующей:

$$\begin{pmatrix} 34,75 & -49,36 \\ -49,36 & 70,79 \end{pmatrix},$$

т.е. в данном случае дисперсия регуляризованного решения составляет приблизительно 5% от дисперсии решения нерегуляризованного.

Заметим, что если число уравнений линейной системы равно числу неизвестных ($n = k$) и матрица системы (100) является положительно определенной, хотя и плохо обусловленной, то, как показано в [24], регуляризованное решение может быть построено путем решения системы

$$(A + \alpha I)m = y.$$

Матрица системы нормальных уравнений в методе наименьших квадратов является, как известно, положительно определенной. Поэ-

тому, если указанным способом строить регуляризованное решение не исходных уравнений (100), а непосредственно системы нормальных уравнений

$$A^T A m = A^T y,$$

то мы получим решение, совпадающее с тем, которое получается путем минимизации сглаживающего функционала для исходной системы при $m_0 = 0$.

§ 3. Регуляризованное решение интегрального уравнения первого рода

Рассмотрим линейризованную задачу, в которой искомая модель принадлежит пространству функций $m(r)$, а наблюдения y определены в виде функциональной зависимости от некоторого параметра x . Связь между $y(x)$ и $m(r)$ определяется интегральным преобразованием вида

$$\int_0^1 G(x, r) m(r) dr = y(x). \quad (103)$$

Примеры обратных геофизических задач, приводящихся к такому уравнению, были даны в § 3 гл. I (примеры 2, 4).

Уравнение (103) является интегральным уравнением Фредгольма первого рода относительно функции $m(r)$. Как известно, задача нахождения решения интегрального уравнения первого рода является некорректно поставленной: решение неустойчиво по отношению к малым вариациям правой части. Для нахождения устойчивого приближенного решения таких уравнений используется метод регуляризации.

Вообще, если решение принадлежит пространству функций, то при построении стабилизирующего функционала должны учитываться априорные ограничения на степень гладкости решения, вытекающие из физической природы задачи. Требование гладкости функции до p -го порядка означает, что функция $m(r)$ должна удовлетворять условию минимизации

$$\Omega(m) = \sum_{k=0}^p \int_0^1 q_k(r) \left[\frac{d^k m}{dr^k} \right]^2 dr, \quad (104)$$

где $q_k(r)$ - некоторые заданные неотрицательные непрерывные функции, $q_p(r) > 0$. Функционал $\Omega(m)$ в форме (104) называется стабилизатором p -го порядка. Для построения регуляризованного

го решения интегрального уравнения (103) можно использовать стабилизатор любого порядка, но практически используются стабилизаторы не выше первого порядка.

Возьмем для построения приближенного решения (103) стабилизатор первого порядка:

$$S_2(m) = \int_{r_1}^{r_2} \left[q_0(r) m^2(r) + q_1(r) \left(\frac{dm}{dr} \right)^2 \right] dr.$$

Регуляризованное решение должно минимизировать функционал

$$\begin{aligned} \mathcal{A}^\omega(m, y) = & \int_a^b \left\{ \int_{r_1}^{r_2} G(x, r) m(r) - y(x) \right\}^2 dx + \\ & + \int_{r_1}^{r_2} \left\{ q_0(r) m^2(r) + q_1(r) \left(\frac{dm}{dr} \right)^2 \right\} dr. \end{aligned} \quad (105)$$

Минимизация функционала (105) определяется из условия равенства нулю его первой вариации, которая записывается в виде

$$\begin{aligned} \int_{r_1}^{r_2} \left\{ -\alpha \left[\frac{d}{dr} \left(q_1(r) \frac{dm}{dr} \right) - q_0(r) m(r) \right] + \int_{r_1}^{r_2} \bar{G}(r, \xi) m(\xi) d\xi - \right. \\ \left. - \delta(r) \right\} \delta m(r) dr + \alpha q_1(r) m'(r) \delta m(r) \Big|_{r_1}^{r_2}, \end{aligned} \quad (106)$$

где $\delta m(r)$ — произвольная вариация функции $m(r)$, такая, что $m(r)$ и $m(r) + \delta m(r)$ принадлежат классу допустимых функций;

$$\bar{G}(r, \xi) = \int_a^b G(x, r) G(x, \xi) dx, \quad \delta(r) = \int_a^b G(x, r) y(x) dx.$$

Условие равенства нулю выражения (106) при произвольной функции $\delta m(r)$ выполняется, если

$$\int_{r_1}^{r_2} \bar{G}(r, \xi) m(\xi) d\xi - \alpha \left\{ \frac{d}{dr} \left(q_1(r) \frac{dm}{dr} \right) - q_0(r) m(r) \right\} = \delta(r), \quad (107)$$

$$q_1(r) m'(r) \delta m(r) \Big|_{r_1}^{r_2} = 0. \quad (108)$$

Если известны значения $m(r)$ на концах промежутка интегрирования (m_1 и m_2), то $\delta m(r_1) = \delta m(r_2) = 0$ и условие (108) выполняется. В этом случае задача нахождения регуляризованного решения сводится к решению интегро-дифференциально-

го уравнения (107) при краевых условиях $m(r_1) = m_1$, $m(r_2) = m_2$.

Если значения $m(r)$ на концах r_1 и r_2 неизвестны, то условие (108) может быть выполнено, если принять

$$m'(r_1) = m'(r_2) = 0. \quad (109)$$

В этом случае в качестве регуляризованного решения можно брать такое решение уравнения (107), которое удовлетворяет крайним условиям (109).

Уравнение (107) с крайними условиями (108) или (109) может быть решено численно, при этом получается устойчивое решение. В частном случае, когда $q_1(r) = 0$, уравнение (107) переходит в интегральное уравнение Фредгольма второго рода, для численного решения которого существуют устойчивые вычислительные схемы.

Определим теперь связь между регуляризованным решением интегрального уравнения и регуляризованным решением линейной системы. Заметим, что если искомую функцию $m(r)$ аппроксимировать кусочно-постоянной, а вместо функции $y(x)$ брать ее значения y_i в дискретных точках x_i , то интегральное уравнение (103) сведется к линейной системе (100). Такая замена интегрального уравнения линейной системой производится при постановке большинства обратных задач геофизики, когда разрез дробится на слои, искомое распределение того или иного физического параметра с глубиной заменяется набором средних значений m_i этого параметра в каждом слое, а наблюдения выполняются в отдельных дискретных точках. Обратим внимание, что в этом случае мощности слоев должны быть фиксированы, т.е. они не должны входить в число искомых параметров разреза.

Нетрудно видеть, что если для регуляризации решения интегрального уравнения (103) взять стабилизатор нулевого порядка ($q_0 = 1$, $q_k = 0$, $k \geq 1$), то при переходе от интегрального уравнения к линейной системе решение (107), (108) сведется к регуляризованному решению линейной системы при $m_0 = 0$. Но в рассматриваемом случае, когда параметры разреза представляют значения изучаемой физической величины в последовательных слоях фиксированной мощности, при построении регуляризованного решения линейной системы можно учесть не только рассмотренное в § 2 априорное ограничение на искомую модель, но и условие, аналогичное условию гладкости функции $m(r)$, заключающееся в том,

что значения m_i в соседних слоях не должны сильно различаться. В простейшем случае, когда мощности всех слоев равны, стабилизирующий функционал можно записать в виде

$$S_{\alpha}(m) = \sum_{i=1}^k (m_i - m_{\alpha i})^2 + q \sum_{i=1}^k (m_i - m_{i-1})^2,$$

где q — некоторый априорно выбранный коэффициент, определяющий, в каком соотношении находятся условия "гладкости" и близости решения к начальному приближению m_{α} . Регуляризованное решение системы (100) при использовании такого стабилизирующего функционала будет решением следующей системы:

$$[A^T A + \alpha (I + q J)] m = A^T y + \alpha m_{\alpha},$$

где J — тридиагональная матрица вида

$$\begin{pmatrix} 1 & -1 & 0 & 0 & \dots & \dots & 0 \\ -1 & 2 & -1 & 0 & \dots & \dots & 0 \\ 0 & -1 & 2 & -1 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & -1 & 2 & -1 \\ 0 & 0 & \dots & \dots & 0 & -1 & 1 \end{pmatrix}$$

Нетрудно видеть, что к решению именно такой системы для случая $m_{\alpha} = 0$, $q_0 = 1$, $q_1 = q$ сведется (107) при условиях (109), если производные заменить конечными разностями.

§ 4. Выбор параметра регуляризации

Как уже говорилось в § 1, при построении регуляризованного решения возникает проблема выбора параметра регуляризации α . Исходя из определения регуляризованного решения как такого, которое минимизирует стабилизирующий функционал при условии

$$\|\Phi(m) - y\|^2 = \delta^2, \quad (110)$$

ясно, что параметр α должен выбираться так, чтобы соответствующее ему решение m_{α} удовлетворяло условию (110), т.е. параметр α должен быть согласован с погрешностью исходных данных. Если бы было возможно для каждой конкретной задачи определить такую функцию $\alpha = \alpha(\delta)$, для которой оператор $R_{\delta}(y; \alpha)$ является регуляризирующим, то тем самым была бы решена задача о выборе параметра α по заданному δ . Но, как правило, решение

такой задачи оказывается затруднительным. Поэтому обычно значение параметра α подбирается в процессе вычислений. Такой подбор может осуществляться разными способами в зависимости от той информации, которая имеется относительно исходных данных.

Если известно значение δ , то можно просто определять α путем подбора так, чтобы удовлетворить условию (110). Обычно это делается следующим образом. Задача о минимизации оглаживающего функционала $M^\alpha(\mathbf{y}, \mathbf{m})$ решается для некоторой последовательности значений $\alpha \{ \alpha_0, \alpha_1, \alpha_2, \dots \}$, составляющих геометрическую прогрессию. В практике чаще всего берут следующую последовательность: $10^0, 10^{-1}, 10^{-2}, \dots$. Для каждого значения α_p минимизируется сглаживающий функционал $M^\alpha(\mathbf{m}, \mathbf{y})$ и по полученному решению \mathbf{m}^α строится невязка $\| \Phi(\mathbf{m}^\alpha) - \mathbf{y} \|$. В качестве искомого значения α берется такое, для которого выполняется равенство (110) с заданной точностью.

Чтобы пользоваться этим способом, необходимо иметь достаточно точные представления о дисперсии ошибок наблюдений. Но часто в практических задачах величина δ либо неизвестна, либо оценена лишь приближенно. В таких случаях можно использовать так называемый квазиоптимальный способ выбора α . Он заключается в том, что в качестве параметра регуляризации выбирается наименьшее из значений, реализующих минимум квазиоптимальной нормы:

$$\mu(\alpha) = \left\| \alpha \frac{d\mathbf{m}^\alpha}{d\alpha} \right\|.$$

Если минимизировать $M^\alpha(\mathbf{m}, \mathbf{y})$ для последовательных значений α_p , составляющих геометрическую прогрессию, то в качестве приближенной оценки $\mu(\alpha)$ можно брать $\| \mathbf{m}^{\alpha_p} - \mathbf{m}^{\alpha_{p-1}} \|$. Оказывается, что при некотором значении α_p расстояние между последовательными решениями \mathbf{m}^{α_p} , минимизирующим $M^\alpha(\mathbf{m}, \mathbf{y})$, становится малым: решение как бы стабилизируется, а при дальнейшем уменьшении α решения \mathbf{m}^{α_p} начинают расходиться. В качестве параметра регуляризации предлагается принимать такое, при котором последовательные решения \mathbf{m}^{α_p} наименее различаются. Если при этом известно (хотя бы приближенно) значение δ , то выбор α можно одновременно контролировать вычислением на каждом шаге значения невязки $\| \Phi(\mathbf{m}^\alpha) - \mathbf{y} \|$.

Выяснилось, что для достаточно широкого круга задач эти два способа определения параметра регуляризации — квазиопти-

мальный и по величине невязки - практически эквивалентны. Это было показано вначале чисто экспериментально, на отдельных частных задачах, а затем обосновано теоретически [42].

В качестве примера подбора параметра регуляризации рассмотрим решение плохо обусловленной системы уравнений:

$$\begin{aligned} 1,11x_1 + x_2 &= 3,11, \\ x_1 + 0,8x_2 &= 2,80. \end{aligned} \quad (III)$$

Пусть "точные" правые части уравнений равны соответственно 3,10 и 2,81, так что "истинное" решение есть $x_1 = 1$, $x_2 = 2$. Правые части системы (III) содержат погрешности (0,01 и -0,01), такие, что $\delta = 0,0141$. Если решать систему (III) с возмущенными правыми частями, то решение в силу плохой обусловленности матрицы системы оказывается очень далеким от истинного ($x_1 = +20$, $x_2 = -19,1$).

Применим для решения системы (III) метод регуляризации. Возьмем стабилизатор в виде $\Omega = \|x\|^2 = x_1^2 + x_2^2$, т.е. в качестве априорного значения неизвестного вектора примем (0,0). Результаты определения вектора x при последовательных значениях α приведены ниже:

$\alpha \dots$	10^0	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
$x_1^\alpha \dots$	1,240	1,509	1,543	1,551	1,593	1,995
$x_2^\alpha \dots$	1,116	1,360	1,388	1,387	1,343	0,899
$\delta_\alpha \dots$	0,837	0,099	0,0176	0,0137	0,0095	0,0083

Из них видно, что минимум норм $\|x^{\alpha p} - x^{\alpha p-1}\|$ достигается при значении $\alpha = 10^{-3}$. И в то же время невязка $\delta_\alpha = \|Ax^\alpha - y\|$, соответствующая этому значению α , оказывается наиболее близкой к заданной.

Таким образом, значение $\alpha = 10^{-3}$ одновременно соответствует обоим критериям выбора параметра регуляризации.

Регуляризованное решение системы (III) при $\alpha = 10^{-3}$ есть $x_1 = 1,551$; $x_2 = 1,387$. Оно отличается от истинного решения, и это неудивительно, так как в данном примере погрешность исходных данных взята достаточно большой. Однако оно является значительно лучшим приближением к истинному решению, чем нерегуляризованное решение. Смещение θ и дисперсия D_{ii} регуляризованного решения, вычисленные по формулам (101) и (102), а также значение среднеквадратичной ошибки $s_i = \sqrt{\theta_i^2 + D_{ii}}$

приведены ниже:

i	$x_{\text{ист}}$	$x_{\text{рег}}$	θ	$\sqrt{D_{ii}^{\text{рег}}}$	$S_i^{\text{рег}}$	$S_i^{\text{н.к}}$
1	1	1,551	0,54	0,32	0,63	13,4
2	2	1,387	-0,60	0,36	0,70	14,9

Видно, что основной вклад в погрешность регуляризованного решения обусловлен смещением. Как следует из формулы (101), смещение можно уменьшить, если выбрать априорное значение x_0 более близким к истинному решению. Для сравнения мы привели и значения среднеквадратичных ошибок решения, получаемого по методу наименьших квадратов без регуляризации. Легко видеть, что они оказываются значительно больше, чем среднеквадратичные ошибки регуляризованного решения, несмотря на то, что регуляризованное решение характеризуется довольно значительным смещением.

§ 5. Использование метода регуляризации при решении обратных задач геофизики

В этом параграфе мы приведем заимствованные из работ [35, 41] решения некоторых модельных задач сейсмологии и МТЗ с помощью метода регуляризации.

П р и м е р 1. Определение структуры среды по дисперсии фазовой скорости основного тона волны Рэлея. Этот численный пример взят из работы [35].

Для модели, представляющей два однородных слоя на полупространстве, были рассчитаны дисперсионные кривые фазовой скорости волны Рэлея для периодов от 5 до 50 с. В эти данные была внесена погрешность, имеющая равномерное распределение, при разных значениях δ . Поскольку оператор решения прямой задачи в данном случае является нелинейным, для того чтобы свести обратную задачу к решению линейной системы, производилась линеаризация (см. § 3 гл. I), и для каждого значения α стабилизирующий функционал $\mathcal{N}^\alpha(m, y)$ минимизировался методом последовательных приближений. Стабилизирующий функционал был взят в виде $S_2(m) = \sum_i q_i (m_i - m_{i0})^2$, где m_i - искомые параметры, m_{i0} - начальное приближение для этих параметров; q_i - весо-

вне множители, зависящие от размерности соответствующего параметра. Выбор параметра регуляризации α производился по условию квазиоптимальности (см. § 4). Принятое начальное приближение довольно сильно отличалось от истинного. Определению подлежало только пять параметров: глубины границ слоев (h_1 и h_2) и скорости поперечных волн σ_i ($i = 1, 2, 3$). Значения плотностей были фиксированы, поскольку скорости поверхностных волн слабо зависят от распределения плотности.

Ниже приводятся результаты определения параметров структуры при разных значениях погрешности исходных данных.

	h_1 , км	h_2 ,	σ_1 ,	σ_2 ,	σ_3 ,	a_i	$\rho_{2,3}$ г/см ³	ρ_3 .
Точный разрез.....	18	40	3,45	3,95	4,70	$\sqrt{3}\sigma_i$	2,9	3,3
Начальное приближе- ние.....	25	54	3,29	3,70	4,565	$\sqrt{3}\sigma_i$	2,85	3,25
Решение при								
$\delta = 0 \dots$	18,01	40,01	3,45	3,95	4,700	$\sqrt{3}\sigma_i$	-	-
$\delta = 0,01 \dots$	18,05	40,05	3,45	3,97	4,69	$\sqrt{3}\sigma_i$	-	-
$\delta = 0,13 \dots$	19,10	49,60	3,45	4,32	4,57	$\sqrt{3}\sigma_i$	-	-

В случае точного задания исходных данных ($\delta = 0$) разрез восстанавливается практически точно. При $\delta = 0,01$ приближение к истинному решению также оказывается очень хорошим. А при $\delta = 0,13$ решение хотя и определяется устойчиво, но некоторые параметры довольно сильно отличаются от истинных. В то же время без регуляризации (при $\alpha = 0$) вообще не удается получить устойчивого решения — итерационный процесс нахождения последовательных приближений расходится.

П р и м е р 2. Обратная задача МТЗ. В работе [41] рассмотрен случай, когда распределение проводимости $\sigma(x)$ отыскивается в классе кусочно-гладких функций. Оператор решения прямой задачи здесь также является нелинейным, и нахождение экстремума сглаживающего функционала производится, как и в предыдущем примере, методом последовательных приближений, при этом на каждом шаге решается линеаризованная задача (линеаризация оператора задачи МТЗ была показана на примере 4 в § 3 гл. I).

При построении стабилизирующего функционала в данной задаче учитывается требование гладкости функции внутри отдельных слоев. В связи с этим стабилизатор был взят в виде

$$S_{\delta} = \sum_i \int_{z_i}^{z_{i+1}} [G'(z)]^2 dz + p \sum_i (h_i - h_{i0})^2 + r(G_{n+1} - G_{n+1}^{(0)})^2,$$

что отвечает гладкости распределения проводимости $G(z)$ внутри слоев и наименьшему уклонению границ слоев и значения проводимости в полупространстве G_{n+1} от априорно заданных.

Нахождение регуляризованного решения с таким стабилизатором сводится к решению систем интегро-дифференциальных и алгебраических уравнений (см. § 2, 3).

В качестве начального приближения принята структура с кусочно-постоянным распределением проводимости (рис. 22, б), причем достаточно далекая от истинной структуры. На рис. 22, а изображены кривые кажущегося сопротивления, рассчитанные для истинной структуры (сплошная линия) и начального приближения (штрихпунктирная). Кривая, соответствующая истинному разрезу, осложнялась случайными ошибками при разных значениях δ . На рис. 22, б приведены истинный разрез (сплошная линия), начальное приближение (штрихпунктирная) и результаты решения при разных δ . Параметр регуляризации выбирался квазиоптимальным способом. Видно, что при малых δ ($\delta < 10^{-4}$) разрез восстанавливается практически точно. При увеличении погрешности входных данных ($\delta = 10^{-3}$ и $\delta = 10^{-2}$) наблюдается некоторое отклонение от истинного разреза. Но при этом рассчитанные для этих приближений кривые кажущегося сопротивления находятся в пределах полосы погрешности исходных данных.

При $\delta = 10^{-2}$ относительная погрешность входных данных по отношению к максимальному значению ρ_k равна 0,3%. Следует заметить, что в практике работ по методу МТЗ не удается получить погрешности кажущегося сопротивления менее 5%. При такой низкой точности исходных данных метод регуляризации при неудачном выборе начального приближения может привести к решению, довольно далекому от истинного. И при этом нет никакой возможности оценить неопределенность полученного решения, так как (хотя и можно оценить дисперсию) смещение определить нельзя, поскольку неизвестно истинное решение.

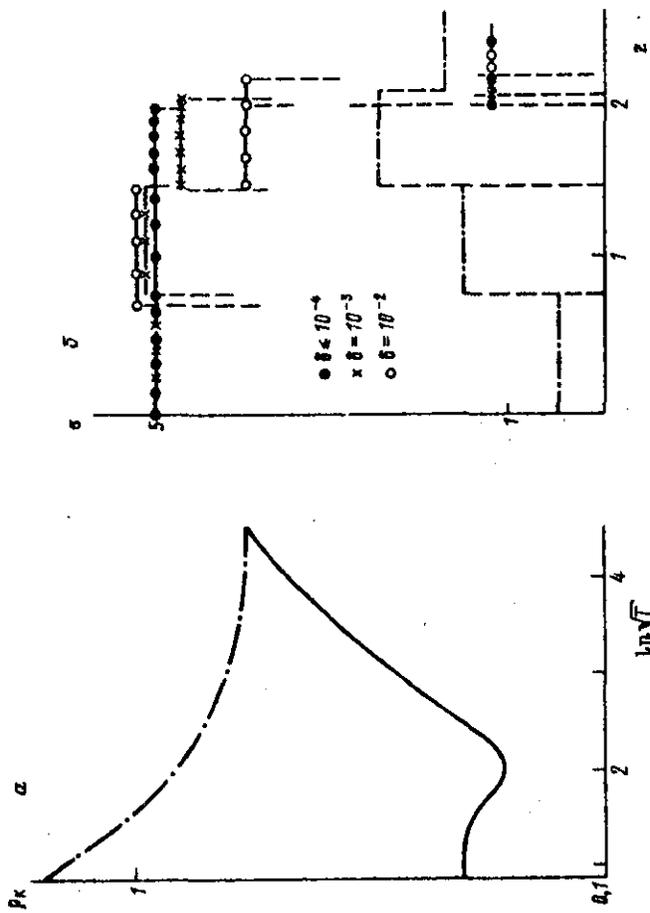


Рис. 22.

§ 6. Метод регуляризации как способ учета априорной информации

Рассмотрим статистическую трактовку метода регуляризации в применении к решению задач, в которых модель задана совокупностью конечного числа параметров, а наблюдения образуют дискретный ряд значений, содержащих погрешности. (Область применения метода регуляризации, вообще говоря, более широкая, так как с его помощью можно решать и такие задачи, в которых искомая модель представлена в виде неизвестной функции $m(r)$, наблюдения y — в виде функции некоторого параметра x , а связь между $m(r)$ и $y(x)$ определяется интегральным преобразованием.)

Как метод регуляризации, так и статистические методы, описанные в гл. II, служат для нахождения приближенного решения в случае, когда исходные данные содержат погрешности. При этом предполагается, что при точных исходных данных решение обратной задачи является единственным. Цель в том, чтобы по возмущенным данным дать наилучшую оценку этого решения. Различие же в подходах к построению оценки в статистических методах и методе регуляризации определяется тем, какие требования предъявляются к приближенному решению.

Статистическим критерием построения наилучшей оценки является требование эффективности, т. е. минимума дисперсии. Дисперсия оценки зависит от погрешностей в исходных данных, и при отсутствии этих погрешностей она должна быть равна нулю, а сама оценка должна совпадать с истинным решением. Если исходить из таких требований, то, как было показано в гл. II, задача построения оценки сводится к минимизации некоторой целевой функции $\lambda(m)$. При построении оценки методом максимального правдоподобия функция $\lambda(m)$ с точностью до слагаемого, не зависящего от m , равна с обратным знаком логарифмической функции правдоподобия.

Но в ряде случаев в силу некорректности задачи полученная таким способом оценка, даже при малых погрешностях исходных данных, может сильно отличаться от истинного решения и иметь при этом большую дисперсию. Целью метода регуляризации и является собственно получение устойчивого приближенного ре-

шения такой задачи. К оценке, получаемой по методу регуляризации, предъявляется требование, чтобы оно было как можно точнее аппроксимировать истинное решение при достаточно малой погрешности исходных данных. Как только что было сказано, оценка, получаемая статистическим методом, может не удовлетворять такому критерию.

