

*для заметок*

*Юрий Михайлович Тимофеев*  
*Александр Викторович Поляков*

МАТЕМАТИЧЕСКИЕ АСПЕКТЫ РЕШЕНИЯ ОБРАТНЫХ ЗАДАЧ  
АТМОСФЕРНОЙ ОПТИКИ

*Учебное пособие*

Зав. редакцией Г. И. Чередниченко  
Редактор Ф. С. Бастиан  
Технический редактор Л. Н. Иванова  
Компьютерный набор и верстка А. В. Поляков

*Лицензия ЛР N 040050 от 15.08.96*

Подписано к печати с оригинал-макета 26.01.2001  
Печать офсетная. Ф-т 60x84/16. Усл. печ. л. 11,62  
Уч.-изд. л. 11,84  
Тираж 300 экз. Заказ № .

РОПИ Издательства С.-Петербургского университета.  
199034, С.-Петербург, Университетская наб., 7/9.

ЦОП типографии Издательства С.-Петербургского университета  
199034, С.-Петербург, наб. Макарова, 6.

**Ю.М. Тимофеев, А.В. Поляков**

**МАТЕМАТИЧЕСКИЕ АСПЕКТЫ  
РЕШЕНИЯ ОБРАТНЫХ ЗАДАЧ  
АТМОСФЕРНОЙ ОПТИКИ**

**Санкт-Петербург**

**2001**

Санкт-Петербургский государственный университет

Ю.М. Тимофеев, А.В. Поляков

**МАТЕМАТИЧЕСКИЕ АСПЕКТЫ  
РЕШЕНИЯ ОБРАТНЫХ ЗАДАЧ  
АТМОСФЕРНОЙ ОПТИКИ**

*Учебное пособие*

Издательство Санкт-Петербургского университета

2001

ББК 26.233  
Т41

Рецензенты: докт. техн. наук, проф. *В.В. Глазунов* (СПГТИ),  
докт. физ.-мат наук, проф. *В.Н. Троян* (СПбГУ)

*Печатается по постановлению  
Редакционно-издательского совета  
Санкт-Петербургского государственного университета*

**Тимофеев Ю.М., Поляков А.В.**

Т41 Математические аспекты решения обратных задач  
атмосферной оптики: Учеб. пособие. — СПб.: Изд-во  
С.-Петерб. ун-та, 2001. — 188 с.  
ISBN 5-288-02810-9

Учебное пособие предназначено для студентов, бакалавров, магистрантов и аспирантов, изучающих методы решения обратных задач атмосферной оптики в рамках специального курса. Основное внимание уделяется вопросам, связанным с некорректностью обратных задач. Рассматриваются различные подходы к решению некорректных обратных задач и анализу получаемого решения. Дается понятие об информативности косвенных измерений и оптимизации дистанционных спектроскопических измерений.

**ББК 26.233**

ISBN 5-288-02810-9

© Ю.М. Тимофеев,  
А.В. Поляков, 2001  
© Издательство  
С.-Петербургского  
университета, 2001

## **Предисловие**

Курс лекций и настоящее пособие «МАТЕМАТИЧЕСКИЕ АСПЕКТЫ РЕШЕНИЯ ОБРАТНЫХ ЗАДАЧ АТМОСФЕРНОЙ ОПТИКИ» завершают важный этап подготовки студентов (бакалавров, магистрантов и аспирантов) физического факультета СПбГУ по специальности «Мониторинг параметров окружающей среды». Предшествующие два курса лекций:

«Теоретические основы атмосферной оптики»;

«Дистанционные методы измерений параметров атмосферы и поверхности»,

входят в программу обучения бакалавров физического факультета. Первый из них дает основные сведения об атмосфере Земли (газовый состав, термический режим, атмосферные аэрозоли и облака), о механизмах трансформации электромагнитного излучения (ослабление, рассеяние, поглощение), начальные сведения из теории переноса излучения, радиационной энергетике планеты Земля и т. д. Второй курс лекций подробно рассматривает различные (пассивные и активные) дистанционные методы измерений параметров атмосферы и поверхности.

Дистанционные методы измерений включают в себя помимо специализированной аппаратуры и физико-математической модели переноса излучения также и алгоритмы интерпретации. В связи с тем что все обратные задачи атмосферной оптики сводятся к решению некорректных (в классическом смысле) задач математической физики, в настоящем пособии излагаются основные сведения о специфике и современных методах их решения. В пособии также кратко рассмотрены вопросы информативности дистанционных косвенных измерений и оптимизации условий измерений.

Заметим, что обратные задачи атмосферной оптики могут быть сформулированы как на языке непрерывных функций (относительно непрерывных высотных зависимостей атмосферных параметров), так и на языке дискретных объектов – конечномерных векторов в евклидовом пространстве (относительно, например, табличного представления тех же атмосферных параметров на сетке высот или давлений). Поскольку целью этого пособия является не строгое последовательное изложение математической теории, а знакомство читателя с основными подходами, методами и результатами в области решения некорректных обратных задач в приложении к

дистанционному зондированию атмосферы, мы сознательно придерживаемся смешанного изложения. Практически все основные факты и результаты, приведенные в пособии, могут рассматриваться с любой из этих двух точек зрения.

Структура пособия ориентирована на постепенное знакомство с предметом.

В первой главе вводятся понятия прямых и обратных задач. Рассматриваются примеры математической формулировки ряда наиболее часто встречающихся обратных задач атмосферной оптики и дана их классификация. Поясняются вопросы линеаризации нелинейных задач атмосферной оптики, дается понятие вариационной производной.

Во второй главе рассматриваются общие понятия и проблемы, связанные с решением некорректных обратных задач. На простых примерах показана неустойчивость подобных задач, рассматриваются простейшие подходы, использованные для их решения.

Третья глава пособия посвящена современным методам решения и исследования некорректных обратных задач. В ней рассматриваются различные способы доопределения задачи, позволяющие получать устойчивые решения, имеющие физический смысл, и связанные с этими способами подходы к решению некорректных задач.

В четвертой главе рассматриваются вопросы, связанные с оптимизацией и информативностью дистанционных измерений.

В Приложении приводятся сведения из различных разделов математики, необходимые для понимания материала пособия.

В конце некоторых разделов мы привели упражнения, которые смогут помочь в понимании рассмотренных вопросов.

Настоящее пособие может быть полезным для студентов, бакалавров, магистрантов и аспирантов, специализирующихся в области физики атмосферы, метеорологии, климатологии, мониторинга параметров окружающей среды, экологии.

При подготовке пособия мы использовали в первую очередь главы 2 и 3 пособия [3], подготовленные В.В. Розановым, и работы В.П. Козлова, а также ряд публикаций российских и зарубежных ученых. Для более глубокого изучения вопросов, связанных с решением обратных задач мы можем рекомендовать книги [6, 18, 19, 24].

## Глава 1

# Общая характеристика обратных задач атмосферной оптики

### 1.1. Введение

*Прямые задачи* ориентированы по ходу *причинно-следственной связи*, т. е. представляют собой задачи отыскания *следствия* известных *причин*. Таковыми являются задачи определения полей во времени и пространстве при заданных источниках, вычисление реакции прибора по известному сигналу на входе и т. д.

*Обратные задачи* связаны с *обращением причинно-следственной связи*, т. е. отысканием неизвестных *причин* по известным *следствиям* – определение характеристик источников поля в некоторых точках или областях пространства по результатам измерения параметров поля, восстановление входного сигнала по реакции на выходе прибора и т. д.

Обратные задачи возникают, как правило, как задачи интерпретации тех или иных наблюдений. Собственно говоря, любая задача измерений тех или иных характеристик физического объекта есть обратная задача в смысле приведенного выше определения. Однако принято применять термин «обратная задача» к достаточно сложным задачам интерпретации, когда либо одновременно и независимым образом измеряется много параметров физического объекта, либо число этих параметров вообще неопределенно велико. *Обратные задачи математической физики* являются одним из классов *некорректных задач*, т. е. задач, не удовлетворяющих тем или иным условиям *корректности*, например корректности по Ж. Адамару. Первоначально термин «некорректные задачи» использовался в широком смысле и обозначал задачи, неправильно сформулированные, т. е. неправильно поставленные. В дальнейшем развитие подходов к решению определенных классов задач, ранее

рассматривавшихся как некорректные, привело к приданию этому термину четкого математического смысла (см. раздел 2.3).

Строго говоря, некорректными могут быть задачи, в которых измеряется неопределенно большое число параметров физического объекта. Однако и при конечном (но достаточно большом) числе параметров могут наблюдаться характерные особенности некорректных задач, связанные с тем, что информация, содержащаяся в измерениях, недостаточна для определения всех параметров. Способы решения в условиях такой информационной недоопределенности, по существу, одинаковы для задач обоих типов.

Примерами некорректных задач являются:

- – обратная задача теории потенциала, связанная с интерпретацией гравиметрических наблюдений в геологоразведке;
- – обратные задачи теплопроводности и диффузии;
- – обратные задачи теории переноса излучения, т. е. исследование состояния объекта по характеристикам выходящего излучения;

– «инструментальные» обратные задачи, состоящие в исключении влияния измерительного прибора в спектроскопии, радиоастрономии и т. д.

Возвращаясь к общим определениям прямых и обратных задач, основанным на направлении причинно-следственного хода событий и явлений, легко видеть, что мы встречаемся с прямыми и обратными задачами не только в физике и математике, но и во всех разделах наук и различных областях нашего бытия. Так, археология занимается решением обратных задач, направленных на реконструкцию условий жизни человека, исторических событий на основе ограниченного объема зарегистрированных «следствий». В медицине все задачи диагностики относятся также к обратным задачам. Наблюдая те или иные параметры состояния, например человека, требуется определить причины этого состояния (болезни). Говоря о таких обратных задачах (как и обратных задачах математической физики), можно обратить внимание на то, что ряд из них относится к временному пространству (реконструкция событий), другие – к обычному пространству.

Понятно, что исход решения обратной задачи связан с объемом имеющейся информации (результатов измерений). Так, например, высокая температура у человека может быть вызвана десятками и

сотнями болезней, и, если врач располагает только информацией о температуре тела больного, маловероятно, что он сможет определить правильно причину, т. е. болезнь пациента.

Важной чертой измерительных обратных задач и ряда других является вероятностная природа величин, наблюдаемых в эксперименте. Причиной этого могут быть либо «шумы» измерительной аппаратуры, искажающие «полезный» (истинный) сигнал, либо природа самого «полезного» сигнала. Этот сигнал может быть случайным, например при измерениях интенсивности потока квантов. Стохастичность является неотъемлемой чертой всякого реального эксперимента и поэтому должна явным образом входить в формулировку обратной задачи.

Статистические подходы и методы решения некорректных обратных задач являются прямым следствием указанной стохастичности эксперимента, а в ряде случаев и стохастичности полезного сигнала.

Из приведенных примеров интуитивно ясно, что решение некоторых обратных задач невозможно, либо неоднозначно, либо, наконец, может быть найдено лишь приближенное решение (т. е. решение определяется с большой погрешностью).

## **1.2. Прямые и обратные задачи атмосферной оптики**

При рассмотрении прямых и обратных задач атмосферной оптики мы имеем дело с функциями и параметрами, которые можно подразделить на три класса (рис 1.2-1).

### *а) Параметры состояния среды $\varphi$*

К ним можно отнести, например, высотные профили температуры и концентраций поглощающих и рассеивающих субстанций (газов, аэрозолей), давление, размеры и форму частиц аэрозоля, их состав и структуру, параметры, описывающие подстилающую поверхность и т. д.

### *б) Характеристики взаимодействия электромагнитного (ЭМ) излучения со средой $B$*

Это могут быть коэффициенты поглощения и рассеяния или определяющие их квантовомеханические параметры (параметры тонкой структуры спектра) – интенсивности, полуширины и т. д.

спектральных линий, излучательные способности подстилающих поверхностей и т. д.



**Рис 1.2-1.** Прямые и обратные задачи атмосферной оптики

в) *Характеристики поля ЭМ излучения  $f$*

К ним относятся интенсивность излучения, потоки излучения, характеристики поляризации излучения. В более общем случае характеристикой излучения являются параметры Стокса.

Отметим также, что при решении прямых и обратных задач атмосферной оптики требуется задать геометрию рассматриваемой среды.

Исторически исследования в области теории переноса излучения (как и в других областях), были начаты с так называемых *прямых задач* (рис. 1.2-1). В таких задачах предполагаются известными параметры физического состояния среды  $\varphi$  и характеристики взаимодействия ЭМ излучения со средой  $B$ . При этом требуется определить те или иные характеристики поля ЭМ излучения  $f$ . В дальнейшем, после всестороннего исследования прямых задач, т. е. тщательного изучения закономерностей формирования полей излучения в атмосферах планет, все больший интерес начали приобретать обратные задачи атмосферной оптики. (Отметим, что в настоящее время продолжается и интенсивное развитие методов решения разнообразных прямых задач атмосферной оптики.) Это было также связано с постепенным накоплением экспериментального материала о различных характеристиках ЭМ излучения атмосфер планет и естественным желанием использовать его для получения новой информации об оптических и физических свойствах атмосфер.

Как следует из рис. 1.2-1, можно сформулировать 2 типа обратных задач. В I типе обратных задач предполагается, что известны (измерены) те или иные характеристики поля излучения  $f$ , заданы характеристики взаимодействия ЭМ излучения со средой  $B$  и требуется определить те или иные параметры физического состояния атмосферы и подстилающей поверхности  $\varphi$ . Именно к этому типу обратных задач атмосферной оптики относятся дистанционные оптические (в широком смысле) методы измерений, особенно интенсивно развиваемые и используемые в последние 20–30 лет.

В обратных задачах II типа также предполагается, что известны (измерены) те или иные характеристики поля излучения  $f$ , но при этом также известны (измерены) параметры физического состояния среды  $\varphi$ . Требуется определить те или иные характеристики взаимодействия ЭМ излучения со средой  $B$ . Такой подход давно и плодотворно используется в различных разделах физики в

лабораторных исследованиях различных количественных характеристик взаимодействия излучения с газовой и конденсированной средами (например определение коэффициентов рассеяния и поглощения). Обратные задачи II типа не нашли очень широкого применения в атмосферной оптике, но именно к ним следует отнести, например, задачу определения коэффициентов континуального поглощения водяного пара по натурным измерениям поглощения солнечного излучения в земной атмосфере и ряд других задач.

Отметим, что этот подход перспективен при исследовании процессов взаимодействия излучения со средами, воспроизведение которых в лабораторных условиях сложно или невозможно, например в разреженных средах или в газовых средах, содержащих ионы или радикалы.

Формально прямая задача может быть записана в виде общего операторного уравнения

$$f(v) = \mathbf{A}(v, \varphi(s), B(v, s)), \quad (1.2-1)$$

где  $f(v)$  и  $\varphi(s)$  принадлежат некоторым функциональным пространствам  $\mathbf{F}$  и  $\Phi$ ,  $v$  и  $s$  вещественные числа (например,  $v$  – частота,  $s$  – пространственная переменная),  $\mathbf{A}$  – оператор, в общем случае нелинейный, действующий из  $\Phi$  в  $\mathbf{F}$ , – оператор прямой задачи (например интегро-дифференциальный оператор уравнения переноса излучения). Оператор  $\mathbf{A}$  устанавливает причинно-следственную связь между  $f(v)$  и  $\varphi(s)$ , т. е. функция  $f(v)$  описывает следствие некоторого процесса, причина которого определяется функцией  $\varphi(s)$  (процесса переноса излучения).  $B(v, s)$  обозначает характеристики взаимодействия излучения со средой. На данном этапе рассмотрения будем считать  $B(v, s)$  известным.

Как следует из вышеприведенного определения прямой задачи, она состоит в определении  $f(v)$ . Так, например, если нам заданы вертикальные профили температуры  $T(z)$  и концентрации поглощающих и излучающих газов  $n_i(z)$ , известны коэффициенты поглощения  $k_i(v, z)$ , с помощью интегральной формы уравнения переноса излучения можно определить интенсивность теплового излучения  $f(v)$  как функцию частоты.

Обратная задача I типа состоит в определении  $\varphi(s)$  при известной  $B(\nu, s)$  по измеренной функции  $f(\nu)$ , т. е. определении причин  $\varphi(s)$  по известному следствию  $f(\nu)$ . Классическим примером обратной задачи I типа является задача термического зондирования атмосферы, т. е. определение вертикального профиля температуры  $T(z)$  по измеренному спектру теплового излучения атмосферы  $I(\nu)$ .

Обратные задачи возникают непосредственно в связи с конкретными дистанционными методами исследования. Поэтому, прежде чем перейти к рассмотрению различных обратных задач атмосферной оптики, приведем в следующем разделе классификацию методов дистанционного измерения различных параметров атмосферы и подстилающей поверхности.

### **1.3. Классификация дистанционных методов измерений атмосферных параметров**

Дистанционные методы измерений параметров окружающей среды принято классифицировать по различным признакам:

- по используемому излучению (природе излучения);
- по основным процессам взаимодействия излучения с исследуемой средой;
- по условиям освещения (времени суток);
- по области спектра;
- по определяемому параметру;
- по геометрии измерений;
- по используемому носителю или расположению прибора.

Прежде всего принято дистанционные методы подразделять на пассивные и активные (по природе используемого излучения). Активные методы, использующие искусственные источники излучения (например лазеры), мы не рассматриваем, поскольку при их использовании не возникает сложностей с решением обратной задачи. К пассивным методам, использующим измерения характеристик естественных полей излучения, относятся:

- методы собственного излучения атмосферы (равновесного и неравновесного);
- методы рассеянного излучения (солнечного);

- методы поглощения (прозрачности), прежде всего солнечного излучения (но и Луны и звезд);
- методы рефракции;
- методы отраженного излучения.

Часто метод рассеянного и метод отраженного излучения объединяют в один – метод рассеянного и отраженного излучения.

Классификация дистанционных методов измерений по основным процессам взаимодействия излучения с исследуемой средой близка к приведенной выше. В этой классификации выделяются методы рассеяния (различных типов – молекулярного (релеевского), аэрозольного, комбинационного и т. д.), поглощения, собственного излучения, рефракции и т. д.

По условиям освещения (времени суток) дистанционные методы можно подразделить на дневные (над освещенным полушарием планеты), ночные и методы, используемые в области терминатора (область перехода между дневным и ночным полушариями). Методы рассеянного и отраженного излучения и метод прозрачности связаны, в основном, с использованием солнечного излучения как источника информации о состоянии среды и применимы для дневного полушария планеты. В методах прозрачности можно также использовать отраженное Луной солнечное излучение и излучение звезд. Метод собственного излучения (свечения) атмосферы можно использовать в любое время суток, по крайней мере, для ИК и МКВ диапазонов. Методы рефракции могут использовать различные источники излучения (Солнце, звезды, искусственные источники).

Существует классификация дистанционных методов в соответствии с используемой областью спектра, которая различает методы, использующие:

- ультрафиолетовую (УФ);
- видимую (ВИД);
- ближнюю инфракрасную (БИК);
- инфракрасную (ИК);
- микроволновую (МКВ);
- сантиметровую (СМ) области спектра.

По геометрии измерений космические методы измерений можно подразделить на методы надирного и касательного (лимбового, на горизонт планеты) зондирования. При надирной геометрии измерений уходящее излучение регистрируется в направлениях вблизи местного вертикала. При этом большинство современных спутниковых

приборов осуществляет угловое сканирование вблизи надира, что позволяет получать горизонтальные поля исследуемых характеристик. Аналогичный результат может быть получен при использовании специальных приемников линейчатого или матричного типов. Диапазон углов сканирования, угловая апертура приборов, тип сканирования и высота космического носителя определяют пространственную область и горизонтальное разрешение дистанционных измерений. Эти характеристики очень важны, так как определяют возможности осуществления глобальных наблюдений, периодичность наблюдений и, следовательно, их использование при решении различных научных и прикладных задач.

Для некоторых пассивных методов можно использовать оба типа геометрии измерений, для методов прозрачности и рефракции – только лимбовую геометрию (касательные трассы).

Дистанционные методы измерений параметров атмосферы можно также классифицировать по определяемому параметру. По этой классификации выделяются дистанционные методы для определения:

- температуры, плотности и давления атмосферы;
- характеристик облаков – бальности, высоты верхней и нижней границ (вертикальной структуры), температуры верхней границы (В.Г.), водности, фазового состава, микрофизики облаков;
- интенсивности осадков;
- содержания поглощающих газов – водяного пара, озона и других малых газовых составляющих (МГС);
- поля ветра;
- характеристик аэрозолей.

С учетом важности ряда характеристик подстилающей поверхности для физики атмосферы следует упомянуть дистанционные методы определения:

- температуры океанов и суши;
- влажности почв;
- оптических характеристик подстилающих поверхностей (отражательных и излучательных способностей).

Наконец, по расположению прибора или используемому носителю дистанционные методы подразделяют на наземные, самолетные, спутниковые (космические) и т. д.

В следующем разделе рассмотрим физико-математическую формулировку обратных задач, соответствующих наиболее широко

применяемым методам дистанционных измерений атмосферных параметров.

#### **1.4. Физико-математическая формулировка обратных задач атмосферной оптики**

После рассмотрения общих особенностей формулировки прямых и различных обратных задач атмосферной оптики перейдем к формулировке конкретных методов дистанционного зондирования атмосферы и подстилающей поверхности и математической постановке соответствующих обратных задач. При этом мы будем пользоваться сведениями из атмосферной оптики, в частности – элементами теории переноса излучения, молекулярной спектроскопии и теории рассеяния (см., например, [14]).

Наше рассмотрение начнем с относительно простых методов, основанных на измерениях прошедшего через атмосферу излучения Солнца. Отметим, что в качестве источников излучения в этих методах можно также использовать Луну, звезды и различные искусственные источники света, например лазеры.

#### **Методы прозрачности**

Физической основой рассматриваемых методов является закон Бера–Бугера–Ламберта (закон ББЛ)

$$J(\nu) = J_0(\nu) \exp(-\tau(\nu, s_1, s_2)). \quad (1.4-1)$$

Здесь  $J(\nu)$  и  $J_0(\nu)$  – интенсивность прошедшего через атмосферу и падающего на рассматриваемый слой атмосферы излучения,  $\tau(\nu, s_1, s_2)$  – оптическая толщина атмосферы,  $\nu$  – частота излучения,  $s$  – координата вдоль трассы распространения излучения,  $s_1$  и  $s_2$  – границы рассматриваемого слоя атмосферы. С учетом различных ослабляющих излучение факторов оптическая толщина может быть представлена в виде

$$\tau(\nu, s_1, s_2) = \int_{s_1}^{s_2} \sum_{i=1}^N k_i(\nu, s) \rho_i(s) ds + \tau_R(\nu, s_1, s_2) + \tau_a(\nu, s_1, s_2), \quad (1.4-2)$$

где  $k_i(v, s)$  – коэффициент поглощения  $i$ -й газовой составляющей,  $\rho_i(s)$  – ее плотность,  $\tau_R(v, s_1, s_2)$  и  $\tau_a(v, s_1, s_2)$  – оптические толщи релеевского и аэрозольного ослабления. Отметим, что выражение (1.4-2) записано для монохроматической интенсивности. В реальных экспериментах следует учесть конечное спектральное разрешение приборов. Таким образом, проводимое нами рассмотрение является идеализированным, так как измерения строго монохроматического излучения невозможны. Однако существуют приборы сверхвысокого спектрального разрешения (например гетеродинные спектрометры), для которых рассмотрение монохроматического случая представляет и практический интерес.

Из выражения (1.4-2) видно, что ослабление излучения обусловлено различными физическими причинами – истинным поглощением различными атмосферными газами, ослаблением за счет молекулярного (релеевского) рассеяния, поглощением и рассеянием на атмосферных аэрозолях. В зависимости от спектральной области измерений, характеристик прибора и целей интерпретации соотношения (1.4-1) и (1.4-2) могут лежать в основе различных дистанционных методов:

- определения характеристик газового состава атмосферы;
- определения температуры атмосферы;
- измерений скорости ветра;
- определения плотности атмосферы;
- измерений различных характеристик атмосферных аэрозолей.

Рассмотрим различные дистанционные методы измерений, предполагая для простоты, что тем или иным методом можно осуществить в выражении (1.4-2) разделение различных компонент ослабления излучения.

### **Определение характеристик газового состава атмосферы методом прозрачности**

Предположим, что нам удалось выделить компоненту ослабления, обусловленную истинным поглощением  $i$ -го поглощающего газа. Тогда, опуская индекс, выражение (1.4-2) можно записать в виде

$$J(\nu) = J_0(\nu) \exp \left( - \int_{s_1}^{s_2} k(\nu, s) \rho(s) ds \right). \quad (1.4-3)$$

Для монохроматической интенсивности выражение (1.4-3) можно записать в форме

$$\tau_\rho(\nu) = - \ln \frac{J(\nu)}{J_0(\nu)} = \int_{s_1}^{s_2} k(\nu, s) \rho(s) ds. \quad (1.4-4)$$

Отношение  $J(\nu)/J_0(\nu)$  в атмосферной оптике называется функцией пропускания и имеет простой физический смысл – оно характеризует долю излучения, например солнечного, прошедшего через рассматриваемый слой, т. е. прозрачность этого слоя. Поэтому рассматриваемые здесь дистанционные методы часто называют методами прозрачности.

Таким образом, если нам известны (измерены) функции  $J_0(\nu)$  и  $J(\nu)$ , то можно определить оптическую толщину – функцию  $\tau_\rho(\nu)$ , связанную уравнением (1.4-4) с искомой плотностью поглощающего газа  $\rho(s)$ . Уравнение (1.4-4) является уравнением Фредгольма I рода. Как мы покажем далее, большинство дистанционных методов зондирования сводятся именно к этим (линейным или нелинейным) уравнениям. Запишем это уравнение в общем виде.

Линейное уравнение

$$f(y) = \int_a^b A(x, y) \varphi(x) dx. \quad (1.4-5)$$

В уравнении (1.4-5)  $f(y)$  и  $\varphi(x)$  – измеренная и искомая функции,  $A(x, y)$  – ядро интегрального уравнения, которое предполагается заданным.

Нелинейное уравнение

$$f(y) = \int_a^b A(x, y, \varphi(x)) \varphi(x) dx. \quad (1.4-6)$$

В последнем случае ядро интегрального уравнения зависит от искомой функции.

Рассмотрим два конкретных примера формулировки обратных задач рассматриваемого типа.

### Наземные измерения прозрачности

Выражение (1.4-4) перепишем для модели плоскопараллельной горизонтально однородной атмосферы

$$\tau_{\rho}(\nu) = \frac{1}{\cos \theta} \int_0^{\infty} k(\nu, z) \rho(z) dz. \quad (1.4-7)$$

В качестве пространственной координаты мы используем здесь высоту над поверхностью Земли  $z$  (рис. 1.4-1). Угол  $\theta$  на рис. 1.4-1 – зенитный угол Солнца.

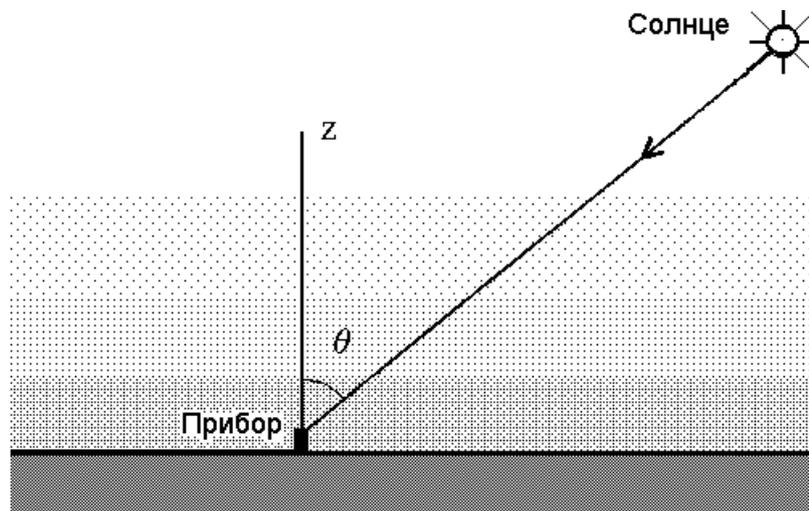


Рис. 1.4-1. Схема наземных измерений прозрачности атмосферы.

Успех решения обратной задачи – определения вертикального профиля плотности поглощающего газа  $\rho(z)$  из уравнения (1.4-7) – зависит от поведения ядра уравнения (в данном случае коэффициента поглощения  $k(\nu, z)$ ) как функции высоты при различных значениях частоты. Пусть, например, коэффициент поглощения не зависит от высоты. Тогда можно сделать очевидные преобразования:

$$\tau_{\rho}(\nu) = \frac{k(\nu, z)}{\cos \theta} \int_0^{\infty} \rho(z) dz = \frac{k(\nu, z)}{\cos \theta} U, \quad U = \int_0^{\infty} \rho(z) dz. \quad (1.4-8)$$

Из соотношений (1.4-8) следует, что в рассматриваемом случае мы не можем получить информацию о вертикальном профиле

плотности поглощающего газа, а только о его интегральном содержании во всей толще атмосферы. Хотя мы не можем решить поставленную ранее обратную задачу, отметим, что определение величины общего содержания  $U$  во многих случаях представляет большой практический интерес. Например, наземные измерения поглощения УФ солнечного излучения позволяют с высокой точностью (1–3%) определять интегральное (общее) содержание озона. Это связано с тем, что в УФ области спектра коэффициенты поглощения озона не зависят от давления атмосферы и слабо зависят от ее температуры. Отметим также, что задача определения общего содержания, в отличие от определения профиля плотности газа, корректна.

В ИК области спектра закономерности молекулярного поглощения атмосферных газов в большинстве случаев другие. Из курсов теоретической атмосферной оптики мы знаем, что коэффициент молекулярного поглощения в общем случае зависит от давления и температуры атмосферы –  $k(\nu, z) = k(\nu, p(z), T(z))$ . Для лоренцевской спектральной линии коэффициент поглощения можно записать в виде

$$k(\nu, z) = \frac{S(T(z))}{\pi} \frac{\alpha_L(p(z), T(z))}{(\nu - \nu_0)^2 + \alpha_L^2(p(z), T(z))}, \quad (1.4-9)$$

где  $S(T(z))$  – интенсивность спектральной линии,  $\alpha_L(p(z), T(z))$  – ее полуширина,  $\nu_0$  – частота центра линии. Напомним, что полуширина лоренцевской линии зависит от давления и температуры

$$\alpha_L(p, T) = \alpha_{0L} \frac{p}{p_0} \left( \frac{T}{T_0} \right)^m, \quad (1.4-10)$$

где  $\alpha_{0L}$  – полуширина линии при фиксированных значениях давления  $p_0$  и температуры  $T_0$ ,  $m$  – параметр.

Рассмотрим поведение коэффициента поглощения в лоренцевской линии как функции давления для различных частот в ее контуре:

– центр линии ( $\nu = \nu_0$ )

$$k(\nu, z) = \frac{S(T(p))}{\pi \alpha_{0L}} \frac{p_0}{p} \left( \frac{T_0}{T} \right)^m; \quad (1.4-11)$$

– крыло линии ( $|v - v_0| \gg \alpha_L$ )

$$k(v, z) = \frac{S(T(p))\alpha_{0L}}{\pi(v - v_0)^2} \frac{p}{p_0} \left( \frac{T}{T_0} \right)^m. \quad (1.4-12)$$

Из приведенных соотношений для центра и крыла линии видно, что зависимость коэффициента поглощения лоренцевской линии от давления и, следовательно, высоты в атмосфере существенно различна. Коэффициент поглощения в крыле линии прямо, а в центре линии – обратно пропорционален давлению. Таким образом, вклад в оптическую толщину молекулярного поглощения различных слоев атмосферы в различных частях лоренцевской линии существенно различен и интуитивно ясно, что решение интегрального уравнения (1.4-7) может позволить получить информацию именно о вертикальном профиле плотности поглощающего газа.

Отметим, что достаточно часто в качестве независимой вертикальной координаты в выражении для оптической плотности молекулярного поглощения используется давление. Исходя из уравнения гидростатики и определения отношения смеси, интегральное уравнение (1.4-7) можно записать в форме

$$\tau_\rho(v) = \frac{\sec \theta}{g} \int_0^{p_0} k(v, p, T) q(p) dp, \quad (1.4-13)$$

где  $q(p)$  – отношение смеси поглощающего газа,  $g$  – ускорение свободного падения,  $p_0$  – давление у поверхности Земли.