Условие малого отличия приближенного решения от истинного при малой погрешности входных данных приводит к тому, что оценка производится на основе минимизации не просто целевой функции $\lambda(m)$, а с добавлением стабилизатора $\Omega(m)$, т.е. $\lambda(m) + \alpha \Omega(m)$. При этом стабилизатор строится на основе учета априорной информации об искомым параметрах. Параметр регуляризации α зависит от погрешности входных данных δ : $\alpha = 0$ при $\delta = 0$, так что в отсутствие погрешностей решения, получаемые обоими методами, совпадают друг с другом и с истинным решением. В случае, если минимум $\lambda(m)$ определяется устойчиво, $\alpha \rightarrow 0$ и опять же решения, полученные с учетом и без учета регуляризации, будут совпадать.

Различие между оценками, полученными статистическими методами и методом регуляризации, возникает тогда, когда задача оказывается некорректной.

В § 9 гл. II мы видели, что и в рамках статистических методов оценку можно сделать устойчивой, вводя те или иные априорные ограничения на параметры. В общем под регуляризацией и понимается наложение таких априорных ограничений на решение, при которых некорректная задача может быть переведена в класс условно корректных. Например, путем фиксации некоторых параметров (см. § 8 гл. II) можно добиться получения устойчивых оценок (с малыми дисперсией и смещением) для остальных параметров. Укрупнение параметров, нахождение определенных связей или ограничений на параметры также является способом регуляризации решения. Если же в качестве априорных ограничений использовать распределение априорной вероятности искомым параметров и проводить оценку путем максимизации апостериорной функции правдоподобия (§ 9 гл. II), то оказывается, что построение таких оценок будет производиться в полном соответствии с методом регуляризации по Тихонову.

Без ограничения общности можно принять, что априорное распределение параметров m — нормальное со средним m_0 и

матрицей ковариации $\sigma_m^2 I$. Апостериорная функция правдоподобия будет при этом иметь вид

$$L_y(m) = \log p(y; m) - \frac{(m - m_a)^T (m - m_a)}{\sigma_m^2}.$$

Максимум апостериорной функции правдоподобия эквивалентен минимуму функции

$$-\log(y; m) + \|m - m_a\|^2 / \sigma_m^2. \quad (II2)$$

Опять же без ограничения общности можно принять, что распределение исходных наблюдений нормальное с ковариационной матрицей $\sigma_y^2 I$. Тогда

$$-\log p(y; m) = (1/\sigma_y^2) (\Phi(m) - y)^T (\Phi(m) - y)$$

и минимизация функции (II2) сводится к минимизации

$$(\Phi(m) - y)^T (\Phi(m) - y) + (\sigma_y^2 / \sigma_m^2) \|m - m_a\|^2. \quad (II3)$$

Таким образом, учет априорной вероятности параметров приводит к тому, что к целевой функции $\lambda(m) = (\Phi(m) - y)^T (\Phi(m) - y)$ добавляется дополнительный член, который может рассматриваться как стабилизатор. Он и по форме совпадает с тихоновским стабилизатором, при этом отношение дисперсий σ_y^2 / σ_m^2 играет роль параметра регуляризации. Следовательно, оценка, получаемая максимизацией апостериорной функции правдоподобия, будет такой же, как и решение, построенное методом регуляризации, при условии, что $\alpha = \sigma_y^2 / \sigma_m^2$. Однако если подходить к оценке параметров из условия максимума апостериорной функции правдоподобия, то необходимо фиксировать на основании каких-то соображений значение σ_y^2 / σ_m^2 и при этом значении искать минимум функции (II3). При построении же регуляризованного решения параметр регуляризации α подбирается из условия

$$(\Phi(m) - y)^T (\Phi(m) - y) = \delta^2, \quad (II4)$$

т.е. предполагается заданная дисперсия исходных наблюдений $\sigma_y^2 = \delta^2 / n$.

Итак, различие в способах построения оценок состоит в том, какая используется информация относительно погрешности наблюдений: либо задается отношение дисперсий наблюдений и параметров σ_y^2 / σ_m^2 и при этом величина дисперсии наблюдений оценивается на основе полученной оценки параметров, либо задается дисперсия наблюдений (или связанная с ней величина δ^2)

и оценивается параметр регуляризации α из условия, что оценка должна удовлетворять уравнению (II4). Правда, при использовании метода регуляризации, как уже говорилось в § 4, есть возможность оценивать параметр регуляризации α при отсутствии информации о величине δ (квазиоптимальный способ выбора α). В этом случае, казалось бы, можно обойтись вообще без информации о погрешностях наблюдений. Но такой способ приводит к удовлетворительным оценкам (близким к решению) только при достаточно малых δ . При погрешностях, соответствующих реальным наблюдениям, регуляризованное решение хотя и есть устойчивое решение, но его точность может оказаться очень низкой.

Остановим еще внимание читателя на различии метода регуляризации и метода поиска экстремума функции отклика с помощью демпфирования (§ 7 гл. II). Дело в том, что по форме эти методы довольно похожи, и может создаться впечатление, что они эквивалентны по существу. Метод демпфирования на каждом шаге итерации использует регуляризацию, но в конечном счете он приводит не к регуляризованному решению, а к такому, которое соответствует минимуму заданной функции отклика. Это обусловлено тем, что на каждом шаге в качестве начального приближения, используемого при построении стабилизатора, берется не одно и то же значение, а решение, полученное на предыдущем шаге. Таким образом метод демпфирования представляет собой устойчивый метод поиска экстремума, который приводит к решению в случае, когда обычные итеративные методы расходятся.

§ I. Постановка задачи

Метод, разработанный Бэйкусом и Гильбертом [50-52, 55] и широко используемый в зарубежных исследованиях, предназначен для решения таких задач, в которых число исходных данных ограничено, но точность их достаточно высока; в то же время модель - пространственное изменение тех или иных физических параметров $m(r)$ - не может быть априорно представлена конечным набором параметров. Таким образом, пространство наблюдений Y является конечномерным, "наблюдения" - y - представляют в этом пространстве точку или довольно узкую область, а пространство моделей M является бесконечномерным. (Разумеется, какому-то модель - для конкретности будем ее считать одномерной функцией $m(r)$ - о любой степени приближения можно представить конечным набором значений m_i в достаточно тонких слоях $r_i < r < r_{i+1}$, но при этом размерность пространства M будет весьма велика и в общем может значительно превышать размерность пространства Y). Очевидно, что подобная задача имеет бесконечное множество решений. И для того чтобы ставить задачу о выборе единственного решения из этого множества, на класс решений следует наложить определенные ограничения.

Задача в такой постановке возникла при попытке использовать данные о периодах собственных колебаний для изучения строения Земли. Каждой модели распределения упругих постоянных в Земле - $\lambda(r)$, $\mu(r)$, $\rho(r)$ - может быть поставлена в соответствие определенная совокупность периодов собственных колебаний, число которых неограничено. Наблюдения же позволяют определить конечный и притом весьма ограниченный набор таких данных. Но поскольку эти данные несут информацию об упругих свойствах Земли, дополняющую наши представления, полу-

чаемые путем интерпретации других сейсмических наблюдений (объемных и поверхностных волн), то имеются основания для такой постановки задачи.

На каждом этапе наших знаний о строении Земли существует определенное представление о распределении тех или иных физических параметров в Земле, обобщающее все предыдущие исследования. Поэтому можно ставить задачу так: что добавляют те или иные наблюдения (пусть их даже весьма ограниченный набор) к имеющимся представлениям? Иначе говоря, какое минимальное изменение следует внести в принятый на данном уровне наших представлений разрез Земли, чтобы при этом можно было удовлетворить используемой совокупности наблюдений? При таком подходе ограничение, накладываемое на модель, заключается в том, что искомая модель должна быть как можно более близка к заданной.

Полученное при таких априорных ограничениях на разрез решение $m(r)$ будет одним из возможных. Но, кроме того, как уже говорилось, можно построить бесконечное множество разрезов, которые бы удовлетворяли заданной совокупности исходных данных. Спрашивается, можно ли охарактеризовать эти решения какими-то общими для всех них свойствами? Очевидно, что хорошо было бы найти такие интегральные характеристики разрезов, которые бы определялись по заданной совокупности наблюдений однозначно и в то же время наилучшим образом отражали бы поведение искомой функции $m(r)$. Собственно говоря, такой подход к нахождению осредненных характеристик разреза положен в основу методов, использующих предварительную параметризацию разреза (см. гл. II): фактически вместо истинного распределения $m(r)$ мы ставим задачу о нахождении осредненных характеристик этого распределения в априори выбранных слоях. При этом выбор параметров (дробление разреза на слои) производится обычно на основании качественных соображений о том, какова разрешающая способность исходных данных в том или ином интервале глубин.

Подход Бэйкуса и Гильберта тем и отличается от методов, использующих предварительную параметризацию разреза, что он дает возможность формализовать определение разрешающей способности исходных данных и нахождение таких осредненных характеристик, которые бы по заданной совокупности данных определялись бы однозначно.

Перейдем теперь к математической формулировке задачи.

Прежде всего отметим, что метод Бэйкуса — Гильберта применим только для решения линеаризованных задач, т.е. таких, в которых связь между наблюдениями y_i ($i = 1, 2, \dots, n$) и моделью $m(r)$ определяется линейным функционалом

$$y_i = \int_V G_i(r) m(r) dV. \quad (II5)$$

Интегрирование здесь производится по той области среды, которая оказывает влияние на используемые в качестве исходных данных характеристики геофизического поля y_i . Как уже говорилось в гл. I, в таком виде может быть определена в общем случае только связь между поправкой к некоторому начальному приближению $m^0(r)$ и разностью между наблюдаемыми характеристиками поля y_i и рассчитанными для этого начального приближения. В дальнейшем для простоты мы будем считать Землю сферически-симметричной и интегрирование в (II5) производить только по r от центра Земли ($r = 0$) до поверхности ($r = R$). При этом под $m(r)$ будем понимать поправку к начальному приближению, а под y_i — разность между наблюдаемой и рассчитанной для начального приближения характеристикой поля. Таким образом, вместо функционала (II5) будем рассматривать в дальнейшем следующий:

$$y_i = \int_0^R G_i(r) m(r) dr, \quad i=1, 2, \dots, n \quad (II6)$$

или в общем виде

$$y = A m, \quad (II7)$$

где A — линейный оператор. В частном случае, когда m принадлежит функциональному пространству, вид этого оператора описывается выражением (II6). Если m принадлежит конечномерному пространству M^k , то A — матрица $n \times k$, преобразующая $M^k \rightarrow Y^n$, причем $k > n$.

Поскольку система уравнений (II7) имеет бесконечное множество решений, то обратная задача ставится так:

- а) найти решение, удовлетворяющее уравнениям (II7) и некоторым ограничениям, накладываемым на искомого модель;
- б) определить такую гладкую функцию $\langle m(r) \rangle$, которая определяется однозначно, и в то же время дает по возможности наилучшее представление о функции $m(r)$.

Это по сути дела два разных подхода к решению задачи (II7). Но и тот и другой известны в литературе под общим названием метода Байкуса - Гильберта.

§ 2. Построение одного из возможных решений

Вначале рассмотрим случай, когда исходные данные y_i не содержат погрешностей. Как уже говорилось, для построения решения, удовлетворяющего системе уравнений (II7), должны быть наложены какие-то ограничения на функцию $m(r)$. Поскольку, как правило, $m(r)$ представляет поправку к некоторому начальному приближению, то естественным является требование минимизации этой поправки, т.е. наибольшей близости искомой модели к начальному приближению (в дальнейшем мы увидим, что могут быть и другие ограничения). Если расстояния в пространстве моделей определяются в квадратичной метрике, то это сводится к требованию

$$\|m\|^2 = \int_0^R m^2(r) dr = \min. \quad (II8)$$

По форме функционал (II8) совпадает со стабилизирующим функционалом (I04) при $q_0 = 1$, $q_n = 0$. Однако в данном случае не следует понимать наложение ограничения (II8) как регуляризацию. Напомним, что метод регуляризации используется в случае, когда задача при точных входных данных имеет единственное решение, а при наличии погрешностей решение становится неустойчивым. В данном случае задача при точных входных данных имеет бесконечное множество решений и наложение ограничения (II8) используется не для нахождения устойчивого приближения к истинному решению, а для выбора одного из возможных решений.

В линейной алгебре при решении неопределенных систем вводится понятие нормального решения. Если система имеет бесконечное множество решений, то нормальным решением называется такое, которое минимизирует норму искомого вектора $\|m\|$. Таким образом, решение задачи (II7) при ограничении (II8) является нормальным решением.

Задача нахождения решения, минимизирующего функционал (II8) и удовлетворяющего уравнениям (II6), является задачей вариационного исчисления и решается с помощью неопределенных множителей Лагранжа. Вместо условного минимума (II8) отыскивается абсолютный минимум функционала

$$\int_0^R \left[m^2(r) - \sum_j \frac{\lambda_j}{2} G_j(r) m(r) \right] dr.$$

Уравнения Эйлера дают

$$m(r) = \sum_j \lambda_j G_j(r), \quad (II9)$$

а коэффициенты λ_j определяются подстановкой выражения (II9) в систему (II6):

$$\sum_j \lambda_j \int_0^R G_j(r) G_i(r) dr = y_i, \quad i=1,2,\dots,n. \quad (I20)$$

Таким образом, задача сводится к решению системы линейных уравнений (I20). Матрица этой системы является симметричной, а в случае, когда все ядра $G_i(r)$ линейно независимы, то и положительно определенной. При этом система (I20) всегда имеет решение.

Используя выражение для оператора прямой задачи в форме (II7), можно записать решение, удовлетворяющее ограничению (II8) в виде

$$m = A^T (AA^T)^{-1} y. \quad (I21)$$

В рассмотренном случае, когда m принадлежит функциональному пространству, элементы матрицы AA^T определяются следующим образом:

$$(AA^T)_{ij} = \int_0^R G_i(r) G_j(r) dr. \quad (I22)$$

Если между какими-то ядрами $G_i(r)$ существует линейная зависимость, то система (I20) оказывается вырожденной и матрицы, обратной AA^T , не существует. Этот случай будет рассмотрен в гл. V. Здесь же мы будем считать, что система (I20) невырожденная.

Теперь перейдем к рассмотрению случая, когда исходные данные содержат ошибки. Вначале опишем способ решения, предложенный для этого случая Вейкусом и Гильбертом [50].

Сначала предположим, что ошибки некоррелированы и максимально возможная ошибка величины y_j равна σ_j . Тогда задача ставится так: среди всех $m(r)$, удовлетворяющих неравенствам

$$y_j - \sigma_j \leq \int_0^R G_j(r) m(r) dr \leq y_j + \sigma_j, \quad (I23)$$

найти такое, которое минимизирует $\int_0^R m^2(r) dr$.

Легко понять, что такая постановка задачи справедлива только для случая, когда распределение ошибок наблюдений можно принять равномерным, т.е. когда пространство y имеет метрику C . Эта задача решается аналогично предыдущей. Модель $m(r)$ выражается в виде линейной комбинации ядер

$$m(r) = \sum_i \lambda_i G_i(r), \quad (124)$$

а коэффициенты λ_i подбираются из линейных неравенств

$$y_i - \epsilon_i \leq \sum_j \int_0^R G_i(r) G_j(r) dr \lambda_j \leq \bar{y}_i + \epsilon_i,$$

причем так, чтобы минимизировать $\int_0^R m^2(r) dr$.

Предположим вначале, что ядра $G_i(r)$ ортонормированы, т.е.

$$\int_0^R G_i(r) G_j(r) dr = \delta_{ij}.$$

Тогда система линейных неравенств сведется к следующей:

$$y_i - \epsilon_i \leq \lambda_i \leq \bar{y}_i + \epsilon_i, \quad (125)$$

$$\int_0^R m^2(r) dr = \sum_j \lambda_j^2.$$

Требование $\sum_j \lambda_j^2 = \min$ приводит к тому, что из возможных значений λ_j , определяемых неравенствами (125), следует выбрать минимальные по абсолютной величине.

Если ядра неортонормальные (что обычно и имеет место), то производится такое преобразование, чтобы ортогонализировать ядра. Иначе говоря, выбирается новая система данных, являющаяся линейной комбинацией исходных, для которых ядра были бы ортонормированы. Это легко сделать путем приведения матрицы системы (120) к диагональной форме. При этом, однако, для новых данных ковариационная матрица ошибок может оказаться недиагональной, и для них уже не будет справедлива система неравенств (123). Таким образом, задача заключается в том, чтобы диагонализировать матрицу $B = AA^T$ ($B_{ij} = \int_0^R G_i(r) G_j(r) dr$), сохранив диагональной матрицу ковариации.

Если же матрица ковариаций была недиагональной, то нужно диагонализировать и ее. Поскольку обе матрицы симметричны и

положительно определенные, то их можно совместно привести к диагональному виду.

Теперь рассмотрим случай, более соответствующий реальным наблюдениям: будем считать, что распределение ошибки величины y_j не равномерное, а нормальное с нулевым средним и дисперсией σ_j^2 . В этом случае мы уже не имеем права считать, что решение прямой задачи, определяемое функционалом $\int_0^a G_j(r)m(r)dr$, должно удовлетворять системе неравенств (123). Здесь следует исходить из того, что ошибка $y_j - \int_0^a G_j(r)m(r)dr$ распределена нормально, и если ошибки независимы, то функция $m(r)$ должна удовлетворять условию минимизации взвешенной суммы квадратов отклонений:

$$\sum_j \frac{1}{\sigma_j^2} \left(y_j - \int_0^a G_j(r)m(r)dr \right)^2.$$

В то же время $m(r)$ должно удовлетворять и минимуму функционала $\int_0^a m^2(r)dr$. Очевидно, что решение такой задачи возможно, если заранее определить, в каком соотношении должны находиться оба минимизируемых функционала. Иначе говоря, такая задача сводится к минимизации линейной комбинации функционалов вида

$$\sum_j \frac{1}{\sigma_j^2} \left(y_j^2 - \int_0^a G_j(r)m(r)dr \right)^2 + \vartheta \int_0^a m^2(r)dr, \quad (126)$$

где значение ϑ должно быть выбрано из некоторых априорных соображений.

Рассмотрим, как можно трактовать ϑ . Если $m(r)$ есть поправка к некоторому априорно выбранному начальному приближению модели, то обычно предполагается, что это начальное приближение получено как решение обратной задачи по каким-то другим данным. Поскольку такое решение всегда определяется по данным, имеющим погрешности, то оно никогда не бывает точным, а имеет некоторую ошибку, дисперсия которой может быть известна. Строго говоря, она должна зависеть от r , но в первом приближении можно принять ее постоянной, равной σ_m^2 . Тогда на основании выводов § 9 гл. II легко видеть, что $\vartheta = 1/\sigma_m^2$ и выражение (126) по своей структуре аналогично функции отклика при оценке по методу наименьших квадратов, построенной с учетом априорной вероятности модели.

Представляя, как и ранее, $m(r)$ в виде линейной комбинации ядер (124), подставляя в выражение (126) и определяя коэффициенты из условия его минимизации, получим

$$\sum_i \lambda_i \left[B_{iq} + \sum_j \frac{\sigma_m^2}{\sigma_j^2} B_{qj} \beta_{ij} \right] = \sum_j \frac{\sigma_m^2}{\sigma_j^2} y_j \beta_{qj}, \quad (127)$$

где $\beta_{ij} = \int_0^R G_i(r) G_j(r) dr$. В случае, когда ядра ортонормированы, решением системы (127) будет

$$\lambda_q = y_q / (1 + \sigma_q^2 / \sigma_m^2).$$

Систему (127) удобнее преобразовать к виду, который в предельном случае $\sigma \rightarrow \infty$ непосредственно переходит в систему (120). Действительно, (127) можно записать в виде

$$\sum_j \frac{\sigma_m^2}{\sigma_j^2} B_{qj} \left[y_j - \sum_i \beta_{ij} \lambda_i - \lambda_j \frac{\sigma_j^2}{\sigma_m^2} \right] = 0, \quad q=1,2,\dots,n. \quad (128)$$

Поскольку матрица B по предположению не является вырожденной, то решение системы (128) совпадает с решением следующей системы:

$$\sum_i \beta_{ij} \lambda_i + \frac{\sigma_j^2}{\sigma_m^2} \lambda_j = y_j, \quad (129)$$

и при $\sigma_m \rightarrow \infty$ уравнения (129) переходят в уравнения (120).

Система (120) по исходному предположению не является вырожденной. Но матрица системы B может оказаться плохо обусловленной, и, таким образом, если правые части y содержат малые погрешности, то можно получить решение, далекое от истинного. Для нахождения устойчивого приближения к решению можно использовать метод регуляризации в той форме, в которой он был описан в § 2 гл. III. Но в данном случае удобнее использовать другой подход.

В конечном счете нам необходимо получить не вектор $\Lambda \{ \lambda_1, \lambda_2, \dots, \lambda_n \}$, являющийся решением системы (120), а решение системы (117), минимизирующее норму $\|m\|$, т.е. нормальное решение. А.Н.Тихоновым было показано [40], что регуляризация нахождения нормального решения сводится к минимизации следующего сглаживающего функционала:

$$\|y - Am\|^2 + \alpha \|m\|^2, \quad (130)$$

так что регуляризованное нормальное решение определяется из системы:

$$(A^T A + \alpha I) m = A^T y,$$

где α - параметр регуляризации.

Учтем теперь, что $m = A^T \Lambda$, тогда Λ определится из системы

$$(A^T A A^T + \alpha A^T) \Lambda = A^T y$$

или

$$(A A^T + \alpha I) \Lambda = y. \quad (131)$$

И поскольку $A A^T = B$, то система (131) совпадает с системой (129) в случае, когда все G_j равны между собой. Если обозначить $G_j = \sigma$ ($j = 1, 2, \dots, n$), то параметр регуляризации α оказывается равен отношению дисперсий s^2/G_m^2 . Этот же результат был получен при сопоставлении метода регуляризации с методом максимальной апостериорной вероятности (см. § 6 гл. II). В предельном случае $\alpha \rightarrow 0$ регуляризованное решение совпадает с тем, которое получается по методу Бэйкуса – Гильберта, когда исходные данные заданы точно.

Трудности, с которыми приходится сталкиваться при решении обратных геофизических задач методом Бэйкуса – Гильберта, связаны с неединственностью решения и необходимостью линеаризации для построения одного из возможных решений. Как мы видели, для нахождения одного из возможных решений вводится некоторое ограничение, но не на саму искомую модель $m(r)$, а на поправку $\Delta m(r)$ к некоторому произвольно выбранному начальному приближению $m_0(r)$. Поэтому следует ожидать, что получаемая в результате решения искомая функция $m(r)$ будет зависеть от начального приближения. Эта зависимость будет тем более существенна, чем меньше информации содержится в используемых данных.

Пример I. Определение скоростей упругих волн и плотности в мантии по данным о периодах собственных колебаний Земли. В качестве примера приведем результаты решения обратной задачи, полученные Бэйкусом и Гильбертом в одной из ранних работ [55], где они впервые применили описанный метод для построения упругой модели Земли, т.е. для нахождения скоростей продольных и поперечных волн $v_p(r)$ и $v_s(r)$ и плотности $\rho(r)$ по данным о периодах собственных колебаний. В качестве исходных данных были взяты периоды 17 колебаний: ${}^0S_{19}$, ${}^0S_{23}$, ${}^0S_{26}$, ${}^0S_{35}$, ${}^0S_{43}$, ${}^0S_{52}$, ${}^0S_{60}$, ${}^0S_{76}$, ${}^0S_{97}$, ${}^0T_{23}$, ${}^0T_{29}$, ${}^0T_{36}$, ${}^0T_{44}$, ${}^0T_{55}$, ${}^0T_{65}$, ${}^0T_{84}$, ${}^0T_{105}$. Кроме того, в качестве ограни-

чений на распределение плотности были добавлены данные о средней плотности Земли и ее моменте инерции.

Решение строилось из двух начальных приближений (рис.23,а), одно из которых соответствовало модели Гутенберга (пунктирные кривые), а для второго скорости в мантии были взяты постоянными, равными их средним значениям, плотность же монотонно возрастала с глубиной (сплошные кривые). Результаты решения, полученные путем последовательных приближений, изображены на рис. 23, б.

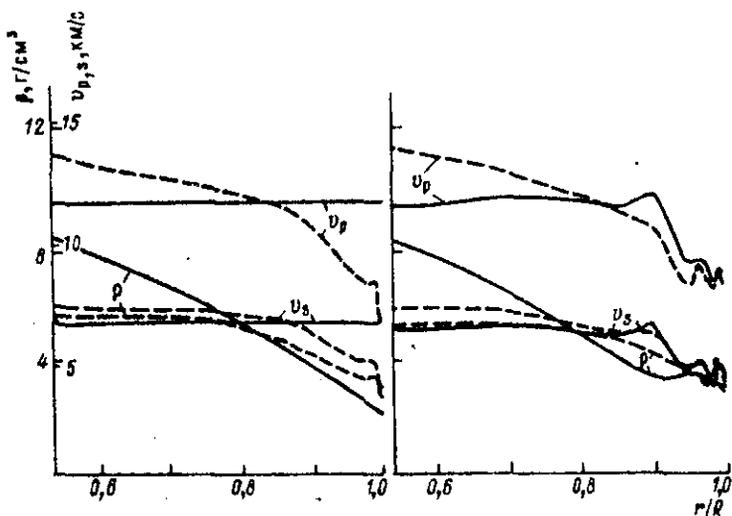


Рис. 23.

В глубинных частях мантии решения сильно различаются, причем эти различия имеют тот же характер, что и различия между начальными приближениями. Но в верхней части мантии решения имеют общие особенности, выраженные в наличии двух зон пониженной скорости. Особенно хорошее совпадение наблюдается для первой зоны, в которой минимальная скорость соответствует глубине 86 км. Здесь оба решения практически совпадают. Во второй зоне (на глубине 320–340 км) уже заметны различия в распределении скоростей, хотя в обоих решениях эта зона выявляется.

Отмеченные особенности решений обусловлены природой использованных данных. При решении этой задачи использовались данные о периодах только основных мод ($n = 0$) с достаточно высокими номерами (от 19 до 77 для сферодальных и от 23 до 105 для крутильных колебаний). Такие колебания, как известно из теории, сосредоточены в верхних слоях Земли и несут информацию соответственно о верхней части разреза.

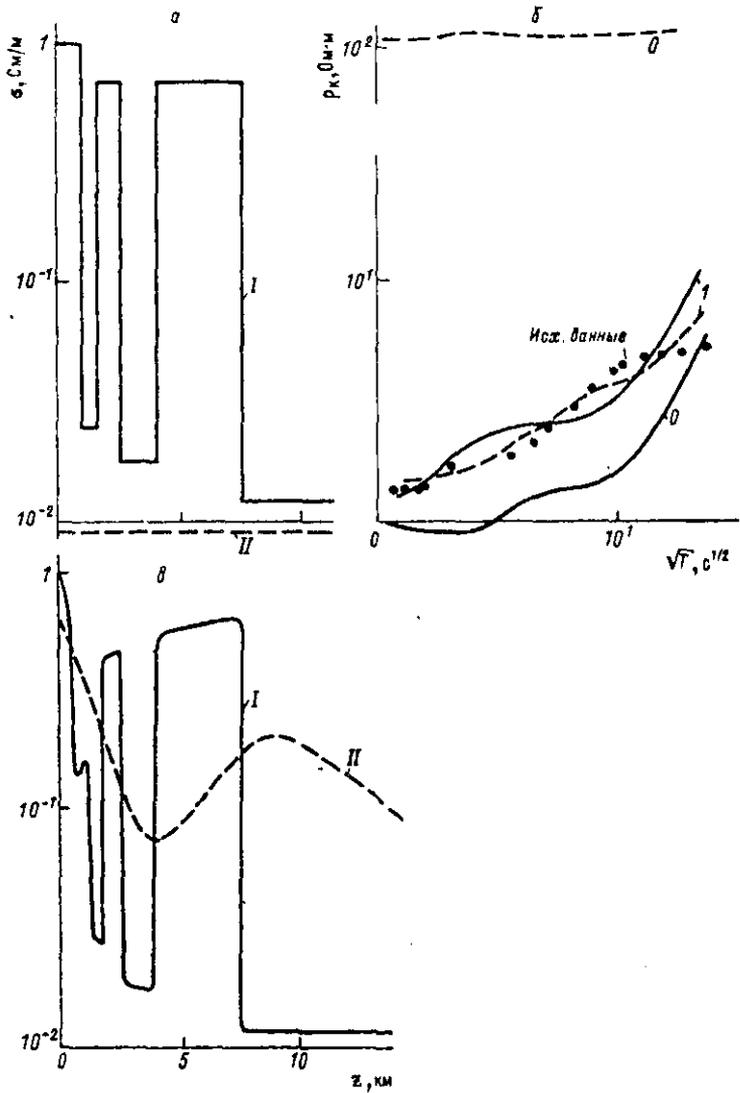
Приведенный пример показывает, что для того чтобы судить о "качестве" решения, надо иметь возможность оценивать разрешающую способность исходных данных, т.е. уметь определять, в каких областях с какой степенью детализации может быть построено решение по заданной совокупности данных. При использовании статистических методов эта задача решается путем построения информативных областей (§ 8 гл. II).

П р и м е р 2. Определение геоэлектрического разреза по данным МТЗ. В качестве исходных данных были взяты значения кажущегося сопротивления ($n = 20$) в интервале периодов $0,81 \text{ с} < T < 530 \text{ с}$, осложненные ошибками наблюдений.

Так же, как и в предыдущем примере, решение строилось из двух различных, представленных на рис. 24, а, начальных приближений. В случае I (на рис. 24 – сплошные кривые) распределение проводимости на глубинах $0 < z < 8 \text{ км}$ аппроксимировалось кусочно-постоянной функцией $\bar{h}_i = (1; 0,5; 1; 1,5; 3,5; \infty)$ км, $\sigma_i = (1; 0,023; 0,63; 0,017; 0,63; 0,012)$ См/м; в случае II (пунктирные кривые на рис. 24) проводимость считалась постоянной во всем полупространстве $z > 0$ и равной $9 \cdot 10^{-3}$ См/м. Начальные приближения были выбраны так, что соответствующие им кривые зондирования (0 на рис. 24, б) оказались весьма далеки от точек, отвечающих исходным данным. На этом же рисунке изображены кривые (1), соответствующие решениям, полученным из этих начальных приближений.

Найденные решения различаются количественно (рис. 24, б), но для них обоих характерны одни и те же качественные закономерности: понижение проводимости на глубинах 3–4 км и 8–10 км и возрастание ее на глубине 5–7 км.

Другие примеры использования метода Бэйкуса – Гильберта для оценки распределения проводимости в Земле можно найти в работах [66, 67].



§ 3. Сглаженное решение и разрешающая способность данных

Приведенные в § 2 примеры иллюстрируют неединственность решения обратной задачи для исходной модели. Но и в линейризованной постановке (II6) обратная задача имеет бесконечное множество решений. Использование ограничения на искомую функцию позволяет выбрать из этого множества одно возможное решение. А полная совокупность решений $m(r)$ характеризуется тем, что все они должны удовлетворять интегральным соотношениям (II6). Каждый из функционалов (II6) можно трактовать как результат осреднения функции $m(r)$ в интервале $(0, R)$ с весом $G_i(r)$, а значения этих результатов осреднения и определяют то общее, чем характеризуются все возможные решения задачи. Но при этом они не дадут никакого представления о характере функции $m(r)$. И хотя бессмысленно ставить задачу о поиске $m(r)$ (так как решение этой задачи неединственно), можно искать другие способы осреднения $m(r)$, которые можно было бы получить по исходным данным однозначно и результаты которых давали бы наилучшее представление о поведении функции $m(r)$, т.е. отвечали бы усреднению $m(r)$ в возможно более узкой окрестности заданной точки r_0 . При условии, что $m(r)$ — достаточно гладкая функция, результат такого сглаживания будет более или менее близок к истинному значению m в точке r_0 . А ширина интервала сглаживания, точнее такого, внутри которого сглаживающая функция значительно больше, чем вне его, будет характеризовать разрешающую способность данных в окрестности точки r_0 .

В рамках метода Бейкуса — Гильберта рассматривается только линейное осреднение функции $m(r)$:

$$\langle m \rangle_{r_0} = \int_0^R \mathcal{A}(r, r_0) m(r) dr, \quad (132)$$

где $\mathcal{A}(r, r_0)$ — сглаживающее ядро, которое предполагается нормированным, т.е. удовлетворяющим условию $\int_0^R \mathcal{A}(r, r_0) dr = 1$. Очевидно, что если $\mathcal{A}(r, r_0) = \delta(r - r_0)$ ($\delta(x)$ — функция Дирака), то сглаженное решение $\langle m \rangle_{r_0}$ точно совпадает с $m(r_0)$. Поэтому для того чтобы $\langle m \rangle_{r_0}$ как можно лучше отражало значение $m(r_0)$, необходимо, чтобы сглаживающее ядро было как можно более близко к δ -функции.

Сглаженное решение определяется через исходные функционалы (116). При этом, поскольку и функционалы (116) и сглаженное решение $\langle m \rangle_{r_0}$ линейно выражаются через $m(r)$, и $\langle m \rangle_{r_0}$ может быть представлено в виде линейной комбинации этих функционалов. Соответственно значения сглаженной функции $\langle m \rangle_{r_0}$, отвечающие заданным наблюдениям y_j будут линейной комбинацией данных. Таким образом,

$$\langle m \rangle_{r_0} = \sum_j a_j(r_0) y_j = \sum_j a_j \int_0^R G_j(r) m(r) dr$$

и сглаживающее ядро выражается в виде линейной комбинации функционалов (116):

$$\mathcal{A}(r, r_0) = \sum_j a_j(r_0) G_j(r). \quad (133)$$

Коэффициенты этой линейной комбинации должны определяться из условия наилучшей "сосредоточенности" функции $\mathcal{A}(r, r_0)$ около точки r_0 , т.е. максимальной близости $\mathcal{A}(r, r_0)$ к δ -функции.

Критерии "дельтообразности" сглаживающего ядра могут быть разными. Одним из подходов к построению функции $\mathcal{A}(r, r_0)$ характеризующейся большими значениями при r , близких к точке r_0 , и быстро убывающей по абсолютной величине при удалении от нее, является минимизация функционала

$$\int_0^R J(r, r_0) \mathcal{A}^2(r) dr, \quad (134)$$

в котором функция $J(r, r_0)$ обращается в нуль в точке $r = r_0$ и монотонно возрастает с возрастанием $|r - r_0|$. Примерами таких функций могут быть

$$\begin{aligned} J(r, r_0) &= (r - r_0)^2, \\ J(r, r_0) &= 2\sigma^2 \{1 - \exp[-(r - r_0)^2 / (2\sigma^2)]\}. \end{aligned} \quad (135)$$

Очевидно, что из нескольких возможных функций $J(r, r_0)$ следует выбрать такую, для которой значение функционала (134) для любой функции, определяемой выражением (133) и удовлетворяющей условию нормировки, будет минимальным. Обычно для простоты вычислений выбирают функцию (135).

Возможны и другие подходы к выбору критерия дельтообразности сглаживающей функции, некоторые из которых будут рассмотрены далее.

Решения $\langle m \rangle_{r_0}$, удовлетворяющее выбранному критерию дельтаобразности сглаживаемого ядра, можно назвать локально усредненным.

Поскольку построение функции $\mathcal{A}(r, r_0)$ сводится к определению коэффициентов a_i в выражении (133), то использование критерия дельтаобразности (134) для функции $\mathcal{A}(r, r_0)$ приводит к задаче на нахождение условного экстремума в пространстве коэффициентов a_i . Обозначим функционал (134) через $s(r_0)$. Учитывая выражение (133) для $\mathcal{A}(r, r_0)$, назовем $s(r_0)$ как функцию неизвестных коэффициентов a_i :

$$s(r_0) = \sum_i \sum_j a_i a_j \int_0^R J(r, r_0) G_i(r) G_j(r) dr. \quad (136)$$

Условие нормировки $\mathcal{A}(r, r_0)$ приводит к уравнению

$$\sum_j a_j(r_0) \int_0^R G_j(r) dr = 1. \quad (137)$$

Таким образом, задача сводится к отысканию экстремума (136) при условии (137).

Перейдем к матричным обозначениям. Введем матрицу \mathbf{S} , элементы которой

$$S_{ij} = \int_0^R J(r, r_0) G_i(r) G_j(r) dr. \quad (138)$$

Неизвестные коэффициенты a_i будем считать компонентами вектора \mathbf{a} , а величины $g_i = \int_0^R G_i(r) dr$ — компонентами вектора \mathbf{g} . Таким образом, следует минимизировать $s = \mathbf{a}^T \mathbf{S} \mathbf{a}$ при условии $\mathbf{g}^T \mathbf{a} = 1$.

Введение неопределенного множителя Лагранжа приводит к задаче на нахождение абсолютного экстремума функции $\mathbf{a}^T \mathbf{S} \mathbf{a} + \lambda (\mathbf{g}^T \mathbf{a} - 1)$, что эквивалентно решению системы уравнений

$$\begin{aligned} 2\mathbf{S}\mathbf{a} + \lambda\mathbf{g} &= 0, \\ \mathbf{g}^T \mathbf{a} &= 1. \end{aligned}$$

Выразим \mathbf{a} из первого уравнения:

$$\mathbf{a} = -(\lambda/2)\mathbf{S}^{-1}\mathbf{g} \quad (139)$$

и, подставив это значение во второе уравнение системы, получим выражение для множителя Лагранжа: $-\lambda/2 = (\mathbf{g}^T \mathbf{S}^{-1} \mathbf{g})^{-1}$. Подставляя его в формулу (139), получаем решение для искомого вектора коэффициентов:

$$a = S^{-1} g (g^T S^{-1} g)^{-1}. \quad (140)$$

Величина $s(r_0)$ определяет степень близости сглаживающего ядра к δ -функции.

Рассмотрим, как можно интерпретировать эту величину. Возьмем в качестве $J(r, r_0)$ функцию $12(r-r_0)^2$. И если $A(r, r_0)$ в интервале $(r_0-l/2, r_0+l/2)$ постоянна и равна $1/l$, то

$$s(r_0) = 12 \int_{r_0-l/2}^{r_0+l/2} (r-r_0)^2 \frac{1}{l} dr = l.$$

Таким образом, в этом случае величина $s(r_0)$ точно равна ширине интервала сглаживания. В случае, когда функция $A(r, r_0)$ имеет другой вид, $s(r)$ можно трактовать как эффективную ширину интервала допустимого разрешения. Если детали искомой функции $m(r)$ имеют масштаб, превышающий величину s , то такие детали могут быть разрешены при использовании рассматриваемой выборки данных. В противном случае выявить их по заданной выборке наблюдений принципиально невозможно.

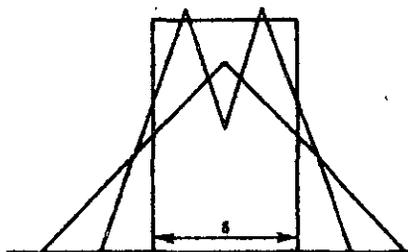


Рис.25.

Величина s характеризует разрешающую способность данных, хотя и не дает представления о характере сглаживания. На рис.25 изображены разные функции $A(r, r_0)$, которым соответствует одно и то же значение s . Поэтому для полного суждения о характере локального усреднения ре-

шения необходимо знать не только соответствующую величину s , но и вид функции $A(r, r_0)$.

Пример I. Проиллюстрируем вид сглаживающих ядер $A(r, r_0)$, получаемых по разным выборкам данных. Пример взят из работы [61], где исследовалась разрешающая способность разных наборов данных о периодах собственных колебаний Земли для оценки плотности. Как уже говорилось, периоды собственных колебаний определяют распределение скоростей упругих волн и

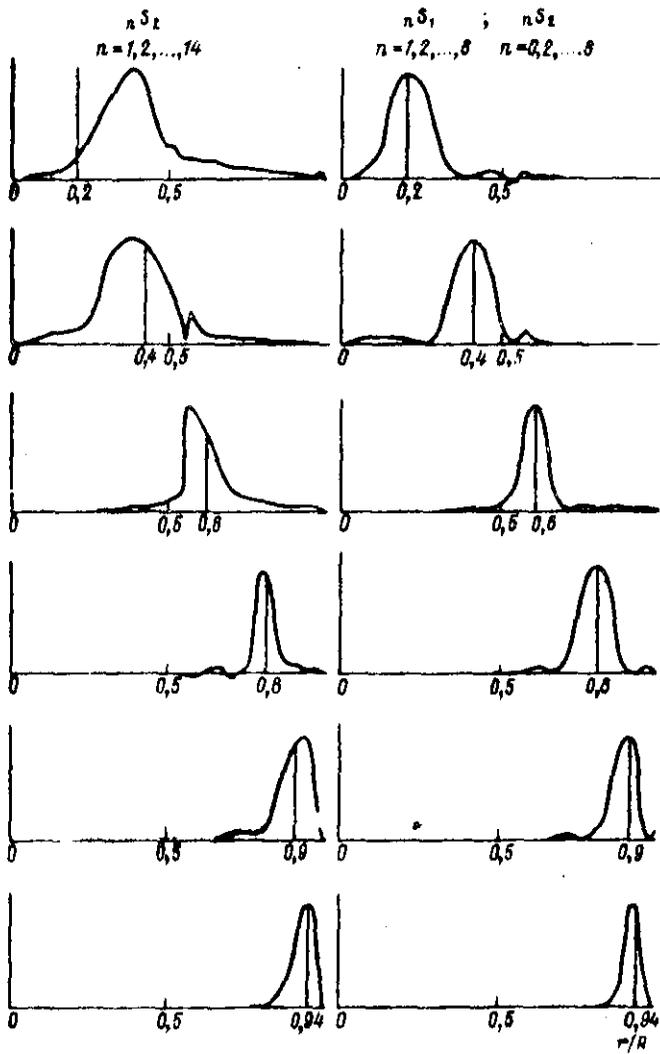


Рис. 26.

плотности в Земле. Если распределение скоростей фиксировать, то обратная задача будет заключаться в определении плотности $\rho(r)$. Совокупности периодов разных собственных колебаний будут характеризоваться разной разрешающей способностью. Это видно из результатов по двум выборкам данных о периодах сферических колебаний. Первая выборка содержала периоды колебаний ${}_n S_2$ при $n = 0, 1, \dots, 14$, вторая - периоды ${}_n S_1$ при $n = 1, 2, \dots, 8$ и ${}_n S_2$ при $n = 0, 1, 2, \dots, 8$. Кроме того, к каждой из выборок были добавлены данные о средней плотности Земли и моменте ее инерции, поскольку они дают ограничения на распределение плотности. На рис. 26 изображены сглаживающие ядра $\mathcal{A}(r, r_0)$ для разных значений r_0 (вертикальные линии). Несмотря на то, что выборки характеризуются приблизительно одинаковым объемом (в первой 17 данных, во второй - 19), видно, что во втором случае разрешающая способность данных в целом значительно выше, чем в первом, особенно в глубинных областях Земли. Такое различие обусловлено природой данных, содержащихся в этих двух выборках.

Выбор того или иного критерия дельтообразности сглаживающего ядра определяется видом ядер $G_i(r)$, т.е. характером используемых данных. В случае, когда $G_i(r)$ - ступенчатые функции, критерий минимума (134) не наилучший. Это иллюстрируется приводимым ниже примером.

Пример 2. Определение распределения скорости по данным о временах пробега волн, отраженных от горизонтальных границ при нормальном падении. Пусть глубины границ z_i известны, и задача состоит в оценке скоростного разреза $v(z)$. Наблюдения можно определить следующими функционалами:

$$t_i = 2 \int_0^{z_i} n(z) dz,$$

где $n(z) = [v(z)]^{-1}$, $i = 1, 2, \dots, n$. Из элементарных физических соображений ясно, что по таким данным можно оценить средние скорости в каждом из слоев: $v_i^{\text{cp}} = \frac{z_i - z_{i-1}}{\int_{z_{i-1}}^{z_i} dz / v(z)}$. Покажем, что этот результат получается и при формальном применении метода Байкуса - Гильберта, если в качестве критерия дельтообразности взята ширина $\delta(z_0)$ интервала, где $\mathcal{A}(z, z_0)$ отлична от нуля. Поскольку

$$A(z, z_0) = \sum_i a_i(z_0) G_i(z),$$

а

$$G_i(z) = 2(1 - H(z - z_i)),$$

где $H(x)$ - функция Хевисайда, то очевидно, что величина $s(z_0)$ будет минимальной, если в качестве нее взять толщину слоя, где находится рассматриваемая точка z_0 : $s(z_0) = z_i - z_{i-1}$, если $z_{i-1} < z_0 < z_i$. Коэффициенты a_i в этом случае будут следующими:

$$a_{i-1} = -\frac{1}{2(z_i - z_{i-1})}, \quad a_i = \frac{1}{2(z_i - z_{i-1})}, \quad a_k = 0, \quad k \neq i-1, i.$$

При этом

$$\langle n \rangle_{z_0} = \int_{z_{i-1}}^{z_i} \frac{n(z)}{z_i - z_{i-1}} dz = \frac{l_i - l_{i-1}}{2(z_i - z_{i-1})} = \frac{1}{v_{\text{ср}}}.$$

А если бы при решении этой задачи определять a_i из условия минимизации (134), то сглаживающая функция была бы отлична от нуля не только в том слое, где находится рассматриваемая точка, но и во всех остальных слоях. На рис. 27 для случая трех

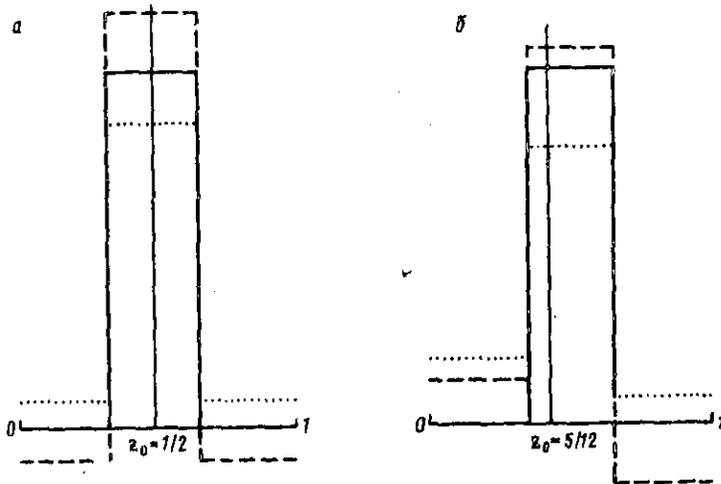


Рис. 27.

равных по толщине слоев $h_1 = h_2 = h_3 = 1/3$ изображены сглаживающие ядра, отвечающие разным критериям дельтаобразности. Сплошной линией изображена функция $\mathcal{A}(x, x_0)$, соответствующая минимуму ширины интервала, где она отлична от нуля, а точечной линией - минимуму (134) при $J(x, x_0) = (x - x_0)^2$. Даже когда рассматриваемая точка x_0 находится в центре слоя (рис. 27, а), сглаживание, удовлетворяющее критерию дельтаобразности $\int (x - x_0)^2 \mathcal{A}^2 dx = \min$, проводится по более широкой области. Если же точка x_0 сдвигается к границе слоя (рис. 27, б) в сглаженном решении возрастает вес соседнего слоя.

Упомянем здесь еще об одном критерии дельтаобразности, который будет более подробно рассмотрен в § 6. Если функция $\mathcal{A}(r, r_0)$ приближается к δ -функции, то очевидно, что интеграл от нее $\Phi(r, r_0) = \int_0^r \mathcal{A}(r', r_0) dr'$ должен приближаться к функции Хэвисайда $H(r - r_0)$. А близость $\Phi(r, r_0)$ к $H(r - r_0)$ можно определить величиной функционала

$$s(r_0) = 12 \int_0^R [\Phi(r, r_0) - H(r - r_0)]^2 dr. \quad (141)$$

Таким образом, можно определять коэффициенты a_i из условия минимизации $s(r_0)$, определяемого выражением (141). На рис. 27 пунктиром изображено сглаживающее ядро, отвечающее минимуму (141) для того же самого примера. Из этого рисунка видно отличие такого критерия от рассмотренных ранее.

Нетрудно показать, что если $\mathcal{A}(r, r_0)$ является П-образной функцией, отличной от нуля в интервале $(r_0 - 1/2, r_0 + 1/2)$, то $s(r_0) = 1$.