Подстановкой в это уравнение выражения для коэффициента поглощения в контуре лоренцевской спектральной линии, окончательно получаем

$$\tau_\rho(v) = \frac{\sec \theta}{g} \int_0^{p_0} \frac{S(T(p))\alpha_L(p, T(p))}{\pi(v - v_0)^2 + \alpha_L^2(p, T(p))} q(p) dp. \quad (1.4-14)$$

### **Космические измерения прозрачности**

Рассмотрим другой пример использования измерений прозрачности атмосферы. На рис. 1.4-2 изображена космическая геометрия измерений солнечного излучения. При движении спутника по орбите Солнце восходит или заходит за горизонт планеты относительно спутника. В эти моменты с помощью спектральных приборов, установленных на спутнике, можно проводить измерения

солнечного излучения как при прохождении его через атмосферу, так и вне ее. В этом случае можно зарегистрировать функцию пропускания атмосферы на касательных (скользящих) трассах, т. е. отношение  $J(\nu)/J_0(\nu)$ . Такие методы зондирования атмосферы часто называют затменными, поскольку они осуществляются во время «затмения» Землей источника излучения (в нашем случае Солнца).

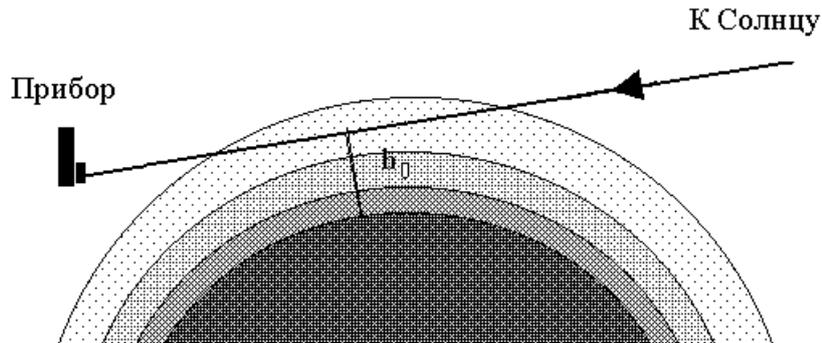


Рис. 1.4-2. Схема космических измерений прозрачности атмосферы.

Закон ББЛ для рассматриваемого случая можно записать в виде

$$J(\nu, h_0) = J_0(\nu) \exp \left( -2 \int_{h_0}^{\infty} w(z, h_0) k(\nu, z) \rho(z) dz \right), \quad (1.4-15)$$

где  $h_0$  – прицельная высота, т. е. минимальное расстояние траектории распространения луча света от поверхности Земли (рис. 1.4-2),  $w(z, h_0)$  – якобиан перехода от координаты  $s$  вдоль трассы распространения излучения к высоте  $z$ . Множитель 2 возникает из-за разбиения трассы на два одинаковых участка интегрирования, что справедливо для модели сферической слоисто-стратифицированной атмосферы.

Отметим, что измерения солнечного излучения можно осуществлять как функцию частоты и траектории распространения излучения, характеризуемой, например, прицельной высотой  $h_0$ . Существенно также отметить, что в космическом эксперименте (в противоположность наземной схеме измерений, когда существует проблема определения внеатмосферной величины солнечного излучения  $J_0(\nu)$ , решаемая специальными методами) имеется

возможность непосредственного измерения внеатмосферного излучения при больших значениях прицельной высоты  $h_0$ .

Переходя в монохроматическом случае к оптической плотности молекулярного поглощения, получим

$$\tau(\nu, h_0) = 2 \int_{h_0}^{\infty} w(z, h_0) k(\nu, z) \rho(z) dz. \quad (1.4-16)$$

Здесь ядром интегрального уравнения является произведение  $w(z, h_0)k(\nu, z)$ , предполагаемое известным, а искомой функцией – плотность поглощающего газа  $\rho(z)$ . Уравнение (1.4-16) является уравнением Вольтерра I рода, так как предел интегрирования переменный. Формально можно получить решение этого уравнения аналитически, дифференцируя оптическую плотность по прицельной высоте  $h_0$

$$\rho(h_0) = - \frac{d\tau(\nu, h_0)}{dh_0} \frac{1}{2w(h_0, h_0)k(\nu, h_0)}. \quad (1.4-17)$$

При этом проблемы устойчивого численного решения уравнения (1.4-17) остаются, так как численное дифференцирование измеренной функции также относится к некорректным (в классическом смысле) задачам.

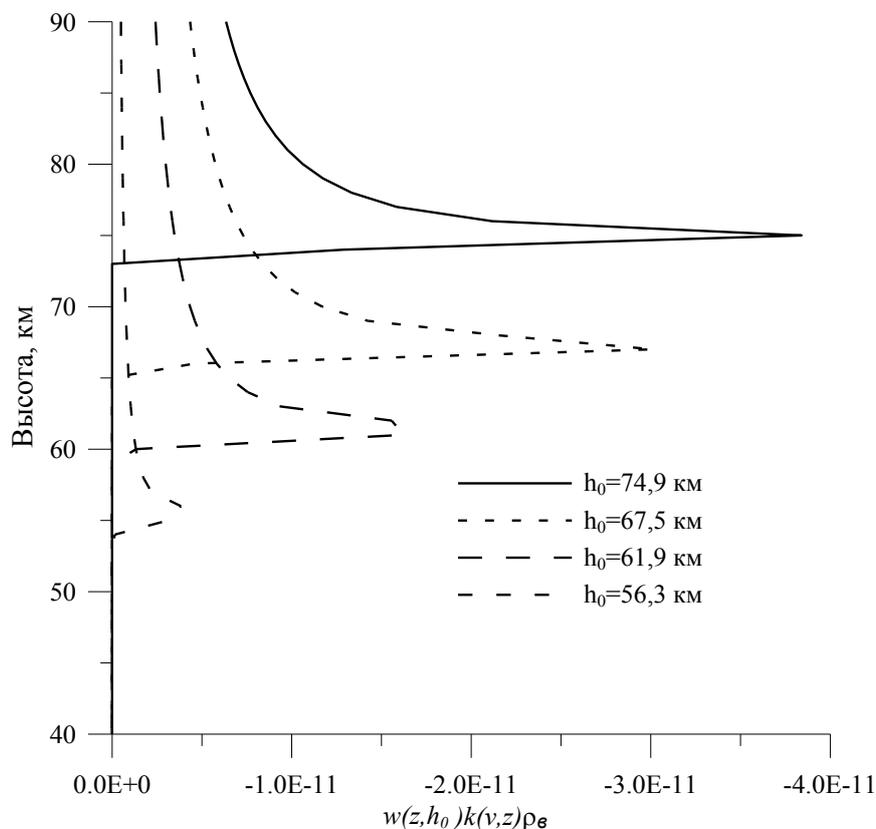
Необходимо отметить, что наше рассмотрение было проведено для интенсивности излучения. Точно так же, как приборы не могут измерять монохроматическое излучение, а измеряют энергию излучения в конечных спектральных интервалах (что учитывается интегрированием по спектру с учетом аппаратной функции прибора), они измеряют энергию излучения в конечных телесных углах. Учет этого осуществляется интегрированием интенсивности излучения в конечных телесных углах с учетом угловой чувствительности приборов.

Как следует из уравнения (1.4-16), космические измерения позволяют сформулировать обратную задачу относительно вертикального профиля плотности поглощающего газа при использовании измерений оптической плотности как функции прицельной высоты  $h_0$  при фиксированной частоте  $\nu$ .

Таким образом, в наземных измерениях информация о газовом составе атмосферы извлекается из спектральной зависимости

оптической плотности, а при космической геометрии измерений – из высотной зависимости.

Возможности реального извлечения информации о вертикальном профиле плотности поглощающего газа связаны с поведением ядра уравнения (1.4-16) как функции высоты при различных прицельных высотах. На рис. 1.4-3 приведены ядра интегрального уравнения (1.4-16) для отношения смеси озона  $q(z) = \rho_{O_3} / \rho_e$  в случае измерений солнечного излучения в УФ области спектра на длине волны 285 нм в области полосы поглощения озона Хартли–Хюггинса.



**Рис. 1.4-3.** Ядра интегрального уравнения прозрачности атмосферы на касательных трассах относительно профиля отношения смеси озона.

В этом случае обратная задача формулируется относительно отношения смеси озона. Из приведенного рисунка видно, что при

различных прицельных высотах ядро интегрального уравнения (1.4-16) локализовано на различных высотах в атмосфере. Такое поведение ядер и является физической основой возможности получения с помощью космических измерений прозрачности атмосферы информации о вертикальном профиле отношения смеси или плотности поглощающего газа. Характер зависимости ядер уравнения (1.4-16) от высоты в атмосфере определяется спецификой геометрии космических измерений, в частности тем, что при заданной прицельной высоте  $h_0$  солнечное излучение не поглощается слоями, расположенными ниже этой прицельной высоты.

### **Определение вертикального профиля температуры методом прозрачности**

Вернемся к выражению оптической плотности молекулярного поглощения для космической геометрии измерений – уравнению (1.4-16).

Ранее мы предполагали, что искомой функцией является плотность поглощающего газа. Предположим, однако, что она известна. Это предположение в земной атмосфере выполняется с хорошей точностью и в широком диапазоне высот для таких атмосферных газов как кислород и углекислый газ. В этом случае можно сформулировать обратную задачу относительно вертикального профиля температуры атмосферы  $T(p)$ . Действительно, как мы отмечали выше, интенсивности спектральных линий (а также полуширины, но, как правило, в меньшей степени) зависят от температуры. Например, для интенсивностей колебательно-вращательных линий линейных молекул в первом приближении можно записать

$$S(T) = S_0 \left( \frac{T}{T_0} \right) \exp \left( k E'' \left( \frac{1}{T_0} - \frac{1}{T} \right) \right), \quad (1.4-18)$$

где  $S_0$  – интенсивность линии при реперной температуре  $T_0$ ,  $E''$  – энергия нижнего колебательного состояния квантового перехода молекулы,  $k$  – постоянная Больцмана.

Как видно из соотношения (1.4-18), в зависимости от значений энергии  $E''$  интенсивность линии по разному зависит от температуры атмосферы. Например, для линии поглощения  $CO_2$ , расположенной

вблизи частоты  $668,60867 \text{ см}^{-1}$ , значение  $E''$  равно  $464,1717 \text{ см}^{-1}$ . Несложные расчеты показывают, что для этой линии производная  $dS(T)/dT$  при  $T = T_0 = 296 \text{ К}$  приблизительно равна  $0,01 S_0 [\text{К}^{-1}]$ . Таким образом, при изменении температуры атмосферы на  $1 \text{ К}$  интенсивность линии и, следовательно, оптическая толщина  $\tau(\nu)$  изменятся на  $1\%$ , что может быть зарегистрировано космическими приборами. Это означает, что в рассматриваемом случае уравнение (1.4-16) является нелинейным интегральным уравнением Вольтерра I рода относительно вертикального профиля температуры атмосферы.

### Определение вертикального профиля плотности атмосферы методом прозрачности

Определение плотности атмосферы может осуществляться на основе различных физических принципов. В связи с этим укажем на два дистанционных метода:

- метод молекулярного поглощения;
- метод ослабления за счет релеевского рассеяния.

В первом случае в основе метода лежит соотношение (1.4-15). С его помощью определяется, например, для космической схемы измерений (рис. 1.4-2) вертикальный профиль плотности  $i$ -й газовой составляющей атмосферы. Определение плотности таких газовых составляющих как  $O_2$  или  $CO_2$ , отношение смеси которых в земной атмосфере известно и постоянно во времени и пространстве, позволяет определить плотность атмосферы на основе соотношения

$$\rho_g = \rho_i / q_i, \quad (1.4-19)$$

где  $\rho_g$  – плотность атмосферы. Подстановка соотношения  $\rho_i(z) = \rho_g(z) q_i(z)$  в уравнение (1.4-16) дает явную форму интегрального уравнения для нахождения плотности атмосферы

$$\tau(\nu, h_0) = 2 \int_{h_0}^{\infty} w(z, h_0) k(\nu, z) q_i(z) \rho_g(z) dz, \quad (1.4-20)$$

где ядро интегрального уравнения, записанного в виде (1.4-5),  $A(z, h_0, \nu) = w(z, h_0) k(\nu, z) q_i(z)$ .

Для иллюстрации второго подхода предположим для простоты, что единственной компонентой ослабления излучения является оптическая толща релеевского (молекулярного) рассеяния. Из курса

атмосферной оптики известно, что коэффициент релеевского рассеяния определяется формулой

$$\alpha_R = \frac{8\pi^3(n^2 - 1)^2}{3N\lambda^4}, \quad (1.4-21)$$

где  $N$  – концентрация молекул воздуха,  $\lambda$  – длина волны,  $n$  – показатель преломления. Учитывая, что  $(n^2 - 1) \sim N$ , можно записать интегральное уравнение для нахождения вертикального профиля  $N(z)$ :

$$\tau_R(\lambda, h_0) = \frac{2C}{\lambda^4} \int_{h_0}^{\infty} w(z, h_0) N(z) dz, \quad (1.4-22)$$

где  $C$  – известная константа.

### **Определение характеристик атмосферных аэрозолей методом прозрачности**

В методах прозрачности сформулированы также различные обратные задачи относительно характеристик атмосферных аэрозолей. Предположим для простоты, что ослабление излучения происходит только за счет рассеяния и поглощения на атмосферных аэрозолях. Тогда в соотношении (1.4-2) оптическая плотность определяется ослаблением излучения на аэрозолях. Для случая ослабления на сферических однородных частицах для полидисперсной системы аэрозольных частиц можно записать следующее выражение для оптической толщи аэрозольного ослабления

$$\begin{aligned} \tau_a(\lambda) &= \sec \theta \int_0^{\infty} \alpha_a(z) dz = \\ &= \sec \theta \int_0^{\infty} dz \int_0^{\infty} \pi r^2 Q_e(r, m, \lambda, z) N_a(r, z) dr, \end{aligned} \quad (1.4-23)$$

где  $Q_e(r, m, \lambda, z)$  – фактор ослабления частиц радиуса  $r$  с комплексным показателем преломления (КПП)  $m$ ,  $N_a(r, z)$  – функция распределения частиц по размерам. Отметим, что в выражении (1.4-23) мы учли зависимость  $Q_e$  и  $N_a$  от высоты в атмосфере. Для случая аэрозольно-однородной атмосферы, т. е. независимых от высоты  $Q_e$  и  $N_a$ , можно

сформулировать следующее интегральное соотношение относительно функции распределения  $N_a(r)$ :

$$\tau_a(\lambda) = \sec \theta N_0 \int_0^{\infty} \pi r^2 Q_e(r, m, \lambda) n(r) dr, \quad (1.4-24)$$

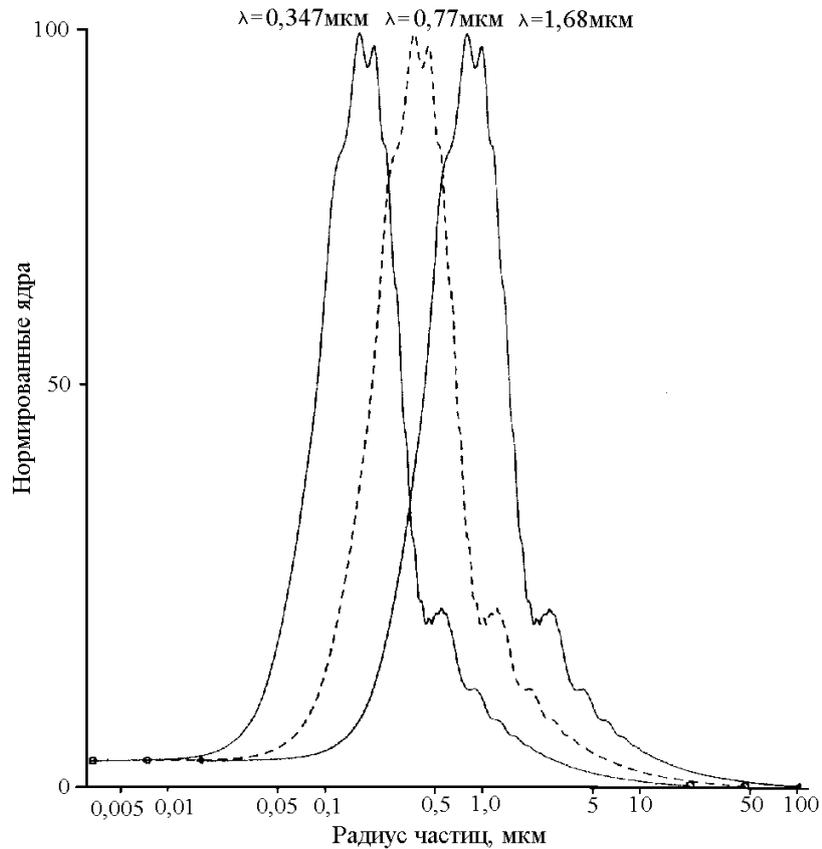
где  $N_0$  – полное число аэрозольных частиц,  $n(r)$  – нормированная функция распределения. Измеряя спектральную зависимость оптической плотности аэрозольного ослабления, можно определять  $n(r)$ . Реальные возможности рассматриваемого метода связаны с поведением ядра этого уравнения как функции радиусов частиц при различных длинах волн. На рис. 1.4-4 приведены ядра относительно функции распределения объема частиц по размерам для случая частиц с комплексным показателем преломления  $m = 1,60 - 0,02i$ .

Из соотношения (1.4-23) следует, что для расчетов ядер уравнения (1.4-24) необходимо знать физико-химические характеристики атмосферного аэрозоля – его КПП  $m$ . Эта ситуация реализуется, например, для водного аэрозоля, т. е. для облаков и осадков. В общем случае, задача дистанционного определения характеристик аэрозоля должна формулироваться как задача одновременного определения  $n(r)$  и КПП  $m$ . Тогда соответствующее интегральное уравнение является нелинейным и его решение представляется очень сложным.

### **Дистанционная рефрактометрия**

Существует также целый класс обратных задач, объединяемых названием дистанционной рефрактометрии. При этом для определения различных атмосферных параметров используется эффект рефракции, т. е. искривления траектории распространения излучения в атмосфере за счет неоднородностей коэффициента преломления, изменения фазы и амплитуды электромагнитного излучения.

Исследуя, например, характер искривления траекторий распространения излучения, можно судить о плотности атмосферы. Так, можно сформулировать обратную задачу определения вертикального профиля плотности атмосферы по измерениям угла рефракции в виде интегрального уравнения Абеля, являющегося частным случаем уравнения Вольтерра I рода.



**Рис. 1.4-4.** Нормализованные ядра  $r^2 Q_e(m, r, \lambda)$  для  $m = 1,60 - 0,02i$  и трех длин волн.

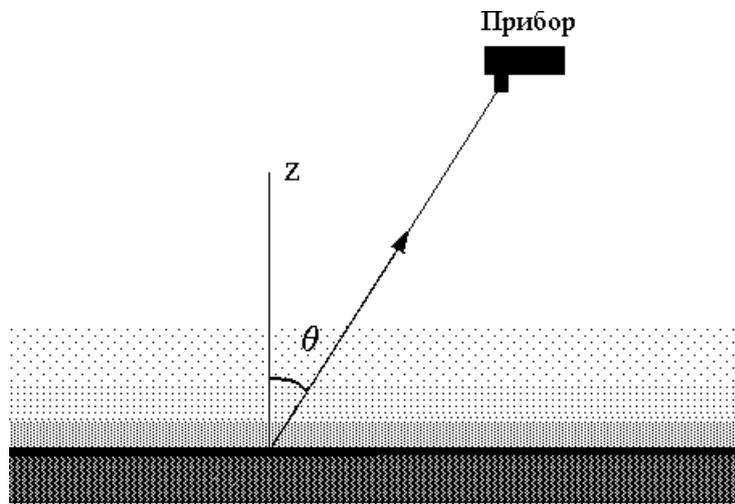
### Методы собственного излучения атмосферы

В настоящее время сформулирован большой класс обратных задач атмосферной оптики для собственного излучения атмосферы. В основе этих методов лежит интегральная форма уравнения переноса излучения.

Для уходящего теплового излучения для плоскопараллельной, горизонтально однородной атмосферы (рис. 1.4-5) можно записать:

$$J(\nu, \theta) = B_\nu(T_s) \exp\left(-\int_0^\infty k_\nu(z) \rho(z) dz\right) + \int_0^\infty B_\nu(T(z)) k_\nu(z) \rho(z) \exp\left(-\int_z^\infty k_\nu(z') \rho(z') dz'\right) dz, \quad (1.4-25)$$

где  $B_\nu(T)$  – функция Планка излучения абсолютно черного тела,  $T(z)$  – вертикальный профиль температуры,  $T_s$  – температура подстилающей поверхности,  $\rho(z)$  – плотность излучающего и поглощающего газа,  $k_\nu(z)$  – его коэффициент поглощения.



**Рис. 1.4-5.** Схема космических измерений уходящего теплового (собственного) излучения системы атмосфера–подстилающая поверхность.

При записи соотношения (1.4-25) мы использовали предположения о выполнимости локального термодинамического равновесия, о пренебрежимой малости эффектов рассеяния теплового излучения и абсолютно черной подстилающей поверхности. Уравнение (1.4-25) является физико-математической основой для формулировки разнообразных обратных задач тепловой области спектра при различной геометрии измерений – наземной и космической. Рассмотрим примеры таких обратных задач.

### **Определение вертикального профиля температуры (собственное излучение атмосферы, надирная геометрия измерений)**

Анализируя выражение для интенсивности уходящего теплового излучения (1.4-25) можно заметить, что  $J(\nu, \theta)$  является функционалом от ряда атмосферных параметров

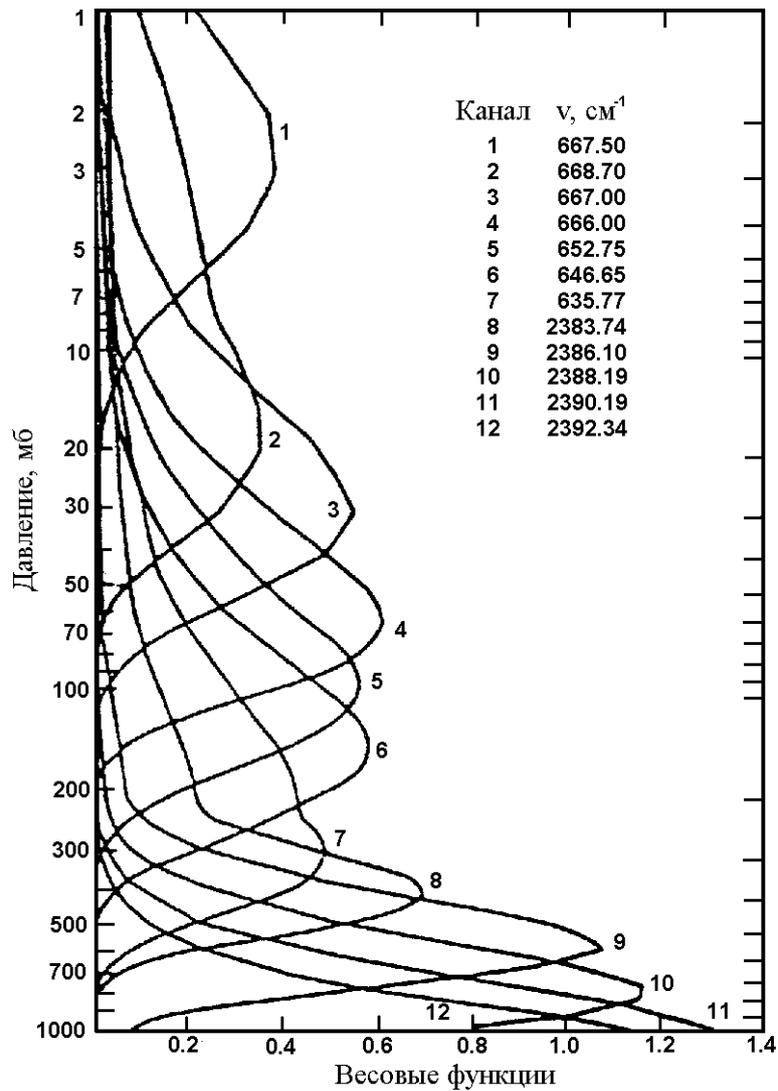
$$J(\nu, \theta) = J(\nu, \theta, T(z), \rho(z), T_s), \quad (1.4-26)$$

в общем случае – вертикальных профилей температуры и содержаний поглощающих и излучающих газов и аэрозолей, характеристик облаков, свойств подстилающей поверхности. Измерения спектральной или угловой зависимости уходящего теплового излучения позволяют рассматривать различные обратные задачи определения этих параметров.

Однако задача получения информации одновременно о нескольких неизвестных функциях  $T(z)$ ,  $\rho(z)$ ,  $T_s$  из измерений одной функции, например  $J(\nu, \theta)$  как функции даже двух переменных – длины волны и угла наблюдения – достаточно сложна. Поэтому в различных обратных задачах тепловой области спектра выбирают условия измерений так, чтобы минимизировать число неизвестных функций. Например, измеряя излучение в «окнах прозрачности» атмосферы, т. е. спектральных областях с малым поглощением излучения, определяют свойства подстилающей поверхности, прежде всего температуру поверхностей океанов или суши. Далее, измеряя уходящее излучение в полосах поглощения кислорода и углекислого газа, содержание которых с хорошей точностью неизменно и известно, можно «исключить» влияние на интенсивность теплового излучения профилей содержания поглощающих газов. Именно поэтому к настоящему времени разработаны методы определения вертикального профиля температуры по измерениям спектральной зависимости уходящего излучения в ИК полосах поглощения углекислого газа при 15 и 4,3 мкм и полосе поглощения кислорода в микроволновой области 0,5 см. Можно использовать для этого и отдельную линию поглощения кислорода при 0,25 см.

Возможность дистанционного определения именно вертикальных профилей температуры, а не средних температур атмосферы связана с тем, что уходящее излучение в спектральных интервалах с различной оптической плотностью генерируется различными высотными слоями

атмосферы. Этот механизм иллюстрирует рис. 1.4-6, где изображены так называемые весовые функции рассматриваемой обратной задачи, характеризующие области формирования уходящего теплового излучения в различных спектральных интервалах 4,3 и 15 мкм полос поглощения  $CO_2$ .



**Рис. 1.4-6.** Весовые функции задачи определения температурного профиля по измерениям теплового излучения в полосах 4,3 и 15 мкм  $CO_2$ . Номера указывают центры каналов измерений.

Под весовыми функциями понимаются производные функций пропускания атмосферы в интегральной форме уравнения переноса, которые «взвешивают» высотные распределения функции Планка или температуры в МКВ области. Для МКВ области они очень близки к ядрам или вариационным производным (см. раздел 1.5), если не учитывать температурную зависимость функций пропускания. Для ИК области ядра отличаются от весовых функций множителем – производной функции Планка по температуре.

Весовые функции наряду с ядрами интегральных уравнений и вариационными производными часто используются для анализа и иллюстраций возможностей дистанционных измерений.

### **Определение вертикального профиля температуры (собственное излучение атмосферы, касательная геометрия измерений)**

Измерения уходящего излучения можно осуществлять, направляя прибор, установленный на спутнике, на горизонт планеты. В этом случае говорят о касательной геометрии измерений или измерении излучения горизонта планеты. При такой схеме измерений прибор должен иметь достаточно высокое угловое разрешение или, иными словами, он должен позволять регистрировать уходящее излучение в узком диапазоне прицельных высот. Геометрия измерений в этом случае напоминает геометрию измерений, изображенную на рис. 1.4-2 для метода прозрачности. Отличие состоит в том, что сканирование атмосферы осуществляется не за счет взаимного движения спутника и источника излучения (Солнца, звезды), а за счет сканирования горизонта самим прибором или использования многоплощадочных приемников, измеряющих уходящее излучение на различных прицельных высотах одновременно.

Уравнение переноса для рассматриваемой геометрии измерений можно записать в виде

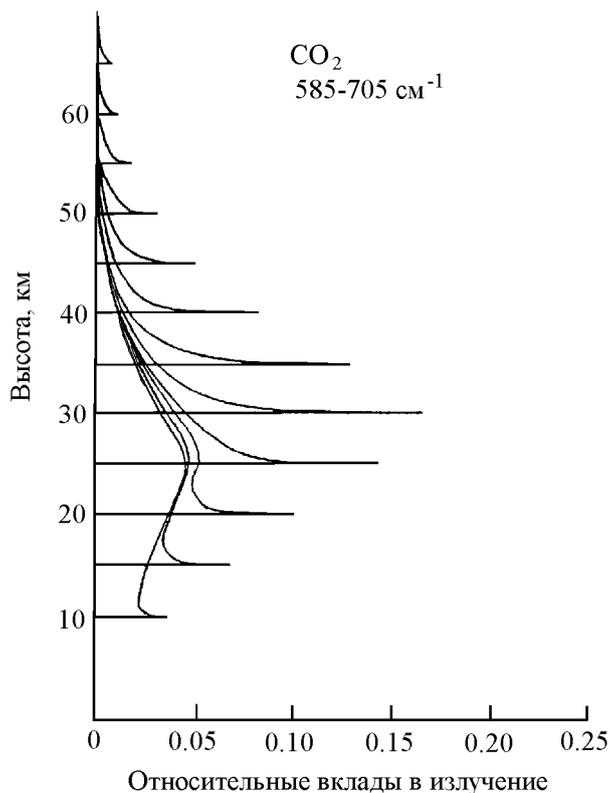
$$\begin{aligned}
J(v, h_0) = & \int_{h_0}^{\infty} B_v(T(z))w(z, h_0) \times \\
& \times \frac{\partial}{\partial z} \left( \exp \left( - \int_z^{\infty} k_v(z') \rho(z') w(z', h_0) dz' \right) + \right. \\
& \left. + \exp \left( - \int_{h_0}^z k_v(z') \rho(z') w(z', h_0) dz' - \int_{h_0}^{\infty} k_v(z') \rho(z') w(z', h_0) dz' \right) \right) dz
\end{aligned} \tag{1.4-27}$$

где  $w(z, h_0)$  – якобиан перехода от интегрирования вдоль трассы формирования излучения к вертикальной координате. Выражение (1.4-27) справедливо для модели сферической слоисто-однородной атмосферы, которая предполагает, что температура атмосферы зависит только от вертикальной координаты. Другими словами, мы пренебрегаем наличием горизонтальных градиентов температуры на трассе формирования уходящего излучения.

На рис. 1.4-7 изображено поведение еще одной характеристики, вклада слоев атмосферы в уходящее излучение, которая наряду с ядрами, весовыми функциями и вариационными производными также используется для анализа задач дистанционного зондирования атмосферы. На рисунке изображены вклады в излучение для задачи, описываемой уравнением (1.4-27) на различных прицельных высотах измерений для одного спектрального интервала в полосе 15 мкм  $CO_2$ .

В рассматриваемой здесь задаче термического зондирования вклады отдельных слоев атмосферы в величины уходящего излучения равны произведению функции Планка на разность функций пропускания на границах рассматриваемого слоя. Они характеризуют долю, вносимую слоем атмосферы в излучение.

Подчеркнем, что для касательной геометрии измерений при решении обратных задач используются измерения угловой зависимости уходящего излучения или, что то же самое, зависимости интенсивности излучения от прицельной высоты измерений. В общем случае можно использовать одновременно как угловые, так и спектральные зависимости. При этом можно избавиться от предположения о сферической слоисто-стратифицированной модели атмосферы и сформулировать более сложную обратную задачу относительно определения температуры атмосферы как функции высоты и горизонтальной координаты.



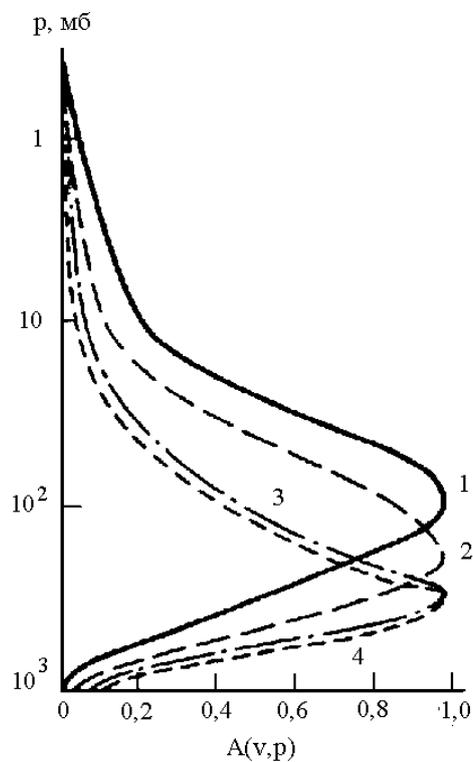
**Рис. 1.4-7.** Вклады слоев атмосферы в тепловое излучение на касательных трассах для различных прицельных высот измерений в спектральном интервале  $585\text{--}705\text{ см}^{-1}$  в полосе  $15\text{ мкм}$ .

### **Определение характеристик газового состава атмосферы (собственное излучение атмосферы, надирная геометрия измерений)**

Зная (определив) вертикальный профиль температуры и используя измерения в полосах поглощения (излучения) других газовых компонент, можно получить информацию о содержании, например, водяного пара, озона и других газов. При этом следует обратить внимание, что искомые характеристики, например вертикальные распределения плотностей поглощающих газов, находятся в показателе экспоненты функции пропускания и под двумя интегралами по пространственной переменной. Сглаживающее

действие этих операторов, конечно, должно затруднять получение информации об этих профилях. Более того, можно показать, что в условиях изотермической атмосферы и абсолютно черной подстилающей поверхности в надирных измерениях уходящего теплового излучения не содержится вообще никакой информации о составе атмосферы. Этот вывод следует фактически из самого определения абсолютно черного излучения как излучения изотермической полости.

На рис. 1.4-8 изображены ядра интегрального уравнения для нахождения вертикального профиля содержания водяного пара, полученного после линеаризации уравнения (1.4-27)



**Рис. 1.4-8.** Нормированные на единицу ядра интегрального уравнения задачи определения вертикального профиля содержания водяного пара по измерениям в надир уходящего теплового излучения.

$$\delta J(\lambda, \theta) = \int_0^{\infty} A(v, z, \theta) \delta \rho(z) dz. \quad (1.4-28)$$

Здесь  $\delta J(\lambda, \theta)$  и  $\delta \rho(z)$  – вариации интенсивности уходящего излучения и содержания водяного пара относительно средних значений. Ядра даны для различных частот в контуре спектральной линии  $H_2O$  в ИК области спектра (1 –  $\alpha_0/10$ , 2 –  $\alpha_0/5$ , 3 –  $\alpha_0/2$ , 4 –  $\alpha_0$ ).

Из рис. 1.4-8 видно специфическое поведение ядер рассматриваемой обратной задачи для ИК области спектра – стремление ядра к нулю вблизи поверхности Земли.

### Методы рассеянного солнечного излучения

Для дистанционного измерения различных атмосферных параметров можно использовать наряду с измерениями прозрачности и собственного излучения атмосферы измерения различных характеристик рассеянного поля излучения. В этом случае математической основой этих методов является интегрально-дифференциальная форма уравнения переноса излучения. Во многих ситуациях при учете реальной картины переноса, например солнечного излучения в неоднородной атмосфере с учетом многократного рассеяния, не удастся записать аналитические выражения для интенсивности рассеянного излучения.

Для иллюстрации этого класса методов рассмотрим простейший случай. Допустим, что поле уходящего рассеянного солнечного излучения определяется однократным молекулярным рассеянием при наличии молекулярного поглощения. Такая упрощенная постановка близка к реальной задаче определения вертикального профиля содержания озона по рассеянному солнечному излучению в УФ области спектра. Для компоненты однократно рассеянного солнечного излучения (в надир) можно записать

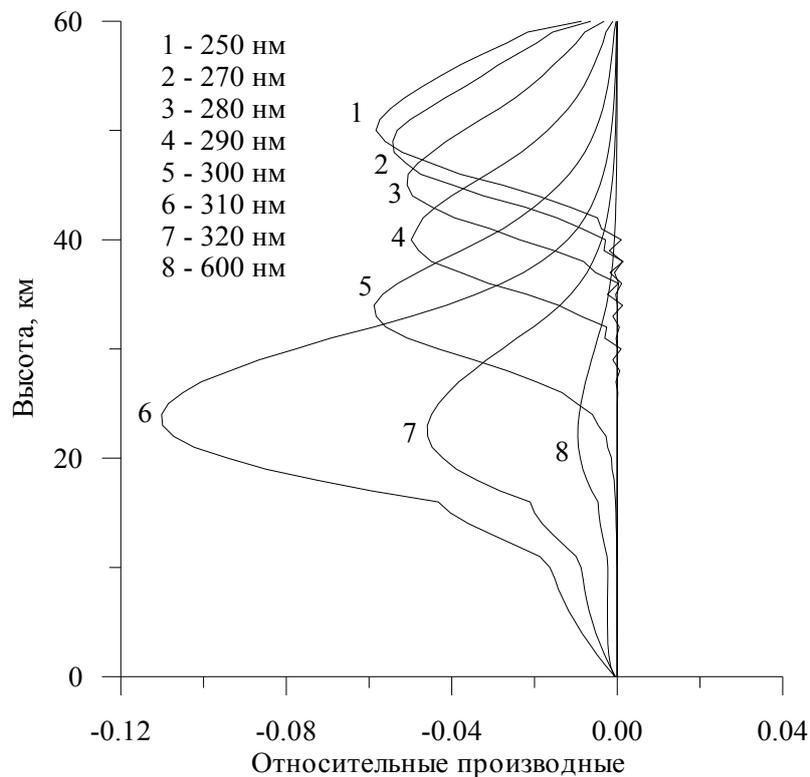
$$J(\lambda, \theta) = \alpha_R P(\theta) \frac{F_0}{4\pi} \times \int_0^{p_0} \exp \left[ - (1 + \sec \theta_S) \left( \frac{1}{g} \int_0^p k_v(p') q_{O_3}(p') dp' + \tau_R(p) \right) \right] dp, \quad (1.4-29)$$

где  $\alpha_R$  – релеевский коэффициент рассеяния,  $P(\theta)$  – релеевская индикатриса рассеяния,  $F_0$  – внеатмосферный поток излучения,  $\theta_s$  – зенитный угол Солнца,  $q_{O_3}(p)$  – отношение смеси озона как функция давления,  $\tau_R(p)$  – оптическая толщина релеевского ослабления.

Линеаризация выражения (1.4-29) позволяет свести рассматриваемую обратную задачу к интегральному уравнению Фредгольма I рода. На рис. 1.4-9 изображены вариационные производные (ядра линеаризованного уравнения, см. раздел 1.5) этой задачи, характеризующие области формирования уходящего рассеянного солнечного излучения в различных спектральных интервалах полосы поглощения озона Хартли–Хюггинса. На рис. 1.4-9 приведены относительные производные, т. е. производные, отнесенные к значению излучения в соответствующем канале измерений.

Из рис. 1.4-9 наглядно видно, что уходящее излучение для различных спектральных интервалов формируется в различных слоях атмосферы: в спектральных интервалах с большими значениями коэффициента поглощения озона – в верхней стратосфере, в спектральных интервалах с малыми значениями – в нижней стратосфере и тропосфере. Именно эта особенность формирования рассеянного солнечного излучения и является физической основой возможности решения рассматриваемой обратной задачи и получения информации о вертикальных профилях содержания озона. Данный косвенный метод активно используется для изучения озоносферы Земли и позволил, в частности, обнаружить такое явление как «озонные дыры» над Антарктидой.

В настоящее время существует много других дистанционных методов измерений, основанных на регистрации поглощения солнечного излучения, его отраженных и рассеянных компонент, рефракции, собственного излучения атмосферы и т. д.



**Рис 1.4-9.** Вариационные производные по относительной вариации профиля плотности озона рассеянного солнечного излучения в УФ и видимой областях спектра. Цифры указывают центры спектральных каналов измерения.

Например, идет активная разработка дистанционных методов зондирования неравновесной атмосферы. В этом случае возникают сложные обратные задачи атмосферной оптики относительно десятков неизвестных функций – кинетической и колебательных температур, содержаний поглощающих и излучающих (неравновесно) атмосферных газов. При этом часто используется радиационно-физический подход, когда при формулировке задачи используется не только уравнение переноса, но и модели возбуждения и дезактивации молекул.

Отметим в заключение, что рассмотренные обратные задачи атмосферной оптики формулируются и исследуются в настоящее

время не только для атмосферы Земли, но и для атмосфер других планет солнечной системы.

Заканчивая рассмотрение различных методов дистанционного зондирования атмосферы и возникающих в них обратных задач, заметим, что во многих случаях соответствующие интегральные уравнения оказываются нелинейными. Даже если некоторые задачи и формулируются первоначально в линейном приближении, растущие требования к точности определения параметров атмосферы приводят к необходимости уточнения их формулировок и, следовательно, к выявлению нелинейных зависимостей. Поэтому представляют интерес методы линеаризации задач атмосферной оптики, к рассмотрению которых мы и перейдем в следующем разделе.

### **1.5. Линеаризация обратных задач атмосферной оптики**

В настоящее время математическая теория решения обратных задач достаточно разработана как для линейных, так и для нелинейных задач (операторов  $\mathbf{A}$ ). В то же время многие методы исследования задачи (оценки погрешностей или пространственного разрешения решения, исследование информативности измерений) применимы лишь к задачам линейным. Поэтому часто приходится прибегать к локальной линеаризации задачи, т. е. приближенной замене нелинейного оператора  $\mathbf{A}$  близким ему в некоторой области пространства  $\Phi$  линейным оператором. Подобная локальная линеаризация также используется в ряде широко применяемых методов решения нелинейных задач.

Существуют различные подходы к линеаризации уравнения (1.4-6). Часто эту линеаризацию можно провести преобразованием исходного выражения с привлечением физических соображений. Например, одной из причин нелинейности обратных задач является присутствие в различных выражениях для функции пропускания экспоненциального члена, описывающего ослабление вдоль трассы распространения излучения (использованный нами выше закон ББЛ)

$$P(v) = \exp \left( - \int_{s_1}^{s_2} \alpha(v, s) \rho(s) ds \right). \quad (1.5-1)$$

Рассмотрим отклонение  $\delta\rho(s)$  искомого профиля плотности ослабляющей субстанции  $\rho(s)$  от среднего профиля  $\bar{\rho}(s)$ ,  $\delta\rho(s) = \rho(s) - \bar{\rho}(s)$  (вариация профиля). Подставив  $\rho(s) = \bar{\rho}(s) + \delta\rho(s)$  в выражение (1.5-1), можно записать:

$$P(v) = \exp\left(-\int_{s_1}^{s_2} \alpha(v,s)\bar{\rho}(s)ds\right) \exp\left(-\int_{s_1}^{s_2} \alpha(v,s)\delta\rho(s)ds\right). \quad (1.5-2)$$

Предположим, что вариация  $\delta\rho(s) \ll \bar{\rho}(s)$  и показатель второй экспоненты мал, что позволяет ограничиться двумя первыми членами разложения в ряд Тейлора. Тогда

$$P(v) \approx \bar{P}(v) - \bar{P}(v) \int_{s_1}^{s_2} \alpha(v,s)\delta\rho(s)ds, \quad (1.5-3)$$

$$\text{где } \bar{P}(v) = \exp\left(-\int_{s_1}^{s_2} \alpha(v,s)\bar{\rho}(s)ds\right).$$

Таким образом, мы получили линейную связь между вариациями функции пропускания  $\delta P(v)$  и вариацией искомого профиля  $\delta\rho(s)$

$$\delta P(v) = P(v) - \bar{P}(v) = -\bar{P}(v) \int_{s_1}^{s_2} \alpha(v,s)\delta\rho(s)ds. \quad (1.5-4)$$

Существенно использование условия  $\delta\rho(s) \ll \bar{\rho}(s)$ . Понятно, что хотя точная формулировка этого условия связана с поведением оператора  $\mathbf{A}$ , для применимости линейного приближения принципиально важна величина вариаций  $\delta\rho(s)$ , рассматриваемых в конкретной постановке задачи. Приведем практический пример, иллюстрирующий важность последнего замечания.

Зависимости функций пропускания атмосферы от содержания водяного пара и углекислого газа формально выглядят одинаково, но содержание углекислоты в тропосфере и стратосфере меняется крайне мало – на доли процента, а содержание водяного пара может изменяться в несколько раз. Поэтому, рассматривая влияние естественных вариаций содержания этих газов на перенос излучения, мы можем с высокой точностью использовать линейное приближение для описания связи концентрации углекислоты и прозрачности

атмосферы. В то же время попытка использования того же линейного приближения этой связи для водяного пара может привести к качественно неверным результатам.

### Функциональный ряд Тейлора и вариационная производная

Существует общий подход к линеаризации уравнения (1.4-6), основанный на разложении функционала в *функциональный ряд Тейлора*.

Введем обозначения. Пусть  $\Phi$  – множество функций, заданных на интервале  $[a, b]$ ,  $F$  – множество функций, заданных на интервале  $[c, d]$  и оператор  $A$  действует из  $\Phi$  в  $F$ . Пусть функции  $\varphi, \delta\varphi \in \Phi$  и  $A(\varphi) = f \in F$ . Зафиксировав функцию  $\varphi$ , будем рассматривать функции  $\varphi + \delta\varphi$ , считая, что функция  $\delta\varphi$  в интересующем нас смысле мала. В такой ситуации принято называть функцию  $\delta\varphi$  вариацией функции  $\varphi$ . В пространстве  $F$  при отображении  $A$  ей соответствует вариация  $\delta f = \delta A(\varphi) = A(\varphi + \delta\varphi) - A(\varphi)$ , т. е. некоторая функция, заданная на интервале  $[c, d]$ . Наша задача – построить линейный оператор, связывающий вариации (функции)  $\delta\varphi$  и  $\delta f$ .

Попробуем неформально объяснить, что мы хотим получить. Для того чтобы построить линейную связь между функциями  $\delta\varphi$  и  $\delta f$ , сначала свяжем линейными соотношениями  $\delta\varphi(x)$  и  $\delta f(y)$  для любой пары аргументов  $x$  и  $y$ . Для этого рассмотрим функцию двух переменных  $A|_{\varphi_0}(x, y)$ , такую, что  $\delta f(y) = A|_{\varphi_0}(x, y)\delta\varphi(x) + o(\delta\varphi(x))$  при любых  $x, y$  при условии, что функция  $\delta\varphi$  в некотором пока не определенном строго смысле отлична от нуля только в точке  $x$  (здесь  $o(\delta\varphi(x))$  – бесконечно малая порядка выше  $\delta\varphi(x)$  при  $\delta\varphi(x) \rightarrow 0$ ).

Тогда в линейном приближении при малых вариациях  $\delta\varphi$  будет выполняться требуемое нам соотношение:

$$\delta f(y) = \int A|_{\varphi_0}(x, y)\delta\varphi(x)dx.$$

Теперь перейдем к строгим формулировкам. В вариационном исчислении рассматривается разложение функционала, т. е. отображения из множества функций в множество вещественных чисел, в функциональный ряд Тейлора. В соответствии с этим вводится

понятие вариационной производной функционала. Зафиксировав значение  $y$ , рассмотрим функционал  $A_y$ , порожденный оператором  $\mathbf{A}$ , зависящий от  $y$  параметрически и определяемый выражением  $A_y(\varphi) = (\mathbf{A}(\varphi))(y) = f(y)$ . Ограничившись линейным членом в разложении в ряд, можно в линейном приближении записать равенство

$$\delta A_y(\varphi) = A_y(\varphi + \delta\varphi) - A_y(\varphi) = \int \frac{\delta A_y(\varphi(s))}{\delta\varphi(t)} \delta\varphi(t) dt. \quad (1.5-5)$$

Принято использовать обозначение  $\frac{\delta A_y(\varphi(s))}{\delta\varphi(x)}$  и называть эту функцию вариационной (функциональной) производной функционала  $A_y$  от функции  $\varphi$  в точке  $x$ .

Формально по определению

$$\frac{\delta A_y(\varphi(s))}{\delta\varphi(x)} = \lim_{\Delta x \rightarrow 0} \frac{\delta A_y(\varphi)}{\int_{x-\Delta x/2}^{x+\Delta x/2} \delta\varphi(\tau) d\tau}, \quad (1.5-6)$$

при условии  $\delta\varphi(\tau) \equiv 0$  вне интервала  $[x - \Delta x/2, x + \Delta x/2]$ , где

$$\delta A_y(\varphi) = \delta f(y) = (\mathbf{A}(\varphi + \delta\varphi) - \mathbf{A}(\varphi))(y) \quad (1.5-7)$$

– вариация функционала (приращение функционала), а интервал интегрирования в знаменателе (1.5-6) «стягивается» к точке  $x$ . Напомним, что  $\mathbf{A}(\varphi + \delta\varphi) - \mathbf{A}(\varphi)$  – функция, заданная на интервале  $[c, d]$ , и выражение  $(\mathbf{A}(\varphi + \delta\varphi) - \mathbf{A}(\varphi))(y)$  обозначает значение этой функции в точке  $y$ .

Формально из определения (1.5-6), (1.5-7) мы можем рассматривать значение вариационной производной как функционал от  $\varphi$ , зависящий от  $x$  и  $y$  как от параметров.

### Практическое вычисление вариационной производной

Формальное определение вариационной производной (1.5-6), (1.5-7) неудобно для непосредственного ее вычисления, поэтому в практических задачах часто используют следующий прием: рассматривают вариации функции  $\varphi$  вида  $\varepsilon\delta\varphi$ , зависящие от параметра  $\varepsilon$ . Тогда, в предположении, что  $\delta\varphi$  фиксирована,

функционал  $A_y(\varphi(s) + \varepsilon\delta\varphi(s))$  можно рассмотреть как функцию переменной  $\varepsilon$ . Оставив только линейный член в разложении этой функции в обычный ряд Тейлора по степеням  $\varepsilon$ , получим приближенное выражение для вариации функционала

$$\delta A_y(\varphi(s)) = \varepsilon \frac{d}{d\varepsilon} A_y(\varphi(s) + \varepsilon\delta\varphi(s)) \Big|_{\varepsilon=0}, \quad (1.5-8)$$

где  $\frac{d}{d\varepsilon}$  – обычная производная. Если путем преобразований удастся привести (1.5-8) к виду, аналогичному (1.5-5), то задача нахождения вариационной производной решена.

Приведем пример вычисления вариационной производной описанным методом для нелинейного функционала

$$A(\varphi(s)) = \exp\left(\int_a^b b(s)\varphi(s)ds\right). \quad (1.5-9)$$

Для рассматриваемого случая

$$A(\varphi(s) + \varepsilon\delta\varphi(s)) = \exp\left(\int_a^b b(s)(\varphi(s) + \varepsilon\delta\varphi(s))ds\right). \quad (1.5-10)$$

Вычислим производную по  $\varepsilon$  от этого выражения

$$\begin{aligned} \frac{d}{d\varepsilon} A(\varphi(s) + \varepsilon\delta\varphi(s)) &= \\ &= \exp\left(\int_a^b b(s)(\varphi(s) + \varepsilon\delta\varphi(s))ds\right) \int_a^b b(s)\delta\varphi(s)ds. \end{aligned} \quad (1.5-11)$$

Полагая в выражении (1.5-11)  $\varepsilon=0$ , получим для вариации функционала (1.5-8)

$$\delta A(\varphi(s)) = \exp\left(\int_a^b b(s)\varphi(s)ds\right) \int_a^b b(s)\delta\varphi(s)ds. \quad (1.5-12)$$

Сравнивая полученное выражение с (1.5-5), можно сделать вывод, что для нелинейного функционала (1.5-9) вариационная производная имеет вид

$$\frac{\delta A(\varphi(s))}{\delta \varphi(t)} = b(t) \exp\left(\int_a^b b(s)\varphi(s)ds\right). \quad (1.5-13)$$

Действительно, подставив выражение (1.5-13) в соотношение (1.5-5), получим формулу (1.5-12).

Как мы уже отмечали, вариационную производную (1.5-13) можно рассматривать как функционал от функции  $\varphi$ , зависящий от  $t$  как от параметра. Наконец отметим, что полученная нами другим способом формула (1.5-3) согласуется с приведенными выкладками.

### **Использование вариационных производных в численном решении задач**

Как правило, решая прикладные задачи, мы сводим нелинейную задачу (запишем это соотношение в общем виде)

$$f(y) = \mathbf{A}(y, \varphi),$$

сформулированную относительно непрерывной функции  $\varphi$ , к приближенному дискретному представлению

$$f(y_i) = \mathbf{A}(y_i, \vec{\varphi}), \quad i = 1, \dots, m, \quad (1.5-14)$$

где  $\vec{\varphi} = (\varphi_1, \dots, \varphi_n)$  – конечномерный вектор размерности  $n$ . Этот переход – редукция задачи на конечномерные пространства – неизбежен, поскольку интерпретация результатов измерений, представляющих собой конечный набор величин, методами вычислительной математики влечет получение результата также в форме конечного набора величин. При таком переходе от непрерывных функций к их конечномерному представлению обычно подразумевается, что с некоторой погрешностью возможен и обратный переход, т. е. конечный набор чисел  $\vec{\varphi} = (\varphi_1, \dots, \varphi_n)$ , полученный нами в результате решения обратной задачи, позволит с некоторой погрешностью восстановить непрерывную функцию  $\varphi(x)$ .

Этому этапу в решении обратных задач очень часто не уделяется необходимого внимания, что может приводить к существенным искажениям получаемых результатов. В частности, из интуитивных соображений понятно, что устанавливаемое им соответствие между конечномерными векторами и непрерывными функциями приводит к

существенным ограничениям на эти функции, т. е. на получаемое решение задачи.

Одним из сопутствующих переходу от непрерывного к дискретному представлению задачи аспектов является переход от непрерывных вариационных производных к их конечномерному аналогу. В принципе последний переход может быть реализован двумя способами:

1) непосредственная редукция вариационных производных как непрерывных функций к конечномерному представлению;

2) рассмотрение конечномерной задачи (1.5-14) как самостоятельной и соответствующее рассмотрение вместо вариационных производных частных производных  $\partial A(v_i)/\partial \varphi_i$  по координатам вектора  $\vec{\varphi}$ .

С нашей точки зрения второй путь, безусловно, предпочтительнее, поскольку в первом случае мы дважды должны выполнить весьма тонкую и влияющую на результат решения задачи процедуру перехода к конечномерному представлению. Во втором случае этот переход происходит один раз, и использование частных производных в рамках решения уже сформулированной конечномерной задачи не может привести ни к каким искажениям решения.

### **Примеры вариационных производных**

Для иллюстрации рассмотрим приведенные на рис. 1.4-9 вариационные производные рассеянного атмосферой солнечного излучения по профилю относительного отклонения от среднего концентрации озона. На рис. 1.4-9 можно видеть, что форма и положение максимумов вариационных производных (разумеется, вычислялся их конечномерный аналог) существенно различаются для различных областей спектра и даже для различных каналов в одной и той же области спектра. Анализ подобных рисунков часто позволяет оценить, насколько те или иные измерения могут оказаться пригодны для измерений конкретного атмосферного параметра. Так, из рис. 1.4-9 видно, что в измерениях излучения в УФ и видимой областях спектра можно получить информацию о содержании озона в диапазоне высот 20—70 км, поскольку на этих высотах расположены максимумы вариационных производных.

## Упражнение

Рассчитайте вариационную производную яркостной температуры нисходящего излучения  $T_{\text{я}}(\nu)$  как функционала от температурного профиля  $T(z)$  в предположении, что функция пропускания  $P_{\nu}(z)$  не зависит от температуры. Напомним, что  $T_{\text{я}}(\nu)$  можно выразить следующим образом:

$$T_{\text{я}}(\nu) = - \int_0^{\infty} T(z) \frac{dP_{\nu}(z)}{dz} dz .$$

## Глава 2

### Общие особенности некорректных обратных задач

#### 2.1. Наглядные примеры, иллюстрирующие особенности решения обратных задач

Туоми [24] привел целый ряд примеров, облегчающих понимание особенностей решения рассматриваемых нами обратных задач.

#### Примеры задач, не имеющих единственного решения

К классу обратных задач, например, относится такая. В зоопарке есть птицы и звери. Следует определить, сколько их в зоопарке по отдельности, если общее число их ног – 300, а голов – 100. Соображения, на которых может основываться решение, очевидны: птицы и звери имеют по одной голове, птицы – по две ноги, звери – по четыре. Если обозначить  $a$  – число зверей,  $b$  – число птиц, то можно сформировать систему двух линейных уравнений с двумя неизвестными

$$\begin{cases} a + b = 100, \\ 4a + 2b = 300. \end{cases}$$

Решение ее тривиально:  $a = b = 50$ .

Сформулируем другую задачу. В зоопарке есть тигры и львы. Если снова считать, что они имеют 100 голов и 300 ног, тогда система уравнений запишется

$$\begin{cases} l + t = 100, \\ 4l + 4t = 300 \end{cases}$$

( $l$  и  $t$  – число львов и тигров соответственно). Эти уравнения, однако, несовместимы, т. е. взаимно противоречат друг другу, так как второе

уравнение можно переписать:  $l + t = 75$ , что противоречит первому. Значит, такая задача не имеет решения, поскольку исходные данные задачи противоречивы.

Пусть в зоопарке есть снова тигры и львы. В сумме они имеют 100 голов и 400 ног. Соответствующие уравнения

$$\begin{cases} l + t = 100, \\ 4l + 4t = 400. \end{cases}$$

Эти уравнения являются совместимыми, но они не дают информации о числе львов и тигров по отдельности. Такая задача имеет много различных решений (например,  $l = 3, t = 97$  или  $l = 50, t = 50$ ), что не позволяет нам ответить на вопрос о количествах тигров и львов.

Приведенные примеры наглядно иллюстрируют тот факт, что система  $N$  уравнений с  $N$  неизвестными может не иметь решения или иметь несколько допустимых решений.

Рассмотрим систему алгебраических уравнений

$$\begin{cases} x + y + z = 2, \\ 2x + y + 3z = 9, \\ 3x + 2y - z = -1. \end{cases}$$

Легко получить ее решение:  $x = 1, y = -1, z = 2$ . Следующая система уравнений:

$$\begin{cases} x + y + z = 2, \\ 2x - y + 3z = 9, \\ 4x + y + 5z = 12, \end{cases} \quad (2.1-1)$$

демонстрирует несовместимость уравнений. Действительно, складывая первые два уравнения и вычитая первое из третьего, получим очевидное противоречие

$$\begin{cases} 3x + 4z = 11, \\ 3x + 4z = 10. \end{cases}$$

Теория систем алгебраических уравнений (см. Приложение) говорит о том, что подобные проблемы возникают, если определитель системы равен нулю.

Действительно, определитель системы (2.1-1)

$$\Delta = \begin{vmatrix} 1 & 1 & 1 \\ 2 & -1 & 3 \\ 4 & 1 & 5 \end{vmatrix} = 0.$$

### Примеры неустойчивости решения

Проблемы неустойчивости решения обратной задачи иллюстрируют следующие примеры. Рассмотрим систему 2 уравнений

$$\begin{cases} x + y = 2, \\ 2x + 2,000001y = 4,000001. \end{cases} \quad (2.1-2)$$

Решение системы очевидно:  $x=1, y=1$ . Однако предположим, что значения в правой части уравнений получены в результате измерений или у нас есть ограничения в вычислительной технике по точности представления исходных данных. Теперь рассмотрим систему уравнений

$$\begin{cases} x + y = 2, \\ 2x + 2,000001y = 4, \end{cases} \quad (2.1-3)$$

отличающуюся от (2.1-2) тем, что правая часть второго уравнения задана с очень маленькой ошибкой (0,000001!).

Система (2.1-3), однако, имеет решения, кардинально отличающиеся от решения системы (2.1-2):  $x=2, y=0$ .

Из общих соображений понятно, что в реальных обратных задачах мы практически никогда не знаем абсолютно точно ядер соответствующих интегральных уравнений, т. е. элементов соответствующей прямой матрицы  $\mathbf{A}$ . Это аналогично тому, что один из коэффициентов системы (2.1-2) задан с ошибкой. Например, рассмотрим систему

$$\begin{cases} x + y = 2, \\ 2x + 2y = 4,000001. \end{cases} \quad (2.1-4)$$

Попытка ее решения приводит к противоречию:  $0 = 0,000001$ . Если, наконец, оставить в уравнении (2.1-2) только целые числа, допустив очень маленькие погрешности (0,000001) в задании коэффициента и правой части уравнения, то наша система превратится в

$$\begin{cases} x + y = 2, \\ 2x + 2y = 4. \end{cases} \quad (2.1-5)$$

В этом случае второе уравнение просто повторяет первое, и система имеет бесконечное множество решений.

### Упражнение

Рассмотрите решения системы, полученной добавлением малой величины 0,000001 последовательно к различным коэффициентам системы (2.1-1). Сравните полученные решения между собой.

## 2.2. Примеры некорректности обратных задач математической физики

### Неразрешимость и неоднозначность решения интегрального уравнения

В разделе 1.5 было показано, что общее операторное нелинейное уравнение (1.4-6), к которому сводятся обратные задачи атмосферной оптики, линеаризацией может быть сведено к линейному интегральному уравнению Фредгольма I рода

$$\int_a^b A(x, y) \varphi(x) dx = f(y), \quad y \in [c, d], \quad (2.2-1)$$

где  $A(x, y)$  – ядро уравнения, определяющее линейный оператор  $\mathbf{A}$  прямой задачи, связывающий неизвестную функцию  $\varphi(x)$  ( $\delta\varphi(s)$  в разделе 1.5) с измеряемой (или связанной с измеряемой  $\delta f(y)$ ) функцией  $f(y)$ .

Прежде всего покажем, что при некоторых видах ядра  $A(x, y)$  решение уравнения (2.2-1) может не существовать вообще или существовать не для всякой правой части  $f(y)$ . Пусть, например,

$$A(x, y) = y + x, \quad (2.2-2)$$

тогда

$$\int_a^b (x+y)\varphi(x)dx = \int_a^b x\varphi(x)dx + y\int_a^b \varphi(x)dx = f_0 + f_1 y,$$

где  $f_0$  и  $f_1$  – числа,  $f_0 = \int_a^b x\varphi(x)dx$ ,  $f_1 = \int_a^b \varphi(x)dx$ . Следовательно, правая часть уравнения (2.2-1)  $f(y)$  в этом случае *должна* быть линейной функцией от  $y$ . Если правая часть не является линейной функцией, то решения уравнения (2.2-1) не существует. Если же правая часть линейна и решение существует, оно не единственно. Для рассмотренного примера, если правая часть  $f(y) = f_0 + f_1 y$ , то любая интегрируемая на  $[a, b]$  функция  $\varphi(x)$  при условиях нормировки

$$\int_a^b \varphi(x)dx = f_1 \text{ и } \int_a^b x\varphi(x)dx = f_0 \quad (2.2-3)$$

является решением уравнения (2.2-1) с ядром  $A(x, y) = y + x$ .

Мы продемонстрировали, что интегральное уравнение Фредгольма I рода может не иметь решений или иметь неединственное решение. Но даже при условии существования и единственности решения оно может обладать одной специфической особенностью, затрудняющей решение. Эта особенность – неустойчивость решения, которая проявляется при наличии погрешности задания правой части уравнения. Одним из проявлений такой неустойчивости является сглаживающее действие ядра уравнения.

### Сглаживающее действие ядра уравнения свертки

Частным случаем уравнения (2.2.1) является уравнение «свертки»

$$\int_{-\infty}^{\infty} A(x-y)\varphi(x)dx, \quad (2.2-4)$$

в котором ядро уравнения зависит только от разности аргументов.

Пусть  $A(x, y)$ , как функция  $x$ , непрерывная и достаточно гладкая функция. Иными словами, существует число  $M$ , ограничивающее одновременно абсолютную величину значений ядра и его производной, т. е.  $|A(x, y)| < M$ ,  $\left| \frac{dA(x, y)}{dx} \right| < M$ ,  $x \in [a, b]$ ,  $y \in [c, d]$ .