§ 4. Разрешающая способность данных, содержащих погрешности

В § 3 сглаженное решение $\langle m \rangle_{r_0}$ было определено как линейная комбинация данных

$$\langle m \rangle_{r_0} = \sum_i a_i(r_0) y_i, \quad (142)$$

и для того, чтобы $\langle m \rangle_{r_0}$ являлось наилучшим локальным осреднением функции $m(r)$ в окрестности точки r_0 , коэффициенты a_i рассчитывались на основании того или иного критерия дельтаобразности сглаживающего ядра. Очевидно, что если данные y_i не точные, а содержат ошибки, то и решение $\langle m \rangle_{r_0}$

тоже будет определено с ошибкой. Дисперсия этой ошибки выражается через ковариации R_{ij} ошибок исходных наблюдений:

$$\rho^2 = D(\langle m \rangle_{r_0}) = \sum_{i,j} a_i a_j R_{ij}$$

или, если ковариационную матрицу ошибок наблюдений обозначить через R , то $\rho^2 = \mathbf{a}^T R \mathbf{a}$.

Чтобы решение (142) наиболее точно отражало сглаженную тем или иным способом модель, необходимо, чтобы ее дисперсия была как можно меньше. Таким образом, при определении коэффициентов a_i следует исходить не только из условия наилучшего разрешения, но и из требования, чтобы точность получаемого решения была достаточно высокой.

Можно легко показать, что минимизировать одновременно s и ρ^2 нельзя, если только матрицы R и S не окажутся случайно подобными: при уменьшении s величина ρ^2 будет расти, и наоборот, уменьшение ρ^2 будет приводить к увеличению s . Это можно понять из рассмотрения геометрической интерпретации решения.

Решение, удовлетворяющее наилучшему разрешению, т.е. минимуму $s = \mathbf{a}^T S \mathbf{a}$ при условии $\mathbf{g}^T \mathbf{a} = 1$, можно представить как точку касания эллипсоида $\mathbf{a}^T S \mathbf{a} = s_{\min}$ в пространстве параметров a_i с гиперплоскостью \mathcal{X} , определяемой уравнением $\mathbf{a}^T \mathbf{g} = 1$. На рис. 28 это проиллюстрировано для случая трех-

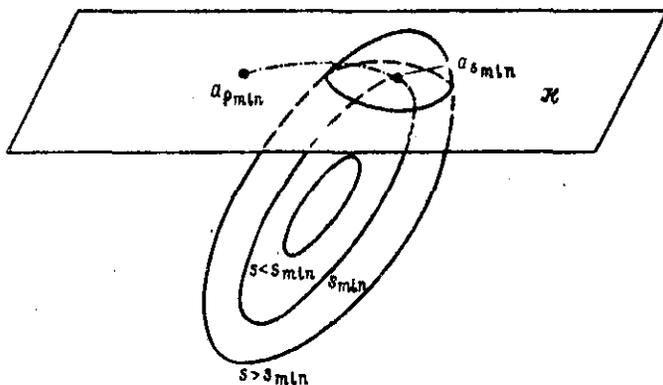


Рис. 28.

мерного пространства ($n = 3$). Аналогично, решение, обладающее наибольшей точностью, будет представлять точку касания другого эллипсоида, определяемого уравнением $\rho^2 = \mathbf{a}^T \mathbf{R} \mathbf{a}$, с той же гиперплоскостью. Поскольку матрицы \mathbf{S} и \mathbf{R} являются положительно определенными, то оба эллипсоида будут строго выпуклыми.

Разным значениям s будут соответствовать подобные эллипсоиды, при этом значению $s < s_{\min}$ будет отвечать такой, который не имеет общих точек с гиперплоскостью \mathcal{X} , а значению $s > s_{\min}$ - такой, который пересекается с ней по некоторому эллипсу. Если решению, соответствующему наибольшей точности ($\rho^2 = \rho_{\min}^2$), отвечает некоторая точка $\mathbf{a}_{\rho_{\min}}$ на гиперплоскости \mathcal{X} , то очевидно, что при передвижении вдоль \mathcal{X} от точки $\mathbf{a}_{s_{\min}}$ к точке $\mathbf{a}_{\rho_{\min}}$ величина s будет монотонно возрастать. То же самое рассуждение можно провести и по отношению к величине ρ^2 , которая будет возрастать при перемещении от точки $\mathbf{a}_{\rho_{\min}}$ к $\mathbf{a}_{s_{\min}}$.

Поэтому в данном случае задачу можно ставить одним из способов: либо задаться величиной разрешения, т.е. фиксировать s и при этом условии минимизировать дисперсию, либо фиксировать дисперсию и минимизировать s .

Пусть задано, например, значение s . Тогда задача сведется к нахождению минимума функции

$$\mathbf{a}^T \mathbf{R} \mathbf{a} + \alpha (\mathbf{a}^T \mathbf{S} \mathbf{a} - s) + \lambda (\mathbf{a}^T \mathbf{g} - 1), \quad (143)$$

где α и λ - неопределенные множители Лагранжа. Дифференцируя (143) по неизвестным параметрам, приходим к системе уравнений

$$\mathbf{R} \mathbf{a} + \alpha \mathbf{S} \mathbf{a} + \frac{\lambda}{2} \mathbf{g} = 0,$$

$$\mathbf{a}^T \mathbf{S} \mathbf{a} = s, \quad (144)$$

$$\mathbf{a}^T \mathbf{g} = 1.$$

Решив эту систему, мы бы нашли $\hat{\alpha} = \alpha(s)$, $\hat{\lambda} = \lambda(s)$, $\hat{\mathbf{a}} = \mathbf{a}(s)$. К сожалению, эта система нелинейная, поэтому ее решение сопряжено с вычислительными трудностями.

Свести поставленную задачу к решению системы линейных уравнений можно, если фиксировать не s , а α [12]. Для удоб-

ства вводится следующая замена: $\alpha = \omega^{-1} \lg \vartheta$ (множитель ω^{-1} выбирается для сохранения размерности), так что задаваться следует значением ϑ и искать s , λ , а как функции ϑ . Система (I44) в этих переменных будет иметь вид

$$\begin{aligned}
 (S \sin \vartheta + \omega R \cos \vartheta) \alpha &= -\frac{\lambda(\vartheta) \omega}{2} g \cos \vartheta, \\
 a^T g &= 1, \\
 s &= a^T S a.
 \end{aligned}
 \tag{I45}$$

Введем обозначения: $W = \omega R \cos \vartheta + S \sin \vartheta$, $\beta(\vartheta) = (\omega \lambda(\vartheta)/2) \cos \vartheta$.

Тогда $W a = -\beta g$, $a^T g = 1$,

откуда $a = W^{-1} g (g^T W^{-1} g)^{-1}$, $\tag{I46}$

$$s = a^T S a = \frac{g^T W^{-1} S W^{-1} g}{(g^T W^{-1} g)^2}, \tag{I47}$$

$$\rho^2 = a^T R a = \frac{g^T W^{-1} R W^{-1} g}{(g^T W^{-1} g)^2}. \tag{I48}$$

Рассмотрим теперь, каков смысл ϑ и как следует выбирать значение этой величины.

Прежде всего заметим, что поскольку множитель ω выбирается так, чтобы привести к одной размерности S и R , то величины $\omega \rho^2$ и s также окажутся одной размерности. Рассмотрим кривую зависимости $\omega \rho^2$ от s , имея в виду, что ρ^2 и s определяются по формулам (I47), (I48), т.е. соответствуют решению системы (I45) (рис. 29). Каждой точке этой кривой будет соответствовать определенное значение ϑ , причем легко видеть, что

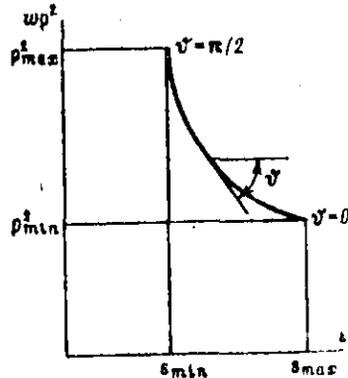


Рис. 29.

точке, соответствующей φ_{\min}^2 (и соответственно s_{\max}), будет отвечать $\vartheta = 0$, а точке, соответствующей φ_{\max}^2 и s_{\min} , — $\vartheta = \pi/2$. Можно показать (мы здесь не приводим этот вывод), что $\operatorname{tg} \vartheta = -\frac{d(\omega\varphi^2)}{ds}$, т.е. ϑ есть угол, составляемый касательной к кривой $\omega\varphi^2(s)$ с горизонталью.

Поскольку в реальных геофизических задачах всегда желательно и повысить точность, и увеличить разрешение, то ясно, что следует выбирать значение ϑ так, чтобы соответствующая ему точка кривой $\omega\varphi^2(s)$ была наиболее близка к точке $(s_{\min}, \varphi_{\min}^2)$ (рис.29). Если размерный множитель ω выбрать так, чтобы кривая $\omega\varphi^2(s)$ была по возможности симметрична относительно биссектрисы прямого угла, то такой точке будет соответствовать значение ϑ , близкое к $\pi/4$.

Сравним решение (146) для коэффициентов α_i с решением (140). В случае, если ошибки не учитываются, $\vartheta = \pi/2$ и выражение (146) переходит в (140). Матрица \mathbf{S} , как и рассмотренная в § 2 матрица \mathbf{B} , может оказаться плохо обусловленной, а это приведет к тому, что обращение такой матрицы будет неустойчивым, и коэффициенты α_i , вычисляемые по формуле (140), будут содержать большие погрешности. Учет ошибки наблюдений и соответственно использование вместо \mathbf{S} матрицы \mathbf{W} приводит к регуляризации решения. Если ошибки наблюдений некоррелированы и имеют одну и ту же дисперсию σ^2 , то $\mathbf{R} = \sigma^2 \mathbf{I}$ и вычисление по формуле (146) эквивалентно вычислению по формуле (140), но при этом вместо обращения плохо обусловленной матрицы \mathbf{S} будет производиться обращение матрицы, полученной из \mathbf{S} добавлением некоторого положительного параметра к ее диагональным элементам. Такая матрица будет уже хорошо обусловлена.

§ 5. Нахождение одного из возможных решений
в случае, когда ядра не являются квадратично интегрируемыми

Все выводы предыдущих параграфов справедливы только в случае, когда ядра G_i квадратично интегрируемы. При определении одного из возможных решений матрица систем: (120) имеет диагональные элементы $\int_0^{\infty} [G_i(r)]^2 dr$; при построении сплаженного решения в рамках критерия дальтообразности, заключающегося в минимизации (134), также используется матрица, диаго-

нальные элементы которой $\int_0^R J(r, r_0) [G_i(r)]^2 dr$. Очевидно, что если ядра $G_i(r)$ таковы, что указанные интегралы не существуют, то весь описанный в предыдущих параграфах аппарат построения решения становится не применимым. В таком случае приходится встречаться при решении обратных кинематических задач лучевой сейсмологии, т.е. при определении скоростного разреза по голографу. В § 3 гл. I было показано, что связь поправки к голографу δt_i с поправкой к скоростному разрезу $\delta v(r)$ определяется следующим функционалом:

$$\delta t_i = -2 \int_{r_m^i}^R \frac{\delta(r) dr}{v_0^2(r) \sqrt{1 - p_i^2 v_0^2 / r^2}}, \quad (149)$$

где r_m^i — значение r в точке минимума i -го луча. В этой точке ядро $G_i(r)$ имеет интегрируемую особенность, но интеграл от квадрата $G_i(r)$ расходится.

Для нахождения решения, удовлетворяющего условию минимизации $\int_0^R [\delta v]^2 dr$, нельзя непосредственно применить метод, описанный в § 2, но если рассматривать ядро функционала (149) как предел функции

$$\tilde{G}_i(r) = \begin{cases} -2(v_0^2 \sqrt{1 - p_i^2 v_0^2 / r^2})^{-1}, & \tilde{r} \leq r \leq R, \\ 0, & r < \tilde{r}; \tilde{r} > r_m^i, \end{cases}$$

при стремлении \tilde{r} к r_m^i , то формально такое решение можно построить. При $\tilde{r} \rightarrow r_m^i$ это решение $\delta v(r)$ во всех точках, кроме $r = r_m^i$, будет равно нулю, а в самих точках $r = r_m^i$ будет иметь интегрируемую особенность такого типа, что $\int_0^R [\delta v]^2 dr = 0$, а интегралы (149) конечны. Для случая $v_0 = \text{const}$ нетрудно показать, что особенность в точках $r = r_m^i$ при стремлении r к r_m^i справа будет типа

$$[\sqrt{r^2 - (r_m^i)^2} \ln(r - r_m^i)]^{-1}. \quad (150)$$

Такое решение действительно будет удовлетворять условию минимизации функционала $\int_0^R (\delta v)^2 dr$, и в то же время соответствующим выбором коэффициентов при (150) можно добиться того, чтобы выполнялись соотношения (149). Однако это решение не является приемлемым с физической точки зрения, а значит, критерий минимизации $\int m^2(r) dr$, использованный для построения

этого решения в случае, когда ядра не являются квадратично интегрируемыми, не годится. Очевидно, что в данном случае необходимо потребовать, чтобы решение было непрерывным и более или менее плавно изменилось в исследуемом интервале. Оказывается, что для построения такого решения после некоторого преобразования условий (116) может быть использован формальный аппарат, описанный в § 2 [60].

Рассмотрим построение одного из возможных решений в случае, когда данные точные, так что искомая функция определяется из условий (116).^{*} Проинтегрируем (116) по частям:

$$\mathcal{X}_i(R) m(R) - \int_0^R \mathcal{X}_i(r) m'(r) dr = y_i, \quad i=1,2,\dots,n, \quad (151)$$

$$\text{где } \mathcal{X}_i(r) = \int_0^r G_i(r) dr.$$

Если бы $m(r)$ была разрывна в интервале $(0, R)$, то в левой части выражения (151) добавились бы дополнительные члены, содержащие значения $m(r)$ в точках разрыва, но если исходить из условия непрерывности функции, то эти члены будут отсутствовать. Непрерывность решения $m(r)$ можно обеспечить, минимизируя $\int_0^R [m'(r)]^2 dr$. Но и такое решение должно быть в то же время наиболее близким к исходному (в данной постановке — к $m(r)$). Поэтому должно быть наложено еще некоторое условие на значение m на конце интервала, т.е. в точке $r=R$. Значение $m(R)$ и производная $m'(r)$ однозначно определяют искомую модель $m(r)$. Условия, наложенные на $m(R)$, могут быть различными. Например, можно принять $m(R) = 0$, если точно известно значение начального приближения в точке $r=R$. Можно минимизировать $m^2(R)$ одновременно с $\int_0^R [m'(r)]^2 dr$, т.е. определять $m(R)$ и $m'(r)$ из условия минимизации функционала

$$\int_0^R [m'(r)]^2 dr + \alpha m^2(R). \quad (152)$$

Здесь α — размерный коэффициент, который выбирается из условий конкретной геофизической задачи.

^{*} Обобщение описанной далее методики на случай, когда данные имеют погрешность, производится точно так же, как это делается для квадратично интегрируемых ядер (см. § 2).

Поскольку y_i линейно зависит от $m(R)$ и $m'(r)$, то метод, описанный в § 2, может быть полностью применен в данном случае.

Определение минимума (15) при условиях (151) сводится к решению следующей системы линейных уравнений:

$$\begin{aligned} m'(r) &= \sum_i \lambda_i \mathcal{K}_i(r), \\ \alpha m(R) &= - \sum_i \lambda_i \mathcal{K}_i(R), \\ \sum_j \lambda_j \left[- \frac{\mathcal{K}_i(R) \mathcal{K}_j(R)}{\alpha} - \int_0^R \mathcal{K}_i(r) \mathcal{K}_j(r) dr \right] &= y_i, \quad i=1, 2, \dots, n. \end{aligned}$$

Пример I. Проиллюстрируем смысл условий, накладываемых на искомую функцию в случае, когда ядра не являются квадратично интегрируемыми. (Этот пример не отвечает геофизической задаче, а относится к области аппроксимации функции, заданной значениями в дискретных точках.)

Пусть значения искомой функции $m(x)$ заданы в дискретных точках: $m(x_i) = m_i$ ($i = 0, 1, \dots, n$). Задача состоит в том, чтобы на интервале (x_0, x_n) определить функцию $m(x)$, проходящую через заданные точки.

Формально значения функции $m(x)$ в точках x_i можно представить в виде линейных функционалов типа (116), в которых ядра будут δ -функциями:

$$\begin{aligned} \int_{x_0}^{x_n} \delta(x - x_i) m(x) dx &= m_i, \quad i=1, 2, \dots, n-1, \\ 2 \int_{x_0}^{x_n} \delta(x) m(x) dx &= m_0. \end{aligned} \quad (153)$$

Если оценивать функцию $m(x)$ из условия минимизации $\int m^2(x) dx$, то получим функцию, равную нулю везде, кроме точек $x = x_i$, а в самих точках x_i равную ее заданным значениям m_i .

Рассмотрим теперь решение, которое получается из условия минимизации $\int [m'(x)]^2 dx$. Проинтегрируем по частям уравнение (153):

$$m_i = m_n - \int_{x_0}^{x_n} H(x - x_0) m'(x) dx, \quad i=0, 1, \dots, n-1.$$

Здесь $H(x)$ — функция Хевисайда. Поскольку в данном случае значение $m(x_n) = m_n$ предполагается заданным, то в минимизируемом функционале (152) следует положить $\alpha = 0$. Решением этой задачи будет

$$m'(x) = \sum_{i=0}^{n-1} \lambda_i H(x - x_i),$$

где

$$\lambda_i = \frac{m_{i+1} - m_i}{x_{i+1} - x_i} - \frac{m_i - m_{i-1}}{x_i - x_{i-1}}, \quad i=1, 2, \dots, n-1,$$

$$\lambda_0 = \frac{m_1 - m_0}{x_1 - x_0}.$$

Иначе говоря, производная искомой функции оказывается кусочно-постоянной, а сама функция — кусочно-линейной. Таким образом, в результате формального применения описанного метода мы получаем кусочно-линейную аппроксимацию функции, заданной своими значениями в отдельных точках.

Теперь рассмотрим решение, которое получается, если кроме непрерывности потребовать еще и гладкости функции $m(r)$. Проинтегрируем выражения (151) по частям, предполагая, что $m'(r)$ непрерывна в рассматриваемом интервале:

$$\mathcal{K}_i(R) m(R) - \mathcal{K}_i(R) m'(R) + \int_0^R \mathcal{K}_i(r) m''(r) dr = y_i,$$

где

$$\mathcal{K}_i(r) = \int_0^r \mathcal{K}_i(r') dr'.$$

Наблюдения y_i являются линейными функционалами $m''(r)$ и линейно зависят от значений $m(R)$ и $m'(R)$. Аналогично предыдущему, можно определять $m(r)$ из условия минимума выражения

$$\int_0^R [m''(r)]^2 dr + \alpha [m'(R)]^2 + \beta [m(R)]^2.$$

Функция $m''(r)$ и значения $m'(R)$ и $m(R)$ определяются из системы линейных уравнений, которые строятся так же, как и в предыдущем случае.

Если такой подход применить к рассмотренной задаче интерполяции функции при условии, что в точке $x = x_n$ задано не только значение функции m , но и значение ее первой производной $m'(x_n)$, то получим, что вторая производная искомой функции является кусочно-линейной, и соответственно сама функция

$m(x)$ в каждом интервале (x_i, x_{i+1}) аппроксимируется кубической параболой, причем в точках x_i будет непрерывна не только первая производная (в соответствии с принятым нами допущением), но и вторая - в силу непрерывности функции $\mathcal{K}_i(x) = \int_{x_0}^x \mathcal{H}(x' - x_i) dx'$.

Таким образом, применение метода Вейкуса - Гильберта в такой форме к задаче интерполяции функции дает интерполяцию кубическим сплайном.

Пример 2. Рассмотрим теперь решение обратной кинематической задачи геометрической сейсмоки, о постановке которой упоминалось в начале этого параграфа.

При решении этой задачи оказывается удобным выбирать в качестве искомой функции не поправку к скорости $\delta v(r)$, а безразмерную величину $m(r) = -\delta v(r)/v_0(r)$, где $v_0(r)$ - исходный скоростной разрез, к которому определяется поправка. При таком выборе искомой функции

$$G_i(r) = \begin{cases} 2 \left(v_0(r) \sqrt{1 - \frac{p_i^2 v_0^2}{r^2}} \right)^{-1}, & r_m^i < r < R, \\ 0, & r < r_m^i, \end{cases}$$

$$\text{а } \mathcal{K}_i(r) = \int_0^r G_i(r') dr' = \tau_i(r).$$

Функция $\tau_i(r)$ имеет следующий смысл: при $r < r_m^i$ она равна нулю, а при $r > r_m^i$ она представляет удвоенное время пробега волны вдоль участка луча от точки $r = r_m$ до точки, имеющей координату r (рис.30). Соответственно $\mathcal{K}_i(R) = t_i$ - время

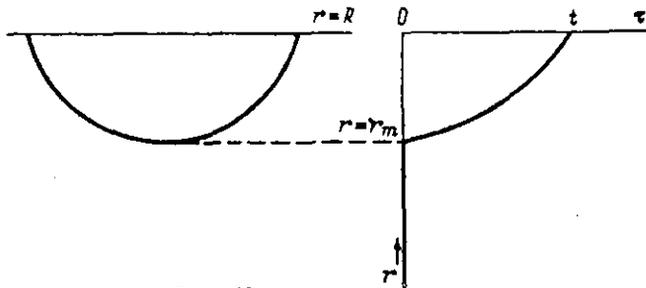


Рис. 30.

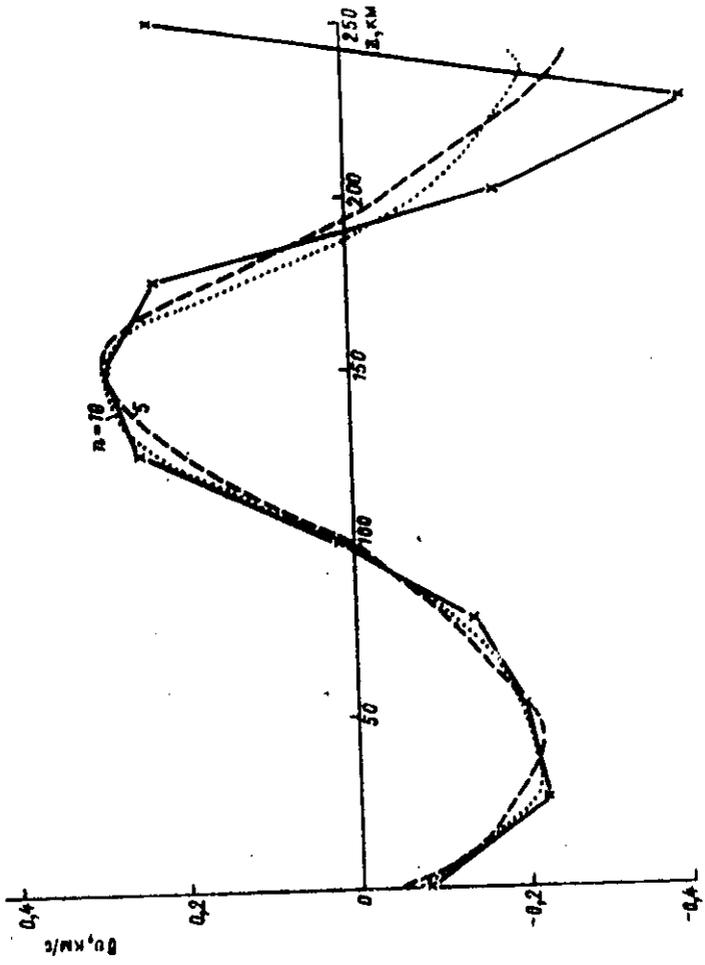


FIG. 31.

пробега волны вдоль i -го луча. Элементы матрицы системы уравнений для определения коэффициентов λ_i будут соответственно иметь вид

$$C_{ij} = -l_i l_j / \alpha - \int_0^a \tau_i(r) \tau_j(r) dr.$$

Вектор Λ , элементами которого являются неизвестные коэффициенты λ_i , определяется выражением $\Lambda = C^{-1} \delta t$, где δt - вектор невязок времен пробега относительно исходной модели.