Это требование всегда выполняется для любой обратной задачи атмосферной оптики. Для количественной оценки влияния ошибок измерений на решение нам потребуется сравнивать «близость» различных пар элементов тех или иных функциональных пространств. В связи с этим воспользуемся понятием *расстояния* в пространствах функций, краткая информация о котором содержится в приложении. Учитывая, что в реальной ситуации  $f(y)$  известна с погрешностью за счет различных «шумов», включим в рассмотрение ошибки измерений  $\varepsilon(y)$ . Рассмотрим вновь уравнение

$$f(y) = \int_a^b A(x, y) \varphi(x) dx + \varepsilon(y). \quad (2.2-5)$$

Предположим, что  $\varphi_1(x)$  – решение уравнения (2.2-5) при  $f(y) = f_1(y)$ . Рассмотрим функцию

$$\varphi_2(x) = \varphi_1(x) + N \sin(\omega x), \quad (2.2-6)$$

где  $N$  и  $\omega$  произвольные константы. Подставляя функцию (2.2-6) в уравнение (2.2-5), видим, что  $\varphi_2(x)$  является решением уравнения

$$\int_a^b A(x, y) \varphi(x) dx = f_2(y) = f_1(y) + N \int_{a'}^{b'} A(x, y) \sin(\omega x) dx.$$

Оценим расстояние между функциями  $f_1(y)$  и  $f_2(y)$ , соответствующим двум решениям  $\varphi_1$  и  $\varphi_2$ . При этом в пространстве решений можно использовать метрику пространства  $\mathbf{C}_{[a,b]}$ , так как наши решения есть физические функции и мы можем считать их непрерывными. В пространстве измерений  $\mathbf{F}$  целесообразно использовать метрику  $\mathbf{L}_2$ , так как мы никогда не знаем конкретных значений ошибок  $\varepsilon(y)$ , а можем охарактеризовать их только в среднеквадратическом смысле, например с помощью дисперсии шумов. Тогда

$$\rho_{L_2}(f_1, f_2) = \sqrt{\int_c^d N^2 \left( \int_a^b A(x, y) \sin(\omega x) dx \right)^2 dy}. \quad (2.2-7)$$

Для оценки этого расстояния преобразуем внутренний интеграл, воспользовавшись формулой интегрирования по частям

$$\int_a^b A(x, y) \sin(\omega x) dx =$$

$$= -\frac{1}{\omega} \cos(\omega x) A(x, y) \Big|_a^b + \frac{1}{\omega} \int_a^b \cos(\omega x) \frac{dA(x, y)}{dx} dx .$$

Поскольку

$$|\cos(\omega x)| \leq 1, |A(x, y)| < M \text{ и } \left| \frac{dA(x, y)}{dx} \right| < M ,$$

то верны оценки

$$\left| \cos(\omega x) A(x, y) \Big|_a^b \right| < 2M(b-a) \text{ и } \left| \int_a^b \cos(\omega x) \frac{dA(x, y)}{dx} dx \right| < M(b-a).$$

Тогда

$$\int_a^b A(x, y) \sin(\omega x) dx < \frac{1}{\omega} (b-a) 3M ,$$

$$\rho_{L_2}(f_1, f_2) < \frac{1}{\omega} N(b-a) 3M \sqrt{d-c} . \quad (2.2-8)$$

Таким образом, при любом  $N$  выбором достаточно большой величины  $\omega$  можно добиться, чтобы функции  $f_1(y)$  и  $f_2(y)$  были сколь угодно близки (в среднеквадратическом смысле, т. е. в метрике  $L_2$ ). В то же время при достаточно большом значении  $N$  отклонение решений  $\varphi_1(x)$  и  $\varphi_2(x)$  может быть очень большим

$$\rho_C(\varphi_1, \varphi_2) = \max_{x \in [a, b]} |N \sin(\omega x)| = N . \quad (2.2-9)$$

Из соотношений (2.2-8) и (2.2-9) следует, что, выбирая достаточно большие  $N$  и  $\omega$ , можно получить два решения, сколь угодно отличающиеся друг от друга, при этом соответствующие им измерения отличаются сколь угодно мало, т. е. решение неустойчиво.

С другой стороны, полученные соотношения показывают, что гладкость ядра уравнения свертки приводит к принципиальной невозможности различить на основании анализа результатов измерений  $f(y)$ , выполненных с конечными погрешностями, решение  $\varphi(x)$ , содержащее высокочастотные компоненты и сглаженное

решение, этих компонент не содержащее. Это свойство и называют сглаживающим действием ядра.

Приведенный пример также демонстрирует важное свойство уравнения (2.2-5) с приближенно (как и бывает всегда на практике) заданным  $f(y)$  и, следовательно, обратных задач атмосферной оптики (ранее мы показали, что они сводятся к интегральным уравнениям Фредгольма или Вольтерра I рода) – *неустойчивость* решения относительно изменения входных данных. Это свойство приводит к тому, что для двух измерений  $f(y)$ , неразличимых в пределах ошибок измерений, можно получить сколь угодно сильно отличающиеся решения  $\varphi$ . Иными словами, из приведенного примера видно, что уравнение (2.2-5) в этих условиях не имеет единственного решения. В связи с этим возникает вопрос, что же следует понимать под решением рассматриваемого уравнения при наличии измерений  $f(y)$  с ошибками?

Анализируя рассмотренный пример, следует подчеркнуть существенное значение ограничения абсолютной величины производной  $dA(x,y)/dx$ . Рассмотрим рис. 2.2-1, который схематически иллюстрирует сглаживающее действие ядра. Интуитивно понятно, что ядро, имеющее характерный «размер» деталей своей структуры  $\Delta x$  (верхняя часть рис 2.2-1), не позволяет обнаружить особенности поведения решения (нижняя часть рис 2.2-1) масштаба мельче чем  $\Delta x$ .

Иными словами, особенности поведения функции, график которой показан тонкой линией в нижней части рисунка, не могут быть обнаружены при измерениях с осредняющим ядром, показанным в верхней части того же рисунка. То есть различить функции, графики которых показаны тонкой и толстой линиями, невозможно. Уменьшение величины  $\Delta x$  при сохранении значений ядра  $A(x,y)$  приведет к росту производной  $dA(x,y)/dx$  и одновременно к улучшению пространственного разрешения (обычно переменная  $x$  ассоциирована с высотой или другой пространственной переменной).

Это рассуждение показывает, что «сглаживающее действие ядра», приводящее к неразличимости высокочастотных компонент решения, вызвано ограничением величины производной ядра и ужесточение этого ограничения приводит к ухудшению

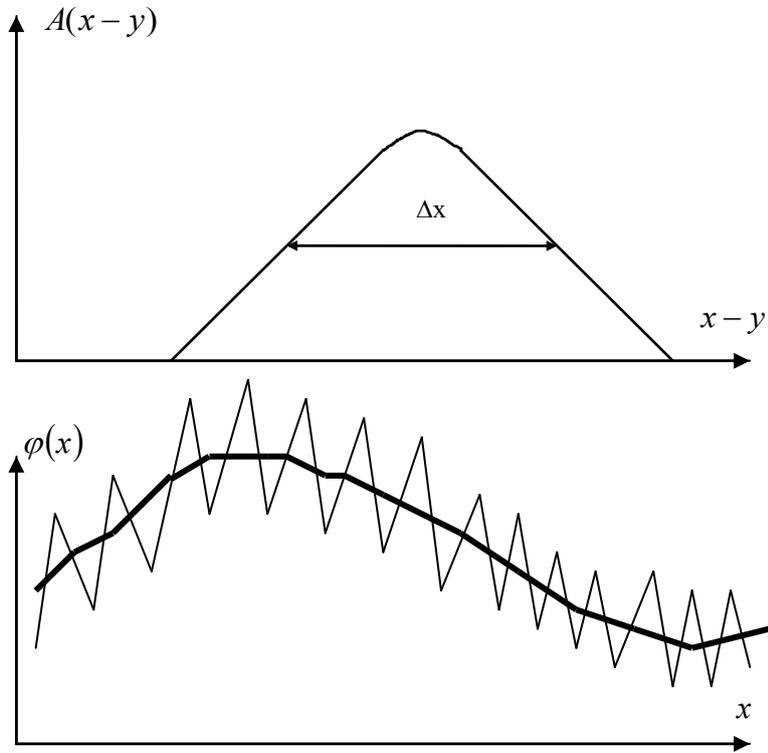


Рис. 2.2-1. Сглаживающее ядро интегрального уравнения.

чувствительности измерений  $f(y)$  к высокочастотным (мелко-масштабным) компонентам решения.

### Редукция спектра с конечным спектральным разрешением

Другой пример, иллюстрирующий некорректность рассматриваемой задачи, связан с задачей спектроскопии – редукцией спектра, записанного на приборе с конечным спектральным разрешением, к монохроматическому спектру, измеряемому гипотетическим «идеальным» прибором. В этом случае соответствующая задача формулируется в виде уравнения свертки

$$\int_{-\infty}^{\infty} A(y-x)\varphi(x)dx = f(y). \quad (2.2-10)$$

При этом естественно также предполагать, что регистрируемая функция  $f(y)$  есть сумма точной правой части  $f(y)$  и стационарного случайного процесса (см. Приложение)  $\varepsilon(y)$  (помехи) с нулевым средним и корреляционной функцией  $\Delta(\eta) = \langle \varepsilon(y) \varepsilon(y + \eta) \rangle$  (здесь  $\langle \rangle$  – знак статистического усреднения). Формальное решение уравнения (2.2-10) легко получить с помощью преобразования Фурье

$$\varphi(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ipx} \frac{\tilde{f}(p)}{\tilde{A}(p)} dp, \quad (2.2-11)$$

где знаком  $\sim$  отмечено преобразование Фурье от соответствующей функции. Если мы учтем шум измерений, т. е. подставим вместо  $f$  в правую часть формулы (2.2-10)  $f^* = f + \varepsilon$ , то соответствующее решение  $\varphi^*(x)$ , очевидно, будет зависеть от реализации  $\varepsilon(y)$  в конкретном эксперименте, и также может рассматриваться как случайная функция. Найдем дисперсию функции  $\varphi^*(x)$ . Можно показать, что соответствующее выражение имеет вид

$$D(\varphi^*) = \langle (\varphi^*)^2 \rangle - \langle \varphi^* \rangle^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{G(p)}{|\tilde{A}(p)|^2} dp, \quad (2.2-12)$$

где  $G(p) = \tilde{\Delta}(p)$  – спектр мощности процесса  $\varepsilon(y)$  (помехи). Отметим, что, поскольку аппаратная функция прибора  $A(y-x)$  достаточно гладкая, ее преобразование Фурье  $\tilde{A}(p)$  не содержит высокочастотных компонент, т. е.

$$|\tilde{A}(p)| \rightarrow 0 \text{ при } |p| \rightarrow \infty. \quad (2.2-13)$$

Поэтому для того чтобы дисперсия решения задачи (2.2-10) оставалась конечной, спектр мощности шумов  $G(p)$  должен достаточно быстро убывать при  $|p| \rightarrow \infty$ . Это накладывает жесткие ограничения на класс процессов  $\varepsilon(y)$ , допустимых в качестве помех. На практике эти условия никогда не выполняются, так как помеха всегда содержит компоненту «белого шума», спектр мощности которого – постоянная величина и, следовательно, при  $|p| \rightarrow \infty$  спектр мощности  $G(p)$  стремится к конечному пределу. Поэтому выражение под интегралом в

формулах (2.2-11) и (2.2-12) неограниченно растет с ростом  $p$  и дисперсию решения по формуле (2.2-12) можно считать бесконечной, а подстановка в формулу (2.2-11) вместо  $\tilde{f}(p)$  найденной из эксперимента функции  $\tilde{f}^*(p)$  невозможна.

### Упражнение

Постройте 3–4 существенно различных решения уравнения (2.2-1) с ядром (2.2-2). (Используйте, например, полиномы различных степеней или тригонометрические функции.)

### 2.3. Формулировка понятия корректности задач математической физики

Как показали примеры, приведенные в двух предыдущих разделах, одной из особенностей обратных задач является их *неустойчивость* к изменению входных данных. Это означает, что малые вариации исходных данных могут приводить к большим (не убывающим с уменьшением вариации исходных данных) вариациям решения. Важность этой особенности очевидна. Входные данные задачи – это результаты того или иного эксперимента, поэтому мы никогда не знаем истинную измеряемую функцию  $f(v)$ , а располагаем сведениями о функции  $f^*(v) = f(v) + \varepsilon(v)$ , где  $\varepsilon(v)$  – «помеха», обусловленная шумами приемника и электроники, связанная с градуировкой измерений, вызванная естественной стохастичностью сигнала и т. д.

Долгое время в математике (классической математике) вообще не рассматривались такие неустойчивые задачи, поскольку преобладала точка зрения французского математика Ж. Адамара, что подобные задачи не имеют физического смысла (ее придерживался также известный специалист в области интегральных уравнений профессор СПбГУ Г.С. Михлин). Тем не менее физикам часто удавалось их успешно решать, применяя те или иные методы стабилизации решения.

## Корректность по Адамару

Еще в начале века Ж. Адамар ввел требования, которым должна удовлетворять любая задача матфизики – требования *корректности* (об используемом ниже понятии метрических пространств см. Приложение).

Рассмотрим отображение (оператор)  $\mathbf{A}$ , сопоставляющее каждой функции  $\varphi(s)$  из метрического пространства  $\Phi$  функцию  $f(v)$  из метрического пространства  $\mathbf{F}$ . Следуя Адамару, будем называть задачу определения решения  $\varphi(s)$  по известной функции  $f(v)$  корректно поставленной на паре метрических пространств  $(\mathbf{F}, \Phi)$ , если удовлетворяются следующие три условия:

- 1) для всякого элемента  $f \in \mathbf{F}$  существует решение  $\varphi$  из пространства  $\Phi$ ;
- 2) решение  $\varphi$  единственно;
- 3) задача устойчива на паре пространств.

Последнее условие означает, что малым вариациям  $f$  соответствуют малые вариации  $\varphi$  (в выбранных в пространствах  $\Phi$  и  $\mathbf{F}$  метриках).

Задачи, не удовлетворяющие одному из этих сформулированных требований, называются некорректно поставленными (в классическом смысле или по Адамару). Подчеркнем, что определение некорректно поставленных задач относится к *рассматриваемой паре метрических пространств*  $(\mathbf{F}, \Phi)$ , так как в других метриках та же задача может быть корректно поставленной.

Как правило, большое число задач интерпретации данных эксперимента, в том числе обратных задач атмосферной оптики, приводит к некорректно поставленным, в классическом смысле, задачам. Добавление слов «в классическом смысле» связано с тем, что позднее А. Н. Тихоновым были предложены новые три условия корректности задач матфизики именно в связи с разработкой современных численных методов решения разнообразных обратных (и неустойчивых) задач матфизики.

## Корректность по Тихонову

Сформулируем условия корректности по Тихонову. При этом кроме метрических пространств  $\Phi$ ,  $F$  и оператора  $A$  задано некоторое замкнутое множество  $M \subset \Phi$ . Назовем задачу корректной по Тихонову, если выполняются следующие условия:

- 1) априори известно, что для некоторого класса данных  $f$  существует решение  $\varphi$ , принадлежащее заданному множеству  $M \subset \Phi$ ;
- 2) решение единственно в классе функций  $M \subset \Phi$ ;
- 3) бесконечно малым вариациям  $f$ , не выводящим решение  $\varphi$  за пределы  $M \subset \Phi$ , соответствуют бесконечно малые вариации решения  $\varphi$ .

Обозначая  $M_A = A(M)$  – образ множества  $M \subset \Phi$ , третье условие можно сформулировать следующим образом:

- 3') решение  $\varphi$  непрерывно зависит от  $f$  на  $M_A$ .

Необходимость рассмотрения, наряду с  $M$  и  $M_A$ , пространств  $\Phi$  и  $F$  обусловлена тем, что в реальных задачах погрешности измерений  $f$  обычно выводят решение задачи из множества  $M$ .

Одним из типичных примеров некорректных, в классическом смысле, задач математической физики является задача решения интегрального уравнения Фредгольма I рода

$$f(y) = \int_a^b A(x, y) \varphi(x) dx, \quad y \in [c, d], \quad (2.3-1)$$

где  $A(x, y)$  – ядро интегрального уравнения,  $f(y)$  – известная функция, а  $\varphi(x)$  – искомая функция.

Как показал Г. М. Марчук (см. также раздел 1.4), в линейном приближении все обратные задачи атмосферной оптики можно свести к решению этого интегрального уравнения. Выше было показано, что это уравнение не удовлетворяет определению корректности по Адамару.

Заметим, что в задачах физики, в том числе в задачах дистанционного зондирования, мы уверены в существовании решения.

Если имеется, например, измеренное уходящее собственное излучение атмосферы и мы знаем, что оно является тепловым, нет сомнений, что оно обусловлено излучением элементарных объемов атмосферы, находящихся при определенной кинетической температуре. Таким образом, существование определенного термического состояния атмосферы, описываемого вертикальным профилем кинетической температуры, не подвергается сомнению.

Для полноты рассмотрения заметим, что реальные ситуации оказываются сложнее последнего рассуждения. Нельзя забывать о том, что математическая модель физического явления всегда является в той или иной степени его приближенным описанием (как, например, Ньютоновская механика, не учитывающая релятивистских эффектов теории относительности). Решая обратные задачи, мы опираемся именно на математическую модель природных явлений. Поэтому, например, правильность утверждения о существовании термического состояния атмосферы, приводящего к наблюдаемому излучению, не всегда означает существование решения интегральных уравнений, которые мы используем для описания переноса излучения. Тем не менее потенциальная возможность уточнения математической модели явления позволяет говорить о принципиальном существовании решения.

## Упражнения

1. Сформулируйте возможно лаконичнее определение корректности по Адамару (нам удалось уложиться в фразу из четырех слов).
2. То же для корректности по Тихонову.

### **2.4. Алгебраизация некорректной задачи и обусловленность**

При численном решении интегрального уравнения Фредгольма I рода

$$f(y) = \int_a^b A(x, y)\varphi(x)dx, \quad y \in [c, d], \quad (2.4-1)$$

его практически всегда тем или иным способом сводят к решению системы линейных алгебраических уравнений

$$\sum_{i=1}^n a_{i,j} \varphi_i = f_j, \quad j = 1, \dots, m, \quad (2.4-2)$$

где  $\varphi_i$  и  $f_j$  – линейные функционалы от функций  $\varphi(x)$  и  $f(y)$ , а именно либо их значения в опорных точках, либо коэффициенты разложения по ортогональной системе функций. Если забыть о некорректности обратных задач, то проблема состоит, казалось бы, только в выборе порядка приближения  $n$  (квадратуры или числа членов разложения), достаточного для получения требуемой точности. Однако некорректность задачи проявляется в плохой обусловленности системы уравнений (2.4-2), т. е. в чрезвычайно сильной зависимости решения от вариаций свободного члена, а также от ошибок коэффициентов  $a_{i,j}$  и от ошибок счета. Как показано Фаддеевым, обусловленность системы (2.4-2) тесно связана с набором собственных чисел  $\{\lambda_k\}_{k=1,n}$  матрицы  $\mathbf{A}'\mathbf{A}$  ( $\mathbf{A}$  – матрица системы (2.4-2),  $\mathbf{A}'$  – транспонированная матрица). Обусловленность (устойчивость) падает с ростом отношения  $\lambda_{\max} / \lambda_{\min}$  (где  $\lambda_{\max}$  и  $\lambda_{\min}$  – максимальное и минимальное собственные числа матрицы  $\mathbf{A}'\mathbf{A}$ ). При этом решение «разбалтывается» в направлении собственных векторов  $\vec{\psi}_k$  матрицы  $\mathbf{A}'\mathbf{A}$ , соответствующих малым собственным числам  $\lambda_k$ . Более того, чувствительность проекции вектора решения  $\vec{\varphi}$  на направление  $\vec{\psi}_k$  к вариациям компонент вектора  $\vec{f}$  и элементов матрицы  $\mathbf{A}$  пропорциональна  $\lambda_k^{-1}$ . Следовательно, для больших  $\lambda_k$  эта чувствительность мала, а для малых  $\lambda_k$  может быть очень велика. Поэтому, в частности, требования к точности вычислений быстро растут по мере роста отношения  $\lambda_{\max} / \lambda_{\min}$  даже при точно заданном векторе  $\vec{f}$ .

Алгебраизация некорректной задачи всегда (при достаточно большом порядке приближения  $n$ ) дает плохо обусловленную систему уравнений, так как можно показать, что если исходная задача некорректна, то  $\lambda_{\max} / \lambda_{\min} \rightarrow \infty$  при  $n \rightarrow \infty$ . Следовательно, выбирая порядок приближения достаточно большим, можно сделать систему уравнений (2.4-2) сколь угодно плохо обусловленной.

## **2.5. Эмпирические подходы к решению некорректных обратных задач**

Ученые в различных областях естествознания решают обратные задачи многие сотни лет, что очевидно из общего определения обратных задач на основе причинно-следственного хода. Так, медики решают их уже тысячелетия. Во многих случаях такие обратные задачи решались достаточно успешно. Это связано с тем, что в процессе конкретной постановки обратной задачи можно существенно «снизить» ее некорректность, т. е. повысить устойчивость решения.

### **Методы, повышающие устойчивость решения**

Например, в задачах инструментальной спектроскопии, если интересоваться не всем истинным спектром поглощения газовой смеси  $P(\lambda)$ , а только положениями центров спектральных линий поглощения и их интегральными интенсивностями, то такая задача (при выполнении определенных требований к эксперименту) может быть решена успешно без использования специальных математических «ухищрений». Таким образом, если решение формулируется относительно определенных функционалов от искомой функции  $\varphi(x)$ , то мы можем и не столкнуться с эффектами неустойчивости задачи, обусловленными ее некорректностью.

Если говорить о более общих методах решения, то, по существу, первая попытка решения некорректного уравнения свертки (2.2-2) была предпринята еще лордом Релеем в 1871 году, предложившим итерационный метод коррекции щелевых искажений в спектроскопии. В дальнейшем были предложены разнообразные итерационные методы решения уравнения свертки, основанные на том, что во многих случаях аппаратная функция прибора имеет вид достаточно узкого импульса по сравнению с регистрируемым спектром. Однако обнаружилось, что в ходе итерационного процесса могут возникать осцилляции решения (вспомните рассмотрение сглаживающего действия ядра из раздела 2.2), и проблема на практике сводится к выбору «оптимального» решения, т. е. выбору числа итераций.

Другой общий метод решения состоит в решении системы уравнений (2.4-2). Если алгебраизация проведена удачно, то иногда

удается получить приемлемое решение, ограничившись небольшим порядком приближения  $n$ , при котором неустойчивость еще не проявляется. Так, в первых работах по термическому зондированию атмосферы Л. Каплана, Д. Уорка атмосфера разбивалась на небольшое число слоев, и в численных экспериментах удавалось достаточно успешно восстанавливать средние температуры слоев (достаточно протяженных). Однако в дальнейшем естественное желание получить более адекватное представление вертикального профиля температуры заставляло исследователей повышать порядок приближения  $n$ , и тут проявлялась неустойчивость обратной задачи. При увеличении  $n$  точность восстановления профиля температуры  $T(z)$  ухудшалась.

Более перспективным подходом является метод аппроксимации искомой функции с помощью того или иного ортогонального базиса, согласованного со свойствами искомого решения. Так, В. И. Татарский и М. С. Малкевич решали задачу термического зондирования атмосферы, используя в качестве базиса эмпирические ортогональные функции ковариационной матрицы температуры. Как показано А. М. Обуховым [11], этот базис является оптимальным в том смысле, что при любом порядке приближения  $n$  он дает наивысшую точность аппроксимации по сравнению с другими базисами той же размерности. Это позволяло использовать невысокий порядок приближения  $n$  и избежать явных эффектов неустойчивости обратной задачи. Однако и при этом подходе желание более точной аппроксимации искомой функции приводило к увеличению порядка аппроксимации  $n$  и проявлениям неустойчивости.

В последнем подходе видно, что устойчивость решения может быть достигнута за счет исключения «высокочастотных» гармоник решения, которые непосредственно «продуцируются» погрешностями измерений. (см. раздел 2.2). Иными словами, наличие шума в измерениях приводит к необоснованному появлению в решении высокочастотных компонент. В связи с этим можно предложить очевидный устойчивый метод решения уравнения свертки.

Вместо формального аналитического, но мало конструктивного решения (2.2-11) можно ограничиться восстановлением спектра сигнала  $f(p)$  в конечной полосе частот  $|p| \leq p_0$  и записать приближенное решение уравнения свертки в виде

$$\varphi^*(x) = \frac{1}{2\pi} \int_{-p_0}^{p_0} e^{ipx} \frac{\tilde{f}(p)}{\tilde{A}(p)} dp. \quad (2.5-1)$$

Дисперсия приближенного решения  $\varphi^*$  определяется соответственно «урезанным» интегралом (2.2-12)

$$D(\varphi^*) \langle \varphi^{*2} \rangle - \langle \varphi^* \rangle^2 = \frac{1}{2\pi} \int_{-p_0}^{p_0} \frac{G(p)}{|\tilde{A}(p)|^2} dp \quad (2.5-2)$$

и остается конечной, если  $p_0$  не слишком велика. Изменяя величину  $p_0$ , можно менять соотношение между «точностью приближения» искомой функции и величиной ошибки. При этом, правда, появляется типичный для обратных задач эффект – при попытке повышения точности приближения (за счет увеличения  $p_0$ ) выше некоторого предела автоматически возрастает ошибка. Для выбора частоты «среза»  $p_0$  привлекались различные соображения, а также предположения об искомой функции.

Имеется еще один способ регулирования степени приближения при использовании решения (2.2-11), являющийся обобщением рассмотренного подхода. Этот способ состоит в домножении подынтегрального выражения (2.2-11) на некоторую, вообще говоря, произвольную функцию  $\tilde{g}(p)$ , достаточно быстро убывающую при  $|p| \rightarrow \infty$ , чтобы соответствующий интеграл (2.2-12) оставался конечным. Тогда решение уравнения свертки можно записать в виде

$$\varphi^*(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{g}(p) e^{ipx} \frac{\tilde{f}(p)}{\tilde{A}(p)} dp. \quad (2.5-3)$$

Заметим, что выражение (2.5-3) принимает вид (2.5-1) при специальном выборе фактора  $\tilde{g}$ . Можно также показать, что использование выражения (2.5-3) эквивалентно нахождению «сглаженной» версии истинного решения

$$\varphi^*(x) = \int_{-\infty}^{\infty} g(x-x') \varphi(x') dx',$$

где  $g(x)$  получается из  $\tilde{g}(p)$  обратным преобразованием Фурье. Степень этого «сглаживания» можно регулировать, меняя параметры

функции  $\tilde{g}(\rho)$  или ее вид. Однако при таких подходах к решению уравнения (2.2-10) настораживает крайняя свобода (произвол) воздействия на получаемое решение.

Стремление избавиться от произвольных факторов и разработать общие методы решения некорректных задач привели к разработке новых подходов и методов решения некорректных задач. Фундаментальное значение для этих подходов имело соображение, что *некорректную задачу можно рассматривать как эффективно недоопределенную*. Действительно, при классическом (по Адамару) понимании некорректности в качестве решения задачи (2.4-1) может фигурировать любая функция  $\varphi$ , удовлетворяющая условию

$$\rho_1(f^*, A\varphi) \leq \varepsilon, \quad (2.5-4)$$

где  $\rho_1(f^*, f)$  – мера отклонения (расстояние) регистрируемой функции  $f^*$  от точной правой части  $f$ , зависящая от условий эксперимента. Поскольку среди этих функций («решений») имеются и «плохие» (даже физически бессмысленные), возникает задача *доопределения* задачи (2.5-4) с тем, чтобы получить единственное решение, по возможности близкое к истинному.

Доопределить задачу (2.5-4) можно различными способами. Однако любой способ доопределения должен основываться на тех или иных представлениях о характере искомого решения, т. е. на *априорной информации* о решении.

### Метод Филлипса

Первой работой данного направления явилась работа Д. Филлипса, в которой из множества функций  $\varphi$ , удовлетворяющих условию (2.5-4), предлагалось выбрать «самую гладкую» функцию, а точнее, функцию, минимизирующую норму производной

$$\int_a^b \left( \frac{d\varphi}{dx} \right)^2 dx = \min. \quad (2.5-5)$$

Обычно удается показать, что искомый минимум достигается на границе области, определяемой неравенством (2.5-4), и, следовательно, мы можем заменить условие (2.5-4) на равенство

$$\rho_1(f^*, A\varphi) = \varepsilon. \quad (2.5-6)$$

Возникает задача на условный экстремум. Будем решать ее методом Лагранжа. (см. Приложение). Для этого минимизируем следующее выражение относительно функции  $\varphi$  :

$$\int_a^b \left( \frac{d\varphi}{dx} \right)^2 dx + \lambda (\rho_1(f^*, A\varphi) - \varepsilon) = \min . \quad (2.5-7)$$

В теории некорректных задач традиционно используется несколько нестандартная форма записи метода Лагранжа. Поскольку задача минимизации решается при фиксированном  $\lambda$ , то для  $\lambda > 0$ , умножив выражение (2.5-7) на константу  $\alpha = \lambda^{-1}$  и отбросив постоянное слагаемое, получим эквивалентную задаче (2.5-7) задачу

$$\rho_1(f^*, A\varphi) + \alpha \int_a^b \left( \frac{d\varphi}{dx} \right)^2 dx = \min . \quad (2.5-8)$$

Решением задачи (2.5-8) является функция  $\varphi_\alpha$ , параметрически зависящая от  $\alpha$ . Подставив полученную зависимость в (2.5-6), можно определить величину  $\alpha$ , играющую теперь роль неопределенного множителя.

Для евклидовой метрики условие (2.5-8) приводит после алгебраизации к системе линейных уравнений. Для алгебраизированной системы (2.4-2) несколько более общая форма условия (2.5-5) была рассмотрена С. Туоми

$$\Omega(\varphi) = \sum_{i,j}^n h_{i,j} \varphi_i \varphi_j = \min , \quad (2.5-9)$$

где матрица  $\mathbf{H} = \{h_{i,j}\}$  положительно определена. В остальном подход Туоми совпадает с подходом Филлипса. Условие (2.5-8) принимает вид

$$(\mathbf{A}\varphi - f)^2 + \alpha \varphi' \mathbf{H} \varphi = \min , \quad (2.5-10)$$

а неопределенный множитель  $\alpha$  определяется подстановкой решения задачи (2.5-10), зависящего от  $\alpha$  как от параметра, в уравнение

$$\|\mathbf{A}\varphi - f\| = \varepsilon .$$

Заметим, что наряду с задачей (2.5-8) рассматривается также задача (2.5-8a) [15]

$$\left(\rho_1(f^*, A\varphi)\right)^2 + \alpha \int_a^b \left(\frac{d\varphi}{dx}\right)^2 dx = \min. \quad (2.5-8a)$$

А.Н. Тихонов ввел понятие *регуляризации* решения некорректной задачи. Под этим понимается построение семейства корректных задач, зависящих от параметра регуляризации  $\alpha$ , которое обладает тем свойством, что при  $\alpha \rightarrow 0$  и при одновременном стремлении к нулю погрешности правой части (измерений) решение корректной задачи стремится к истинному решению исходной некорректной задачи.

Хотя мы ничего не говорили о метрике  $\rho_1(f^*, f)$ , в которой выше оценивалась невязка измерений, описанные выше способы решения некорректной задачи во многих случаях исходили из предположения, что эксперимент позволяет контролировать точную верхнюю границу ошибки  $|f^*(y) - f(y)|$ . Использование этого предположения на практике приводит к слишком заглаженным решениям, что связано, по-видимому, с тем, что фактическая (случайная) ошибка обычно меньше своего максимального значения.

Кроме того, это предположение не отвечает природе реального эксперимента, неадекватно его характеризует и не позволяет корректно оценить ошибку решения задачи (ошибку восстановления искомой функции). Практически всегда погрешности экспериментальных измерений представляют собой случайную величину, подчиненную какому-либо непрерывному закону распределения вероятности. Наиболее часто погрешности, вызванные шумами аппаратуры, распределены по нормальному закону, хотя в ряде случаев (например квантовый шум при наблюдениях звезд) закон распределения погрешностей может быть и другим. Поэтому в дальнейшем были развиты подходы, рассматривающие проблему решения некорректной задачи с учетом статистического характера ошибок эксперимента и с использованием других типов априорной информации, в том числе статистической.

## Упражнение

Все рассмотренные методы решения неустойчивых обратных задач тем или иным образом основаны на некотором предположении об искомом решении. Каким обычно оказывается это предположение?

## 2.6. Оценки погрешности решения обратной задачи

Исследуем матричный аналог интегрального уравнения

$$\vec{f} = \mathbf{A}\vec{\varphi} + \vec{\varepsilon}, \quad (2.6-1)$$

где  $\vec{f}$  и  $\vec{\varphi}$  – вектора в евклидовых пространствах  $R^n$  и  $R^m$  соответственно,  $\vec{\varepsilon}$  –  $m$ -мерный вектор ошибок измерений. Переход от интегрального уравнения (1.4-6) к матричному аналогу, как мы уже упоминали, можно осуществить, либо используя квадратурную формулу, либо разлагая искомое решение в конечный ряд по какому-либо ортогональному базису.

Относительно ошибок измерений будем предполагать, что они являются гауссовыми с нулевым средним значением и заданной ковариационной матрицей. Для простоты изложения будем предполагать, что  $m = n$  (число измерений равно числу неизвестных) и все матрицы невырождены.