Совершенно аналогично определяются функции $G_i(x)$ и $X_i(x)$ в случае, когда изучаемая среда представляет полупространство $x > 0$, а скорость $v = v(x)$. Для такого случая на рис. 31 представлены решения обратной задачи для следующей модели. "Истинный" скоростной в интервале $0 < x < 250$ км разрез взят в виде кусочно-линейной функции x , а в качестве начального приближения принята линейная функция $v_0(x) = (8 + 0,02x)$ км/с. Разность $\delta v(x)$ между "истинным" разрезом и начальным приближением (сплошная кривая) является также кусочно-линейной функцией. В начальной части ($x < 175$ км) эта функция изменяется довольно плавно, а при больших x ее изменения взяты значительно более резкими. Поскольку при построении решения мы исходили из условия плавного изменения функции $\delta v(x)$, следует ожидать лучшего совпадения решения с моделью при $x < 175$ км и худшего при больших значениях x . Для нахождения решения использовались две выборки данных: $n = 5$ и $n = 10$. Значения δt_i используемые в качестве исходных данных, соответствовали лучам, выходящим в интервале расстояний $0 < x < 1000$ км, в первом случае с шагом $\Delta x = 200$ км, а во втором - с шагом $\Delta x = 100$ км. Параметр α был принят равным 0,1.

Как и ожидалось, при больших значениях x между решением и исходной моделью $\delta v(x)$ отмечается заметное расхождение, но при малых x решение очень хорошо аппроксимирует модель даже для малой выборки ($n = 5$).

§ 6. Построение сглаженного решения

в случае, когда ядра не квадратично интегрируемы

Если ядра G_i не являются квадратично интегрируемыми, то критерий дельтаобразности сглаживающего ядра, заключающийся в

минимизации функционала (134), уже не годится, так как интегралы \mathcal{S}_{ij} , определяемые формулой (138), не существуют. В этом случае имеет смысл минимизировать функционал (141), который, как было показано, также характеризует ширину интервала разрешения.

Поскольку $G(r)$ являются интегрируемыми функциями, то и определенная в § 3 функция $\Phi(r, r_0)$ не имеет сингулярностей и, таким образом, подынтегральное выражение в (141) является конечным.

Рассмотрим, как производится построение сглаженного решения и оценка разрешающей способности на основе критерия дельтообразности, заключающегося в минимизации функционала (141). Поскольку $\mathcal{K}(r, r_0) = \sum_i a_i(r_0) G_i(r)$, то

$$\Phi(r, r_0) = \sum_i a_i \int_0^r G_i(r') dr' = \sum_i a_i \mathcal{K}_i(r). \quad (154)$$

Подставим выражение (154) в (141):

$$s(r_0) = 12 \left[\sum_{i,j} a_i a_j \int_0^R \mathcal{K}_i \mathcal{K}_j dr - 2 \sum_i a_i \int_0^R \mathcal{K}_i(r) dr + R - r_0 \right].$$

В векторной форме это выражение может быть написано следующим образом:

$$s(r_0) = 12(\mathbf{a}^T \mathbf{P} \mathbf{a} - 2\mathbf{a}^T \mathbf{h} + R - r_0),$$

где элементы матрицы \mathbf{P} и вектора \mathbf{h} равны соответственно

$$P_{ij} = \int_0^R \mathcal{K}_i(r) \mathcal{K}_j(r) dr,$$

$$h_i = \int_0^R \mathcal{K}_i(r) dr.$$

Величина s должна минимизироваться при условии $\mathbf{a}^T \mathbf{g} = 1$, так что искомое решение должно минимизировать функционал

$$12(\mathbf{a}^T \mathbf{P} \mathbf{a} - 2\mathbf{h}^T \mathbf{a} + R - r_0) + \lambda(\mathbf{a}^T \mathbf{g} - 1)$$

или определяться из системы уравнений

$$24(\mathbf{P} \mathbf{a} - \mathbf{h}) + \lambda \mathbf{g} = 0,$$

$$\mathbf{a}^T \mathbf{g} = 1.$$

Выражая \mathbf{a} из первого уравнения и подставляя его во второе, получаем

$$\mathbf{a} = \mathbf{P}^{-1} \mathbf{h} + \frac{1 - \mathbf{g}^T \mathbf{P}^{-1} \mathbf{h}}{\mathbf{g}^T \mathbf{P}^{-1} \mathbf{g}} \mathbf{P}^{-1} \mathbf{g}. \quad (155)$$

Если наблюдения содержат ошибки, то так же, как и в рассмотренном в § 4 случае, минимизировать следует не \mathcal{E} , а линейную комбинацию $\mathcal{E} \sin \vartheta + \omega \rho^2 \cos \vartheta$, что приводит к решению, имеющему вид (155), но в нем матрица \mathbf{P} и вектор \mathbf{h} должны быть заменены следующими:

$$\begin{aligned}\tilde{\mathbf{P}} &= \mathbf{P} \sin \vartheta + \omega \mathbf{R} \cos \vartheta, \\ \tilde{\mathbf{h}} &= \mathbf{h} \sin \vartheta.\end{aligned}\quad (156)$$

В случае, когда наблюдения некогерентны и их дисперсии равны, переход к выражениям (156) сводится просто к добавлению постоянного слагаемого $\alpha = \omega \sigma^2 \operatorname{ctg} \vartheta$ к диагональным элементам матрицы \mathbf{P} . Такая процедура оказывается существенной в случае, когда матрица \mathbf{P} плохо обусловлена. Поскольку \mathbf{P} является положительно определенной, то добавление постоянного слагаемого к ее диагональным элементам приводит к регуляризации при обращении плохо обусловленной матрицы (см. § 2 гл. III).

Пример 1. Как и в § 5, проиллюстрируем смысл полученных результатов на задаче об интерполяции функции, заданной своими значениями в дискретных точках.

Итак, пусть заданы значения m_i исследуемой функции $m(x)$ в точках $x_i \in (0, \mathcal{X})$ ($i = 1, 2, \dots, n$), причем $x_1 < x_2 < \dots < x_n$. Ядра $G_i(x)$ в этом случае, как уже отмечалось, будут δ -функциями $\delta(x - x_i)$, а $\mathcal{K}_i(x)$ - функциями Хависайда $H(x - x_i)$. Элементы матрицы \mathbf{P} определяются следующим образом:

$$P_{ij} = \mathcal{X} - \max\{x_i, x_j\},$$

а компоненты векторов $\mathbf{h}(x_0)$ и \mathbf{g} имеют вид

$$h_i(x_0) = \mathcal{X} - \max\{x_0, x_i\}, \quad g_i = 1.$$

Вычисляя вектор коэффициентов a_i по формуле (155), получим, что если рассматриваемая точка x_0 , в которой определяется сглаженная функция $\langle m \rangle_{x_0}$, находится в интервале (x_{i-1}, x_i) , то

$$a_{i-1} = \frac{x_i - x_0}{x_i - x_{i-1}}, \quad a_i = \frac{x_0 - x_{i-1}}{x_i - x_{i-1}}, \quad a_k = 0, \quad k \neq i-1, i.$$

И следовательно, $\langle m \rangle_{x_0}$ определяется линейной интерполяцией со значениям m_i и m_{i-1} :

$$\langle m \rangle_{x_0} = a_{i-1} m_{i-1} + a_i m_i = m_{i-1} + \frac{m_i - m_{i-1}}{x_i - x_{i-1}} (x_0 - x_{i-1}).$$

Величина s при этом определится как

$$s = 12 \frac{(x_i - x_0)(x_0 - x_{i-1})}{x_i - x_{i-1}}, \quad (157)$$

т.е. в точках $x_0 \neq x_i$, как и следовало ожидать, $s = 0$, поскольку в этих точках отсутствует неоднозначность в определении значений функции $m(x)$, так что локально усредненное значение совпадает с самим значением функции. В промежутках же между этими точками функция $m(x)$ может быть произвольной. Вся информация о функции $m(x)$ в промежуточных точках $x_0 \neq x_i$ дается интегралом (132), где ядро \mathcal{A} даже не определяет в обычном смысле локальное сглаживание в окрестности x_0 , так как является суммой двух δ -функций, сосредоточенных в точках x_{i-1} и x_i . Построенное таким способом сглавленное решение будет более или менее соответствовать функции $m(x)$ только в случае, если функция достаточно плавно меняется между заданными точками.

Ясно, что при таком "сглаживающем" ядре нет смысла говорить о ширине интервала сглаживания, хотя величина s и характеризует некоторым образом степень усреднения функции. Из формулы (157) следует, что наибольшее значение s принимает в середине интервала между точками, где оно равно $3(x_i - x_{i-1})$. Интересно сравнить этот результат с тем, который был получен в § 3 для случая, когда сглаживающее ядро является Π -образной функцией, симметричной относительно рассматриваемой точки и отличной от нуля в интервале длиной l : в этом случае $s = l$. Если же усреднение проводится не по всему интервалу равномерно, что имеет место при сглаживании Π -образной функцией, а только по значениям в двух крайних точках, то $s = 3l$.

П р и м е р 2. Обратимся теперь к задаче нахождения скоростного разреза $v(x)$ по данным о временах пробега волн в конечном числе точек, для которой в § 5 были построены решения, удовлетворяющие критерию гладкости $\delta v(x)$, и посмотрим, что будет в этом случае представлять сглавленное решение.

Прежде всего отметим то, что сглавленное решение имеет смысл только для поправочной функции $m(x) = -\delta v(x)/v_0(x)$, а не для окончательного распределения скорости $v(x)$. Иначе говоря, если $\langle m \rangle_{x_0}$ есть результат локального сглаживания

функции $m(x)$, то $v_0(1 - \langle m \rangle_{x_0})$ не является сглаженным решением для $v(x)$. Поэтому далее будут приводиться результаты только для функции $m(x)$.

На рис. 32, а приведены сглаженные решения $\langle m \rangle_{x_0}$ для тех же выборок исходных данных: $n = 5$, $n = 10$, которые были использованы в примере § 5, и истинное распределение $m(x)$ (сплошная кривая). При рассмотрении примера I настоящего параграфа упоминалось, что сглаженное решение должно более или менее подходяще аппроксимировать искомую функцию только в тех областях, где эта функция меняется достаточно плавно. Относительно

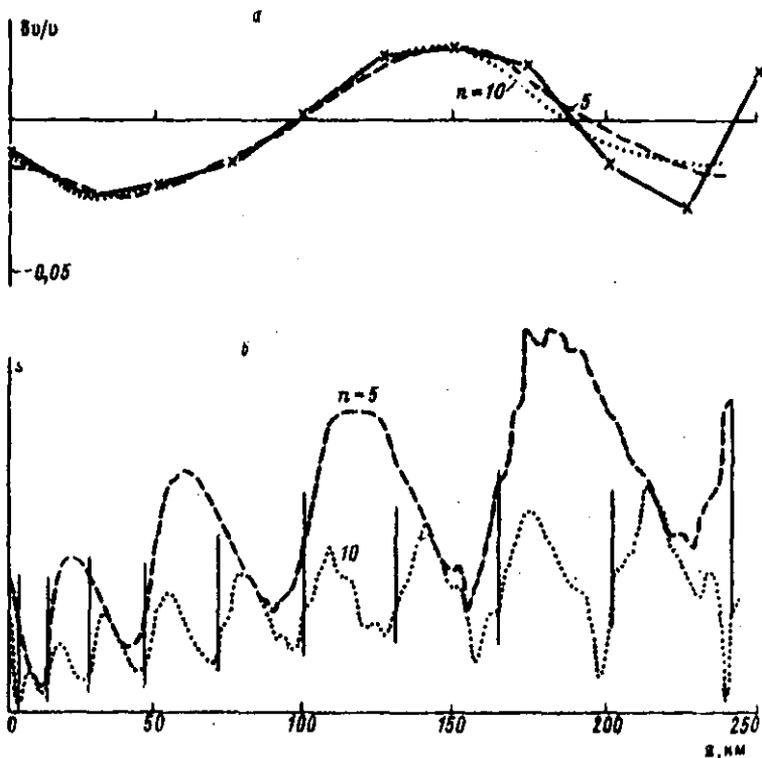


Рис. 32

тельно плавное изменение функции $m(z)$ в рассматриваемом случае имеет место при $0 < z < 175 \div 200$ км. Поэтому сглаженное решение $\langle m \rangle_{z_0}$ и оказывается близким к исходной модели $m(z_0)$ в этом интервале глубин. Напомним, что решение, построенное на основе критерия наибольшей плавности функции $\delta v(z)$, было в хорошем согласии с исходной моделью на таких же глубинах.

На рис.32,б приведены кривые $s(z)$, определяющие степень сглаженности решения для двух использованных выборок данных. Вертикальные линии на этом рисунке указывают глубины, к которым приурочены минимумы лучей.

Видно, что на этих глубинах величина s уменьшается, так что именно там использованные данные дают наилучшее разрешение.

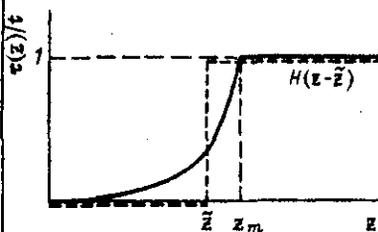


Рис. 33.

Тот факт, что минимумы $s(z)$ оказываются несколько смещены относительно значений z_m^i в сторону меньших значений z , нетрудно понять, если обратиться к рис.33: приближение функции $\tau(z)$ к ступенчатой функции будет наилучшим, если ступень начинается со значения $z = \tilde{z}$, несколько меньшего, чем z_m .

Обратим внимание на то, что в рассмотренной задаче оба решения – локально сглаженное и то, которое удовлетворяет условию минимизации (152) (пример 2 из § 5), – оказываются близки между собой и хорошо отражают поведение искомой функции в той области, где она достаточно плавная. В следующей главе будет показано, что решение неопределенной системы линейных уравнений (такой, в которой число уравнений меньше числа неизвестных), удовлетворяющее условию минимизации нормы искомого вектора, обладает и свойством наилучшего разрешения, но в рамках критерия, несколько отличного от тех критериев дельтосоразности сглаживающей функции, которые рассматривались в настоящей главе. Таким образом, понятно, что два подхода к построению решения в рамках метода Байкура – Гильберта должны приводить к близким результатам.

§ I. Обобщенная постановка
линеаризованных обратных задач

В предыдущих главах были рассмотрены подходы к решению обратных задач, различающихся соотношением количества исходных данных и количества определяемых параметров модели. Когда данных больше, чем искомым параметров, последние определяются приближенно методами оптимизации, базирующимися на теории статистического оценивания. Если количество данных мало по сравнению с числом параметров модели (представление модели в форме неизвестной функции $m(r)$ эквивалентно представлению ее бесконечно большим числом параметров), то для построения решения может быть использован метод Байкуса - Гильберта. В этой главе мы покажем, что для линеаризованных обратных задач как тот, так и другой методы являются частными случаями метода псевдообращения (или обобщенного обращения), разработанного в линейной алгебре для решения несовместных и неопределенных систем уравнений [2, 45, 65, 68].

Очевидно, что все линеаризованные обратные задачи сводятся к решению линейной системы

$$A m = y, \quad (158)$$

где $m = \{m_1, m_2, \dots, m_k\}$; $y = \{y_1, y_2, \dots, y_n\}$; A - матрица размера $n \times k$, причем на соотношение между n и k не наложено никаких ограничений.

Покажем это сначала для случая, когда $k < n$ и для оценки параметра m_i используется метод наименьших квадратов при условии, что наблюдения неавноточны и коррелированы. Оценка параметра в этом случае производится путем минимизации следующей функции отклоня:

$$\lambda(m) = (y - A m)^T R^{-1} (y - A m),$$

где R - ковариационная матрица ошибок наблюдений. Поскольку

матрица R^{-1} является симметричной и положительно-определенной, $\lambda(m)$ можно преобразовать к сумме квадратов. Если выбрать новые исходные данные $y' = R^{-1/2} y$ и соответственно место матрицы A взять матрицу $A' = R^{-1/2} A$, то $\lambda(m)$ может быть записана в следующем виде:

$$\lambda(m) = (y' - A'm)'(y' - A'm).$$

А минимизация такой квадратичной формы эквивалентна решению системы уравнений $A'm = y'$ по методу наименьших квадратов. Решение при этом получается таким, которое удовлетворяет (158) приближенно.

В случае, когда число исходных данных ограничено, а модель не может быть представлена конечным числом параметров, для решения обратной задачи используется метод Бэйкуса - Гильберта, который представляет процедуру построения решения с минимальной нормой. Как уже упоминалось в гл. IV, такая процедура может быть вообще использована для решения задач, в которых искомого параметров больше, чем исходных данных, и решений существует бесконечное множество, а связь между наблюдениями y и моделью m определяется линейным оператором $y = Am$, причем m может принадлежать как пространству конечного числа измерений, так и пространству функций. В качестве решения обратной задачи в этом случае принимается нормальное решение системы (158).

Таким образом, в общем решение линеаризованной обратной задачи сводится к решению системы уравнений (158). При этом в зависимости от соотношения между k , n и рангом матрицы A система может вообще не иметь решения, иметь единственное решение или иметь их бесконечное множество. В случае, когда ранг матрицы $r = k$ ($k \leq n$), то, если решение существует, оно единственно. Условие $r = n$ гарантирует существование решения: при $n = k$ оно будет единственным, а при $n < k$ будет существовать бесконечное множество решений. Поэтому задачей решения системы (158) является в общем случае построение одного из возможных решений либо, если точного решения не существует, построение приближенного решения.

Решение строится путем преобразования пространства Y^n в M^k с помощью некоторого линейного оператора N :

$$\hat{m} = Ny. \quad (159)$$

Очевидно, что определяемое так "решение" \hat{m} может не быть точным решением (158), если число уравнений превышает число неизвестных, но во всяком случае оно должно удовлетворять (158) приближенно. В то же время, если число неизвестных больше числа уравнений и может существовать бесконечное множество решений (158), то \hat{m} должно быть одним из многих решений, причем желательно, чтобы оно наилучшим образом отражало свойства всех решений. Сказанное определяет требования, предъявляемые к оператору N [58]:

1. Очевидно, что если полученное решение $\hat{m} = Ny$ подставить в систему (158), то результат не должен сильно отличаться от y . Это значит, что $AN \sim I_n$ (I_n - единичная матрица порядка n). Близость AN к I_n определяет, насколько адекватна выбранная модель. Если, например, параметров, описывающих модель, взято слишком мало, то может оказаться, что какие-то особенности наблюдений не могут быть выявлены в рамках такой модели.

2. Поскольку $y = Am$, то $\hat{m} = NAm$, и для того чтобы \hat{m} наиболее хорошо отражало истинное решение m , которое может быть одним из множества возможных, необходимо, чтобы матрица NA была близка к единичной матрице I_d . Близость NA к I_d определяет разрешающую способность наблюдений. Если $NA = I_d$, то существует только одно решение (158), которое и определяется по (159). Если NA отличается от I_d , то это значит, что существует бесконечное множество решений (158), и \hat{m} описывает общие свойства этих решений. Если параметры m_i являются значениями искомой функции $m(r)$ в последовательных точках r_i , то \hat{m} будет отвечать сглаженной функции $\langle m \rangle$, при этом чем более NA отличается от I_d , тем более сглаженное решение мы будем получать.

3. Наблюдения всегда содержат ошибки. Поэтому и решение будет получаться с некоторой ошибкой. Если ковариационная матрица наблюдений R_y , то ковариационная матрица оценок параметров $R_m = NR_y N^T$. Очевидно, что дисперсия решения не должна быть слишком велика, иначе решение теряет смысл. Поскольку матрица R_y может считаться заданной, то требование малости дисперсии решения накладывает определенные ограничения на оператор N . Другими словами, это есть требование регуляризации решения.

Нахождение решения системы (158) является, собственно говоря, задачей линейной алгебры. Но поскольку это решение будет рассматриваться как решение обратной геофизической задачи, представляется важным его истолкование и сопоставление с результатами решения, получаемыми другими методами, описанными в предыдущих главах. Кроме того, в аспекте именно геофизической задачи помимо построения решения еще подлежат рассмотрению следующие вопросы: а) определение разрешающей способности данных; б) определение информационной содержательности исходного материала; в) определение оптимальной параметризации, т.е. таких параметров, которые по данной выборке могли бы быть определены однозначно.

§ 2. Сингулярное разложение

Метод построения решения (159) предложен Е. Муром и Р. Пенроузом [65, 68]. Матрица \mathbf{M} , называемая псевдообратной или обобщенной обратной к \mathbf{A} , с помощью которой осуществляется нахождение такого решения, строится на основе сингулярного разложения прямоугольной матрицы \mathbf{A} . Сингулярное разложение заключается в том, что любая прямоугольная матрица \mathbf{A} , ранг которой $r \leq k, n$, может быть представлена в виде следующего произведения:

$$\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T, \quad (160)$$

где $\mathbf{\Lambda}$ - диагональная матрица размера $r \times r$; \mathbf{U} и \mathbf{V} - прямоугольные матрицы размера соответственно $n \times r$ и $k \times r$, столбцами которых являются векторы \mathbf{u}_i и \mathbf{v}_j ($i, j \leq r$), определяемые из системы уравнений

$$\begin{aligned} \mathbf{A} \mathbf{u}_i &= \lambda_i \mathbf{u}_i, \\ \mathbf{A}^T \mathbf{v}_i &= \lambda_i \mathbf{v}_i. \end{aligned} \quad (161)$$

Диагональные элементы λ_i матрицы $\mathbf{\Lambda}$ называются сингулярными числами матрицы \mathbf{A} . Систему (161) можно иначе записать в виде

$$\begin{aligned} \mathbf{A} \mathbf{V} &= \mathbf{U} \mathbf{\Lambda}, \\ \mathbf{A}^T \mathbf{U} &= \mathbf{V} \mathbf{\Lambda} \\ \text{или} \quad \mathbf{A}^T \mathbf{A} \mathbf{V} &= \mathbf{V} \mathbf{\Lambda}^2, \\ \mathbf{A} \mathbf{A}^T \mathbf{U} &= \mathbf{U} \mathbf{\Lambda}^2. \end{aligned}$$

Таким образом, v_i являются собственными векторами, отвечающими ненулевым собственным значениям симметричной матрицы $A^T A$, а u_i — собственными векторами матрицы AA^T . В связи с этим матрицы U и V ортонормированы, т.е. удовлетворяют соотношениям $U^T U = I_r$, $V^T V = I_r$.

Псевдообратная матрица H определяется следующим образом:

$$H = VA^{-1}U^T. \quad (162)$$

В случае, когда матрица A квадратная и ее ранг $r = k = n$, $A = V\Lambda V^T$ и соответственно $H = VA^{-1}V^T = A^{-1}$, т.е. в этом случае псевдообратная матрица просто совпадает с обратной.

Операция псевдообращения может быть обобщена и на случай, когда число искомым параметров становится бесконечным, т.е. M является функциональным пространством. Используем обозначения, принятые в предыдущей главе:

$$\int_0^a G_i(r) m(r) dr = y_i, \quad i = 1, 2, \dots, n. \quad (163)$$

В этом случае λ_i^2 и $v_i(r)$ будут соответственно собственными значениями и собственными функциями интегрального уравнения

$$\int_0^a K(r, r') v_i(r') dr' = \lambda_i^2 v_i(r),$$

где ядро $K(r, r') = \sum_i G_i(r) G_i(r')$.

В дальнейшем мы будем считать, что размерность пространства M может быть как конечной, так и бесконечной.

§ 3. Свойства решения, получаемого путем псевдообращения

Рассмотрим свойства решения, определяемого с помощью оператора псевдообращения (162), в разных случаях.

Как уже было сказано, в случае $r = k = n$ матрица A имеет обратную, система (158) имеет единственное решение $m = A^{-1}y$ и обобщенное совпадает с ним.

В случае $r = k < n$ решения, точно удовлетворяющего (158), не существует (система несовместная). Для построения приближенного решения в таких задачах используется метод наименьших квадратов. Это решение минимизирует средний квадрат невязки

$$e^T e = (Am - y)^T (Am - y).$$

Покажем, что решение, определяемое оператором псевдообращения (162), совпадает с решением, получаемым по методу наименьших квадратов $\mathbf{m}_{n,k}$. Оно, как известно, определяется из системы нормальных уравнений

$$(\mathbf{A}^T \mathbf{A}) \mathbf{m}_{n,k} = \mathbf{A}^T \mathbf{y}.$$

Если $r = k$, то матрица $\mathbf{A}^T \mathbf{A}$ имеет обратную и

$$\mathbf{m}_{n,k} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}. \quad (164)$$

Используем для \mathbf{A} сингулярное разложение (160). Тогда

$$\mathbf{m}_{n,k} = (\mathbf{V} \mathbf{\Lambda} \mathbf{U}^T \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T)^{-1} \mathbf{A}^T \mathbf{y}.$$

Учитывая ортонормированность матриц \mathbf{U} и \mathbf{V} , получим

$$\mathbf{m}_{n,k} = \mathbf{V} \mathbf{\Lambda}^{-2} \mathbf{V}^T \mathbf{V} \mathbf{\Lambda} \mathbf{U}^T \mathbf{y} = \mathbf{V} \mathbf{\Lambda}^{-1} \mathbf{U}^T \mathbf{y} = \mathbf{m}_y.$$

Таким образом, в этом случае решение (162) совпадает с тем, которое получается по методу наименьших квадратов.

Теперь рассмотрим случай $k > n = r$, когда система (158) будет иметь бесконечное множество решений. Спрашивается, чем будет характеризоваться то из решений, которое определяется псевдообратным оператором \mathbf{H} ? Оказывается, оно будет таким, которое, удовлетворяя (158), одновременно минимизирует $\|\mathbf{m}\|^2 = \mathbf{m}^T \mathbf{m}$, т.е. является нормальным решением.

Поскольку $r = n$, а векторы \mathbf{v}_i представляют n собственных векторов матрицы $\mathbf{A}^T \mathbf{A}$, отвечающих ненулевым собственным значениям, должны существовать еще $k - n$ собственных векторов \mathbf{v}_{0i} , отвечающих нулевым собственным значениям матрицы $\mathbf{A}^T \mathbf{A}$. Вместе с n собственными векторами \mathbf{v}_i ($i = 1, 2, \dots, n$) они образуют полную ортонормированную систему, из которой может быть разложен любой вектор \mathbf{m} в пространстве \mathcal{M}^k .

Таким образом, общее решение (158) можно представить в виде

$$\mathbf{m} = \mathbf{V} \boldsymbol{\alpha} + \mathbf{V}_0 \boldsymbol{\alpha}_0, \quad (165)$$

где \mathbf{V}_0 - матрица размера $k \times (k - n)$, столбцы которой - собственные векторы \mathbf{v}_{0i} . Построение решения \mathbf{m} сводится, следовательно, к нахождению векторов $\boldsymbol{\alpha}$ и $\boldsymbol{\alpha}_0$.

В силу ортонормированности собственных векторов имеет место следующие соотношения:

$$\mathbf{V}^T \mathbf{V} = \mathbf{I}_n, \quad \mathbf{V}_0^T \mathbf{V}_0 = \mathbf{I}_{k-n}, \quad \mathbf{V}^T \mathbf{V}_0 = 0, \quad \mathbf{V}_0^T \mathbf{V} = 0. \quad (166)$$

Поскольку $r=n$, система собственных векторов u_i в пространстве Y^n является полной и вектор наблюдений y может быть разложен по этой системе векторов:

$$y = J\beta. \quad (167)$$

Учитывая сингулярное разложение матрицы A и разложения m и y по собственным векторам, запишем систему уравнений (158) в виде

$$U\Lambda V^T(V\alpha + V_0\alpha_0) = U\beta. \quad (168)$$

Преобразуя это выражение с учетом (166), получим

$$U\Lambda\alpha = U\beta,$$

откуда $\alpha = \Lambda^{-1}\beta$.

Таким образом, в общем решении вектор α фиксирован, а α_0 может выбираться произвольно. Величина $m^T m$ выражается через α и α_0 следующим образом:

$$m^T m = (V\alpha + V_0\alpha_0)^T (V\alpha + V_0\alpha_0) = \alpha^T \alpha + \alpha_0^T \alpha_0.$$

Из этого выражения видно, что минимальное значение $m^T m$ будет достигнуто тогда, когда $\alpha_0 = 0$. Из формулы (164) следует, что $\beta = U^T y$.

Таким образом, решение с минимальной нормой может быть записано в виде

$$m = V\Lambda^{-1}U^T y = Ny,$$

откуда видно, что оно совпадает с тем, которое определяется псевдообратным оператором (162). Именно такое решение, как уже говорилось в гл. IV, и дает метод Бойкуса - Гильберта.

Теперь рассмотрим наиболее общий случай, когда $r < k, n$, причем соотношение между k и n может быть любым. Если в этом случае $k < n$, то решений по методу наименьших квадратов может быть бесконечное множество. Это следует из того, что система нормальных уравнений оказывается вырожденной и матрица этой системы $A^T A$ не имеет обратной. Если же $k > n$, то система (158) может не иметь точного решения.

Так же, как и в рассмотренном ранее случае, разложим решение по системе собственных векторов v_i и v_{0i} :

$$m = V\alpha + V_0\alpha_0. \quad (169)$$

В виде такого же разложения можно представить и вектор наблюдений:

$$y = U\beta + U_0\beta_0, \quad (170)$$

где u_{i0} — столбцы матрицы U_0 — представляют собой собственные векторы матрицы AA^T , отвечающие нулевым собственным значениям. Если подставить выражения (169) и (170) в систему (158) и для матрицы A использовать ее сингулярное разложение, то легко видеть, что точного решения системы в этом случае не существует.

Рассмотрим решение, минимизирующее средний квадрат невязки $e^2 = (Am - y)^T(Am - y)$. Подставляя в это выражение m и y из выражений (169) и (170), получим

$$e^2 = (U\Lambda\alpha - U\beta - U_0\beta_0)^T(U\Lambda\alpha - U\beta - U_0\beta_0) = \beta^T\beta + \beta_0^T\beta_0 + \alpha^T\Lambda^2\alpha - 2\alpha^T\Lambda\beta.$$

Отсюда видно, что минимум квадрата невязки достигается в случае $\alpha = \Lambda^{-1}\beta$. Но это условие определяет только α , но не α_0 . Если на решение еще наложить условие минимума нормы $\|m\|$, то получим $\alpha_0 = 0$.

Таким образом, решение, отвечающее и минимуму нормы, и минимуму квадрата невязки будет иметь вид

$$m = V\Lambda^{-1}\beta.$$

Из формулы (170) нетрудно видеть, что $\beta = U^T y$, так что $m = V\Lambda^{-1}U^T y = Ny$.

Таким образом, в этом случае обобщенный обратный оператор дает приближенное решение по методу наименьших квадратов, обладающее минимальной нормой.

§ 4. Определение оптимальной параметризации

В гл. I были сформулированы основные требования к параметризации модели: с одной стороны, параметров должно быть взято не слишком много, чтобы не получилось бесконечного множества решений, а с другой — при слишком малом числе параметров модель может оказаться неадекватной наблюдениям, т.е. некоторые особенности наблюдений не смогут быть объяснены в рамках принятой модели. Поэтому вопрос о выборе оптимальной параметризации является весьма важным при постановке обратной геофизической задачи.

В рамках рассмотренных ранее методов — оптимизационного и метода Вейкуса — Гильберта — не удастся дать однозначного

ответа на этот вопрос, хотя и имеются некоторые возможности улучшить первоначально выбранную параметризацию, исходя из получаемых результатов. При решении обратной задачи по методу максимального правдоподобия одновременно с оценкой параметров определяются и ковариационная и корреляционная матрицы оценок. Анализ корреляционной матрицы позволяет выявить те параметры, между оценками которых существует сильная корреляционная зависимость и которые нет смысла оценивать раздельно. Вместо нескольких сильно коррелированных параметров можно выбрать один, являющийся некоторой их комбинацией, и таким образом неразрешенные параметры заменить тем, который может быть однозначно оценен по заданной выборке данных. К сожалению, этот метод не дает конструктивного способа детализировать параметризацию, т.е. разукрупнять параметры. Единственная возможность заключается в анализе корреляционных свойств получаемых невязок. Если отклонения наблюдаемых значений от теоретически рассчитанных для полученного решения имеют дисперсию значительно большую, чем та, которая оценивается по наблюдениям, или они, например, оказываются коррелированными, тогда как ошибки наблюдений независимы, то это дает основания полагать, что параметры выбраны неудачно и их следует разукрупнять.

В методе Бэйкуса - Гильберта разрешающая способность наблюдений оценивается как по величине $\sigma(r_0)$, характеризующей ширину интервала сглаживания, так и по виду сглаживающих ядер $\mathcal{A}(r, r_0)$. Анализ этих функций позволяет заключить, с какой степенью детализации исходные данные позволяют определять искомую модель в том или ином интервале глубин. Если параметризация заключается в разбиении разреза на слои и принятии в качестве параметров средних значений искомой физической величины в каждом слое, то для грубой оценки толщины слоев можно использовать значения σ на соответствующих глубинах.

Теперь рассмотрим, как осуществляется выбор оптимальной параметризации в рамках метода псевдообращения. В качестве исходной выберем такую параметризацию, при которой параметры оказываются неразрешенными, так что существует бесконечное множество решений системы (158). Это может быть в случае, когда либо число неизвестных параметров k больше числа уравнений, либо ранг матрицы A меньше k . Последний случай означает, что

некоторые из параметров связаны: они не могут быть определены раздельно. Задача состоит в том, чтобы выбрать такие комбинации параметров, которые могут быть определены однозначно.

Если ранг матрицы A равен r , то очевидно, что однозначно можно определить только r параметров. Определим r -мерный вектор m^* , элементы которого будем считать новыми параметрами, подлежащими определению. Это легко сделать, если обратиться к представлению решения m в форме (165). Из выражения (165) видно, что произвол при построении решения заключается в выборе вектора α_0 , а вектор α определяется однозначно. Он представляет собой вектор в подпространстве $M^r \in M^k$, построенном на собственных векторах v_i . Таким образом, если решение искать не в пространстве M^k , а в подпространстве M^r , то оно оказывается единственным.

Нетрудно видеть, что связь вектора новых параметров $m^* = \alpha$ с исходным вектором m определяется следующим образом: $m^* = V^T m$. Эти параметры однозначно определяются по наблюдениям:

$$m^* = A^{-1} U^T y.$$

Интересно выяснить, какой физический смысл имеют новые параметры m^* . В случае, когда $m = m(z)$, т.е. искомая модель представляет собой зависимость рассматриваемой физической величины от глубины, параметры m_i^* являются результатом сглаживания этой функции по глубине с весовой функцией $v_i(z)$. (Как уже говорилось в § 2, в этом случае $v_i(z)$ являются собственными функциями интегрального уравнения.) Очевидно, что для того чтобы новая параметризация имела физический смысл, необходимо, чтобы такое сглаживание было компактным, т.е. производилось по ограниченной области глубин. Самым идеальным случаем будет такой, какой изображен на рис.34, когда каждая из собственных функций $v_i(z)$ усредняет искомую модель по определенной области глубин, которые не пересекаются между собой. В этом случае новая параметризация будет заключаться в разбиении разреза на слои и выборе в качестве параметров определенных значений $\langle m \rangle_i = \frac{1}{H_i} \int_{z_i}^{z_i + H_i} m(z) dz$ в каждом слое. К сожалению, в большинстве практических задач собственные функции $v_i(z)$ (или собственные векторы v_i) имеют значительно более

сложный вид, и задача заключается в нахождении такой комбинации этих векторов, которая давала бы по возможности компактное разрешение относительно исходной модели. В некоторых случаях это

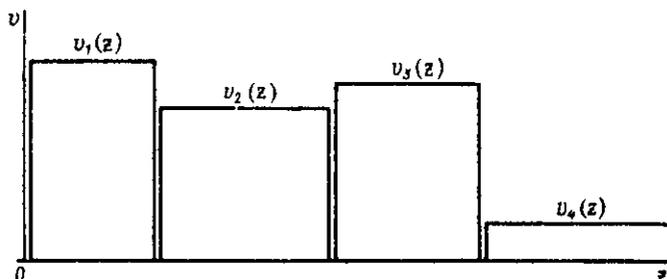


Рис. 34.

удается сделать, в других это оказывается невозможным. Это иллюстрируется приводимыми ниже примерами.

Пример 1. Пусть собственные векторы имеют вид

$$\mathbf{v}_1^T = 1/\sqrt{14} (1, 1, 2, 2, 2),$$

$$\mathbf{v}_2^T = 1/\sqrt{21} (-3, -3, 1, 1, 1).$$

Легко видеть, что следующее преобразование приводит к компактности:

$$\mathbf{v}_1^* = 1/\sqrt{7} (v_1 - 2\sqrt{3/2} v_2); \quad \mathbf{v}_2^* = 1/\sqrt{7} (3v_1 - \sqrt{3/2} v_2),$$

$$\mathbf{v}_1^{*T} = 1/\sqrt{2} (1, 1, 0, 0, 0); \quad \mathbf{v}_2^{*T} = 1/\sqrt{3} (0, 0, 1, 1, 1).$$

Рассмотрим другой случай, когда собственные векторы имеют вид

$$\mathbf{v}_1^T = 1/\sqrt{14} (1, 2, 2, 2, 1); \quad \mathbf{v}_2^T = 1/\sqrt{21} (-3, 1, 1, 1, -3).$$

Здесь уже не удастся добиться компактного разрешения. Преобразование, аналогичное предыдущему, приводит к тому, что

$$\mathbf{v}_1^{*T} = 1/\sqrt{2} (1, 0, 0, 0, 1); \quad \mathbf{v}_2^{*T} = 1/\sqrt{3} (0, 1, 1, 1, 0).$$

И хотя в полученных векторах, отличны от нуля разные компоненты, в первом из векторов они не являются соседними. Если такой вектор осуществляет переход от параметров m_1, m_2, m_3, m_4, m_5 , представляющих значения какой-то величины в последо-

тальных слоях, то преобразованными параметрами будут средние значения этих величин в 1-м и 5-м слоях и во 2, 3 и 4-м. В следующем параграфе будет дан критерий того, возможно или нет осуществление компактного разрешения параметров.

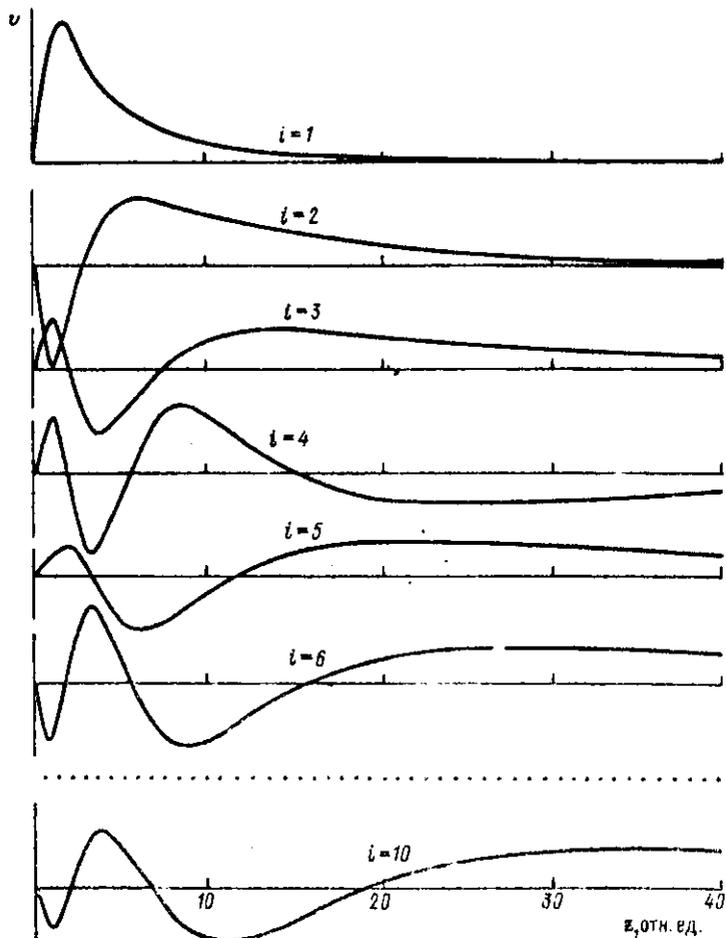


Рис. 35.

Пример 2. Рассмотрим вид собственных функций $v_i(z)$ в следующей задаче гравиметрии. Пусть имеется узкая вертикальная интрузия, характеризующаяся аномальной плотностью, и задача состоит в оценке распределения аномальной плотности по вертикали $\Delta\rho(z)$. Такую аномалию можно моделировать тонким стержнем, расположенным вдоль оси x . Измерения аномалии силы тяжести Δg производятся на поверхности Земли ($x=0$) вдоль оси x . Оператор (163) в этом случае имеет вид

$$\Delta g_i = \gamma S \int_0^{\infty} \frac{\Delta\rho(z) z dz}{(z^2 + x_i^2)^{3/2}},$$

где γ — гравитационная постоянная; S — площадь поперечного сечения стержня. Согласно сказанному ранее, неизвестную функцию $\Delta\rho(z)$ можно заменить набором параметров $\Delta\rho_i^* = \int_0^{\infty} \Delta\rho(z) v_i(z) dz$ ($i = 1, 2, \dots, n$), которые однозначно определяются по исходным данным. Функции $v_i(z)$ для случая, когда наблюдения проводятся в точках $x_i = 1, 2, \dots, 10$ ($n=10$), изображены на рис. 35. Из него видно, что новые параметры $\Delta\rho_i^*$ не дают компактного описания функции $\Delta\rho(z)$, за исключением, пожалуй, только параметра $\Delta\rho_1^*$, представляющего результат усреднения $\Delta\rho(z)$ с весовой функцией $v_1(z)$.

§ 5. Разрешающая способность данных

Среди перечисленных в § 1 требований, предъявляемых к оператору решения обратной задачи N , было требование близости $R = NA$ к единичной матрице I_k . Матрица R названа Р. Витгинсом [70] матрицей разрешения. Смысл такого названия вытекает из того, что эта матрица определяет, в какой степени в получаемом решении разрешены параметры искомой модели, поскольку решение \hat{m} связано с моделью m преобразованием, осуществляемым именно этой матрицей: $\hat{m} = Rm$.

О разрешении параметров имеет смысл говорить только в случае, когда число параметров больше числа наблюдений, или во всяком случае, больше ранга матрицы A . Иначе матрица $A^T A$ имеет обратную, и оператор N может быть записан в виде $N = (A^T A)^{-1} A^T$. При этом

$$R = NA = (A^T A)^{-1} A^T A = I_k,$$

и, таким образом, параметры оказываются полностью разрешены, т.е. могут быть определены раздельно.

Другая ситуация возникает, когда ранг матрицы A меньше числа параметров. В этом случае матрица $A^T A$ не имеет обратной, и R не будет совпадать с единичной матрицей. В то же время задача построения решения требует, чтобы матрица разрешения R была как можно более близка к единичной.

Если близость R к I_n определить по методу наименьших квадратов, то оказывается, что оператор псевдообращения дает наилучшую матрицу разрешения. Покажем это.

Используем сингулярное разложение матрицы A :

$$R = NU\Lambda V^T = BV^T$$

и будем вместо матрицы N искать матрицу B , такую, чтобы каждая из строк R была бы наиболее близка в метрике L_2 к вектору δ_j^T , имеющему компоненты δ_{ji} (δ_{ji} - символ Кронекера, $i = 1, 2, \dots, k$). Иначе говоря, каждая строка матрицы B , которую мы будем рассматривать как вектор δ_j^T , должна являться решением по методу наименьших квадратов следующего уравнения:

$$\delta_j^T V^T = \delta_j^T$$

или полученного из него в результате транспонирования:

$$V\delta_j = \delta_j.$$

Таким решением будет вектор

$$\delta_j = (V^T V)^{-1} V^T \delta_j = V^T \delta_j,$$

откуда следует, что $B = V$. Таким образом, $R = VV^T$.

В случае же, когда N является оператором псевдообращения,

$$R = V\Lambda^{-1}U^T U\Lambda V^T = VV^T,$$

т.е. матрица разрешения, образованная таким оператором, действительно является наиболее близкой к единичной в указанном смысле.

Изложим другой подход к построению оператора N , дающего наилучшее разрешение, не используя при этом сингулярного разложения матрицы A . Наилучшее разрешение достигается тогда, когда матрица $Q = NA - I_n$ близка к нулевой. Близость Q к нулевой матрице можно определить следующим образом. Рассмотрим произведение $Q^T a$, где a - произвольный вектор. Очевидно,

что вектор $Q^T a$ должен иметь минимальную норму, т.е. величина

$$J = (Q^T a)^T Q^T a = a^T Q Q^T a$$

должна быть минимальной.

Выполним следующее преобразование:

$$\begin{aligned} J &= a^T Q Q^T a = a^T (N A A^T N^T - N A - A^T N^T + I_k) a = \\ &= a^T (N^T - (A A^T)^{-1} A)^T (A A^T) (N^T - (A A^T)^{-1} A) a + \\ &\quad + a^T (I_k - A^T (A A^T)^{-1} A) a. \end{aligned}$$

Если матрица $A A^T$ имеет обратную и является, таким образом, положительно определенной, то первое слагаемое в правой части не отрицательно и достигает минимального значения (равного нулю) только тогда, когда $N^T = (A A^T)^{-1} A$ или

$$N = A^T (A A^T)^{-1}. \quad (171)$$

Как уже было показано в гл. IV, такой оператор определяет нормальное решение системы линейных уравнений. И легко видеть, что он совпадает с оператором псевдообращения. Действительно,

$$A^T (A A^T)^{-1} = V \Lambda U^T (U \Lambda^{-2} U^T) = V \Lambda^{-1} U^T.$$

Представление оператора N в форме (171) удобно в случае, когда число уравнений n меньше числа неизвестных и когда ранг матрицы A равен n .

Отметим, что рассмотренные критерии близости матрицы R к I_k не учитывают компактности разрешения. Действительно, эти критерии приводят к возрастанию диагональных элементов R , но при этом они не чувствительны к тому, какие из остальных элементов должны быть малы. Понятно, что решение \hat{w} будет наилучшим образом отражать истинную модель w , если элементы каждой строки матрицы R будут уменьшаться при удалении от диагонали. В этом случае можно говорить о компактном разрешении.

Если собственные векторы обладают компактностью, то и матрица R дает компактное разрешение. Матрицы разрешения для приведенных в § 4 примеров собственных векторов будут иметь вид

$$R_1 = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \end{pmatrix}, R_2 = \begin{pmatrix} 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \end{pmatrix}.$$

Нетрудно показать, что матрица \mathbf{R} является инвариантной к любому преобразованию векторов \mathbf{v}_i , сохраняющему их ортонормальность. Действительно, пусть такое преобразование осуществляется матрицей $\mathbf{C} (r \times r)$:

$$\mathbf{v}^* = \mathbf{v}\mathbf{C},$$

и при этом $\mathbf{v}^{*'}\mathbf{v}^* = \mathbf{C}'\mathbf{v}'\mathbf{v}\mathbf{C} = \mathbf{C}'\mathbf{C} = \mathbf{I}_r$.

Поскольку матрица \mathbf{C} квадратная, то $\mathbf{C}' = \mathbf{C}^{-1}$ и $\mathbf{C}\mathbf{C}' = \mathbf{I}_r$. Новая матрица разрешения \mathbf{R}^* будет в этом случае иметь вид

$$\mathbf{R}^* = \mathbf{v}^*\mathbf{v}^{*'} = \mathbf{v}\mathbf{C}\mathbf{C}'\mathbf{v}' = \mathbf{v}\mathbf{v}',$$

т.е. она совпадает с матрицей \mathbf{R} .

Таким образом, по виду матрицы разрешения \mathbf{R} можно судить о том, существует ли возможность построить такие параметры, которые бы определились однозначно и в то же время давали бы компактное описание модели.

Для примера, рассмотренного в § 4, \mathbf{R} будет являться функцией двух переменных z и z_0 . На рис.36 изображен вид функции $R(z_0, z)$ для нескольких произвольно выбранных значений z_0 . Видно, что только для малых значений z_0 может быть достигнуто компактное разрешение. Иначе говоря, использованная выборка исходных данных (значения $\Delta g(x)$ в точках $x = 1, 2, \dots, 40$) позволяет разрешить особенности распределения плотности лишь в верхней части аномального тела, при этом масштаб деталей, которые могут быть разрешены, составляет $\sim 2-3$ отн. ед.

Описанный подход к построению оператора, позволяющего получить решение, характеризующееся наилучшим разрешением, несколько отличен от того, который использовался Дж.Бэйкусом и Ф.Гильбергом для построения локально сглаженного решения. Во-первых, он не учитывает условия нормировки сглаживающего ядра (т.е. строк матрицы \mathbf{R}), а во-вторых, как уже упоминалось, не удовлетворяет требованию компактности разрешения. Таким образом, различие заключается в выборе критерия дельтаобразности сглаживающего ядра. Если в качестве такого критерия принять тот, который описан в настоящем параграфе, т.е. близость строк матрицы разрешения (или сглаживающего ядра) к единичной матрице в метрике L_2 , то решением, удовлетворяющим такому критерию, оказывается нормальное решение.

Таким образом, решение, получаемое по методу Бэйкуса - Гильберта как одно из возможных (решение с минимальной нор-

мой), оказывается одновременно и таким, которое отвечает наилучшему локальному сглаживанию, но только в рамках описанного здесь критерия дельтообразности сглаживающего ядра.