Оценим ошибку определения  $\vec{\delta}$  искомого вектора  $\vec{\varphi}$ , обусловленную погрешностями измерений  $\vec{\varepsilon}$ . Воспользуемся обычной евклидовой метрикой

$$\rho_{R^m}(\vec{\varphi}_1, \vec{\varphi}_2) = \sqrt{\sum_{j=1}^m (\varphi_{1j} - \varphi_{2j})^2}. \quad (2.6-2)$$

Запишем задачу (2.6-1), учтя явно наличие погрешностей решения  $\delta$

$$\mathbf{A}(\vec{\varphi} + \vec{\delta}) = \vec{f} + \vec{\varepsilon}.$$

Так как  $\mathbf{A}(\vec{\varphi} + \vec{\delta}) = \mathbf{A}\vec{\varphi} + \mathbf{A}\vec{\delta}$  и  $\mathbf{A}\vec{\varphi} = \vec{f}$ , то  $\mathbf{A}\vec{\delta} = \vec{\varepsilon}$ , откуда

$$\vec{\delta} = \mathbf{A}^{-1}\vec{\varepsilon}. \quad (2.6-3)$$

Тогда для норм векторов  $\vec{f}$  и  $\vec{\delta}$  можно записать следующие неравенства (будем для краткости записи опускать значок  $R^m$  у знака нормы):

$$\|\vec{f}\| \leq \|\mathbf{A}\| \|\vec{\varphi}\|; \quad \|\vec{\delta}\| \leq \|\mathbf{A}^{-1}\| \|\vec{\varepsilon}\|, \quad (2.6-4)$$

где  $\mathbf{A}^{-1}$  – обратная  $\mathbf{A}$  матрица, т. е.  $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ ,  $\|\mathbf{A}\|$  – норма матрицы  $\mathbf{A}$  (см. Приложение). Перемножив два последних неравенства, получим

$$\|\vec{f}\| \|\vec{\delta}\| \leq \|\mathbf{A}\| \|\mathbf{A}^{-1}\| \|\vec{\varphi}\| \|\vec{\varepsilon}\|. \quad (2.6-5)$$

Разделив обе части неравенства (2.6-5) на произведение  $\|\vec{f}\| \|\vec{\varphi}\|$ , получим следующее соотношение для оценки относительной ошибки определения вектора  $\vec{\varphi}$  :

$$\frac{\|\vec{\delta}\|}{\|\vec{\varphi}\|} \leq \|\mathbf{A}\| \|\mathbf{A}^{-1}\| \frac{\|\vec{\varepsilon}\|}{\|\vec{f}\|}. \quad (2.6-6)$$

Это неравенство показывает, что относительная ошибка решения уравнения (2.6-1), т. е.  $\|\vec{\delta}\|/\|\vec{\varphi}\|$ , может в  $\|\mathbf{A}\| \|\mathbf{A}^{-1}\|$  раз превосходить относительную ошибку измерений  $\|\vec{\varepsilon}\|/\|\vec{f}\|$ . Величину, характеризующую степень усиления ошибки измерений при решении уравнения (2.6-1), называют *коэффициентом усиления ошибки* или *числом обусловленности*. Таким образом, для рассматриваемого случая число обусловленности  $r$  определяется соотношением

$$r = \|\mathbf{A}\| \|\mathbf{A}^{-1}\|. \quad (2.6-7)$$

Легко увидеть, что для числа обусловленности справедливо неравенство

$$r = \|\mathbf{A}\| \|\mathbf{A}^{-1}\| \geq \|\mathbf{A}\mathbf{A}^{-1}\| = \|\mathbf{I}\| = 1, \quad (2.6-8)$$

т. е. всегда  $r \geq 1$ . Это означает, что относительная ошибка решения не может оказаться меньше относительной ошибки измерений, т. е. при решении задачи (2.6-1) относительная погрешность измерений, «переходя» в относительную погрешность решения может только возрастать, но не уменьшаться.

Матрицы, для которых  $r \gg 1$ , называются плохо обусловленными матрицами. Следует отметить, что полученный результат не связан с каким-либо конкретным методом решения уравнения (2.6-1) и характеризует лишь свойства входящей в это уравнение матрицы  $\mathbf{A}$ .

## 2.7. Анализ устойчивости обратной задачи в ортогональном базисе

Используем для дальнейшего анализа уравнения (2.6-1) метод разложения искомого решения по ортогональному базису. Для упрощения дальнейшего изложения преобразуем уравнение (2.6-1) так, чтобы оно содержало симметричную матрицу. Одним из способов приведения уравнения (2.6-1) к уравнению с симметричной матрицей является его умножение слева на матрицу  $\mathbf{A}'$ . Тогда

$$\mathbf{A}'\vec{f} = \mathbf{A}'\mathbf{A}\vec{\varphi} + \mathbf{A}'\vec{\varepsilon}. \quad (2.7-1)$$

В уравнении (2.7-1) матрица  $\mathbf{A}'\mathbf{A}$  симметрична, так как по правилам транспонирования матриц получаем

$$(\mathbf{A}'\mathbf{A})' = \mathbf{A}'(\mathbf{A}')' = \mathbf{A}'\mathbf{A}.$$

Введя обозначения  $\vec{f} = \mathbf{A}'\vec{f}$ ,  $\vec{\varepsilon} = \mathbf{A}'\vec{\varepsilon}$ ,  $\tilde{\mathbf{A}} = \mathbf{A}'\mathbf{A}$ , запишем преобразованное уравнение в том же виде, что и (2.6-1):

$$\vec{f} = \tilde{\mathbf{A}}\vec{\varphi} + \vec{\varepsilon}. \quad (2.7-2)$$

Далее знак «~» для простоты будем опускать, пользуясь формой записи (2.6-1), однако помня, что матрица  $\mathbf{A}$  является симметричной и нормальной (см. Приложение).

Воспользуемся рядом свойств скалярного произведения векторов в евклидовом пространстве и свойствами матриц, связанными с их собственными числами и векторами (см. Приложение).

Рассмотрим собственные числа  $\lambda_i$  и собственные вектора  $\vec{\psi}_i$ ,  $i=1, \dots, m$ , матрицы  $\mathbf{A}$ . Поскольку матрица  $\mathbf{A}$  была определена нами как произведение матрицы на транспонированную, то все ее собственные числа неотрицательны, а поскольку мы предполагаем ее невырожденность, то и положительны. Здесь и всюду в дальнейшем будем предполагать, что собственные числа упорядочены по убыванию, т. е.  $\lambda_i \geq \lambda_j$ , если  $i < j$ .

Как мы заметили, матрица  $\mathbf{A}$  нормальна, следовательно, все ее собственные вектора попарно ортогональны, т. е. их скалярное произведение равно нулю.

$$(\vec{\psi}_i, \vec{\psi}_j) = 0, \quad i \neq j. \quad (2.7-3)$$

Без потери общности будем считать собственные вектора  $\vec{\psi}_i, i = 1, \dots, m$ , нормированными:

$$(\vec{\psi}_i, \vec{\psi}_i) = \|\vec{\psi}_i\|^2 = 1, \quad j = 1, \dots, m. \quad (2.7-4)$$

Таким образом, собственные вектора  $\vec{\psi}_i, i = 1, \dots, m$ , матрицы  $\mathbf{A}$  попарно ортогональны и нормированы, т. е. они образуют ортонормальный базис в пространстве  $R^m$ .

Рассмотрим разложение вектора  $\vec{\varphi}$  из пространства  $\mathbf{R}^m$  по этому базису

$$\vec{\varphi} = \sum_{k=1}^m a_{\varphi k} \vec{\psi}_k. \quad (2.7-5)$$

Подставив разложение (2.7-5) в уравнение (2.7-2) и воспользовавшись тем, что  $\mathbf{A}\psi_k = \lambda_k \psi_k$ , получим

$$\vec{f} = \sum_{k=1}^m \lambda_k a_{\varphi k} \vec{\psi}_k + \vec{\varepsilon}. \quad (2.7-6)$$

Из уравнения (2.7-6) видно, что решение уравнения (2.7-2) теперь сводится к нахождению коэффициентов разложения  $a_{\varphi k}$ , а точность решения будет определяться точностью определения этих коэффициентов.

Для нахождения коэффициентов  $a_{\varphi k}$  умножив соотношение (2.7-6) скалярно на вектора  $\vec{\psi}_j, j = 1, \dots, m$ , получим систему уравнений относительно  $a_{\varphi k}$

$$(\vec{f}, \vec{\psi}_j) = \sum_{k=1}^m \lambda_k a_{\varphi k} (\vec{\psi}_k, \vec{\psi}_j) + (\vec{\varepsilon}, \vec{\psi}_j). \quad (2.7-7)$$

Вследствие ортонормальности базиса  $\{\vec{\psi}_i\}_{i=1, \dots, m}$  матрица системы (2.7.7) диагональна и

$$a_{fj} = \lambda_j a_{\varphi j} + a_{\varepsilon j}, \quad j = 1, \dots, m, \quad (2.7-8)$$

где  $a_{fj} = (\vec{f}, \vec{\psi}_j)$ ,  $a_{\varepsilon j} = (\vec{\varepsilon}, \vec{\psi}_j)$ . Из уравнений (2.7-8) получим окончательное выражение для коэффициентов разложения  $a_{\varphi j}$

$$a_{\varphi j} = \frac{a_{fj}}{\lambda_j} - \frac{a_{\varepsilon j}}{\lambda_j}, \quad j = 1, \dots, m. \quad (2.7-9)$$

Подставив полученное выражение для коэффициентов (2.7-9) в разложение (2.7-6), получим формулу для решения уравнения (2.7-2)

$$\bar{\varphi} = \sum_{k=1}^m \frac{a_{fk}}{\lambda_k} \bar{\psi}_k - \sum_{k=1}^m \frac{a_{\varepsilon k}}{\lambda_k} \bar{\psi}_k. \quad (2.7-10)$$

Из соотношения (2.7-10) следует, что величина ошибки определения  $\bar{\varphi}$  будет зависеть от соотношения между первым и вторым членами при каждом  $k$ .

Поскольку матрица  $\mathbf{A}$  описывает физический процесс взаимодействия непрерывной среды (атмосферы) с излучением, гладкость ее собственных векторов  $\bar{\psi}_j$  убывает с возрастанием индекса  $j$ , или, что то же самое, с убыванием соответствующих им собственных чисел  $\lambda_j$ .<sup>1</sup>

Вектор  $\bar{f}$  представляет достаточно гладкую функцию, и поэтому компоненты его разложения по используемому базису, *неотягощаемые ошибкой измерений*, убывают по абсолютной величине с ростом номера достаточно быстро. Поэтому, как правило, для его разложения (с хорошей точностью) требуется только несколько первых собственных векторов (иными словами, коэффициенты разложения  $a_{fk}$  убывают, приближаясь к нулю при больших  $k$ ). Иначе обстоит дело с разложением случайной ошибки измерений  $\bar{\varepsilon}$ . Как уже говорилось, шум измерений обычно содержит высокочастотные компоненты, поэтому для его представления необходимо использовать все собственные вектора  $\bar{\psi}_k$ , причем коэффициенты разложения  $a_{\varepsilon k}$  с ростом  $k$  стремятся к конечному пределу. (См. раздел 2.2.)

Принимая во внимание, что с ростом  $k$  величины собственных чисел убывают, видим, что, начиная с некоторого номера  $k$  (зависящего от величины ошибки измерений и скорости убывания собственных чисел, т. е. свойств матрицы  $\mathbf{A}$ ), основной вклад в решение (2.7-10) будет давать второе слагаемое, обусловленное

---

<sup>1</sup> Более подробное объяснение этого факта будет дано в следующих главах пособия.

ошибками измерений (в идеальном эксперименте при  $\vec{\varepsilon} = 0$  второе слагаемое отсутствует). Это приведет к тому, что в решении появятся осциллирующие компоненты с большой амплитудой, т. е. проявляется неустойчивость решения обратной задачи.

Из приведенного анализа следует, что чем меньше собственные числа присутствуют в спектре матрицы  $\mathbf{A}$  (по сравнению с максимальным числом  $\lambda$ ), тем сильнее будет проявляться неустойчивость решения обратной задачи.

Покажем, что введенное ранее число обусловленности  $r$ , равное коэффициенту усиления случайной ошибки измерений, определяется спектром матрицы  $\mathbf{A}$ . Для этого, воспользовавшись тем, что норма матрицы равна ее максимальному собственному числу и связью собственных чисел прямой и обратной матрицы (см. Приложение), получим для числа обусловленности

$$r = \frac{\lambda_{\max}}{\lambda_{\min}}. \quad (2.7-11)$$

Таким образом, анализируя спектр собственных значений матрицы  $\mathbf{A}$  (и не решая обратной задачи), можно понять трудности решения рассматриваемой обратной задачи. Если  $r$  не очень велико (мала размерность задачи  $m$ , «очень хорошее» ядро, т. е. матрица  $\mathbf{A}$ ), то неустойчивость или усиление ошибок измерений будут невелики. Наоборот, при больших значениях  $r$  решение обратной задачи традиционными, классическими методами будет невозможно.

Для примера приведем значения  $\lambda_j$  и  $r$  для конкретной задачи – микроволнового термического зондирования атмосферы с использованием измерений уходящего излучения в полосе  $O_2$  вблизи длины волны 0,5 см. Как будет показано в конце пособия, можно построить оптимальный спектральный прибор с заранее заданным числом спектральных каналов  $n$ , так, что отношение собственных чисел для оператора, соответствующего этому прибору, будет равно отношению собственных чисел ковариационной матрицы излучения, приведенных в таблице 2.7-1. Тогда на основе таблицы 2.7-1 мы можем оценить обусловленность матриц, соответствующих приборам с различным числом каналов  $n$ , измеряющих, соответственно,  $n$  параметров, описывающих высотный профиль температуры.

Таблица 2.7-1

Собственные значения  $\lambda_j$  для 1500 каналов измерений в спектральной области 1–200 ГГц.

j	$\lambda_j$
1	254144,99
2	8975,13
5	210,82
10	2,26
15	0,20
20	0,0067

Из приведенной таблицы видно, что уже отношение первого собственного числа к двадцатому  $\lambda_1 / \lambda_{20} = 3,8 \times 10^7$ . Иными словами, при использовании 20 каналов измерения коэффициент усиления ошибки равен указанной величине. При дальнейшем увеличении  $j$  это отношение очень быстро растет. Следовательно, даже использование всей информации, содержащейся в измерениях в 1500 спектральных каналах, приведет к неустойчивости решения при попытке получения информации более чем о 20 компонентах температурного профиля.

Таким образом, как и в предыдущем разделе, мы показали, что решение обратной задачи непосредственно связано со свойствами матрицы  $\mathbf{A}$ . Наиболее существенным свойством матрицы  $\mathbf{A}$ , определяющим устойчивость решения обратной задачи, является величина линейной зависимости ее строк. Если каждую  $i$ -ю строку матрицы  $\mathbf{A}$  рассматривать как вектор  $\vec{a}_i$ , то можно сказать, что система векторов  $\vec{a}_i$ ,  $i=1, \dots, m$ , является линейно зависимой, если выполняется равенство

$$\sum_{k=1}^m l_k \vec{a}_k = 0, \quad (2.7-12)$$

где не все коэффициенты  $l_i$  равны нулю. Условие (2.7-12) означает, что один из векторов, например  $\vec{a}_k$ , коэффициент  $l_k$  при котором отличен от нуля, можно выразить через другие вектора  $\vec{a}_i$ . Если в матрице имеются линейно зависимые строки, то определитель такой матрицы равен нулю и такую матрицу называют *вырожденной*. Равенство нулю определителя матрицы влечет за собой обращение в нуль по крайней мере одного из ее собственных чисел, и,

следовательно, число обусловленности вырожденной матрицы обращается в бесконечность. Для близкой к вырожденной матрицы в соответствии с полученной оценкой (2.7-10) относительные ошибки определения вектора  $\vec{\varphi}$  могут достигать сколь угодно больших величин. Для вырожденной матрицы без использования специальных приемов нам просто не удастся решить обратную задачу, поскольку матрица  $\mathbf{A}^{-1}$  в этом случае не существует.

## **2.8. Информационный подход В. П. Козлова к анализу решения обратной задачи**

Анализ «информационных возможностей» эксперимента начнем с простейшей задачи, а именно с задачи различения двух близких состояний объекта измерений. Еще лорд Рэлей рассмотрел эту задачу в инструментальной спектроскопии с целью оценки предельных возможностей спектральных оптических приборов (проблема разрешающей способности). В основе подхода Рэля лежит идея сравнения (с помощью данного спектрального прибора) двух стандартных «объектов», например в случае спектрометра – спектрального дублета и изолированной линии. При статистическом подходе это соответствует формулировке В.П. Козлова в терминах теории статистических решений.

Пусть  $\varphi_1(x)$  и  $\varphi_2(x)$  – две функции, описывающие фиксированные состояния объекта измерений, и последний находится в одном из этих состояний. В рассматриваемом примере процессу измерений соответствует уравнение

$$\int_{-\infty}^{\infty} A(x, y)\varphi(x)dx = f(y). \quad (2.8-1)$$

В результате эксперимента «над объектом» мы имеем наблюдаемую функцию  $f^*(y) = f(y) + \varepsilon(y)$ , где  $f(y)$  – функция, связанная с истинным состоянием  $\varphi(x)$  уравнением (2.8-1), а  $\varepsilon(y)$  – случайная функция (шум), характеризующая погрешности измерений. Козлов постулировал, что  $\varepsilon(y)$  является нормальным стационарным случайным процессом с нулевым средним и спектром мощности  $G(p)$ . Задача состоит в том, чтобы, имея измерения  $f^*(y)$ , решить, в каком

из двух возможных состояний –  $\varphi_1(x)$  или  $\varphi_2(x)$  – объект измерений находится в действительности. Будем характеризовать различимость состояний максимально достижимой вероятностью правильных решений  $P_r$  при условии, что сравниваемые состояния *априори* равновероятны и используется оптимальная решающая процедура. Величины  $P_r$  принимают значения между 1/2 и 1. 1/2 соответствует абсолютно неразличимым состояниям, 1 – абсолютно различимым. Козлов показал, что  $P_r$  зависит от данных задачи (ядра  $A(x, y)$ , функций  $\varphi_1(x)$  и  $\varphi_2(x)$  и статистики помех) *только через величину  $\rho$* , которая и характеризует непосредственно степень различимости состояний  $\varphi_1$  и  $\varphi_2$ :

$$P_r = 1/2 + 1/2\Phi(\rho/2), \quad (2.8-2)$$

$$\rho^2 \equiv \|\varphi_1 - \varphi_2\|^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|\tilde{A}(\varphi_1 - \varphi_2)|^2}{G(p)} dp, \quad (2.8-3)$$

где  $\Phi(x)$  – интеграл вероятности, а знак «~» означает преобразование Фурье от выражения, стоящего под знаком модуля.

Величина  $\rho$ , определяемая по формуле (2.8-3) как мера различимости состояний  $\varphi_1$  и  $\varphi_2$  с помощью «прибора», описываемого оператором  $\mathbf{A}$  и спектром мощности шума  $G(\rho)$ , обладает весьма важными свойствами. Прежде всего случай  $\rho(\varphi_1, \varphi_2) = 0$  соответствует полной неразличимости (эквивалентности) состояний в рассматриваемой обратной задаче. Другой предельный случай  $\rho(\varphi_1, \varphi_2) = \infty$  соответствовал бы «бесконечно большой» (абсолютной) различимости, чего на практике обычно не бывает. Удобно принять 1 в качестве пороговой величины, определяющей предел различимости состояний  $\varphi_1$  и  $\varphi_2$ . Подставив в выражение (2.8-2)  $\rho = 1$ , получим, что вероятность правильной идентификации состояния  $\varphi$  для этого случая составляет 63.7%, т. е. на 13.7% превышает вероятность случайного угадывания (50%). При уменьшении  $\rho$  эта вероятность быстро убывает.

Величину  $\rho$  естественно назвать информационным расстоянием в пространстве состояний исследуемого объекта (например атмосферы). Можно показать, что  $\rho$  обладает всеми свойствами

расстояния: неотрицательностью, симметричностью:  $\rho(\varphi_1, \varphi_2) = \rho(\varphi_2, \varphi_1)$ , и удовлетворяет неравенству треугольника:  $\rho(\varphi_1, \varphi_2) \leq \rho(\varphi_1, \varphi) + \rho(\varphi_2, \varphi)$ . Расстояние  $\rho$  характеризует содержащуюся в результатах эксперимента информацию о состояниях  $\varphi_1$  и  $\varphi_2$  с точки зрения их различимости.

Из приведенного анализа следует, что для каждого «истинного решения»  $\varphi_0$  задачи (2.8-1), характеризующего истинное состояние объекта измерений, существует бесконечно много функций  $\varphi$ , соответствующих состояниям, не отличимым от истинного по результатам реального эксперимента, в данных которого присутствует шум  $\varepsilon(y)$ . Все эти функции удовлетворяют неравенству

$$\rho(\varphi, \varphi_0) \leq 1 \quad (2.8-4)$$

(или более жесткому требованию  $\rho(\varphi, \varphi_0) \leq 1/2$ , выбор константы в достаточной степени произволен).

Другими словами, функции  $\varphi$  заполняют некоторую сферу единичного радиуса в функциональном пространстве решений задачи (2.8-1), которую естественно назвать «сферой неопределенности». Действительно, любая из функций  $\varphi$ , удовлетворяющая неравенству (2.8-4) (или  $\rho(\varphi, \varphi_0) \leq 1/2$ ), может быть принята за решение обратной задачи (2.8-1), так как эксперимент не дает информации, достаточной, чтобы отличить эту функцию от любой другой, также удовлетворяющей неравенству (2.8-4). По смыслу метрики (2.8-3) сфера (2.8-4) характеризует максимальную точность (или минимальную погрешность), которая вообще может быть достигнута при решении обратной задачи (2.8-1). Поэтому, в частности, всякая попытка дальше уточнить решение чисто математическими средствами, без привлечения дополнительной информации об искомом решении, аналогична попытке создания «информационного вечного двигателя», производящего информацию из ничего.

Однако практическая ценность метода решения уравнения (2.8-1), приводящего к произвольной функции  $\varphi$  из сферы (2.8-4), невелика. Если только исходная задача некорректна, указанная сфера содержит, главным образом, функции  $\varphi$ , которым не отвечают никакие реальные физические состояния исследуемого объекта. Для определения физически разумного решения (выделения «физической

части» сферы неопределенности) необходимо указать признаки, отличающие «физические» решения от «нефизических» или, как мы уже отмечали ранее, нужна априорная информация о возможных состояниях объекта или, что то же, о функциях, допустимых в качестве решения уравнения (2.8-1). Примеры введения такой априорной информации были даны нами в разделе 2.5.

Завершив на этом общее рассмотрение проблем, возникающих в связи с обратными задачами, перейдем к рассмотрению современных численных методов их решения.

## Глава 3

### Современные численные методы решения обратных задач

#### **3.1. Общие замечания о методах решения некорректных обратных задач**

Возможны различные подходы к решению обратных задач. Одной из основных причин существования различий между подходами является различная степень приближения, с которой мы можем описать физические явления, лежащие в основе рассматриваемого косвенного метода измерений. Иными словами, точность, с которой нам известен оператор  $A$  прямой задачи.

Допустим, что непосредственное вычисление оператора прямой задачи  $A$  чересчур сложно либо невозможно, например, из-за отсутствия информации о параметрах взаимодействия излучения со средой. Но мы располагаем большим объемом прямых измерений интересующих нас параметров атмосферы  $\varphi$  и соответствующих им результатов измерений радиационных величин  $f$ . Тогда для построения алгоритма решения обратной задачи могут быть использованы методы математической статистики.

Кроме того, даже если прямой оператор задачи известен с достаточной точностью, возможны два подхода к самой постановке обратной задачи. Первый из них (условно назовем его детерминистским) предполагает поиск единственного решения и характеристик его погрешности. С точки зрения другого подхода, учитывая стохастический характер поведения как атмосферных параметров, так и погрешностей измерений, следует искать не единственное решение, а распределение вероятности решения. Каждый из подходов правомерен со своей точки зрения, и, более того, в ряде случаев формальные алгоритмы, полученные на основе разных подходов, практически не отличаются.

Дать в этом пособии исчерпывающий обзор математических методов решения обратных задач, используемых в различных областях знания, не представляется возможным, хотя многие из этих методов иногда используются при решении обратных задач атмосферной оптики. Поэтому мы рассмотрим наиболее известные методы, обычно используемые для решения некорректных обратных задач атмосферной оптики.

Преобладающая часть применяемых в атмосферной оптике алгоритмов обращения основана на очевидной идее минимизации невязки, т. е. на поиске решения  $\varphi$ , обеспечивающего наилучшую близость результатов измерений  $f^*$  и соответствующих рассчитанных значений  $A(\varphi)$ . Одним из простейших алгоритмов, реализующих этот подход, является метод наименьших квадратов (МНК), который мы и рассмотрим.

### **3.2. Использование метода наименьших квадратов при решении обратных задач**

Решение системы уравнений (2.4-2), аппроксимирующей исследуемое интегральное уравнение Фредгольма I рода, формально может быть получено с помощью метода наименьших квадратов (МНК). Этот метод широко используется в различных приложениях [9] и, как будет показано ниже, служит основой многих более совершенных подходов к решению обратных задач.

Рассмотрим МНК в общей постановке. Запишем систему уравнений в векторной форме

$$\mathbf{A}\vec{\varphi} = \vec{f}, \quad (3.2-1)$$

где  $\mathbf{A}$  – конечномерный линейный оператор (соответствующая матрица), действующий из  $\mathbf{R}^n$  в  $\mathbf{R}^m$ . Измерения предоставляют нам вектор

$$\vec{f}^* = \vec{f} + \vec{\varepsilon}, \quad (3.2-2)$$

где  $\vec{\varepsilon}$  – ошибки измерений. Мы не предполагаем условия  $m = n$  и используем достаточно общие предположения о поведении ошибки измерений  $\vec{\varepsilon}$ :

$$E(\vec{\varepsilon}) = \vec{0} \text{ и } E(\vec{\varepsilon} \vec{\varepsilon}') = \Sigma,$$

где  $\Sigma$  – ковариационная матрица вектора ошибки измерений  $\vec{\varepsilon}$ . Нашей задачей является оценка вектора  $\vec{\varphi}$  на основе измерений  $\vec{f}^*$ .<sup>2</sup>

Идея, лежащая в основе метода, заключается в поиске оценки  $\vec{\varphi}$  искомого решения, обеспечивающей наименьшее в среднеквадратичном смысле различие между вектором измерений  $\vec{f}^*$  и рассчитанными на основе уравнения (3.2-1) величинами  $\mathbf{A}\vec{\varphi}$ . При этом для сравнения векторов в качестве нормы (расстояния) в пространстве измерений  $\mathbf{R}^m$  будем использовать задаваемый симметричной положительно определенной матрицей  $\mathbf{S}$  функционал

$$\rho_s(\mathbf{A}\vec{\varphi}, \vec{f}^*)^2 = \|\mathbf{A}\vec{\varphi} - \vec{f}^*\|_s^2 = (\mathbf{A}\vec{\varphi} - \vec{f}^*, \mathbf{S}(\mathbf{A}\vec{\varphi} - \vec{f}^*)), \quad (3.2-3)$$

где  $\mathbf{S} = \Sigma^{-1}$ . Использование выражения (3.2-3) для сравнения измеренных величин означает введение весов, соответствующих погрешностям измерений. Например, если измеряются две величины ( $m = 2$ ) и погрешности их измерения независимы и равны  $\sigma_1$  и  $\sigma_2$ , то матрица  $\Sigma$  диагональна и выражение (3.2-3) в координатной форме принимает вид

$$\frac{\left(\sum_{j=1}^n a_{1,j} \varphi_j - f_1^*\right)^2}{\sigma_1^2} + \frac{\left(\sum_{j=1}^n a_{2,j} \varphi_j - f_2^*\right)^2}{\sigma_2^2}.$$

Итак, исходное соотношение для оценки решения выглядит следующим образом:

$$\vec{\varphi} = \arg \min_{\vec{\varphi} \in \mathbf{R}^n} \|\vec{f}^* - \mathbf{A}\vec{\varphi}\|_s^2. \quad (3.2-4)$$

Символ  $\arg \min$  обозначает значение указанного под ним аргумента, реализующее минимум следующего за этим символом выражения.

Иными словами, в качестве решения задачи (3.2-1), (3.2-2) будем искать вектор  $\vec{\varphi}$ , который дает минимальное значение функции

---

<sup>2</sup> Заметим, что неправомерно говорить о решении системы уравнений

$$\mathbf{A}\vec{\varphi} = \vec{f}^* = \vec{f} + \vec{\varepsilon}$$

в обычном смысле теории систем линейных уравнений. Если количество неизвестных превышает количество уравнений, как это часто бывает на практике, такого решения не существует.

$$L(\tilde{\varphi}) = (\tilde{f}^* - \mathbf{A}\tilde{\varphi}, \mathbf{S}\tilde{f}^* - \mathbf{S}\mathbf{A}\tilde{\varphi}). \quad (3.2-5)$$

Из математического анализа известно необходимое условие экстремума функции векторного аргумента

$$\frac{d}{d\tilde{\varphi}} L(\tilde{\varphi}) = 0. \quad (3.2-6)$$

В другой форме записи условие (3.2-6) представляет собой систему  $m$  линейных уравнений, в каждом из которых участвует производная только по одной из компонент вектора  $\tilde{\varphi} = (\tilde{\varphi}_1, \dots, \tilde{\varphi}_n)$

$$\frac{\partial L(\tilde{\varphi})}{\partial \tilde{\varphi}_i} = 0, \quad i = 1, \dots, m. \quad (3.2-7)$$

Выполним преобразования, необходимые для решения уравнения (3.2-6). Подставим скалярное произведение (3.2-5) в уравнение (3.2-6), предварительно раскрыв скобки:

$$\begin{aligned} \frac{dL(\tilde{\varphi})}{d\tilde{\varphi}} &= \frac{d}{d\tilde{\varphi}} (\tilde{f}^*, \mathbf{S}\tilde{f}^*) - \frac{d}{d\tilde{\varphi}} (\mathbf{A}\tilde{\varphi}, \mathbf{S}\tilde{f}^*) - \\ &- \frac{d}{d\tilde{\varphi}} (\tilde{f}^*, \mathbf{S}\mathbf{A}\tilde{\varphi}) + \frac{d}{d\tilde{\varphi}} (\mathbf{A}\tilde{\varphi}, \mathbf{S}\mathbf{A}\tilde{\varphi}) = 0. \end{aligned} \quad (3.2-8)$$

Воспользовавшись правилами дифференцирования скалярного произведения по векторному аргументу (см. Приложение), из уравнения (3.2-8) получим

$$-2\mathbf{A}'\mathbf{S}\tilde{f}^* + 2\mathbf{A}'\mathbf{S}\mathbf{A}\tilde{\varphi} = 0, \quad (3.2-9)$$

откуда, если существует обратная матрица  $(\mathbf{A}'\mathbf{S}\mathbf{A})^{-1}$ ,

$$\tilde{\varphi} = (\mathbf{A}'\mathbf{S}\mathbf{A})^{-1} \mathbf{A}'\mathbf{S}\tilde{f}^*. \quad (3.2-10)$$

Поскольку матрица  $\mathbf{S}$  положительно определена, для существования  $(\mathbf{A}'\mathbf{S}\mathbf{A})^{-1}$  необходимо и достаточно, чтобы ранг матрицы  $\mathbf{A}$  был равен  $n$ . Очевидным необходимым условием этого является неравенство  $m \geq n$ , т. е. для существования решения задачи (3.2-1), (3.2-2) по МНК необходимо, чтобы количество измерений  $m$  (размерность вектора  $\tilde{f}^*$ ) было не меньше количества неизвестных  $n$  (размерности вектора  $\tilde{\varphi}$ ). В случае равенства  $m = n$  и существования матрицы  $\mathbf{A}^{-1}$  из выражения для решения (3.2-10) получим

$$\begin{aligned}\tilde{\varphi} &= (\mathbf{A}'\mathbf{S}\mathbf{A})^{-1}\mathbf{A}'\mathbf{S}\tilde{f}^* = \mathbf{A}^{-1}\mathbf{S}^{-1}\mathbf{A}'^{-1}\mathbf{A}'\mathbf{S}\tilde{f}^* = \\ &= \mathbf{A}^{-1}(\mathbf{A}'\mathbf{S})^{-1}\mathbf{A}'\mathbf{S}\tilde{f}^* = \mathbf{A}^{-1}\tilde{f}^*,\end{aligned}\quad (3.2-11)$$

т. е. для квадратной невырожденной матрицы  $\mathbf{A}$  независимо от конкретного вида матрицы  $\mathbf{S}$  МНК вырождается в решение системы линейных уравнений вида (3.2-1).