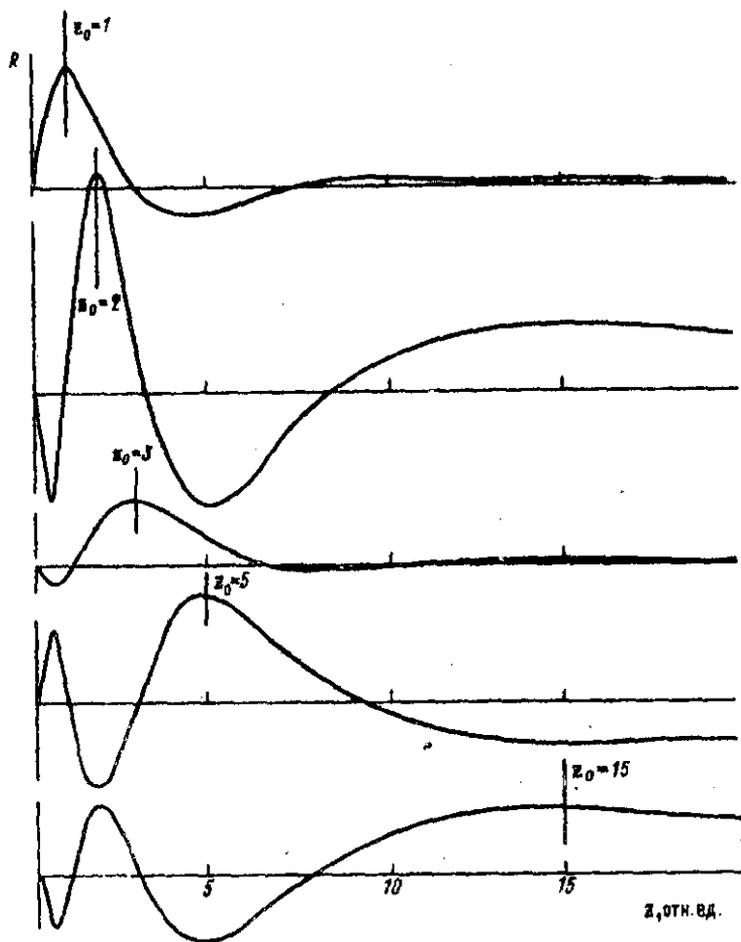


Рис. 36.

Для рассмотренного в § 4 примера на рис. 37 приведены сглаживающие функции $\mathcal{R}(x, z_0)$ (для $z_0 = 1$), построенные на основе двух разных критериев дельтообразности. Благодаря тому, что критерий минимизации эффективной ширины интервала сглаживания $s(z_0) = \int (x - z_0)^2 \mathcal{R}^2(x, z_0) dx$ (сплошная кривая) учитывает компактность разрешения, сглаживающее ядро $\mathcal{R}(x, z_0)$, соответствующее этому критерию, является более сосредоточенным



Рис. 37.

в окрестности точки x_0 , но его максимум оказывается сдвинутым относительно x_0 . В то же время ядро, построенное в соответствии с критерием близости к δ -функции, описанным в настоящем параграфе (пунктирная кривая), имеет максимум в точке x_0 , но оно характеризуется существенно отличными от нуля значениями в более широкой области.

§ 6. Распределение информации

Рассмотрим теперь вопрос об информационной содержательности исходных данных. Каждое из наблюдений несет определенную информацию о параметрах модели. При этом информация, представляемая разными наблюдениями, может не быть независимой. Простейшим примером этого является случай, когда какие-то из строк матрицы A равны или отличаются постоянным множителем. Какое-то из наблюдений может оказаться линейной комбинацией некоторых других. В общем, если ранг матрицы меньше числа уравнений, то имеющийся набор n данных можно заменить r

независимыми линейными комбинациями этих данных. Способ такой замены фактически был приведен в § 3. Он заключается в разложении вектора наблюдений \mathbf{y} по собственным векторам \mathbf{u}_i и \mathbf{u}_0 : $\mathbf{y} = \mathbf{U}\boldsymbol{\beta} + \mathbf{U}_0\beta_0$. Общее решение \mathbf{m} представляется суммой $\mathbf{V}\boldsymbol{\alpha} + \mathbf{V}_0\alpha_0$, в которой $\boldsymbol{\alpha}_0$ - произвольный вектор, а $\boldsymbol{\alpha}$ определяется из системы уравнений $\mathbf{A}\boldsymbol{\alpha} = \boldsymbol{\beta}$. Так что для построения общего решения системы (158) достаточно знать r значений β_j , являющихся компонентами вектора $\boldsymbol{\beta} = \mathbf{U}^T \mathbf{y}$. Эти r компонент можно рассматривать как "преобразованные" наблюдения. Каждая из них представляет собой результат усреднения исходных данных \mathbf{y} с весами, определяемыми собственными векторами \mathbf{u}_i . Легко видеть, что здесь имеется полная аналогия с преобразованием параметров.

Аналогично матрице разрешения вводится матрица распределения информации $\mathbf{S} = \mathbf{A}\mathbf{M}$. Как уже говорилось в § 1, эта матрица должна быть по возможности близка к единичной матрице \mathbf{I}_n . Покажем, что оператор обобщенного линейного обращения удовлетворяет и такому условию.

Запишем выражение для матрицы распределения информации, используя сингулярное разложение матрицы \mathbf{A} :

$$\mathbf{S} = \mathbf{A}\mathbf{M} = \mathbf{U}\mathbf{A}\mathbf{V}^T \mathbf{C} = \mathbf{U}\mathbf{C}$$

Определим неизвестную матрицу \mathbf{C} так, чтобы \mathbf{S} была наилучшим приближением к \mathbf{I}_n по методу наименьших квадратов. Это значит, что вектор-столбец матрицы \mathbf{C} должен быть решением по методу наименьших квадратов следующей системы уравнений:

$$\mathbf{U}\mathbf{c}_j = \boldsymbol{\delta}_j,$$

где $\boldsymbol{\delta}_j$ n -мерный вектор, имеющий компонентами δ_{jk} ($k = 1, 2, \dots, n$). Таким решением является

$$\mathbf{c}_j = (\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T \boldsymbol{\delta}_j = \mathbf{U}^T \boldsymbol{\delta}_j,$$

так что $\mathbf{C} = \mathbf{U}^T$ и, таким образом, $\mathbf{S} = \mathbf{U}\mathbf{U}^T$.

Если же в качестве оператора \mathbf{M} использовать оператор псевдообращения, то получим

$$\mathbf{S} = \mathbf{U}\mathbf{A}\mathbf{V}^T \mathbf{V}\mathbf{A}^{-1} \mathbf{U}^T = \mathbf{U}\mathbf{U}^T.$$

Итак, действительно, оператор псевдообращения дает наилучшее распределение информации.

Пример. Проиллюстрируем характер распределения информации на следующем простом примере. Пусть исходные данные -

годограф волны, отраженной от плоской горизонтальной границы, в некотором ограниченном интервале расстояний: $t_i = t(x_i)$, $0 \leq x_i \leq l$. Известны начальные приближения для мощности слоя h_0 и скорости в слое v_0 . Требуется оценить h и v по годографу.

Линеаризуем задачу, используя формулы, приведенные в § 3 гл. I (для случая $\alpha = 0$):

$$\delta t_i = \frac{2h_0}{v_0 \sqrt{1 + (x_i/2h_0)^2}} \frac{\delta h}{h_0} - \frac{2h_0}{v_0} \sqrt{1 + (x_i/2h_0)^2} \frac{\delta v}{v_0}.$$

Здесь δt_i - разность между наблюдаемым временем пробега в точке x_i и рассчитанным для начального приближения h_0, v_0 . В качестве искоемых параметров будем рассматривать безразмерные величины $\delta h/h_0$ и $\delta v/v_0$.

Пусть наблюдения проводятся в пяти точках профиля: $x/h_0 = 0; 0,25; 0,5; 0,75; 1,0$.

Таким образом, для определения двух неизвестных параметров имеется пять уравнений. Матрица распределения информации S в этом случае оказывается такой:

$$S = \begin{pmatrix} 0,45 & 0,40 & 0,25 & 0,06 & -0,15 \\ 0,40 & 0,35 & 0,24 & 0,09 & -0,08 \\ 0,25 & 0,24 & 0,21 & 0,17 & 0,12 \\ 0,06 & 0,09 & 0,17 & 0,28 & 0,40 \\ -0,15 & -0,08 & 0,12 & 0,40 & 0,71 \end{pmatrix}.$$

Элементы этой матрицы имеют следующий смысл: рассчитанные для полученной модели $(\delta h, \delta v)$ значения δt_i будут представлять результат усреднения исходных наблюдений с весами, равными элементам i -й строки этой матрицы.

Теперь посмотрим, каков будет характер собственных векторов m_i в данной задаче. Собственные векторы оказываются следующими:

$$m_1^T = [0,43 \quad 0,43 \quad 0,44 \quad 0,45 \quad 0,48], \\ m_2^T = [-0,51 \quad -0,41 \quad -0,12 \quad 0,26 \quad 0,70].$$

Как уже упоминалось, вся информация об искоемых параметрах содержится в двух величинах:

$$\beta_1 = \sum_{i=1}^5 u_{1i} \delta t_i \quad \text{и} \quad \beta_2 = \sum_{i=1}^5 u_{2i} \delta t_i.$$

Поскольку элементы первого из векторов почти одинаковы, то β_1 есть результат усреднения всех исходных данных δt_i с практи-

чески равными весами, и, таким образом, β_1 грубо характеризует среднее значение исходных данных. Попытаемся теперь интерпретировать β_2 . Если бы элементы u_2 составляли арифметическую прогрессию так, чтобы средний элемент равнялся нулю (например, $-0,6 \quad -0,3 \quad 0 \quad 0,3 \quad 0,6$), то осредненные с такими весами данные характеризовали бы средний наклон кривой $\delta t(x)$. Элементы вектора u_2 отличаются от такой прогрессии, хотя и не намного. Поэтому приближенно можно β_2 трактовать как "средний" наклон $\delta t(x)$ в интервале наблюдений.

§ 7. Регуляризация оператора обращения

Перейдем теперь к рассмотрению третьего требования, предъявляемого к оператору обращения \mathbf{H} . Наблюдения \mathbf{y} всегда содержат ошибки: даже если ошибки измерений сведены к нулю, всегда остаются ошибки округления при численных расчетах. Они вызывают соответствующие ошибки в оценке вектора параметров \mathbf{m} .

Если $\hat{\mathbf{m}} = \mathbf{H}\mathbf{y}$, то матрица ковариации оценок параметров $\mathbf{R}_{\hat{\mathbf{m}}}$ определяется соотношением

$$\mathbf{R}_{\hat{\mathbf{m}}} = \mathbf{H}\mathbf{R}_y\mathbf{H}^T,$$

где \mathbf{R}_y — ковариационная матрица ошибок наблюдений. Видно, что даже в случае малых ошибок в \mathbf{y} $\mathbf{R}_{\hat{\mathbf{m}}}$ может неограниченно возрастать, если возрастает оператор \mathbf{H} .

Оператор обобщенного линейного обращения определяется как $\mathbf{V}\mathbf{A}^{-1}\mathbf{U}^T$, где \mathbf{A}^{-1} — диагональная матрица, элементы которой — величины, обратные ненулевым сингулярным числам λ_j . Но часто спектр сингулярных чисел содержит очень малые значения λ , что приводит к большой дисперсии оценок параметров. Действительно, если для простоты рассмотреть случай, когда $\mathbf{R}_y = \sigma^2\mathbf{I}$, то $\mathbf{R}_{\hat{\mathbf{m}}} = \sigma^2\mathbf{H}\mathbf{H}^T = \sigma^2\mathbf{V}\mathbf{A}^{-2}\mathbf{V}^T$, откуда дисперсия параметра m_i определяется формулой

$$\sigma_{m_i}^2 = R_{\hat{\mathbf{m}}} \text{ii} = \sigma^2 \sum_j \left(\frac{V_{ij}}{\lambda_j} \right)^2.$$

Из этого выражения видно, что при наличии в спектре сингулярных чисел малых значений ошибки могут стать сколь угодно большими.

Таким образом, задача состоит в том, чтобы видоизменить оператор обращения \mathbf{H} так, чтобы уменьшить дисперсию оценки.

Эта задача тесно связана с задачей решения плохо обусловленных систем уравнений, с которыми часто приходится сталкиваться при оценке параметров по методу наименьших квадратов. Если сингулярные числа матрицы A малы, то тем более будут малы собственные значения матрицы $A^T A = V \Lambda^2 V^T$, что приводит к плохой обусловленности системы нормальных уравнений. Напомним, что при решении таких систем оценка относительной ошибки решения $\| \delta m \| / \| m \|$ определяется неравенством

$$\frac{\| \delta m \|}{\| m \|} \leq c \frac{\| \delta y \|}{\| y \|},$$

где c — так называемое число обусловленности матрицы системы, равное отношению наибольшего и наименьшего собственных значений [39]. Поэтому если собственные значения матрицы λ_i сильно различаются, то решение оказывается неустойчивым.

Существуют разные способы "загрубления" оператора M . Один, наиболее естественный и часто используемый, заключается в отбрасывании малых сингулярных чисел матрицы A . Малые значения λ_i округляются до нуля, тем самым уменьшается число ненулевых сингулярных значений и понижается ранг матрицы A . Пороговое значение определяется допустимой дисперсией решения. Если для m_i можно допустить дисперсию $\sigma_{m_i}^2$, то в спектре сингулярных чисел следует оставить только такие q значений, для которых еще выполняется неравенство

$$\sum_{j=1}^q \left(\frac{V_{ij}}{\lambda_j} \right)^2 < \frac{\sigma_{m_i}^2}{\sigma^2}.$$

Другой способ предложен Карривалином [34] и основан на стохастическом подходе. Идея этого подхода заключается в следующем.

Будем вместо уравнений $Am = y$ рассматривать уравнения

$$Am + e = y, \quad (172)$$

где e — случайная помеха, которая включает в себя как ошибки измерения величин y_i , так и ошибки округления. Наличие случайной помехи приводит к тому, что для m можно получить только оценку, удовлетворяющую некоторому критерию.

Вектор наблюдений y , определяемый уравнением (172), оказывается, как видно из этого уравнения, случайной величиной.

Задача ставится так: по выборочному значению $y = \hat{y}$ получить выборочную оценку $\hat{m} = L\hat{y}$, такую, чтобы минимизировать среднее по всем возможным реализациям y значение квадрата отклонения $m - Ly$. Для помехи ϵ считается известной ее ковариационная матрица $R_\epsilon = E(\epsilon\epsilon^T)$.

Минимизация среднего квадрата отклонения проводится аналогично тому, как это было сделано в § 5 при минимизации матрицы Q . Выберем произвольный вектор a , определим скалярное произведение отклонения $m - Ly$ и этого произвольного вектора и минимизируем среднее значение квадрата этой величины:

$$E\{[(m - Ly)^T a]^2\} = E\{a^T (m - Ly)(m - Ly)^T a\}. \quad (173)$$

Обозначим $E(mm^T) = W$. Эта матрица имеет смысл априорной ковариационной матрицы параметров. Очевидно, что параметры можно считать некоррелированными с помехой, так что $E(m\epsilon^T) = 0$. Тогда ковариационная матрица наблюдений может быть записана в виде

$$S = E(yy^T) = E[(Am + \epsilon)(Am + \epsilon)^T] = AWA^T + R_\epsilon.$$

Преобразуем выражение (173) аналогично тому, как это было сделано в § 5:

$$\begin{aligned} E\{a^T (mm^T - my^T L^T - Ly m^T + Ly y^T L^T) a\} &= \\ = a^T [W + LSL^T - E(my^T)L^T - LE(y m^T)] a &= \\ = a^T [L^T - S^{-1}E(y m^T)]^T S [L^T - S^{-1}E(y m^T)] a + & \\ + a^T [W - E(y m^T)^T S^{-1}E(y m^T)] a. & \end{aligned}$$

Матрица S является положительно определенной, поэтому минимум достигается в том случае, когда первое слагаемое обращается в нуль, т.е. $L^T = S^{-1}E(y m^T)$ или $L = E(m y^T) S^{-1}$. Поскольку

$$E(m y^T) = E[m(Am + \epsilon)^T] = WA^T,$$

$$\text{то} \quad L = WA^T (AWA^T + R_\epsilon)^{-1}.$$

Таким образом, для заданного выборочного значения \hat{y} — наилучшая оценка в указанном ранее смысле — определяется следующим образом:

$$\hat{m} = WA^T (AWA^T + R_\epsilon)^{-1} \hat{y}.$$

Если все параметры независимы и имеют одинаковую дисперсию (одинаковый вес), то $W = \sigma_m^2 I$ и $L = A^T (AA^T + R_\epsilon / \sigma_m^2)^{-1}$.

Наконец рассмотрим случай некоррелированной помехи и будем считать, что все наблюдения имеют одну и ту же дисперсию. В этом случае ковариационная матрица помехи $\mathbf{R}_e = \sigma^2 \mathbf{I}$ и

$$\mathbf{L} = \mathbf{A}^T [\mathbf{A}\mathbf{A}^T + (\sigma^2/\sigma_m^2)\mathbf{I}]^{-1}. \quad (174)$$

Точно в таком виде было получено выражение для оператора обращения в § 2 гл. IV, где рассматривался вопрос об оценке модели $m(\mathbf{r})$ по методу Вайкуса - Гильберта в случае, когда исходные данные имеют погрешности при тех же предположениях о независимости параметров и некоррелированности помехи. Но там вывод проводился для условий, когда параметров больше, чем исходных данных, и из всех возможных решений выбиралось такое, которое минимизировало норму искомого вектора. Приведенный же здесь вывод свободен от такого ограничения, так что результат его оказывается справедливым при любом соотношении числа искомого параметров и исходных данных. Нетрудно показать, что для случая, когда параметров меньше, чем наблюдений, решение, полученное с помощью оператора (174), является регуляризованным решением по методу наименьших квадратов.

Используем для \mathbf{A} сингулярное разложение:

$$\mathbf{L} = \mathbf{V}\mathbf{A}\mathbf{U}^T [\mathbf{U}\mathbf{A}^2\mathbf{U}^T + (\sigma^2/\sigma_m^2)\mathbf{I}_n]^{-1}.$$

Для преобразования правой части этого выражения используем соотношение $\mathbf{I} = \mathbf{A}^{-1}\mathbf{A}$, где \mathbf{A} - произвольная невырожденная матрица. Очевидно, что матрица $\mathbf{A}^2 + (\sigma^2/\sigma_m^2)\mathbf{I}_n$ является невырожденной, поэтому единичную матрицу \mathbf{I}_n можно представить в виде произведения

$$\mathbf{I}_n = [\mathbf{A}^2 + (\sigma^2/\sigma_m^2)\mathbf{I}_n]^{-1} [\mathbf{A}^2 + (\sigma^2/\sigma_m^2)\mathbf{I}_n].$$

Добавление единичной матрицы в качестве множителя не изменяет произведения матриц, поэтому можно записать:

$$\begin{aligned} \mathbf{L} &= \mathbf{V}\mathbf{A}\mathbf{I}_n\mathbf{U}^T [\mathbf{U}\mathbf{A}^2\mathbf{U}^T + (\sigma^2/\sigma_m^2)\mathbf{I}_n]^{-1} = \\ &= \mathbf{V}\mathbf{A}[\mathbf{A}^2 + (\sigma^2/\sigma_m^2)\mathbf{I}_n]^{-1} [\mathbf{A}^2\mathbf{U}^T + (\sigma^2/\sigma_m^2)\mathbf{U}^T] [\mathbf{U}\mathbf{A}^2\mathbf{U}^T + (\sigma^2/\sigma_m^2)\mathbf{I}_n]^{-1} = \\ &= \mathbf{V}\mathbf{A}[\mathbf{A}^2 + (\sigma^2/\sigma_m^2)\mathbf{I}_n]^{-1} [\mathbf{U}^T\mathbf{U}\mathbf{A}^2\mathbf{U}^T + (\sigma^2/\sigma_m^2)\mathbf{U}^T] [\mathbf{U}\mathbf{A}^2\mathbf{U}^T + (\sigma^2/\sigma_m^2)\mathbf{I}_n]^{-1} = \\ &= \mathbf{V}\mathbf{A}[\mathbf{A}^2 + (\sigma^2/\sigma_m^2)\mathbf{I}_n]^{-1}\mathbf{U}^T. \end{aligned} \quad (175)$$

Произведение матриц $\mathbf{A}[\mathbf{A}^2 + (\sigma^2/\sigma_m^2)\mathbf{I}_n]^{-1}$ является диагональной матрицей с элементами $\lambda_i(\lambda_i^2 + \sigma^2/\sigma_m^2)^{-1}$.

Таким образом, оператор L отличается от оператора псевдообращения только тем, что вместо матрицы A^{-1} с диагональными элементами λ_i^{-1} должна быть использована матрица, имеющая диагональные элементы $\lambda_i (\lambda_i^2 + \sigma^2 / \sigma_m^2)^{-1}$. Дисперсия помехи действует при этом как ограничитель на малые сингулярные числа. При малых λ_i вместо очень больших значений λ_i^{-1} мы будем иметь величины, стремящиеся к нулю как $\lambda_i (\sigma_m^2 / \sigma^2)$, а при больших значениях λ_i эти элементы будут практически совпадать с элементами λ_i^{-1} обратной матрицы A^{-1} .

Регуляризация решения системы нормальных уравнений в методе наименьших квадратов заключается в том, что вместо решения $m = (A^T A)^{-1} A^T y$ строится регуляризованное решение (см. § 2 гл. III), которое при $m_0 = 0$ будет иметь вид

$$m_{\text{рег}} = (A^T A + \alpha I)^{-1} A^T y.$$

Если использовать для представления матрицы A сингулярное разложение, то получим

$$\begin{aligned} m_{\text{рег}} &= (VA^2 V^T + \alpha I)^{-1} VAU^T y = \\ &= (VA^2 V^T + \alpha I)^{-1} VI_r AU^T y = \\ &= (VA^2 V^T + \alpha I)^{-1} (VA^2 + \alpha V)(\Lambda^2 + \alpha I_r)^{-1} AU^T y = \\ &= (VA^2 V^T + \alpha I)^{-1} (VA^2 V^T + \alpha V)(\Lambda^2 + \alpha I_r)^{-1} AU^T y = \\ &= V(\Lambda^2 + \alpha I_r)^{-1} AU^T y. \end{aligned}$$

Учитывая, что произведение диагональных матриц коммутативно, можно записать оператор построения регуляризованного решения в виде

$$M_{\text{рег}} = VA(\Lambda^2 + \alpha I)^{-1} U^T,$$

что совпадает с оператором L , определенным выражением (175). Параметр регуляризации α при этом, как и следовало ожидать, равен отношению дисперсии помехи к априорной дисперсии параметров.

§ 8. Учет взаимной корреляции наблюдений и связи между параметрами

До сих пор мы неявно предполагали, что искомые параметры m_i имеют одинаковый вес, а используемые наблюдения y_i независимы и их ошибки характеризуются одной и той же диспер-

сией. Это позволяет строить общее решение систем: $\mathbf{A}\mathbf{m} = \mathbf{y}$ так, чтобы оно с одной стороны минимизировало норму \mathbf{m} , т.е. величину $\mathbf{m}^T \mathbf{m}$, а с другой — средний квадрат невязки наблюдений, т.е. $\mathbf{e}^T \mathbf{e} = (\mathbf{A}\mathbf{m} - \mathbf{y})^T (\mathbf{A}\mathbf{m} - \mathbf{y})$.

В реальных геофизических задачах такая ситуация возникает крайне редко. Прежде всего параметры имеют разный вес, т.е. по-разному влияют на наблюдения. Очевидно, что в этом случае следует минимизировать не $\mathbf{m}^T \mathbf{m} = \sum_{i=1}^k m_i^2$, а взвешенную сумму квадратов $\sum_{i=1}^k w_i m_i^2$, где w_i — вес соответствующего параметра. Веса w_i должны вводиться еще и для того, чтобы привести параметры к безразмерному виду, поскольку в геофизических задачах они могут иметь разную размерность (мощности слоев, скорости и плотности в них и т.д.). В общем можно принять, что параметрам характеризуются некоторой ковариационной матрицей \mathbf{W} , которая в случае отсутствия связи между параметрами будет диагональной:

$$\mathbf{W} = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \sigma_k^2 \end{pmatrix},$$

при этом σ_i^2 можно рассматривать как дисперсию i -го параметра, соответственно $w_i = 1/\sigma_i^2$ представляет вес данного параметра.

Таким образом, в общем случае при поиске решения следует минимизировать не $\mathbf{m}^T \mathbf{m}$, а $\mathbf{m}^T \mathbf{W}^{-1} \mathbf{m}$.

Аналогично, в случае, когда наблюдения характеризуются ковариационной матрицей ошибок \mathbf{R}_y , то в соответствии с выводами гл. II оценка параметров должна производиться из условия минимизации величины $(\mathbf{A}\mathbf{m} - \mathbf{y})^T \mathbf{R}_y^{-1} (\mathbf{A}\mathbf{m} - \mathbf{y})$.

Чтобы привести эти случаи к тому, для которого был построен аппарат обобщенной линейной инверсии, достаточно выполнить преобразование параметров и наблюдений следующим образом. Отметим прежде всего, что матрицы ковариаций \mathbf{R}_y и \mathbf{W} являются симметричными и положительно определенными, и, таким образом, существуют матрицы $\mathbf{R}_y^{-1/2}$ и $\mathbf{S}^{-1/2}$. Произведем преобразования:

$$\mathbf{m}' = \mathbf{W}^{-1/2} \mathbf{m}, \quad \mathbf{y}' = \mathbf{R}_y^{-1/2} \mathbf{y}, \quad \mathbf{A}' = \mathbf{R}_y^{-1/2} \mathbf{A} \mathbf{W}^{1/2}.$$

Тогда новые параметры \mathbf{m}' будут связаны с преобразованными наблюдениями \mathbf{y}' уравнениями типа (158): $\mathbf{A}' \mathbf{m}' = \mathbf{y}'$. Решение этой

системы уравнений, получаемое по методу псевдообращения, будет минимизировать

$$\mathbf{m}'^T \mathbf{m}' = \mathbf{m}'^T (\mathbf{W}^{-1/2})^T \mathbf{W}^{-1/2} \mathbf{m} = \mathbf{m}'^T \mathbf{W}^{-1} \mathbf{m}$$

$$\text{и } \boldsymbol{\varepsilon}'^T \boldsymbol{\varepsilon}' = \boldsymbol{\varepsilon}'^T \mathbf{R}_y^{-1} \boldsymbol{\varepsilon}' .$$

Таким образом, простая замена матрицы \mathbf{A} в исходных уравнениях приводит к тому, что решение будет отвечать взвешенному, т.е. такому, для которого дисперсии всех параметров одинаковы, а параметры независимы.

Обычно бывает известна матрица ковариации ошибок наблюдений \mathbf{R}_y , а ковариационная матрица параметров может быть построена с точностью до постоянного множителя: σ_m^2 . При этом ковариационная матрица "новых" наблюдений будет единичной \mathbf{I} , а ковариационная матрица "новых" параметров - $\sigma_m^2 \mathbf{I}$. Величину σ_m^2 можно рассматривать как априорную дисперсию параметров (см. § 7).

По построенному решению $\hat{\mathbf{m}}'$, в свою очередь, можно оценить остаточную дисперсию:

$$s^2 = (\mathbf{A}' \hat{\mathbf{m}}' - \mathbf{y}')^T (\mathbf{A}' \hat{\mathbf{m}}' - \mathbf{y}')$$

и тем самым проверить, правильно ли была выбрана дисперсия исходных наблюдений. Если окажется, что $s^2 \ll 1$, то исходная дисперсия была, по-видимому, взята слишком большой по сравнению с внутренней согласованностью данных. Если $s^2 \gg 1$, то либо оцененная по наблюдениям дисперсия была сильно занижена, либо модель неадекватна исходным наблюдениям. В последнем случае отклонения ε_i должны содержать систематическую ошибку, которую обычно бывает легко обнаружить.

ЗАКЛЮЧЕНИЕ

После прочтения этой книги у читателей естественно возникает вопрос: а какой же из описанных методов предпочтительнее для решения той или иной конкретной обратной задачи? Ответ на этот вопрос зависит от того, какие используются данные и какие имеются априорные сведения о среде.

Чтобы решить, какой следует использовать метод, надо прежде всего провести анализ имеющейся априорной информации с точки зрения того, можно ли линеаризовать поставленную задачу, иначе говоря, достаточно ли априорных данных, чтобы построить хорошее начальное приближение для искомой модели среды. Следует подчеркнуть, что к настоящему времени большинство обратных задач геофизики допускает линеаризацию. Если рассматриваются глобальные геофизические задачи, связанные с исследованием распределения той или иной физической величины в Земле в целом или в отдельных глубинных областях, то благодаря тому, что строение Земли уже достаточно изучено, начальное приближение может быть всегда взято более или менее близким к искомому решению. В региональных исследованиях, особенно при изучении верхней толщи, нельзя основываться на подобных соображениях, так как строение среды на небольших глубинах может сильно меняться от района к району. Однако обычно до проведения геофизических исследований выполняются разведочные работы (геологическая съемка, скважинные исследования и т.п.), которые предоставляют необходимую априорную информацию.

Конечно, встречаются и такие случаи, когда априорных сведений недостаточно для того, чтобы линеаризовать задачу. В этом случае необходим предварительный анализ исходных данных с тем, чтобы иметь возможность наиболее разумно параметризовать модель. При этом у исследователя должны существовать определенные представления о том, в какой степени те или иные параметры среды влияют на используемые характеристики геофизического поля. Например, если для решения обратной задачи исполь-

зуются данные МТЗ в ограниченном интервале периодов, причем данных не очень много, а их разброс таков, что кривая $\varphi_k(\tau)$, аппроксимирующая эти данные, является монотонной и плавной в этом интервале периодов, то не имеет смысла аппроксимировать геоэлектрический разрез большим числом слоев, поскольку параметры этих слоев окажутся сильно связанными. Если подобные представления о связи используемых данных с моделью среды отсутствуют или являются недостаточно определенными, то можно рекомендовать производить выбор параметров путем построения информативных областей, как было описано в § 8 гл. II. Анализ информативных областей дает возможность более или менее обоснованно подойти к выбору параметров.

Оценку выбранных параметров следует производить методом максимального правдоподобия или нелинейным методом наименьших квадратов, сводящимся к минимизации некоторой целевой функции. При этом если исходных данных много (по отношению к числу оцениваемых параметров) и оценку параметров удается получить с высокой точностью, то погрешность полученной оценки можно характеризовать ковариационной матрицей параметров. Если же решение оказывается неустойчивым (или даже неоднозначным), то для характеристики погрешности необходимо в простейшем случае параметров оконтурить область всех допустимых решений, т.е. таких, которые бы удовлетворяли исходным данным в соответствии с некоторым пороговым условием.

Перейдем теперь к рассмотрению более распространенного случая, когда априорные данные таковы, что допускают линеаризацию задачи. Априорная модель может быть представлена либо набором параметров, что соответствует слоистой среде, либо в виде непрерывной или кусочно-непрерывной функции $m(r)$. Искомые при этом будут поправки к заданной априорно модели. В случае, когда априорная модель представлена набором параметров m_0 , решение можно искать либо в рамках такой же модели, т.е. в виде поправок Δm_i к исходным значениям параметров, либо в форме функциональных зависимостей $m_i(r)$ в каждом из слоев.

В первом случае, если число исходных данных больше числа искомых параметров, решение определяется по методу наименьших квадратов. В общем же, если матрица системы, нормальных

уравнений оказывается вырожденной или число данных меньше числа параметров, для нахождения решения может быть использован метод псевдообращения, который включает в себя метод наименьших квадратов как частный случай.

Второй случай возможен тогда, когда есть основания считать, что используемые данные могут дать разрешение, лучшее, чем то, в соответствии с которым построена априорная модель. При этом для построения решения может использоваться метод Бэйкуса - Гильберта. Опять же, в общем случае, если такое решение оказывается невозможно построить из-за того, что некоторые из данных несут одну и ту же информацию о среде, следует использовать для построения решения метод псевдообращения, частным случаем которого является метод Бэйкуса - Гильберта. Сказанное справедливо, если и априорная модель, и поправка к ней представляются в виде непрерывной функции или задаются очень большим числом параметров, превышающим число исходных данных.

Таким образом, видно, что наиболее общим в линейризованных задачах является метод псевдообращения, поскольку к нему можно свести построение решения как для модели заданной конечным числом параметров, так и для случая, когда модель представляется в виде неизвестной функции координат. При этом метод псевдообращения не накладывает никаких ограничений на соотношение между числом параметров и числом исходных данных и позволяет строить решение как в случае, когда параметры полностью неразрешены (система нормальных уравнений оказывается вырожденной), так и в случае, когда разные данные характеризуются одной и той же информативностью, что приводит к вырожденности матрицы B (§ 2 гл. IV) или матрицы S (§ 3 гл. IV). Этим объясняется то, что метод псевдообращения завоевывает все большее признание среди геофизиков, занимающихся решением интерпретационных задач. Заметим, что хотя этот метод известен в линейной алгебре достаточно давно, он долгое время, несмотря на его широкие возможности, не пользовался даже среди математиков тем вниманием, которого заслуживает. Можно полагать, что в скором времени метод псевдообращения займет главенствующее положение среди методов решения обратных задач.

Мы здесь нигде не упомянули о методе регуляризации. Как следует из изложенного в гл. III, его использование определяется тем, насколько устойчиво решение по какому-либо из указанных ранее методов. Если решение оказывается неустойчивым, то при любом подходе следует использовать регуляризацию.

Отметим, что в рамках всех описанных методов имеется возможность не только строить решение по заданной выборке данных наблюдений, но и анализировать погрешность полученного решения, а также разрешающую способность и информативность исходных данных, и на основании такого анализа определять, как следует изменить данные (дополнить выборку, увеличить точность наблюдений), чтобы получить точность и разрешение, требуемые для целей того или иного исследования.

УКАЗАТЕЛЬ ЛИТЕРАТУРЫ

- 1* Азбель И.И., Кейлис-Борок В.И., Яновская Т.Б. Методика совместной интерпретации голографов и амплитудных кривых при изучении верхней мантии. - В кн.: Вычислительная сейсмология, вып.2, М., 1966, с.3-45.
2. Алберт А. Регрессия, псевдоинверсия и рекуррентное оценивание, М., 1977, 223 с.
3. Аоки М. Введение в методы оптимизации. М., 1977, 343 с.
- 4* Ашманов С.А. Линейное программирование. М., 1981, 304 с.
5. Бембель Р.М., Гольдин С.В. Об использовании априорной информации при обработке сейсмических голографов. - Труды ЗапСибНИГНИ, 1970, вып.36, с.5-27.
6. Бембель Р.М., Гольдин С.В. О некоторых способах регуляризации метода наименьших квадратов при обработке геофизических наблюдений. - Труды ЗапСибНИГНИ, 1970, вып.36, с.86-102.
- 7* Бердичевский М.И., Адамов М.С. Интерпретация аномалий переменного электромагнитного поля Земли. М., 1981, 327 с.
8. Брандт Э. Статистические методы анализа наблюдений. М., 1975, 312 с.
9. Вальд А. Статистические решающие функции. - В кн.: Позиционные игры. М., 1967, с.300-522.
10. Ван дер Варден Б.Л. Математическая статистика. М., 1960, 434 с.
11. Волков И.Е., Яновская Т.Б. О характере распределения амплитуд P -волн. - В кн.: Вопросы геофизики, вып.22, Л., 1972, с.137-147.
12. Гельфанд И.М., Цетлин М.Л. Метод обротов. - Ученые зап. кавк. ун-та, 1962, т.17, № 1, с.3-25.
- 13* Гольдин С.В. Интерпретация данных сейсмического метода отраженных волн. М., 1979, 344 с.
- 14* Гольдман Ф.М. Статистические модели интерпретации. М., 1971, 328 с.
15. Гольдман Ф.М. Статистическая теория интерпретации геофизических полей. - Изв. АН СССР, сер. Физика Земли, 1975, № 1, с.49-53.
- 16* Ибрагимов И.А., Хасьминский Р.З. Асимптотическая теория оценивания. М., 1979, 527 с.

17. Калинина Т.Б., Гольцман Ф.М. Статистический алгоритм интерпретации и его применение для решения обратной задачи магниторазведки. – Изв. АН СССР, сер. Физика Земли, 1967, № 7, с.50–60.
18. Калинина Т.Б., Гольцман Ф.М. Аналитическое сопоставление методов геофизической интерпретации. – Изв. АН СССР, сер. Физика Земли, 1971, № 5, с.32–42.
19. Калинина Т.Б. Учет априорной информации при определении параметров намагниченных объектов. – В кн.: Метод. разведочной геофизики. Л., 1974, вып.22, с.10–19.
20. Калиткин Н.И. Численные методы. М., 1978. 512 с.
21. Кендалл Д.Дж., Стюарт А. Статистические выводы и связи. М., 1973. 512 с.
- 22* Ковтун А.А. Использование естественного электромагнитного поля при изучении электропроводности Земли. Л., 1980. 196 с.
- 23* Крамер Г. Математические методы статистики. М., 1975. 648 с.
24. Лаврентьев М.М. О некоторых некорректных задачах математической физики. Новосибирск, 1962. 92 с.
25. Левшин А.Л. Интерпретация данных о дисперсии поверхностных волн с помощью билогарифмических палеток теоретических дисперсионных кривых. – Изв. АН СССР, сер. геофизич., 1960, № 2, с.318–322.
- 26* Левшин А.Л. Поверхностные и каналовые сейсмические волны. М., 1973. 176 с.
27. Матвеев Б.К. Интерпретация электромагнитных зондирований. М., 1974. 232 с.
28. Некоторые прямые и обратные задачи сейсмологии. М., 1968, № 4, с.3–14.
29. Порохова Л.Н., Васильковский Б.В. Исследование погрешностей параметров при решении обратной задачи электромагнитных зондирований. – В кн.: Вопросы геофизики. Вып.22.Л., 1972, с.232–243.
30. Порохова Л.Н., Мальгинов В.В. Комплексная интерпретация геофизических зондирований. – Вестн. Ленингр. ун-та, 1982, № 46, с.113–115.
31. Порядков И.И., Яновский В.И. Применение статистических методов к интерпретации фоновых кривых КТЗ. – В кн.: Математические методы в геоэлектрике. М.: 1982, с.96–110.

32. Пылаев А.М. Руководство по интерпретации вертикальных электрических зондирований. М., 1948. 168 с.
- 33.* Рао С.Р. Линейные статистические методы и их приложения. М., 1968. 548 с.
- 34.* Саваренский Е.Ф., Гласко В.Б. Вопросы интерпретации экспериментальных данных на примере слоистых сред и концепция регуляризации А.Н.Тихонова. - Геофизический журнал, 1980, т.3, № 1, с.30-40.
35. Определение параметров земной коры Русской платформы по дисперсии поверхностных волн / Е.Ф.Саваренский, В.Б.Гласко, Г.Н.Мартакуо, Л.Н.Паликова, Н.П.Хмелевский. - Изв. АН СССР, сер. Физика Земли, 1974, № 5, с.86-95.
- 36.* Старостенко В.И. Устойчивые численные методы в задачах гравиметрии. Киев, 1978. 228 с.
- 37.* Статистическая интерпретация геофизических данных / Под ред. Ф.М.Гольцмана. Л., 1981. 225 с.
38. Страхов В.И. О решении некорректных задач магнито- и гравиметрии, представляемых интегральными уравнениями типа свертки. - Изв. АН СССР, сер. Физика Земли, 1967, № 4, с.36-54; № 5, с.33-53.
- 39.* Стрэнг Г. Линейная алгебра и ее приложения. М., 1980. 454 с.
- 40.* Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. М., 1974. 223 с.
41. Тихонов А.Н., Гласко В.Б., Кулик Н.И. Регуляризирующие алгоритмы для нелинейных задач и обратная задача магнитотеллурического зондирования. - В кн.: Вычислительные методы и программирование. М., 1973, с.158-173.
42. Тихонов А.Н., Гласко В.Б., Криксин Ю.А. К вопросу о квазиоптимальном выборе регуляризованного приближения. - Докл. АН СССР, 1979, т.248, № 3, с.531-535.
- 43.* Уайлд Дж.Д. Методы поиска экстремума. М., 1967. 267 с.
44. Уилкс С. Математическая статистика. М., 1967. 623 с.
45. Форсайт Дж., Малькольм И., Моулер К. Матричные методы математических вычислений. М., 1960. 280 с.
46. Халфин Л.А. Информационная теория интерпретации геофизических наблюдений. - Докл. АН СССР, 1958, т.122, № 6, с.1007-1010.

- 47.* Химмельблау Д. Прикладное нелинейное программирование. М., 1975. 534 с.
- 48.* Худсон Д. Статистика для физиков. М., 1970. 296 с.
49. Яновская Т.Б. Вычисление скоростных разрезов верхней мантии как обратная математическая задача. - Изв. АН СССР, сер. геофизич., 1963, № 8, с.1171-1177.
- 50.* Backus G., Gilbert F. Numerical application of formalism for geophysical inverse problems. - Geophys. J. Roy. Astr. Soc., 1967, vol.13, p.247-276.
- 51.* Backus G., Gilbert F. The resolving power of gross Earth data. - Geophys. J. Roy. Astr. Soc., 1968, vol.16, p.169-205.
52. Backus G., Gilbert F. Uniqueness in inversion of inaccurate gross Earth data. - Phil. Trans. Roy. Soc., 1969, vol. 266A, N 1173, p.123-192.
53. Cuer M. Application of linear programming to the inverse gravity or magnetic problem basic numerical technique. - In: Lecture notes in physics. Vol. 85: Applied inverse problems. Berlin - Heidelberg - New York, 1978, p.103-110.
54. Franklin J.N. Well-posed stochastic extensions of ill-posed linear problems. - J. Math. Analysis Appl., 1970, vol.31, p.682-716.
55. Gilbert F., Backus G. Approximate solutions to the inverse normal mode problems. - Bull. Seism. Soc. Amer., 1968, vol. 58, N 1, p.103-131.
56. Goffing L. Constrained optimization of gravimetric density models by use of the simplex algorithm. - Earth evolution sciences, 1981, vol.1, p.31-33.
57. Hooke R., Jeeven T.A. "Direct search" solution of numerical and statistical problems. - J. Assoc. Comp. Math., 1961, vol.8, N 2, p.221-229.
- 58.* Jackson D.D. Interpretation of inaccurate, insufficient and inconsistent data. - Geophys. J., 1972, vol.28, p.97-110.
59. Jefreys H. On travel times in seismology. - Publ. du bureau central seism. International, série A, fasc.14, 1936, 97 p.
- 60.* Olsson E., Gilbert F. Inversion and inference for teleseismic data. - Math. Comp. Phys., 1972, vol.12, p.231-266.

61. Jupp D.L.B., Vozoff K. Resolving anisotropy in layered media by joint inversion. - Geophys. Prosp., 1977, N 25, p.460-470.
62. Kellis-Borok V.-I., Yanovskaya T.B. Inverse problem of seismology (structural review). - Geophys. J. Roy. Astr.Soc., 1967, vol.13, p.223-234.
63. Levenberg K.A. Method for the solution of certain nonlinear problems in least squares. - Quart. Appl. Math., 1944, vol.2, p.164-168.
64. Marquardt D.W. An algorithm for least-squares estimation of nonlinear parameters. - J. Soc. Ind. Appl. Math., 1963, vol.11, p.431-441
65. Moore E.H. General analysis. I. - Memoirs Phil. Soc., 1935, vol.1, p.1-231.
66. Park r R.L. The inverse problem of electrical conductivity in the mantle. - Geophys. J. Roy. Astr. Soc., 1970, vol.22, p.121-138.
67. Parker R... Inverse theory with grossly inadequate data. - Geophys. J. Roy. str. Soc., 1972, vol.29, p.123-138.
68. Penrose R. A generalized inverse for matrices. - Proc.Camb. Phil. Soc., 1955, vol.51, p.406-413.
69. Press F. Earth models consistent with geophysical data. - Phys. Earth and Planet. Inter., 1970, vol.3, p.3-22.
70. Wiggins R.A. The general linear inverse problem: implication of surface waves and free oscillations on Earth structure. - Rev. Geophys. and Space Phys., 1972, vol.10, N 1, p.251-285.

ОГЛАВЛЕНИЕ

Предисловие	3
Г л а в а I. ВВЕДЕНИЕ	7
§ 1. Понятие об обратных задачах геофизики	7
§ 2. Математическая постановка обратных задач ...	9
§ 3. Линеаризация оператора прямой задачи	14
§ 4. Некорректность обратных задач	19
§ 5. Параметризация модели	25
§ 6. Классификация методов решения обратных задач	27
Г л а в а II. ОПТИМИЗАЦИОННЫЕ МЕТОДЫ	36
§ 1. Общие принципы статистического оценивания ..	36
§ 2. Статистическая постановка обратных геофизиче- ских задач	50
§ 3. Функции распределения данных геофизических наблюдений	55
§ 4. Оценивание параметров методом максимального правдоподобия	62
§ 5. Оценивание параметров методом наименьших квадратов	70
§ 6. Оценка погрешности решения	77
§ 7. Некоторые методы поиска экстремума	88
§ 8. Выбор параметров	97
§ 9. Оценка параметров при наличии априорной ин- формации	105
§ 10. Равномерное распределение ошибок наблюдений	113
Г л а в а III. МЕТОД РЕГУЛЯРИЗАЦИИ	117
§ 1. Понятие регуляризирующего оператора	117
§ 2. Построение регуляризованного решения линей- ной системы	120
	209

§ 3. Регуляризованное решение интегрального уравнения первого рода	123
§ 4. Выбор параметра регуляризации	126
§ 5. Использование метода регуляризации при решении обратных задач геофизики	129
§ 6. Метод регуляризации как способ учета априорной информации	133
Г л а в а IV. МЕТОД БЭЖКУСА - ГИЛЬБЕРТА	137
§ 1. Постановка задачи	137
§ 2. Построение одного из возможных решений	140
§ 3. Сглаженное решение и разрешающая способность данных	149
§ 4. Разрешающая способность данных, содержащих погрешности	156
§ 5. Нахождение одного из возможных решений в случае, когда ядра не являются квадратично интегрируемыми	160
§ 6. Построение сглаженного решения в случае, когда ядра не квадратично интегрируемы	167
Г л а в а V. МЕТОД ПСЕВДООБРАЩЕНИЯ	173
§ 1. Обобщенная постановка линейаризованных обратных задач	173
§ 2. Сингулярное разложение	176
§ 3. Свойства решения, получаемого путем псевдообращения	177
§ 4. Определение оптимальной параметризации	180
§ 5. Разрешающая способность данных	185
§ 6. Распределение информации	190
§ 7. Регуляризация оператора обращения	193
§ 8. Учет взаимной корреляции наблюдений и связи между параметрами	197
Заключение	200
Указатель литературы	204

ИБ № 1746

Татьяна Борисовна Яновская, Людмила Николаевна Порохова

ОБРАТНЫЕ ЗАДАЧИ ГЕОФИЗИКИ
Учебное пособие

Редактор Т.В.Мизникова

Технический редактор Е.И.Веселова

Корректоры К.Я.Евнина, Н.П.Подгородецкая

Подписано в печать 12.08.83. М-24343 Формат 60x84 1/16.
Бумага тип. № 2. Печать офсетная. Усл.печ.л. 12,32.
Усл. кр.-отг. 12,49. Уг.-взд.л. 11,06. Заказ 281
Тираж 1683 экз. Цена 35 коп.
Издательство ЛГУ им.А.А.Жданова
199164. Ленинград, В-164, Университетская наб., 7/9

Фотофотетная лаборатория ЛГУ
наб.Макарова, 6

ИЗДАТЕЛЬСТВО ЛЕНИНГРАДСКОГО УНИВЕРСИТЕТА В 1984 ГОДУ
ВЫПУСТИТ В СВЕТ УЧЕБНОЕ ПОСОБИЕ:

Б у л д ы р е в В.В., П а в л о в Б.С. Линейная алгебра и функции многих переменных. - 18 печ.лстов.

В пособии (выпускаемом с рекомендательным грифом Минвуза СССР), состоящем из двух тесно связанных частей "Линейная алгебра" и "Функции многих переменных", рассматриваются вопросы теории линейных преобразований, теории операторов в линейных пространствах, теории интеграла Лебега, тензоры, задачи об экстремуме, дифференциальная геометрия.

Для студентов физических и математических факультетов университетов и педагогических вузов.

Заказы на книгу направляйте по адресу: 191186, Ленинград, Невский пр., д.28, Магазин № I, "Дом книги", отдел "Книга - почтой".