Проанализируем влияние вида матрицы  $\mathbf{S} = \Sigma^{-1}$  на решение МНК при  $m > n$ . Пусть ошибки измерений  $(\varepsilon_1, \dots, \varepsilon_m) = \vec{\varepsilon}$  – независимо распределенные случайные величины и измерения равноточные. Тогда  $\mathbf{S} = \sigma^{-2}\mathbf{I}$ , где  $\sigma^2$  – дисперсия ошибки измерений, а  $\mathbf{I}$  – единичная матрица. В этом случае, как легко получить из представления решения (3.2-10), оценка решения по МНК имеет вид

$$\begin{aligned}\tilde{\varphi} &= (\mathbf{A}'\sigma^{-2}\mathbf{I}\mathbf{A})^{-1}\mathbf{A}'\sigma^{-2}\mathbf{I}\tilde{f}^* = \\ &= \sigma^2(\mathbf{A}'\mathbf{A})^{-1}\sigma^{-2}\mathbf{A}'\tilde{f}^* = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\tilde{f}^*.\end{aligned}\quad (3.2-12)$$

Для неравноточных, но по-прежнему некоррелированных измерений матрица  $\Sigma$  также диагональна, но на диагонали ее стоят различные по величине дисперсии ошибок  $\sigma_i^2$ . Записав для этого случая в координатной форме минимизируемое выражение (3.2-5), получим

$$L(\vec{\varphi}) = \sum_{i=1}^n \left( \frac{\sum_{j=1}^m a_{i,j} \varphi_j - f_i^*}{\sigma_i} \right)^2, \quad (3.2-13)$$

откуда видно, что величины, обратные дисперсиям ошибок измерений  $\sigma_i^2$ , играют роль весов для компонент невязки.

Наконец, если погрешности измерений не являются взаимно независимыми, т. е. вектор погрешностей  $\vec{\varepsilon}$  имеет нетривиальное многомерное распределение вероятностей, матрица  $\mathbf{S} = \Sigma^{-1}$  не является диагональной.

Матрица  $\mathbf{A}'\Sigma^{-1}\mathbf{A}$  играет существенную роль в теории оценок решения задачи (3.2-1), (3.2-2) и называется *матрицей Фишера*  $\mathbf{F}$

$$\mathbf{F} = \mathbf{A}'\Sigma^{-1}\mathbf{A}. \quad (3.2-14)$$

Отметим, что при построении решения задачи (3.2-1), (3.2-2) в форме (3.2-10) и (3.2-12) не использовалось предположение о гауссовом характере случайной ошибки измерений.

### 3.3. Анализ характеристик решения обратной задачи методом наименьших квадратов

#### Смещение и ковариационная матрица ошибок оценки решения по МНК

Рассмотрим две основные статистические характеристики оценки решения задачи (3.2-1), (3.2-2): смещение и ковариационную матрицу ошибок оценки решения. По определению *смещение*  $\vec{b}$  – это отклонение математического ожидания оценки решения  $\vec{\tilde{\varphi}}$  (осреднение производится по ансамблю реализаций вектора случайных ошибок  $\vec{\varepsilon}$ ) от точного значения  $\vec{\varphi}$ , соответствующего уравнению (3.2-1)

$$\vec{b} = E(\vec{\tilde{\varphi}}) - \vec{\varphi}, \quad (3.3-1)$$

где  $E$  – знак математического ожидания. Смещение, в частности, показывает, стремится ли к истинной средня величина оценки решения  $\vec{\tilde{\varphi}}$  при осреднении результатов его многократных косвенных измерений.

Величину смещения можно получить, подставив в определение (3.3-1) выражение для оценки решения  $\vec{\tilde{\varphi}}$  (3.2-10) и учтя исходные соотношения (3.2-1) и (3.2-2):

$$\begin{aligned} \vec{b} &= E(\mathbf{F}^{-1} \mathbf{A}' \boldsymbol{\Sigma}^{-1} (\mathbf{A} \vec{\varphi} + \vec{\varepsilon})) - \vec{\varphi} = \\ &= E(\mathbf{F}^{-1} \mathbf{F} \vec{\varphi} + \mathbf{F}^{-1} \mathbf{A}' \boldsymbol{\Sigma}^{-1} \vec{\varepsilon}) - \vec{\varphi} = \\ &= \vec{\varphi} + \mathbf{F}^{-1} \mathbf{A}' \boldsymbol{\Sigma}^{-1} E(\vec{\varepsilon}) - \vec{\varphi} = 0. \end{aligned} \quad (3.3-2)$$

При выводе (3.3-2) использовалось только предположение, что случайная ошибка имеет нулевое среднее, т. е.  $E(\vec{\varepsilon}) = 0$ . Таким образом, среднее значение оценки решения МНК совпадает с точным значением  $\vec{\varphi}$  (т. е. решением, соответствующим  $\vec{\varepsilon} = 0$ ). Полученное свойство решения носит название *несмещенности* оценки.

Ковариационная матрица оценки решения  $\tilde{\mathbf{D}}$ , в отличие от смещения, характеризует случайную погрешность решения за счет случайной погрешности единичного измерения  $\vec{f}^*$ . По определению ковариационной матрицы случайного вектора:

$$\tilde{\mathbf{D}} = E((\tilde{\vec{\varphi}} - \vec{\varphi})(\tilde{\vec{\varphi}} - \vec{\varphi})'). \quad (3.3-3)$$

Подставим, как и при вычислении смещения, в это выражение оценку решения  $\tilde{\vec{\varphi}}$  по МНК (формула (3.2-10)). Тогда

$$\tilde{\mathbf{D}} = E((\mathbf{F}^{-1}\mathbf{A}'\Sigma^{-1}\vec{f}^* - \vec{\varphi})(\mathbf{F}^{-1}\mathbf{A}'\Sigma^{-1}\vec{f}^* - \vec{\varphi})'). \quad (3.3-4)$$

Используя основные уравнения задачи (3.2-2) и (3.2-1), получим

$$\begin{aligned} \tilde{\mathbf{D}} &= E((\mathbf{F}^{-1}\mathbf{A}'\Sigma^{-1}\mathbf{A}\vec{\varphi} + \mathbf{F}^{-1}\mathbf{A}'\Sigma^{-1}\vec{\varepsilon} - \vec{\varphi}) \times \\ &\times (\mathbf{F}^{-1}\mathbf{A}'\Sigma^{-1}\mathbf{A}\vec{\varphi} + \mathbf{F}^{-1}\mathbf{A}'\Sigma^{-1}\vec{\varepsilon} - \vec{\varphi})') = \\ &= E((\mathbf{F}^{-1}\mathbf{F}\vec{\varphi} + \mathbf{F}^{-1}\mathbf{A}'\Sigma^{-1}\vec{\varepsilon} - \vec{\varphi})(\mathbf{F}^{-1}\mathbf{F}\vec{\varphi} + \mathbf{F}^{-1}\mathbf{A}'\Sigma^{-1}\vec{\varepsilon} - \vec{\varphi})') = \\ &= E((\vec{\varphi} + \mathbf{F}^{-1}\mathbf{A}'\Sigma^{-1}\vec{\varepsilon} - \vec{\varphi})(\vec{\varphi} + \mathbf{F}^{-1}\mathbf{A}'\Sigma^{-1}\vec{\varepsilon} - \vec{\varphi})') = \\ &= E((\mathbf{F}^{-1}\mathbf{A}'\Sigma^{-1}\vec{\varepsilon})(\mathbf{F}^{-1}\mathbf{A}'\Sigma^{-1}\vec{\varepsilon})'). \end{aligned}$$

Преобразуем последнее полученное выражение, используя правило транспонирования произведения матриц, а также симметричность матриц  $\mathbf{F}$  и  $\Sigma^{-1}$ . Тогда

$$\begin{aligned} \tilde{\mathbf{D}} &= E(\mathbf{F}^{-1}\mathbf{A}'\Sigma^{-1}\vec{\varepsilon}\vec{\varepsilon}'(\mathbf{F}^{-1}\mathbf{A}'\Sigma^{-1})') = \\ &= \mathbf{F}^{-1}\mathbf{A}'\Sigma^{-1}E(\vec{\varepsilon}\vec{\varepsilon}')(\mathbf{F}^{-1}\mathbf{A}'\Sigma^{-1})' = \\ &= \mathbf{F}^{-1}\mathbf{A}'\Sigma^{-1}\Sigma\Sigma^{-1}(\mathbf{F}^{-1}\mathbf{A}')' = \\ &= \mathbf{F}^{-1}\mathbf{A}'\Sigma^{-1}\mathbf{A}\mathbf{F}^{-1} = \mathbf{F}^{-1}\mathbf{F}\mathbf{F}^{-1} = \mathbf{F}^{-1}. \end{aligned} \quad (3.3-5)$$

Здесь мы воспользовались также определением матрицы  $\Sigma = E(\vec{\varepsilon}\vec{\varepsilon}')$ . Таким образом, матрица ошибок оценки решения совпадает с обратной матрицей Фишера. Диагональные элементы матрицы  $\tilde{\mathbf{D}}$  являются дисперсиями ошибок оценки решения МНК, а внедиагональные характеризуют степень взаимокорреляции ошибок оценки компонент искомого вектора.

Еще одно важное свойство оценок МНК состоит в том, что среди всех несмещенных линейных оценок оценки МНК обладают наименьшими дисперсиями (теорема Гаусса–Маркова).

### Анализ оценки решения по МНК в ортогональном базисе собственных векторов

Формулы (3.2-10) и (3.2-12) показывают, что для получения оценок решения необходимо обратить матрицы  $\mathbf{F}$  или  $\mathbf{A}'\mathbf{A}$ . Как мы уже отмечали, в большинстве обратных задач атмосферной оптики матрица  $\mathbf{A}'\mathbf{A}$  является плохо обусловленной (вырожденной или почти вырожденной), что связано с линейной зависимостью строк матрицы  $\mathbf{A}$ . Плохая обусловленность матрицы, т. е. большая величина числа обусловленности  $r = \lambda_{\max} / \lambda_{\min}$ , очевидным образом связана с собственными числами матрицы. Как правило, матрицы  $\mathbf{F} = \mathbf{A}'\mathbf{\Sigma}^{-1}\mathbf{A}$  и  $\mathbf{A}'\mathbf{A}$  являются плохо обусловленными («почти вырожденными») матрицами и отношение их максимального и минимального собственных чисел очень велико. Поэтому и построенные в рамках МНК решения (3.2-10) и (3.2-12) обычно неустойчивы относительно ошибок измерений, а диагональные элементы матрицы ошибок  $\tilde{\mathbf{D}}$  (дисперсии ошибок решения) могут принимать очень большие значения.

Для количественной оценки этих эффектов воспользуемся разложением решения по ортонормальному базису собственных векторов матрицы  $\mathbf{A}'\mathbf{A}$  (матрица нормальна, поэтому ее собственные вектора ортогональны)

$$\vec{\varphi} = \sum_{i=1}^m a_i \vec{\psi}_i. \quad (3.3-7)$$

Собственные вектора  $\vec{\psi}_i$  и собственные значения  $\lambda_i$  удовлетворяют уравнению

$$(\mathbf{A}'\mathbf{A})\vec{\psi}_i = \lambda_i \vec{\psi}_i, \quad i = 1, \dots, n.$$

Запишем правую часть (3.3-7) в векторно-матричном виде. Для этого введем вектор  $\vec{a} = (a_1, \dots, a_n)$  (вектор коэффициентов разложения оценки решения  $\vec{\varphi}$  по базису  $\{\vec{\psi}_i\}$ ) и унитарную матрицу  $\mathbf{P}$  (матрицу перехода к базису  $\{\vec{\psi}_i\}$ ), столбцами которой являются вектора  $\vec{\psi}_i, i = 1, \dots, n$ . Тогда разложение (3.3-7) можно представить в виде

$$\vec{\varphi} = \mathbf{P}\vec{a}. \quad (3.3-8)$$

Подставляя разложение (3.3-8) в уравнение (3.2-1), получим

$$\vec{f} = \mathbf{A}\mathbf{P}\vec{a}. \quad (3.3-9)$$

Оценку решения  $\vec{a}$  задачи (3.3-9), (3.2-2) по МНК можно получить, воспользовавшись формулой (3.2-10), в которой матрицу  $\mathbf{A}$  надо заменить на матрицу  $\mathbf{A}\mathbf{P}$ :

$$\vec{a} = (\mathbf{P}'\mathbf{A}'\Sigma^{-1}\mathbf{A}\mathbf{P})^{-1}\mathbf{P}'\mathbf{A}'\Sigma^{-1}\vec{f}^* = (\mathbf{P}'\mathbf{F}\mathbf{P})^{-1}\mathbf{P}'\mathbf{A}'\Sigma^{-1}\vec{f}^*. \quad (3.3-10)$$

Ковариационную матрицу ошибок для  $\vec{a}$  также получим заменой  $\mathbf{A}$  на  $\mathbf{A}\mathbf{P}$  в соотношениях (3.3-5):

$$\tilde{\mathbf{D}}_{\vec{a}} = (\mathbf{P}'\mathbf{F}\mathbf{P})^{-1}. \quad (3.3-11)$$

Пусть измерения равноточные и ошибки не коррелированы, т. е.  $\Sigma = \sigma^2\mathbf{I}$ ,  $\Sigma^{-1} = \sigma^{-2}\mathbf{I}$ . Тогда выражение (3.3-11) можно записать в базисе  $\{\vec{\psi}_i\}$  в виде

$$\tilde{\mathbf{D}}_{\vec{a}} = \sigma^2\mathbf{L}^{-1}, \quad (3.3-12)$$

где  $\mathbf{L}$  – диагональная матрица, на диагонали которой стоят собственные числа матрицы  $\mathbf{A}'\mathbf{A}$

$$\mathbf{L} = \begin{vmatrix} \lambda_1 & 0 & \cdot & 0 \\ 0 & \lambda_2 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \lambda_n \end{vmatrix}.$$

Вывод формулы (3.3-12) достаточно элементарен:

$$(\mathbf{A}'\mathbf{A})\vec{\psi}_i = \lambda_i\vec{\psi}_i \Rightarrow (\mathbf{A}'\mathbf{A})\mathbf{P} = \mathbf{P}\mathbf{L},$$

$$\mathbf{P}'(\mathbf{A}'\mathbf{A})\mathbf{P} = \mathbf{P}'\mathbf{P}\mathbf{L} = \mathbf{L},$$

$$\tilde{\mathbf{D}}_{\vec{a}} = (\mathbf{P}'\mathbf{A}'\Sigma^{-1}\mathbf{A}\mathbf{P})^{-1} = (\mathbf{P}'\mathbf{A}'\sigma^{-2}\mathbf{I}\mathbf{A}\mathbf{P})^{-1} = \sigma^2(\mathbf{P}'\mathbf{A}'\mathbf{A}\mathbf{P})^{-1} = \sigma^2\mathbf{L}^{-1}.$$

Таким образом, матрица ошибок  $\tilde{\mathbf{D}}_{\vec{a}}$  имеет вид

$$\tilde{\mathbf{D}}_{\vec{a}} = \begin{vmatrix} \frac{\sigma^2}{\lambda_1} & 0 & \cdot & 0 \\ 0 & \frac{\sigma^2}{\lambda_2} & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \frac{\sigma^2}{\lambda_n} \end{vmatrix}. \quad (3.3-13)$$

Учитывая тот факт, что  $\lambda_n \rightarrow 0$  с ростом  $n$ , диагональные элементы при больших значениях  $n$  могут быть сколь угодно большими, т. е. погрешности определения коэффициентов при собственных векторах  $\vec{\psi}_i$  при больших  $n$  могут быть очень велики, что и приведет к появлению осцилляций в решении. Такие сильно осциллирующие решения не имеют реального физического смысла. Как мы отмечали ранее, единственный путь получения разумного, устойчивого решения задачи (3.2-1), (3.2-2) – использование дополнительной априорной информации о решении  $\vec{\varphi}$ . Существуют разные возможности введения априорной информации. Далее перейдем к их рассмотрению.

#### **3.4. Методы введения априорной информации**

Как мы неоднократно отмечали, введение априорной (дополнительной) информации об искомом решении  $\vec{\varphi}$  является необходимым элементом при практическом решении задачи (3.2-1), (3.2-2) или интегрального уравнения (2.3-1). Существуют два основных подхода к заданию априорной информации – детерминистский и вероятностный. В первом случае априорная информация заключается в указании подпространства (класса), в котором ищется решение (обозначим его, например,  $\mathbf{R}_0$ ), содержащегося в пространстве решений (например,  $\mathbf{R}^n$ ). Такое точное указание класса допустимых решений будем называть информацией типа **T**. Другой подход основан на введении распределения вероятностей в пространстве  $\mathbf{R}^n$ . В этом случае указывается вероятность принадлежности решения элементам некоторого

достаточно широкого класса подмножеств пространства  $\mathbf{R}^n$ . Такое задание априорной информации будем называть типом **B**.

В обратных задачах атмосферной оптики обычно используются следующие конкретные типы задания априорной информации:

$$\mathbf{T1} : \mathbf{R}_0 = L(\vec{u}_1, \dots, \vec{u}_k)$$

– линейная оболочка векторов  $\{\vec{u}_i\}_{i=1, \dots, k}$ , или, иными словами, линейное подпространство размерности  $k < m$  в  $R^n$ , натянутое на вектора  $\{\vec{u}_i\}_{i=1, \dots, k}$ ;

$$\mathbf{T2} : \mathbf{R}_0 = \{\vec{\varphi} : (\vec{\varphi}, \mathbf{H}\vec{\varphi}) \leq C\},$$

где  $\mathbf{H}$  – симметричная положительно определенная матрица, а  $C$  – вещественное число, т. е. в качестве допустимых решений возможны все те вектора  $\vec{\varphi}$ , длина которых в метрике, задаваемой матрицей  $\mathbf{H}$ , не превосходит заданной величины – гипершар радиуса  $\sqrt{C}$ ;

**B**: – для решения в пространстве  $\mathbf{R}^n$  задано нормальное (гауссово) распределение вероятностей своими двумя первыми моментами

$$E(\vec{\varphi}) = \vec{\varphi}_0 \text{ и } E\left((\vec{\varphi} - \vec{\varphi}_0)(\vec{\varphi} - \vec{\varphi}_0)'\right) = \mathbf{D}.$$

В соответствии с тремя перечисленными типами априорных ограничений можно выделить три группы численных методов решения некорректных обратных задач:

- – проекционные;
- – регуляризации;
- – статистических оценок.

### 3.5. Проекционные методы решения обратных задач

Название этой группы вызвано тем, что фактически ищется проекция решения на подпространство меньшей размерности. Наиболее распространенный и разработанный проекционный метод – метод разложения по «усеченному» базису. Термин «усеченный» базис подразумевает, что рассматривается какой-либо базис,

связанный с решаемой задачей, например базис из собственных векторов матрицы  $\mathbf{A}'\mathbf{A}$  или так называемый эмпирический ортогональный базис, т. е. базис, образованный собственными векторами ковариационной матрицы вертикальных профилей искомых параметров атмосферы. Затем вектора этого базиса с большими номерами, соответствующие малым собственным числам, отбрасываются и решение ищется в виде разложения по оставшимся первым собственным векторам, т. е. по базису, размерность которого меньше размерности исходного пространства. Поскольку малые собственные числа соответствуют осциллирующим компонентам решения, неустойчивым относительно ошибок измерений, метод приводит к стабилизации решения.

Следует отметить, что по существу мы уже воспользовались разновидностью проекционного метода при переходе от общего интегрального уравнения Фредгольма I рода к его конечномерному представлению (см, например, раздел 3.2.). Здесь проектирование происходило из бесконечномерного пространства непрерывных функций  $C_{[a,b]}$  в евклидово пространство  $\mathbf{R}^n$ , а базисом в  $\mathbf{R}^n$  были ступенчатые функции вида

$$\psi_j = \begin{cases} 1, & t_j \leq t < t_{j+1}, \\ 0, & t < t_j, t \geq t_{j+1}, \end{cases} \quad j = 1, \dots, n.$$

При этом при достаточно большой размерности  $n$ , необходимой для достаточно точного описания реальной физической функции, применение таким образом проекционного метода не позволяет избавиться от неустойчивости обратной задачи.

Таким образом, решение задачи (3.2-1), (3.2-2) ищется в виде

$$\vec{\varphi} = \sum_{i=1}^k b_i \vec{u}_i, \quad (3.5-1)$$

где  $b_i$  – неизвестные коэффициенты разложения,  $\{\vec{u}_i\}_{i=1, \dots, k}$  – система базисных векторов в  $\mathbf{R}_0 = \mathbf{R}^k$ ,  $k < n$ .

Рассмотрим базис собственных векторов матрицы  $\mathbf{A}'\mathbf{A}$ . Возьмем в качестве  $\{\vec{u}_i\}$  первые  $k$  собственных векторов  $\{\vec{\psi}_i\}_{i=1, \dots, k}$ . Тогда решение сведется к поиску неизвестных коэффициентов  $b_i$  в разложении (3.5-1) при  $\vec{u}_i = \vec{\psi}_i$ ,  $i = 1, \dots, k$ . Рассуждая по аналогии,

например, с выводом соотношения (3.3-10), получим выражение для нахождения вектора коэффициентов  $b_i$ :

$$\vec{b} = (\mathbf{A}'_{(k)} \Sigma^{-1} \mathbf{A}_{(k)})^{-1} \mathbf{A}'_{(k)} \Sigma^{-1} \vec{f}^*, \quad (3.5-2)$$

где  $\mathbf{A}_{(k)}$  – проекция матрицы  $\mathbf{A}$  на подпространство  $R_0$ , определяемая в соответствии с (3.3-9) соотношением

$$\mathbf{A}_{(k)} = \mathbf{A} \mathbf{P}_{(k)}, \quad (3.5-3)$$

а  $\mathbf{P}_{(k)}$  – матрица размерности  $n \times k$ , столбцы которой представляют собой собственные вектора  $\vec{\psi}_i$ ,  $i = 1, \dots, k$ .

Ковариационная матрица ошибок для коэффициентов  $b_i$  по аналогии с матрицей (3.3-13) будет иметь вид

$$\tilde{\mathbf{D}}_{\vec{b}} = \begin{vmatrix} \frac{\sigma^2}{\lambda_1} & 0 & \cdot & 0 \\ 0 & \frac{\sigma^2}{\lambda_2} & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \frac{\sigma^2}{\lambda_k} \end{vmatrix}. \quad (3.5-4)$$

В связи с тем, что в этой матрице (по сравнению с матрицей (3.3-13)) отброшены последние  $m - k$  малых собственных чисел, дисперсии оставшихся коэффициентов  $b_i$  могут иметь «приемлемые» значения, а матрица  $\mathbf{A}'_{(k)} \mathbf{A}_{(k)}$  будет достаточно хорошо обусловленной.

Мы приводили примеры использования такого подхода (см. главу 2) в первых работах по термическому зондированию атмосферы. Их достоинство – относительная простота численной реализации, а также несмещенность получаемых оценок (разумеется, относительно подпространства  $\mathbf{R}_0$ ). Проблемы, возникающие при использовании этого метода, типичны для некорректных обратных задач: для улучшения аппроксимации искомого решения  $\vec{\varphi}$  требуется увеличение размерности пространства  $\mathbf{R}_0$   $k$ , а для улучшения обусловленности матрицы  $\mathbf{A}'_{(k)} \mathbf{A}_{(k)}$  ее уменьшение.

Например, аппроксимация с хорошей точностью ( $\approx 0,5$  К) вертикальных профилей температуры  $T(z)$  требует использования 15–20 собственных векторов эмпирического базиса, но при этой размерности соответствующая матрица  $\mathbf{A}'_{(k)}\mathbf{A}_{(k)}$  уже является плохо обусловленной (см. приведенный в конце главы 2 пример). Вопрос выбора оптимальной размерности подпространства  $\mathbf{R}_0$  остается при этом подходе открытым.

Частичное разрешение этого противоречия возможно при использовании специального базиса, обладающего достаточно хорошими аппроксимационными свойствами и в то же время имеющего достаточно небольшую размерность с точки зрения обусловленности соответствующей системы уравнений. Первым шагом в этом направлении, как мы указывали ранее, является использование в разложении (3.5-1) так называемого эмпирического ортогонального базиса искомого вектора, т. е. собственных векторов матрицы  $\mathbf{K}_{\varphi\varphi}$ :

$$\mathbf{K}_{\varphi\varphi}\vec{\psi}_i = \lambda_i\vec{\psi}_i, \quad (3.5-5)$$

где  $\mathbf{K}_{\varphi\varphi}$  – ковариационная матрица искомого вектора, рассматриваемого как случайный вектор, построенная на основе статистики предыдущих, например, прямых наблюдений  $\vec{\varphi}$ . Так, например, в задаче термического зондирования атмосферы (ТЗА) используются эмпирический ортогональный базис из собственных векторов ковариационной матрицы температуры  $\mathbf{K}_{TT}$ , построенной по данным прямых радиозондовых измерений вертикального профиля температуры  $T(z)$ . Применение в задаче ТЗА ковариационных матриц температуры, построенных для определенных районов и сезонов, позволяет еще уменьшить размерность задачи, т. е. число членов разложения (3.5-1), не теряя в точности аппроксимации.

Дальнейшее развитие этого подхода было осуществлено В.П. Козловым, который предложил использовать базис, составленный из собственных векторов *информационной матрицы Козлова* (см. главу 4), отобранных с учетом величины соответствующих им собственных чисел.

Использование ранее накопленной статистики поведения искомого вектора обосновано тем, что мы выделяем среди решений нашей некорректной обратной задачи решения, «похожие» на

результаты предшествующих прямых измерений. Оптимальность эмпирического ортогонального собственного базиса теоретически обоснована А.М. Обуховым, который показал, что этот базис дает наилучшую аппроксимацию случайного вектора при конечном числе членов разложения (3.5-1) по сравнению с любым другим ортогональным базисом. На практике, однако, применение этого базиса не всегда возможно в связи с отсутствием достаточного объема предыдущих измерений искомого вектора и невозможностью вычисления соответствующих эмпирических ковариационных матриц. Например, в настоящее время имеются в достаточном объеме только данные измерений вертикальных профилей температуры и влажности и соответствующие ковариационные эмпирические матрицы (см., например, книгу В.Е. Зуева и В.С. Комарова [2]). В существенно меньшей степени это относится к вертикальному профилю содержания озона. (Нам известны только эмпирические ковариационные матрицы, построенные В. Деллером и Д. Шпенкухом для г. Потсдама). Для многих других физических параметров атмосферы ковариационные матрицы вообще отсутствуют. Естественно также, что описанный подход неприменим при исследовании атмосфер других планет.

В случае отсутствия эмпирических ковариационных матриц искомого параметра можно в принципе использовать различные формы модельных ковариационных матриц, вид которых задается из общих физических соображений. Так, в различных работах используются ковариационные матрицы вида

$$\mathbf{K}_{\varphi\varphi} = \begin{vmatrix} \sigma_{\varphi}^2(z_1) & 0 & \dots & 0 \\ 0 & \sigma_{\varphi}^2(z_2) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_{\varphi}^2(z_n) \end{vmatrix}. \quad (3.5-7)$$

$\mathbf{K}_{\varphi\varphi}$  в этом случае – диагональная матрица, не учитывающая возможные взаимные корреляции вариаций искомого параметра на различных уровнях, а описывающая только величину его естественных вариаций. Другой вид модельных ковариационных матриц

$$\mathbf{K}_{\varphi\varphi} = \{k_{ij}\}_{\substack{i=1,n \\ j=1,n}}, \quad k_{ij} = \sigma_{\varphi}(z_i) \sigma_{\varphi}(z_j) \exp\left(-\frac{|z_i - z_j|}{r_{\varphi}}\right) \quad (3.5-8)$$

или

$$k_{ij} = \sigma_{\varphi}(z_i) \sigma_{\varphi}(z_j) \exp\left(-\frac{|z_i - z_j|^2}{r_{\varphi}^2}\right), \quad (3.5-9)$$

где  $r_{\varphi}$  – радиус корреляции.

Ковариационные матрицы (3.5-8) и (3.5-9) учитывают взаимокорреляции вариаций параметра  $\varphi$  на различных уровнях в атмосфере. Эти корреляции экспоненциально (но с разной скоростью) убывают при увеличении удаленности уровней  $z_i$  и  $z_j$ . Радиус корреляции  $r_{\varphi}$  выбирается из общих физических соображений или на основе анализа ограниченного эмпирического материала. Интересно отметить, что М.С. Бирюлина [1] показала, что использование при решении обратной задачи модельной ковариационной матрицы вида (3.5-9) эквивалентно одному из методов регуляризации А. Н. Тихонова (см. подробнее далее).

К проекционным методам можно отнести и ряд других известных и используемых подходов. Так, широко распространенным в вычислительной практике способом приближенного решения уравнения  $\mathbf{A}\vec{\varphi} = \vec{f}^*$  является *метод подбора*. Он состоит в том, что для элементов  $\vec{\varphi}$  некоторого заранее задаваемого конечного множества возможных решений  $\mathbf{M}$  ( $\mathbf{M} \subset \mathbf{F}$ ) вычисляется оператор  $\mathbf{A}\vec{\varphi}$ , т. е. решается прямая задача. В качестве приближенного решения берется такой элемент  $\vec{\varphi}^{\sim}$  из множества  $\mathbf{M}$ , на котором невязка  $\rho(\mathbf{A}\vec{\varphi}^{\sim}, \vec{f}^*)$  достигает минимума, т. е.

$$\vec{\varphi}^{\sim} = \arg \min_{\vec{\varphi} \in \mathbf{M}} \rho(\mathbf{A}\vec{\varphi}, \vec{f}^*).$$

Совершенно очевидно, что успех этого подхода зависит в существенной степени от выбранного (конечного) подкласса возможных решений  $\mathbf{M}$ . Идеально этот подкласс должен содержать точное решение  $\vec{\varphi}_T$ , что никогда, естественно, гарантировать невозможно. Богатый практический опыт решения задачи термического зондирования атмосферы показал, что метод подбора («метод библиотеки») в ряде случаев целесообразно использовать для нахождения *оптимального начального приближения* для решения в дальнейшем нелинейной обратной задачи.

Другая используемая на практике методика решения обратных задач, относящаяся к проекционным методам, основана на *аналитической аппроксимации* искомой функции  $\bar{\varphi}$ . Аналитическая форма искомого решения, содержащая, как правило, очень небольшое (2–5) число неизвестных параметров, выбирается из физических соображений. В этом случае проектирование из пространства, например, непрерывных функций  $C_{[c,d]}$  осуществляется в очень узкий подкласс (подмножество) аналитических функций определенного вида. При этом размерность задачи может уменьшаться очень сильно и, как правило, соответствующая система уравнений для нахождения небольшого числа неизвестных параметров оказывается хорошо обусловленной, и серьезных проблем неустойчивости решения, обусловленных погрешностями измерений  $\bar{\varepsilon}$  и задания прямого оператора задачи  $\mathbf{A}$ , не возникает.

Описанный метод реально эффективен, если, действительно, искомое решение «должно» иметь используемую аналитическую форму. Это требует соответствующих доказательств, что в большинстве случаев не делается. Преимущество данного подхода (если «аналитика» выбрана удачно) состоит в том, что практически никогда нельзя получить «абсурдного» решения. Метод достаточно часто применяется (и с практической точки зрения оправданно) для малоинформативных обратных задач, т. е. обратных задач атмосферной оптики, в которых информация в измерениях излучения об искомой функции невелика. (Об информативности косвенных измерений см. главу 4).

В качестве конкретного примера использования метода аналитической аппроксимации укажем на обратную задачу по определению функции распределения аэрозольных (или облачных) частиц по размерам  $f(r)$  на основе измерений аэрозольных оптических толщ ослабления как функции длины волны  $\tau_a(\lambda)$  или рассеянной солнечной радиации как функции угла рассеяния. В этом случае искомая функция  $f(r)$  аппроксимируется, например, формулой логарифмически нормального закона

$$f(r) = \frac{1}{\sigma_e r \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_e^2} \left(\ln \frac{r}{r_0}\right)^2\right), \quad (3.5-10)$$

где  $r$  – радиус частиц аэрозоля,  $\sigma_e$  и  $r_0$  – искомые параметры распределения.

### 3.6. Методы регуляризации

#### Теоретические основы методов регуляризации

Рассматриваемые методы решения обратных задач используют априорную информацию вида **T2**

$$\mathbf{T2} : \mathbf{R}_0 = \{ \bar{\varphi} : (\bar{\varphi}, \mathbf{H}\bar{\varphi}) \leq C \}$$

с различным видом симметричной положительно определенной матрицы (оператора)  $\mathbf{H}$ . Фактически мы начали рассматривать методы регуляризации в главе 2 (методы Филлипса и Туоми). Теоретическое обоснование методов регуляризации было проведено А. Н. Тихоновым.

Необходимо отметить, что большинство работ по методам регуляризации написано в терминологии функциональных пространств и во многом ориентируется на аналитическую математику. В нашем случае подобный подход не удобен, поскольку проблемы, связанные с некорректными обратными задачами, возникают именно в численных приложениях. Поэтому мы излагаем основные сведения о методах регуляризации применительно к конечномерным операторам и соответствующим им матрицам. Все нижеизложенное легко может быть переформулировано и для функциональных пространств.

В основе рассматриваемого в этом разделе подхода лежит понятие сглаживающего функционала. Минимизация функционала нормы невязки

$$L(\bar{\varphi}) = (\vec{f}^* - \mathbf{A}\bar{\varphi}, \mathbf{S}\vec{f}^* - \mathbf{S}\mathbf{A}\bar{\varphi}) = \|\vec{f}^* - \mathbf{A}\bar{\varphi}\|_{\mathbf{S}}^2 \quad (3.6-1)$$

при наличии априорной информации вида **T2** эквивалентна его минимизации при условии

$$(\bar{\varphi}, \mathbf{H}\bar{\varphi}) = C. \quad (3.6-2)$$

Дело в том, что поскольку, как мы уже обсуждали, наличие в измерениях шума приводит к появлению в решении высокочастотных

компонент, то реализующий безусловный минимум функционала (3.6-1) вектор будет лежать вне шара **T2**. Тогда, поскольку функционал  $L(\vec{\varphi})$  выпуклый, можно показать, что минимум функционала на гипершаре реализуется на его границе, т. е. имеет место равенство (3.6-2). Минимизация нормы невязки (3.6-1) с ограничениями (3.6-2) методом неопределенных множителей Лагранжа (см. Приложение) приводит к решению задачи (3.6-3), параметрически зависящей от неопределенного множителя Лагранжа  $\lambda$

$$\vec{\varphi} = \arg \min_{\varphi} \left( \|\vec{f}^* - \mathbf{A}\vec{\varphi}\|_S^2 + \lambda \left( \|\vec{\varphi}\|_H^2 - C \right) \right), \quad (3.6-3)$$

или к эквивалентной ей задаче минимизации функционала  $\tilde{L}$

$$\tilde{L}(\vec{\varphi}) = \|\vec{f}^* - \mathbf{A}\vec{\varphi}\|_S^2 + \alpha \|\vec{\varphi}\|_H^2 \rightarrow \min. \quad (3.6-4)$$

Заметим, что один и тот же функционал  $\tilde{L}$  был нами получен сначала в предположении о равенстве невязки величине погрешности измерений (раздел 2.5, формула (2.5-6)) при минимизации нормы искомого вектора, а теперь – при «симметричных» предположениях о равенстве нормы искомого вектора константе и минимизации невязки. Опыт показывает, что в первом случае решение получается излишне заглаженным, а во втором – занижается невязка. Можно предположить, что наилучшее решение лежит между двумя этими крайними случаями. Отсюда следует целесообразность одновременной минимизации обоих слагаемых в (3.6-4), при этом параметр  $\alpha$  играет роль «веса», регулирующего соотношение между требованиями минимизации невязки и гладкости решения.

Вернемся к рассмотрению методов регуляризации. Основная идея этих методов состоит в минимизации функционала (3.6-4). В отличие от выражения (3.6-1) формула (3.6-4) содержит два слагаемых. Второе слагаемое носит название *сглаживающего функционала*. Сглаживающий функционал можно записать в виде

$$\|\vec{\varphi}\|_H^2 = (\vec{\varphi}, \mathbf{H}\vec{\varphi}). \quad (3.6-5)$$

Получим явный вид решения, доставляющий минимум функционалу (3.6-4). Теперь, в отличие от МНК, рассмотренного в разделе 3.2, вместо функционала (3.6-1) необходимо минимизировать функционал (3.6-4), который можно преобразовать к следующему виду:

$$\begin{aligned} \tilde{L}(\tilde{\varphi}) = & (\tilde{f}^*, \Sigma^{-1} \tilde{f}^*) - (\tilde{f}^*, \Sigma^{-1} \mathbf{A} \tilde{\varphi}) - (\mathbf{A} \tilde{\varphi}, \Sigma^{-1} \tilde{f}^*) + \\ & + (\mathbf{A} \tilde{\varphi}, \Sigma^{-1} \mathbf{A} \tilde{\varphi}) + \alpha (\tilde{\varphi}, \mathbf{H} \tilde{\varphi}). \end{aligned} \quad (3.6-6)$$

Этот функционал отличается от нормы невязки (3.6-1) только слагаемым  $\alpha (\tilde{\varphi}, \mathbf{H} \tilde{\varphi})$ . Используя необходимое условие экстремума и приведенные в приложении правила дифференцирования скалярных функций, для поиска минимума  $\tilde{L}$  продифференцируем выражение (3.6-6) по  $\tilde{\varphi}$  и приравняем результат нулю:

$$\frac{d}{d\tilde{\varphi}} \tilde{L}(\tilde{\varphi}) = -2\mathbf{A}' \Sigma^{-1} \tilde{f}^* + 2\mathbf{F} \tilde{\varphi} + 2\alpha \mathbf{H} \tilde{\varphi} = \bar{0}. \quad (3.6-7)$$

Вынося  $\tilde{\varphi}$  за скобки, получим

$$(\mathbf{F} + \alpha \mathbf{H}) \tilde{\varphi} = \mathbf{A}' \Sigma^{-1} \tilde{f}^*, \quad (3.6-8)$$

откуда получим выражение для оценки решения

$$\tilde{\varphi} = (\mathbf{F} + \alpha \mathbf{H})^{-1} \mathbf{A}' \Sigma^{-1} \tilde{f}^* = (\mathbf{A}' \Sigma^{-1} \mathbf{A} + \alpha \mathbf{H})^{-1} \mathbf{A}' \Sigma^{-1} \tilde{f}^*. \quad (3.6-9)$$

Из выражения (3.6-9) следует, что для вычисления оценки решения необходимо обратить матрицу  $(\mathbf{F} + \alpha \mathbf{H})$ , отличающуюся от плохо обусловленной матрицы Фишера  $\mathbf{F}$  дополнительным слагаемым  $\alpha \mathbf{H}$ , представляющим собой, по определению, положительно определенную матрицу. Добавление матрицы  $\mathbf{H}$  придает устойчивость решению, так как в этом случае малые собственные числа исходной матрицы  $\mathbf{F}$  суммируются с собственными числами положительной симметричной матрицы  $\mathbf{H}$ , которые всегда положительны и строго больше нуля.

### Анализ решения, полученного методом регуляризации

Для дальнейшего анализа решения (3.6-9) выберем в качестве базиса в пространстве  $\mathbf{R}^n$  собственные вектора матрицы

$$\mathbf{C} = \mathbf{H}^{-1} \mathbf{F} = \mathbf{H}^{-1} \mathbf{A}' \Sigma^{-1} \mathbf{A}. \quad (3.6-10)$$

Заметим, что матрица  $\mathbf{C}$  положительно (неотрицательно) определена как произведение симметричных положительно (неотрицательно) определенных матриц. Но матрица  $\mathbf{C}$  не обязательно симметрична.

Можно показать, что собственные вектора  $\vec{\psi}_i$  матрицы  $\mathbf{C}$  обладают очень важным для анализа свойством двойной ортогональности:

$$(\vec{\psi}_i, \mathbf{F}\vec{\psi}_j) = 0, \quad i \neq j, \quad (3.6-11)$$

$$(\vec{\psi}_i, \mathbf{H}\vec{\psi}_j) = 0, \quad i \neq j, \quad (3.6-12)$$

т. е. они соответствуют независимым (вследствие ортогональности векторов в метрике  $\|\cdot\|_{\mathbf{H}}$ ) компонентам решения. Одновременно эти же компоненты независимы с точки зрения измерений, так как ортогональны в метрике матрицы Фишера.

Выберем нормировку для собственных векторов следующим образом:

$$(\vec{\psi}_i, \mathbf{F}\vec{\psi}_i) = \lambda_i, \quad i = 1, \dots, n. \quad (3.6-13)$$

Тогда, как легко видеть,

$$(\vec{\psi}_i, \mathbf{H}\vec{\psi}_i) = \left( \vec{\psi}_i, \mathbf{H} \frac{1}{\lambda_i} \mathbf{C} \vec{\psi}_i \right) = \frac{1}{\lambda_i} (\vec{\psi}_i, \mathbf{F}\vec{\psi}_i) = 1, \quad i = 1, \dots, n, \quad (3.6-14)$$

где  $\{\lambda_i\}_{i=1,n}$  – собственные числа матрицы  $\mathbf{C}$ . Система векторов  $\{\vec{\psi}_i\}_{i=1,n}$  является базисом пространства  $\mathbf{R}^n$ , хотя и не ортонормальной в исходной метрике.

Получим оценку решения  $\vec{\varphi}$  в виде разложения по базису  $\vec{\psi}_i$

$$\vec{\varphi} = \sum_{k=1}^n a_k \vec{\psi}_k. \quad (3.6-15)$$

Подставляя разложение (3.6-15) в равенство (3.6-8), получим систему независимых уравнений

$$(\mathbf{F} + \alpha \mathbf{H}) \sum_{j=1}^n a_j \vec{\psi}_j = \mathbf{A}' \mathbf{\Sigma}^{-1} \vec{f}^*. \quad (3.6-16)$$

Умножим скалярно обе части этого равенства справа на  $\vec{\psi}_k$ , тогда

$$\sum_{j=1}^n a_j (\mathbf{F}\vec{\psi}_j, \vec{\psi}_k) + \alpha \sum_{j=1}^n a_j (\mathbf{H}\vec{\psi}_j, \vec{\psi}_k) = (\mathbf{A}' \mathbf{\Sigma}^{-1} \vec{f}^*, \vec{\psi}_k). \quad (3.6-17)$$

Воспользовавшись соотношениями (3.6-11)–(3.6-16), получим

$$a_k \lambda_k + \alpha a_k = (\mathbf{A}' \Sigma^{-1} \vec{f}^*, \vec{\psi}_k). \quad (3.6-18)$$

Отсюда легко получить искомые коэффициенты разложения решения  $a_k$ :

$$a_k = \frac{1}{\lambda_k + \alpha} (\mathbf{A}' \Sigma^{-1} \vec{f}^*, \vec{\psi}_k), \quad (3.6-19)$$

и оценка решения (3.6-15) записывается в виде

$$\tilde{\vec{\varphi}} = \sum_{k=1}^n \frac{1}{\lambda_k + \alpha} (\mathbf{A}' \Sigma^{-1} \vec{f}^*, \vec{\psi}_k) \vec{\psi}_k. \quad (3.6-20)$$

Поскольку  $\vec{f}^* = \vec{f} + \vec{\varepsilon}$ , где  $\vec{f} = \mathbf{A} \vec{\varphi}$ , то

$$\tilde{\vec{\varphi}} = \sum_{k=1}^n \frac{a_{fk}}{\lambda_k + \alpha} \vec{\psi}_k + \sum_{k=1}^n \frac{a_{\varepsilon k}}{\lambda_k + \alpha} \vec{\psi}_k, \quad (3.6-21)$$

где введены следующие обозначения:

$$a_{fk} = (\mathbf{A}' \Sigma^{-1} \vec{f}, \vec{\psi}_k), \quad (3.6-22)$$

$$a_{\varepsilon k} = (\mathbf{A}' \Sigma^{-1} \vec{\varepsilon}, \vec{\psi}_k). \quad (3.6-23)$$

Второе слагаемое в разложении решения (3.6-21) соответствует вкладу случайной ошибки измерений. Однако теперь (по сравнению с соотношением (2.7-10)) с увеличением номера  $k$  сильному увеличению вклада случайной ошибки за счет близости  $\lambda_k$  к нулю препятствуют наличие в знаменателе второго слагаемого в формуле (3.6-21) параметра  $\alpha$ , имеющего конечное значение. При малых  $\lambda_k$  множитель при соответствующей компоненте разложения случайной ошибки в записи (3.6-21) теперь ограничен сверху:  $1/(\lambda_k + \alpha) < 1/\alpha$  в отличие от разложения (2.7-10), где множитель  $1/\lambda_k$  может оказаться сколь угодно большим.

Таким образом, множитель  $1/(\lambda_k + \alpha)$  можно рассматривать в качестве регуляризирующего множителя, роль которого состоит в подавлении случайных помех.

Ковариационная матрица ошибок коэффициентов разложения  $a_k$  может быть записана в виде

$$\tilde{\mathbf{D}}_{\bar{a}} = \left( \mathbf{P}'_{\psi} \mathbf{F} \mathbf{P}_{\psi} + \alpha \mathbf{P}'_{\psi} \mathbf{H} \mathbf{P}_{\psi} \right)^{-1}, \quad (3.6-24)$$

где  $\mathbf{P}_{\psi}$  – матрица, столбцы которой есть собственные вектора матрицы  $\mathbf{C} = \mathbf{H}^{-1} \mathbf{F}$ . Воспользовавшись снова свойствами (3.6-11)–(3.6-14), ковариационную матрицу ошибок можно записать

$$\tilde{\mathbf{D}}_{\bar{a}} = \left\| \begin{array}{cccc} \frac{\sigma^2}{\lambda_1 + \alpha \sigma^2} & 0 & \dots & 0 \\ 0 & \frac{\sigma^2}{\lambda_2 + \alpha \sigma^2} & \dots & \cdot \\ \cdot & \dots & \cdot & 0 \\ 0 & \dots & 0 & \frac{\sigma^2}{\lambda_n + \alpha \sigma^2} \end{array} \right\|. \quad (3.6-25)$$

Из полученного выражения видно, что теперь даже в случае вырождения матрицы  $\mathbf{A}' \mathbf{A}$ , когда  $\lambda_k = 0$  при больших номерах  $k$ , дисперсия коэффициента  $\tilde{a}_k$  все равно ограничена величиной  $1/\alpha$  (сравните также с выражением (3.3-13)).

## Различные методы регуляризации

Выбор конкретного вида матрицы  $\mathbf{H}$  определяет конкретный метод регуляризации.

### 1. Решение с минимальной нормой или регуляризация нулевого порядка

В этом случае  $\mathbf{H}$  – единичная матрица  $\mathbf{I}_n$  размерности  $n$ . Тогда с учетом (3.6-9) искомое решение запишется:

$$\tilde{\varphi} = \left( \mathbf{A}' \Sigma^{-1} \mathbf{A} + \alpha \mathbf{I}_n \right)^{-1} \mathbf{A}' \Sigma^{-1} \vec{f}^*. \quad (3.6-26)$$

При  $\alpha = 0$  решение (3.6-26) становится решением по методу наименьших квадратов (см. формулу (3.2-10)). Смысл этого подхода – в поиске решения, минимально отличающегося от нуля.

### 2. Решение, обладающее максимальной гладкостью, или регуляризация первого порядка

Дополнительная априорная информация об искомом решении в этом случае заключается в требовании, чтобы в качестве решения была получена наиболее гладкая функция, т. е. функция, обладающая

минимальной первой производной. Здесь  $\mathbf{H}$  представляет матрицу, аппроксимирующую в конечных разностях дифференциальный оператор второго порядка

$$\tilde{\varphi} = (\mathbf{A}' \Sigma^{-1} \mathbf{A} + \alpha \mathbf{H})^{-1} \mathbf{A}' \Sigma^{-1} \vec{f}^*. \quad (3.6-27)$$

### Выбор параметра регуляризации

В формулах (3.6-26) и (3.6-27) параметр  $\alpha$  – параметр регуляризации, позволяющий «регулировать» влияние двух членов функционала  $\tilde{L}(\varphi)$  на решение задачи. Величина  $\alpha$  зависит, вообще говоря, от константы  $C$  в условии **T2**. Выбор оптимального значения  $\alpha$  представляет собой значительные трудности при практическом решении обратных задач методом регуляризации. Один из известных способов его выбора состоит в том, что решения уравнений (3.6-26) или (3.6-27) рассматриваются как параметрически зависящие от  $\alpha$ :  $\tilde{\varphi} = \tilde{\varphi}(\alpha)$ . В качестве решения выбирается такое  $\tilde{\varphi}(\alpha_{opt})$ , при котором достигается минимальное значение невязки:

$$\|\vec{f}^* - \mathbf{A} \tilde{\varphi}(\alpha)\|_{\Sigma^{-1}}^2 = \min. \quad (3.6-28)$$

Такой подход требует построения семейства решений при различных  $\alpha$  и нахождения по соотношению (3.6-28)  $\alpha_{opt}$ . При этом в ряде случаев важное значение имеет выбор сетки по  $\alpha$  (последовательности значений  $\alpha$ ). Определенное противоречие в вычислительной эффективности этой методики состоит в том, что для наиболее точного определения минимума функционала (3.6-28) требуется очень подробная сетка по  $\alpha$ , что приводит к большому объему вычислений, т. е. к низкой эффективности решения обратной задачи. Это обстоятельство играет значительную роль в оперативных системах дистанционного зондирования атмосферы, предназначенных для обработки огромных объемов информации за ограниченное время. Для уменьшения объема вычислений в ряде работ в качестве последовательности параметра регуляризации  $\alpha$  предлагается использовать, например, отрезок геометрической прогрессии  $\alpha_k = \alpha_0 q^k$ ,  $k = 0, 1, \dots, n$ ,  $q > 0$ .

Если измерения равноточные и их ошибки независимы, а дисперсия шума  $\sigma^2$ , то при выборе параметра  $\alpha$  условие (3.6-28) можно заменить на

$$\|\vec{f}^* - \mathbf{A}\tilde{\varphi}(\alpha)\|^2 = \sigma^2 m, \quad (3.6-29)$$

где  $m$  – размерность пространства измерений. Наконец, отметим, что иногда для нахождения  $\alpha$  используют так называемый критерий «сходимости в себе». В этом случае  $\alpha$  и соответствующее решение  $\tilde{\varphi}(\alpha_{opt})$  выбирается на основе критерия

$$\max |\tilde{\varphi}(\alpha_n) - \tilde{\varphi}(\alpha_{n+1})| = \min, \quad (3.6-30)$$

т. е. в окрестности того значения  $\alpha$ , где наблюдается стабилизация искомого решения.

Бросается в глаза очевидное логическое противоречие перечисленных методов выбора параметра регуляризации, использующих критерии (3.6-28), (3.6-29), (3.6-30) и исходной постановки задачи (3.6-4) с ограничением **T2**. Дело в том, что описываемые подходы к определению величины параметра  $\alpha$ , возникшие на заре развития методов решения некорректных задач, являются фактически методом «борьбы» с неустойчивостью процедур обращения плохо обусловленной матрицы. Эта неустойчивость вызвана принципиально конечным представлением чисел в вычислительных системах и может возрасти при неудачном выборе алгоритма обращения матрицы. За счет ошибок вычислений при малых  $\alpha$  оценка решения  $\tilde{\varphi}(\alpha)$  перестает быть решением исходной задачи (3.6-4) с ограничением **T2** и является просто «вычислительным шумом». При увеличении  $\alpha$  число обусловленности обращаемой матрицы уменьшается, алгоритмы начинают работать и оценка  $\tilde{\varphi}(\alpha)$  приближается к своему значению, соответствующему исходной задаче. Таким образом, описанные методы выбора параметра  $\alpha$  являются попросту проверкой достоверности вычислений путем подстановки полученного решения в исходные соотношения. В результате они позволяют получить решение, свойства которого определяются используемыми алгоритмами обращения матриц и вычислительной системой, чему вряд ли можно приписать какой-либо физический смысл.

Интуитивно из вида соотношения (3.6-4) понятно, что величина параметра  $\alpha$  является «весом», регулирующим соотношение между требованиями к искомой функции и невязкой измерений. Поэтому выбор ее должен опираться на соотношение между требованиями к искомой функции, выраженными константой  $C$  в записи ограничений **T2**, и величиной погрешности измерений.

Как мы заметили в начале раздела, традиционное изложение методов регуляризации ведется в терминах аналитической математики. Приведем для примера в этой терминологии построение регуляризованного решения интегрального уравнения Фредгольма I рода на конечном промежутке  $[a, b]$  в общем случае (пример заимствован из книги [15]).

Рассмотрим интегральное уравнение

$$\int_a^b A(x, y)\varphi(x)dx = f(y), \quad (3.6-31)$$

где  $y \in [c, d]$ ,  $f(y) \in L_2[c, d]$ .

Будем пользоваться регуляризатором (стабилизатором) первого порядка. (см. выше). Регуляризованное решение  $\varphi_\alpha(x)$  должно минимизировать функционал

$$\begin{aligned} L^\alpha(\varphi, f) = & \int_c^d \left( \int_a^b A(x, y)\varphi(x)dx - f(y) \right)^2 dy + \\ & + \alpha \int_a^b \left( q_0(x)\varphi^2(x) + q_1(x) \left( \frac{d\varphi}{dx} \right)^2 \right) dx. \end{aligned} \quad (3.6-32)$$

Необходимым условием минимума функционала является равенство нулю его первой вариации, которую можно записать в следующем виде:

$$\begin{aligned} & \int_a^b \left( -\alpha \left( \frac{d}{dx} \left( q_1(x) \frac{d\varphi}{dx} \right) - q_0(x)\varphi(x) \right) + \right. \\ & \left. + \int_a^b \bar{k}(x, t)\varphi(t)dt - b(x) \right) v(x)dx + \alpha q_1(x)\varphi'(x)v(x) \Big|_a^b. \end{aligned} \quad (3.6-33)$$

Здесь  $v(x)$  – произвольная вариация функции  $\varphi(x)$  такая, что  $\varphi(x)$  и  $\varphi(x)+v(x)$  принадлежат классу допустимых функций. В выражении (3.6-33) используются обозначения

$$\bar{k}(x,t) = \int_c^d k(l,x)k(l,t)dl,$$

$$b(x) = \int_c^d (l,x)f(l)dl.$$

Равенство нулю выражения (3.6-33) выполняется, если

$$\int_a^b \bar{k}(x,t)\varphi(t)dt - \alpha \left( \frac{d}{dx} \left( q_1(x) \frac{d\varphi(x)}{dx} \right) - q_0(x)\varphi(x) \right) = b(x) \quad (3.6-34)$$

и

$$q_1(x) \varphi'(x) v(x) \Big|_a^b = 0. \quad (3.6-35)$$

Таким образом, задача нахождения регуляризованного решения в рассматриваемом случае сводится к нахождению решения интегродифференциального уравнения (3.6-34) при определенных граничных условиях. Если при решении интегрального уравнения (3.6-31) пользоваться регуляризатором (стабилизатором) порядка  $p$ , то уравнение Эйлера для функционала  $L^\alpha(\varphi,t)$  будет иметь вид

$$\int_a^b \bar{k}(x,t)\varphi(t)dt - \alpha \sum_{r=0}^p (-1)^r \frac{d^r}{dx^r} \left( q_r(x) \frac{d^r \varphi(x)}{dx^r} \right) = b(x) \quad (3.6-36)$$

с соответствующими (краевыми) граничными условиями.

### 3.7. Статистические методы решения обратных задач

Как мы уже отмечали ранее, важной чертой рассматриваемых нами обратных задач является вероятностная природа величин, наблюдаемых в реальном эксперименте. Это обусловлено наличием в измерениях «случайных помех», искажающих «полезный сигнал». В рассматриваемых обратных задачах атмосферной оптики и искомые функции можно рассматривать как случайные функции (реализации).

Естественно, что эта стохастичность должна явным образом входить в формулировку обратной задачи.

### **Постановка задачи в стохастической форме**

Из методов, основанных на стохастическом подходе к решению обратных задач, наиболее широко применяется метод статистической регуляризации. Метод был независимо предложен приблизительно одновременно В. Ф. Турчиным, Вествотером–Стрендом и Роджерсом [16, 20, 23]. В этом подходе априорная информация об искомой функции вносится путем задания некоторого распределения вероятностей, что позволяет оценить распределение вероятностей оценки решения. Существуют различные способы задания априорного распределения вероятностей, поэтому существуют и различные варианты метода статистической регуляризации. Хотя по сравнению с введением жестких ограничений на искомую функцию введение ограничений вероятностного характера требует в ряде случаев большей априорной информации, оно обладает рядом преимуществ.

1) Задача решения некорректного уравнения в приложениях выступает как задача обработки экспериментальных данных, для которой введение вероятностных характеристик является неизбежным, ибо ошибка, по крайней мере в определенной части, носит случайный характер и может быть охарактеризована лишь вероятностным образом. Поэтому вероятностный способ задания априорной информации приводит к использованию единообразного математического аппарата и является более естественным. Это особенно ясно видно, когда встает вопрос об ошибке решения обратной задачи (восстановления искомого профиля).

2) Вероятностный способ задания априорной информации позволяет в полной мере использовать предыдущий опыт, учитывая его в априорном распределении.

3) Наконец, когда такого опыта нет, вероятностный способ позволяет сформулировать весьма слабые предположения об искомой функции, принципиально невыразимые на языке теории множеств, например о принадлежности решения к «слоистому ансамблю» [16].

Чтобы изложить сущность метода статистической регуляризации, необходимо сформулировать рассматриваемую задачу как задачу математической статистики и ввести соответствующие понятия. Будем рассматривать алгебраизированный аналог интегрального уравнения

$$\sum_{j=1}^n a_{i,j} \varphi_j = f_i, \quad i = 1, \dots, m, \quad (3.7-1)$$

где коэффициенты  $a_{i,j}$  образуют матрицу  $\mathbf{A}$ . При этом не имеет значения, существует ли точное решение уравнения (3.7-1), и число уравнений  $m$  может в любую сторону отличаться от числа неизвестных  $n$ . Величины  $\varphi_j$  могут быть любыми линейными функционалами от  $\varphi(x)$ , однако для метода статистической регуляризации обычно выбирают в качестве  $\varphi_j$  либо значения  $\varphi(x)$  в опорных точках  $x_j$ , либо средневзвешенные значения на интервалах  $[x_j, x_{j+1}]$  и т. д. Вектора  $\vec{\varphi}$  и  $\vec{f}$  принадлежат евклидовым пространствам  $\mathbf{R}^n$  и  $\mathbf{R}^m$ .

Пусть измеряются  $m$  величин  $f_i^*$ . Из-за ошибок измерений величины  $f_i^*$  отличаются от их точных значений  $f_i$ , даваемых уравнением (3.7-1). Это обстоятельство можно учесть с помощью записи

$$\sum_{j=1}^n a_{i,j} \varphi_j + \varepsilon_i = f_i^*, \quad i = 1, \dots, m, \quad (3.7-2)$$

где  $\varepsilon_i$  – ошибки измерений величин  $f_i$ , образующие случайный вектор  $\vec{\varepsilon}$  в пространстве  $\mathbf{R}^m$ . Обозначим через  $P_\varepsilon(\vec{\varepsilon})$  плотность вероятности этого вектора. Напомним, что величина  $P(\vec{\varepsilon})d\vec{\varepsilon}$ , где  $d\vec{\varepsilon} = (d\varepsilon_1, \dots, d\varepsilon_m)$ , определяет вероятность того, что вектор  $\vec{\varepsilon}$  находится в элементарном объеме («гиперпараллелепипеде») пространства  $\mathbf{R}^m$ , являющемся прямым произведением интервалов

$$[\varphi_1, \varphi_1 + d\varphi_1] \times [\varphi_2, \varphi_2 + d\varphi_2] \times \dots \times [\varphi_m, \varphi_m + d\varphi_m].$$

Очевидно, что измеряемые величины  $f_i^*$  также являются случайными и зависят как от истинных значений  $\varphi_j$ , так и ошибок  $\varepsilon_i$ . Условная плотность вероятности (см. Приложение) вектора  $\vec{f}^*$  при заданном векторе  $\vec{\varphi}$  есть

$$P(\vec{f}^* / \vec{\varphi}) = P_\varepsilon(\vec{f}^* - \mathbf{A}\vec{\varphi}). \quad (3.7-3)$$

Заметим, что функция  $P(\vec{f}^* / \vec{\varphi})$   $m+n$  переменных  $f_i^*, i=1, \dots, m$ , и  $\varphi_j, j=1, \dots, n$ , полностью характеризует процесс измерений, т. е. как связь между  $\vec{\varphi}$  и  $\vec{f}$ , так и статистические свойства ошибок измерений.

Рассматриваемую задачу можно сформулировать следующим образом. Неизвестные величины  $\varphi_j$  характеризуют состояние физической среды (например атмосферы). Некий случайный процесс дает нам величины  $f_i^*$ , причем условная плотность вероятности  $P(\vec{f}^* / \vec{\varphi})$  известна. Что можно по величинам  $f_i^*$  сказать о величинах  $\varphi_j$ ? Ответ на подобного рода вопросы и есть основная задача математической статистики. Другая формулировка задачи может быть такова – сколь вероятны те или иные значения величин  $\varphi_j$ ? Иначе говоря, какова плотность вероятности  $P(\vec{\varphi} / \vec{f}^*)$  вектора  $\vec{\varphi}$  при условии, что в результате измерений получен вектор  $\vec{f}^*$ ?

Чтобы ответить на этот вопрос, необходимо ввести априорную плотность вероятности  $P(\vec{\varphi})$ . Для наглядности можно представить себе, что опыт по измерению  $\vec{f}^*$  является одним из серии подобных опытов, которые производятся при различных состояниях среды  $\vec{\varphi}$ , выбираемых случайно в соответствии с плотностью вероятности  $P(\vec{\varphi})$ . Тогда апостериорная вероятность определяется по формуле Байеса (см. Приложение):

$$P(\vec{\varphi} / \vec{f}^*) = \frac{P(\vec{\varphi})P(\vec{f}^* / \vec{\varphi})}{\int P(\vec{\varphi})P(\vec{f}^* / \vec{\varphi}) d\varphi}. \quad (3.7-4)$$

Без задания априорного распределения вероятности апостериорная вероятность  $P(\vec{\varphi} / \vec{f}^*)$  не может быть определена. Из формулы Байеса (3.7-4) мы видим, что апостериорная вероятность есть произведение априорной вероятности  $P(\vec{\varphi})$  и условной вероятности  $P(\vec{f}^* / \vec{\varphi})$  (знаменатель не зависит от  $\vec{\varphi}$  и вводится для нормировки). Поэтому апостериорная информация о  $\vec{\varphi}$  есть сумма априорной и извлеченной из опыта информации. Если один из сомножителей гораздо информативнее другого, то апостериорная вероятность близка к наиболее информативному сомножителю и почти не зависит от

второго. Этот случай встречается очень часто, причем информативным фактором, разумеется, является условная вероятность  $P(\vec{f}^* / \vec{\varphi})$ , являющаяся результатом измерений, иначе незачем было бы ставить эксперимент. Пусть, например,  $P(\vec{f}^* / \vec{\varphi})$ , как функция  $\vec{\varphi}$ , имеет резкий максимум в некоторой области пространства  $V_0$  и быстро стремится к нулю за пределами этой области, а априорная вероятность  $P(\vec{\varphi})$  слабо меняется в этой области. Тогда от конкретного вида априорного распределения  $P(\vec{\varphi})$  апостериорная вероятность  $P(\vec{\varphi} / \vec{f}^*)$  зависит мало, и поэтому мы можем положить  $P(\vec{\varphi}) = \text{const}$  в любой области  $V_1$ , включающей  $V_0$ .

Можно показать, что обычное, традиционное (например по МНК) решение системы уравнений (3.7-1) без применения методов регуляризации вместе с обычной оценкой ошибки по методам классической теории ошибок есть байесовская стратегия для априорной плотности вероятности  $P(\vec{\varphi}) = \text{const}$ . Таким образом, использование априорной информации не есть исключительная привилегия регуляризационных методов. Принципиально оно имеет место всегда. Регуляризационные методы отличаются тем, что они используют *нетривиальную* априорную информацию, и при этом априорная информация рассматривается в явном виде, что позволяет придать ей наглядное физическое истолкование.

### Нормальное распределение ошибок измерений

Конкретизируем функцию плотности вероятности ошибок. Сделаем обычные допущения, что ошибки  $\varepsilon_i$  при разных  $i$  независимы (т. е. не коррелируют) и распределены по нормальному закону с математическим ожиданием (средним), равным нулю. Среднеквадратическую ошибку измерений обозначим через  $\sigma_i$ . Тогда, по определению, плотность распределения случайной величины  $\varepsilon_i$  описывается выражением

$$P(\varepsilon_i) = \sigma_i \sqrt{2\pi} \exp\left(-\frac{1}{2} \left(\frac{\varepsilon_i}{\sigma_i}\right)^2\right).$$

Поскольку случайные величины  $\varepsilon_i$  при разных  $i$  независимы, то плотность совместного их распределения равна произведению плотностей

$$\begin{aligned} P(\vec{\varepsilon}) &= \prod_{i=1}^m \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma_i}\right)^2\right) = \\ &= (2\pi)^{-m/2} \prod_{i=1}^m \frac{1}{\sigma_i} \exp\left(-\frac{1}{2}\sum_{i=1}^m \left(\frac{\varepsilon_i}{\sigma_i}\right)^2\right). \end{aligned} \quad (3.7-5)$$

Тогда условная вероятность (3.7-3) примет вид

$$P(\vec{f}^* / \vec{\varphi}) = \prod_{i=1}^m (2\pi\sigma_i^2)^{-1/2} \exp\left(-\frac{1}{2\sigma_i^2}\left(f_i^* - \sum_{j=1}^n a_{i,j}\varphi_j\right)^2\right). \quad (3.7-6)$$

В более общем случае недиагональной ковариационной матрицы ошибок  $\Sigma$  выражение (3.7-6) примет вид (3.7-7)

$$P(\vec{f}^* / \vec{\varphi}) = \frac{1}{(2\pi)^{m/2} \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}\|f^* - \mathbf{A}\vec{\varphi}\|_{\Sigma^{-1}}^2\right), \quad (3.7-7)$$

где  $|\Sigma|$  – определитель матрицы  $\Sigma$ . Напомним, что значения  $\vec{f}^*$  в формулах (3.7-6), (3.7-7) есть измеренные с погрешностью  $\vec{\varepsilon}$  значения, а  $\vec{\varphi}$  точное решение. Тогда разность под знаком нормы в экспоненте в выражении (3.7-7) есть величина  $\vec{\varepsilon}$  – случайные погрешности измерений.

### Решение для нормального априорного распределения

Использование формулы Бейеса, соотношений (3.7-6) и (3.7-7) при различных предположениях об априорной плотности вероятности  $P(\vec{\varphi})$  позволяет получить различные варианты метода статистической регуляризации. Рассмотрим случай задания априорной информации об искомом решении  $\vec{\varphi}$  в виде плотности вероятности, соответствующей многомерному нормальному распределению

$$P(\vec{\varphi}) = \frac{1}{(2\pi)^{m/2} |\mathbf{D}|^{1/2}} \exp\left(-\frac{1}{2}(\vec{\varphi}, \mathbf{D}^{-1}\vec{\varphi})\right), \quad (3.7-8)$$

где показатель экспоненты есть скалярное произведение,  $\mathbf{D}$  – ковариационная матрица, диагональные элементы которой есть дисперсии соответствующих компонент вектора  $\vec{\varphi}$ , а внедиагональные элементы – коэффициенты ковариации. Напомним, что нормальное распределение (3.7-8) полностью характеризуется двумя характеристиками: средним значением  $\vec{\varphi}$  и матрицей  $\mathbf{D}$ .

Пусть проведены измерения вектора  $\vec{f}^*$ . Напомним формулу Байеса (3.7-4)

$$P(\vec{\varphi} / \vec{f}^*) = \frac{P(\vec{\varphi})P(\vec{f}^* / \vec{\varphi})}{\int P(\vec{\varphi})P(\vec{f}^* / \vec{\varphi})},$$

где  $P(\vec{\varphi} / \vec{f}^*)$  и  $P(\vec{f}^* / \vec{\varphi})$  представляют собой условные плотности вероятностей. В первом случае – плотности вероятности распределения вектора  $\vec{\varphi}$  при условии, что задано значение вектора  $\vec{f}^*$ . Во втором – плотности вероятности  $\vec{f}^*$  при условии, что известен вектор  $\vec{\varphi}$ . Выражение  $P(\vec{\varphi} / \vec{f}^*)$ , как мы уже отмечали выше, называют апостериорной плотностью вероятности, т. е. плотностью вероятности, описывающей распределение  $\vec{\varphi}$  после того, как получены результаты измерения  $\vec{f}^*$ .

На основе соотношения (3.7-4) можно построить наиболее вероятное решение уравнения (3.7-1). В качестве оценки решения  $\vec{\varphi}$  можно взять вектор, обеспечивающий максимум выражения (3.7-4), т. е. «наиболее вероятный» вектор для апостериорной плотности вероятности. Заметим, что из соображений симметрии в данном случае максимум апостериорного распределения вектора  $\vec{\varphi}$  совпадает с его математическим ожиданием.

Для решения этой задачи необходимо прежде всего найти условную плотность вероятности  $P(\vec{f}^* / \vec{\varphi})$ . Из выражения (3.7-1) следует, что искомая плотность

$$P(\vec{f}^* / \vec{\varphi}) = P_{\varepsilon}(\vec{f}^* - \mathbf{A}\vec{\varphi}). \quad (3.7-9)$$

Поскольку

$$\vec{f}^* - \mathbf{A}\vec{\varphi} = \vec{\varepsilon}, \quad (3.7-10)$$

то условная плотность вероятности  $P(\vec{f}^* / \vec{\varphi})$  с точностью до сдвига на вектор  $\mathbf{A}\vec{\varphi}$  совпадает с плотностью вероятности ошибок измерений  $P(\vec{\varepsilon})$ . Учитывая соотношение (3.7-9), выражение (3.7-4) с точностью до нормировки правой части можно записать

$$P(\vec{\varphi} / \vec{f}^*) = P_{\varepsilon}(\vec{f}^* - \mathbf{A}\vec{\varphi})P(\vec{\varphi}). \quad (3.7-11)$$

Для нахождения максимума (3.7-11) учтем, что для  $P(\vec{\varphi})$  справедливо соотношение (3.7-8), а для функционала  $P_{\varepsilon}(\vec{f}^* - \mathbf{A}\vec{\varphi})$  – (3.7-7). Таким образом, для решения поставленной задачи требуется найти максимум по  $\vec{\varphi}$  следующего выражения:

$$P(\vec{\varphi} / \vec{f}^*) = C \exp\left(-\frac{1}{2}\|\vec{f}^* - \mathbf{A}\vec{\varphi}\|_{\Sigma^{-1}}^2 - \frac{1}{2}\|\vec{\varphi}\|_{\mathbf{D}^{-1}}^2\right) \rightarrow \max, \quad (3.7-12)$$

где константа  $C$  не зависит от вектора  $\vec{\varphi}$ . Для поиска решающего экстремальную задачу (3.7-12) вектора  $\vec{\varphi}$  воспользуемся необходимым условием экстремума

$$\frac{d}{d\vec{\varphi}} P(\vec{\varphi} / \vec{f}^*) = 0. \quad (3.7-13)$$

Согласно правилам дифференцирования и используя выражение нормы через скалярное произведение

$$\|\vec{f}^* - \mathbf{A}\vec{\varphi}\|_{\Sigma^{-1}} = (\vec{f}^* - \mathbf{A}\vec{\varphi}, \mathbf{S}\vec{f}^* - \mathbf{S}\mathbf{A}\vec{\varphi}) \equiv L(\vec{\varphi}),$$

где  $\mathbf{S} = \Sigma^{-1}$ , получим

$$\exp\left(-\frac{1}{2}L(\vec{\varphi}) - \frac{1}{2}(\vec{\varphi}, \mathbf{D}^{-1}\vec{\varphi})\right) \frac{d}{d\vec{\varphi}}(L(\vec{\varphi}) + (\vec{\varphi}, \mathbf{D}^{-1}\vec{\varphi})) = 0. \quad (3.7-14)$$

Сокращая на заведомо отличный от нуля множитель, окончательно получим следующее уравнение для нахождения вектора  $\vec{\varphi}$ , обеспечивающего максимум  $P(\vec{\varphi} / \vec{f}^*)$

$$\frac{d}{d\vec{\varphi}}(L(\vec{\varphi}) + (\vec{\varphi}, \mathbf{D}^{-1}\vec{\varphi})) = 0, \quad (3.7-15)$$

которое совпадает с уже рассмотренным ранее при решении задачи (3.6-4) при  $\mathbf{S} = \Sigma^{-1}$ ,  $\mathbf{H} = \mathbf{D}^{-1}$ ,  $\alpha = 1$ . Следовательно, воспользовавшись результатами раздела 3.6, оценку решения можно записать в виде

$$\tilde{\varphi} = (\mathbf{F} + \mathbf{D}^{-1})^{-1} \mathbf{A}' \Sigma^{-1} \tilde{f}^* = (\mathbf{A}' \Sigma^{-1} \mathbf{A} + \mathbf{D}^{-1})^{-1} \mathbf{A}' \Sigma^{-1} \tilde{f}^*. \quad (3.7-16)$$

Практическое использование выражения (3.7-16) для нахождения оценки решения не всегда является удобным, так как размерность матрицы  $\mathbf{D}$  может быть большой, а сама матрица плохо обусловленной. Можно записать эквивалентное (3.7-16) соотношение, воспользовавшись равенством

$$(\mathbf{F} + \mathbf{D}^{-1})^{-1} \mathbf{A}' \Sigma^{-1} = \mathbf{D} \mathbf{A}' (\mathbf{A} \mathbf{D} \mathbf{A}' + \Sigma)^{-1}. \quad (3.7-17)$$

С учетом (3.7-17) оценка решения может быть записана в виде

$$\tilde{\varphi} = \mathbf{D} \mathbf{A}' (\mathbf{A} \mathbf{D} \mathbf{A}' + \Sigma)^{-1} \tilde{f}^*, \quad (3.7-18)$$

где обращение матрицы  $\mathbf{D}$  не требуется. На практике в зависимости от конкретной задачи может быть использовано наиболее удобное из выражений (3.7-16) или (3.7-18).

Для доказательства справедливости соотношения (3.7-17), умножив его слева на  $(\mathbf{F} + \mathbf{D}^{-1})$  и справа на  $(\mathbf{A} \mathbf{D} \mathbf{A}' + \Sigma)$ , получим

$$\mathbf{A}' \Sigma^{-1} (\mathbf{A} \mathbf{D} \mathbf{A}' + \Sigma) = (\mathbf{F} + \mathbf{D}^{-1}) \mathbf{D} \mathbf{A}'.$$

После этого, раскрыв скобки, получим

$$\mathbf{A}' \Sigma^{-1} \mathbf{A} \mathbf{D} \mathbf{A}' + \mathbf{A}' \Sigma^{-1} \Sigma = \mathbf{F} \mathbf{D} \mathbf{A}' + \mathbf{D}^{-1} \mathbf{D} \mathbf{A}'$$

и наконец, сократив произведения  $\Sigma^{-1} \Sigma$  и  $\mathbf{D}^{-1} \mathbf{D}$  и выполнив замену  $\mathbf{A}' \Sigma^{-1} \mathbf{A} = \mathbf{F}$ , получим тривиальное тождество

$$\mathbf{F} \mathbf{D} \mathbf{A}' + \mathbf{A}' = \mathbf{F} \mathbf{D} \mathbf{A}' + \mathbf{A}'.$$

Таким образом, соотношение (3.7-17) эквивалентно тождественному равенству, т. е. также является тождественным равенством.

### **Анализ решения, полученного методом статистической регуляризации**

Можно проанализировать устойчивость полученного решения, используя тот же подход, что и для решения в форме (3.6-9). При этом в матрице  $\mathbf{C} = \mathbf{H}^{-1} \mathbf{F}$  необходимо заменить  $\mathbf{H}^{-1}$  на  $\mathbf{D}$ . В случае решения в форме (3.7-16) регуляризирующий множитель имеет вид

$$\frac{\lambda_i}{\lambda_i + 1}. \quad (3.7-19)$$

В соответствии с рассмотренной выше статистической трактовкой оценку решения  $\tilde{\varphi}$  естественно считать случайной функцией, как в связи с наличием случайных ошибок  $\bar{\varepsilon}$ , так и в связи с самой статистической (вероятностной) природой атмосферных параметров. Для оценки дисперсии ошибки определения  $\tilde{\varphi}$  запишем решение в общем виде

$$\tilde{\varphi} = \mathbf{R} \vec{f}^*, \quad (3.7-20)$$

где  $\mathbf{R}$  – решающий оператор. Например, в соответствии с выражениями (3.7-16) и (3.7-18) решающий оператор может быть записан в следующих формах:

$$\mathbf{R} = (\mathbf{F} + \mathbf{D}^{-1})^{-1} \mathbf{A}' \Sigma^{-1} \quad (3.7-21a)$$

или

$$\mathbf{R} = \mathbf{D} \mathbf{A}' (\mathbf{A} \mathbf{D} \mathbf{A}' + \Sigma)^{-1}. \quad (3.7-21b)$$

Получим выражение для ковариационной матрицы ошибок оценки решения

$$\tilde{\mathbf{D}} = E((\tilde{\varphi} - \bar{\varphi})(\tilde{\varphi} - \bar{\varphi})'), \quad (3.7-22)$$

где  $E$  – математическое ожидание. Подставив в выражение (3.7-22) оценку вектора  $\tilde{\varphi}$  (3.7-20) и учитывая, что  $\vec{f}^* = \mathbf{A} \bar{\varphi} + \bar{\varepsilon}$ , получим

$$\tilde{\mathbf{D}} = E((\mathbf{R}(\mathbf{A} \bar{\varphi} + \bar{\varepsilon}) - \bar{\varphi})(\mathbf{R}(\mathbf{A} \bar{\varphi} + \bar{\varepsilon}) - \bar{\varphi})'). \quad (3.7-23)$$

Учтем тот факт, что все величины, содержащие сомножители  $E(\bar{\varphi}, \bar{\varepsilon}')$  дадут при усреднении ноль, так как мы предполагаем, что вектор  $\bar{\varphi}$  и ошибки  $\bar{\varepsilon}$  некоррелированы. В результате матрицу  $\tilde{\mathbf{D}}$  можно записать в следующем виде:

$$\begin{aligned} \tilde{\mathbf{D}} = & E(\bar{\varphi} \bar{\varphi}') - E(\bar{\varphi} \bar{\varphi}' \mathbf{A}' \mathbf{R}') - E(\mathbf{R} \mathbf{A} \bar{\varphi} \bar{\varphi}') + \\ & + E(\mathbf{R} \mathbf{A} \bar{\varphi} \bar{\varphi}' \mathbf{A}' \mathbf{R}') + E(\mathbf{R} \bar{\varepsilon} \bar{\varepsilon}' \mathbf{R}'). \end{aligned}$$

Учитывая, что по определению  $E(\bar{\varphi} \bar{\varphi}') = \mathbf{D}$ , т. е. равно априорной ковариационной матрице, получим

$$\tilde{\mathbf{D}} = \mathbf{D} - \mathbf{R} \mathbf{A} \mathbf{D} - \mathbf{D} \mathbf{A}' \mathbf{R}' + \mathbf{R} \mathbf{A} \mathbf{D} \mathbf{A}' \mathbf{R}' + \mathbf{R} \Sigma \mathbf{R}'. \quad (3.7-24)$$

Объединим сумму четырех первых слагаемых в правой части выражения (3.7-24) и запишем окончательное выражение для матрицы  $\tilde{\mathbf{D}}$ :

$$\tilde{\mathbf{D}} = (\mathbf{I} - \mathbf{R}\mathbf{A})\mathbf{D}(\mathbf{I} - \mathbf{R}\mathbf{A})' + \mathbf{R}\Sigma\mathbf{R}', \quad (3.7-25)$$

где  $\mathbf{I}$  – единичная матрица. Из выражения (3.7-25) видно, что дисперсия ошибки решения состоит из двух слагаемых: первое слагаемое называют *смещением*, и оно соответствует систематической ошибке в оценке вектора  $\bar{\varphi}$ , второе слагаемое описывает случайную компоненту решения. Отметим, что систематическая ошибка через решающий оператор  $\mathbf{R}$  также зависит от ковариационной матрицы случайной ошибки измерений.

Подставив тот или иной вид решающего оператора  $\mathbf{R}$  в соотношение (3.7-25), можно получить конкретный вид матрицы ошибок  $\tilde{\mathbf{D}}$ . Например, если воспользоваться выражением (3.7-21б)  $\mathbf{R} = \mathbf{D}\mathbf{A}'(\mathbf{A}\mathbf{D}\mathbf{A}' + \Sigma)^{-1}$  для решающего оператора, то получим

$$\begin{aligned} \tilde{\mathbf{D}} = & (\mathbf{I} - \mathbf{D}\mathbf{A}'(\mathbf{A}\mathbf{D}\mathbf{A}' + \Sigma)^{-1}\mathbf{A})\mathbf{D}(\mathbf{I} - \mathbf{A}'(\mathbf{A}\mathbf{D}\mathbf{A}' + \Sigma)^{-1}\mathbf{A}\mathbf{D}) + \\ & + \mathbf{D}\mathbf{A}'(\mathbf{A}\mathbf{D}\mathbf{A}' + \Sigma)^{-1}\Sigma(\mathbf{A}\mathbf{D}\mathbf{A}' + \Sigma)^{-1}\mathbf{A}\mathbf{D}. \end{aligned}$$

Раскроем скобки в первом слагаемом и приведем подобные:

$$\begin{aligned} \tilde{\mathbf{D}} = & \mathbf{D} - 2\mathbf{D}\mathbf{A}'(\mathbf{A}\mathbf{D}\mathbf{A}' + \Sigma)^{-1}\mathbf{A}\mathbf{D} + \\ & + \mathbf{D}\mathbf{A}'(\mathbf{A}\mathbf{D}\mathbf{A}' + \Sigma)^{-1}\mathbf{A}\mathbf{D}\mathbf{A}'(\mathbf{A}\mathbf{D}\mathbf{A}' + \Sigma)^{-1}\mathbf{A}\mathbf{D} + \\ & + \mathbf{D}\mathbf{A}'(\mathbf{A}\mathbf{D}\mathbf{A}' + \Sigma)^{-1}\Sigma(\mathbf{A}\mathbf{D}\mathbf{A}' + \Sigma)^{-1}\mathbf{A}\mathbf{D}. \end{aligned}$$

Вынося за скобки слева  $\mathbf{D}\mathbf{A}'(\mathbf{A}\mathbf{D}\mathbf{A}' + \Sigma)^{-1}$  и справа  $(\mathbf{A}\mathbf{D}\mathbf{A}' + \Sigma)^{-1}\mathbf{A}\mathbf{D}$  из всех слагаемых в правой части, кроме первого, получим

$$\begin{aligned} \tilde{\mathbf{D}} = & \mathbf{D} + \mathbf{D}\mathbf{A}'(\mathbf{A}\mathbf{D}\mathbf{A}' + \Sigma)^{-1}(-2(\mathbf{A}\mathbf{D}\mathbf{A}' + \Sigma) + \mathbf{A}\mathbf{D}\mathbf{A}' + \Sigma) \times \\ & \times (\mathbf{A}\mathbf{D}\mathbf{A}' + \Sigma)^{-1}\mathbf{A}\mathbf{D}. \end{aligned}$$

Таким образом, окончательно можно записать

$$\tilde{\mathbf{D}} = \mathbf{D} - \mathbf{D}\mathbf{A}'(\mathbf{A}\mathbf{D}\mathbf{A}' + \Sigma)^{-1}\mathbf{A}\mathbf{D}. \quad (3.7-26)$$

Выражение для матрицы ошибок, также как и выражение для оценки решения, может быть получено и в другой форме, в которой обращается матрица размерностью числа искомого параметров  $n$ . Для этого воспользуемся известным из алгебры матриц равенством

$$(\mathbf{D}^{-1} + \mathbf{A}'\Sigma^{-1}\mathbf{A})^{-1} = \mathbf{D} - \mathbf{D}\mathbf{A}'(\mathbf{A}\mathbf{D}\mathbf{A}' + \Sigma)^{-1}\mathbf{A}\mathbf{D}. \quad (3.7-27)$$

Используя его, выражение для матрицы ошибок (3.7-26) можно записать в виде

$$\tilde{\mathbf{D}} = (\mathbf{A}' \boldsymbol{\Sigma}^{-1} \mathbf{A} + \mathbf{D}^{-1})^{-1}. \quad (3.7-28)$$

Это выражение может оказаться более удобным, если в задаче число измерений существенно превосходит число искомых параметров.

Формулы решения обратной задачи методом статистической регуляризации, полученные нами на основе максимизации апостериорной вероятности, можно также получить как наилучшую (в среднеквадратическом смысле) линейную оценку.

### 3.8. Решающий оператор наилучшей линейной оценки

В предыдущем разделе мы рассмотрели построение оценок решения обратной задачи при наличии вероятностной информации в виде гауссовской плотности вероятности, описывающей как априорное распределение искомого решения, так и случайную ошибку измерений. На практике, однако, в ряде случаев можно сомневаться, что реальная плотность вероятности подчиняется нормальному закону.

#### Решение в случае задания двух первых моментов априорного распределения

Рассмотрим построение решения в случае задания априорного распределений вектора  $\vec{\varphi}$  и распределения случайной ошибки  $\vec{\varepsilon}$  двумя первыми моментами (см. Приложение). Подчеркнем, что при этом мы не накладываем никаких требований на вид априорного распределения, в частности, не требуем его нормальности. Пусть для искомого профиля заданы ковариационная матрица  $\mathbf{D} = E((\vec{\varphi} - E\vec{\varphi})(\vec{\varphi} - E\vec{\varphi})')$  и средний профиль  $E(\vec{\varphi})$  (для простоты без потери общности предположим, что  $E(\vec{\varphi}) = 0$ ). Для случайной ошибки также заданы ковариационная матрица  $\boldsymbol{\Sigma}$  и среднее, равное нулю:

$$\boldsymbol{\Sigma} = E(\vec{\varepsilon}\vec{\varepsilon}'), \quad E(\vec{\varepsilon}) = 0. \quad (3.8-1)$$

Будем искать решение задачи  $\mathbf{A}\vec{\varphi} = \vec{f} + \vec{\varepsilon} = \vec{f}^*$  в виде линейной комбинации измерений  $\vec{f}^*$  (линейная оценка)

$$\tilde{\vec{\varphi}} = \mathbf{R} \vec{f}^*. \quad (3.8-2)$$

Точность решения при некотором выбранном решающем операторе  $\mathbf{R}$  можно охарактеризовать матрицей ошибок решения

$$\tilde{\mathbf{D}} = E\left(\left(\tilde{\vec{\varphi}} - \vec{\varphi}\right)\left(\tilde{\vec{\varphi}} - \vec{\varphi}\right)'\right) \quad (3.8-3)$$

или, если вместо  $\tilde{\vec{\varphi}}$  подставить выражение (3.8-2),

$$\tilde{\mathbf{D}} = E\left(\left(\vec{\varphi} - \mathbf{R}\vec{f}^*\right)\left(\vec{\varphi} - \mathbf{R}\vec{f}^*\right)'\right). \quad (3.8-4)$$

Выберем матрицу  $\mathbf{R}$  так, чтобы обеспечить минимальное значение среднеквадратической ошибки решения  $\tilde{\vec{\varphi}}$ . Учитывая, что среднеквадратическое значение ошибки определяется суммой диагональных элементов матрицы  $\tilde{\mathbf{D}}$ , естественно в качестве  $\mathbf{R}$  выбрать матрицу, минимизирующую функционал

$$L(\mathbf{R}) = Tr E\left(\left(\vec{\varphi} - \mathbf{R}\vec{f}^*\right)\left(\vec{\varphi} - \mathbf{R}\vec{f}^*\right)'\right), \quad (3.8-5)$$

где  $Tr$  – след матрицы, т. е. сумма ее диагональных элементов. Воспользовавшись необходимым условием экстремума, будем искать матрицу  $\mathbf{R}$ , реализующую минимум (3.8-5), решая систему уравнений

$$\frac{d}{d\mathbf{R}} Tr \tilde{\mathbf{D}} = 0. \quad (3.8-6)$$

Подставим в систему (3.8-6) выражение для  $\tilde{\mathbf{D}}$  по соотношению (3.7-24), тогда наша система примет вид

$$\frac{d}{d\mathbf{R}} Tr(\mathbf{D} - \mathbf{R}\mathbf{A}\mathbf{D} - \mathbf{D}\mathbf{A}'\mathbf{R}' + \mathbf{R}\mathbf{A}\mathbf{D}\mathbf{A}'\mathbf{R}' + \mathbf{R}\Sigma\mathbf{R}') = 0. \quad (3.8-7)$$

Используя приведенные в приложении правила дифференцирования скалярной функции матричного аргумента, выражение (3.8-7) можно записать в виде

$$-\mathbf{D}\mathbf{A}' - \mathbf{D}\mathbf{A}' + 2\mathbf{R}\mathbf{A}\mathbf{D}\mathbf{A}' + 2\mathbf{R}\Sigma = 0. \quad (3.8-8)$$

Отсюда легко получить

$$\mathbf{R} = \mathbf{D}\mathbf{A}'(\mathbf{A}\mathbf{D}\mathbf{A}' + \Sigma)^{-1}. \quad (3.8-9)$$

Сравнивая решающие операторы (3.8-9) и (3.7-21b) видим, что решение с минимальной среднеквадратичной ошибкой и решение, полученное в предположении гауссовости априорных плотностей вероятности распределения  $\vec{\varphi}$  и случайной помехи  $\vec{\varepsilon}$ , совпадают.

## Метод минимума априорной информации

Остановимся на еще одном подходе к решению рассматриваемой задачи, которое в литературе называется *методом минимума априорной информации*. При отсутствии информации о степени коррелированности различных компонент искомого вектора  $\vec{\varphi}$  и степени коррелированности ошибок измерений  $\vec{\varepsilon}$  матрицы  $\mathbf{D}$  и  $\mathbf{\Sigma}$  задаются в диагональной форме

$$\mathbf{D} = \sigma_{\varphi}^2 \mathbf{I}_n, \quad \mathbf{\Sigma} = \sigma_{\varepsilon}^2 \mathbf{I}_m, \quad (3.8-10)$$

где  $\sigma_{\varphi}^2$  и  $\sigma_{\varepsilon}^2$  – дисперсии компонент искомого вектора и вектора ошибок измерений соответственно. Подставляя матрицы (3.8-10) в выражение для решающего оператора (3.8-9), а затем в формулу (3.8-2), получим оценку искомого вектора  $\vec{\varphi}$  в виде

$$\vec{\tilde{\varphi}} = \mathbf{A}'(\mathbf{A}\mathbf{A}' + \gamma \mathbf{I}_n)^{-1} \vec{f}^*, \quad (3.8-11)$$

где  $\gamma = \sigma_{\varepsilon}^2 / \sigma_{\varphi}^2$ . Если воспользоваться выражением для решающего оператора в виде (3.7-21a), то

$$\vec{\tilde{\varphi}} = (\mathbf{A}'\mathbf{A} + \gamma \mathbf{I}_n)^{-1} \mathbf{A}' \vec{f}^*. \quad (3.8-12)$$

Заметим, что выражение для решения (3.8-12) совпадает с выражением (3.6-26) для решения по методу регуляризации Тихонова. Различие заключается в том, что если в регуляризации Тихонова параметр регуляризации  $\alpha$  подбирается, что, как мы отмечали, недостаточно обосновано, то метод статистической регуляризации дает нам оптимальное значение параметра. Как видно из соотношения  $\gamma = \sigma_{\varepsilon}^2 / \sigma_{\varphi}^2$ , величина параметра регуляризации определяется именно из требований к искомой функции и оценки погрешности измерений.

Отметим также, что при  $\gamma = 0$  соотношение (3.8-12) переходит в формулу (3.2-12), соответствующую использованию метода наименьших квадратов. Таким образом, оценки решения (3.8-11) и (3.8-12) можно рассматривать как регуляризованное решение методом наименьших квадратов.

### 3.9. Регрессионный подход к решению обратных задач атмосферной оптики

Рассмотренные ранее формулировка обратной задачи и различные методы ее решения предполагали, что нам известны не только результаты измерений  $\vec{f}^* = \vec{f} + \vec{\varepsilon}$ , но и оператор (матрица) прямой задачи  $\mathbf{A}$ . Задание оператора  $\mathbf{A}$  связано с формулировкой физико-математической модели переноса излучения (как правило, на основе той или иной формы уравнения переноса излучения) и привлечения данных о количественных характеристиках взаимодействия излучения со средой (коэффициентов поглощения, рассеяния и т. д.). При этом мы предполагали, что оператор  $\mathbf{A}$  известен точно. На практике задание оператора  $\mathbf{A}$  практически всегда осуществляется с определенными ошибками. В связи с этим более реалистичская постановка обратной задачи должна быть сформулирована следующим образом:

$$(\mathbf{A} + \Delta\mathbf{A})\vec{\varphi} = \vec{f} + \vec{\varepsilon}, \quad (3.9-1)$$

где  $\Delta\mathbf{A}$  – ошибки задания оператора прямой задачи. Анализ влияния этих ошибок на решение обратных задач атмосферной оптики, а также физических причин их появления будет осуществлен позднее. Ошибки могут быть настолько велики, что лучше предполагать, что оператор прямой задачи нам неизвестен, или же мы можем вообще не рассматривать оператор прямой задачи.

#### Постановка задачи в рамках регрессионного подхода

В основе решения обратных задач при отсутствии информации об операторе  $\mathbf{A}$  лежит предположение, что искомые величины  $\vec{\varphi}$  могут быть найдены с помощью некоторого линейного преобразования измеряемых величин<sup>3</sup>

$$\vec{\varphi} = \mathbf{R}\vec{f}, \quad (3.9-2)$$

---

<sup>3</sup> Поскольку при рассмотрении регрессионного подхода мы не выделяем погрешности измерений  $\vec{\varepsilon}$ , будем в этом разделе для краткости использовать обозначение  $\vec{f}$  для измеренных значений, опуская звездочку.

где  $\mathbf{R}$  – некоторая неизвестная и, следовательно, подлежащая определению матрица (решающий оператор). Таким образом, первым этапом построения решения на основе регрессионного подхода является построение решающего оператора  $\mathbf{R}$ . Решение этой задачи возможно при наличии статистического ансамбля пар векторов  $\vec{\varphi}$  и  $\vec{f}$ , причем каждая пара этих векторов в общем ансамбле должна удовлетворять условию пространственно-временной синхронизации. Это означает, что мы должны прежде всего сформировать статистический ансамбль одновременных измерений характеристик излучения  $\vec{f}$  и независимых (например, локальных) измерений искомого параметра физического состояния атмосферы  $\vec{\varphi}$ , согласованных в пространстве. Так, например, в задаче термического зондирования атмосферы из космоса это требует привлечения (кроме измерений уходящего излучения  $\vec{f}$ ) радиозондовых измерений профиля температуры (вектор  $\vec{\varphi}$ ) в «подспутниковых» точках в момент самих спутниковых измерений. Отметим, что при организации таких измерений и при формировании соответствующего статистического ансамбля пар измерений  $\{\vec{f}_s, \vec{\varphi}_s\}$  (здесь  $s$  – номер измерений) нетривиальной проблемой является формулировка требований к пространственно-временной согласованности двух типов измерений.<sup>4</sup>

### Построение решающего оператора

Будем предполагать, что такой статистический ансамбль сформирован. Используя методы математической статистики, на основе имеющегося ансамбля данных можно построить:

а) выборочные средние значения

$$\bar{\vec{\varphi}} = E(\vec{\varphi}) = \frac{1}{l} \sum_{s=1}^l \vec{\varphi}_s, \quad \bar{\vec{f}} = E(\vec{f}) = \frac{1}{l} \sum_{s=1}^l \vec{f}_s; \quad (3.9-3)$$

б) выборочные ковариационные матрицы

---

<sup>4</sup> Возможен другой способ построения такого ансамбля. В этом случае вначале формируется ансамбль  $\varphi_s$ . Затем на основе, например, уравнения переноса излучения проводится *расчет* ансамбля  $\vec{f}_s$ . При этом мы фактически должны знать какое-то приближение к оператору  $\mathbf{A}$ .

$$\begin{aligned}
\mathbf{K}_\varphi &= E(\vec{v}, \vec{v}') = \left\{ k_{\varphi_{i,j}} \right\}_{\substack{i=1,n \\ j=1,n}}; \\
k_{\varphi_{i,j}} &= \frac{1}{l-1} \sum_{s=1}^l v_{si} v_{sj} = \frac{1}{l-1} \sum_{s=1}^l (\varphi_{si} - \bar{\varphi}_i)(\varphi_{sj} - \bar{\varphi}_j); \\
\mathbf{K}_f &= E(\vec{u}, \vec{u}') = \left\{ k_{f_{i,j}} \right\}_{\substack{i=1,m \\ j=1,m}}; \\
k_{\varphi_{i,j}} &= \frac{1}{l-1} \sum_{s=1}^l u_{si} u_{sj} = \frac{1}{l-1} \sum_{s=1}^l (f_{si} - \bar{f}_i)(f_{sj} - \bar{f}_j); \\
\mathbf{K}_{\varphi f} &= E(\vec{v}, \vec{u}') = \left\{ k_{\varphi f_{i,j}} \right\}_{\substack{i=1,n \\ j=1,m}}; \\
k_{\varphi f_{i,j}} &= \frac{1}{l-1} \sum_{s=1}^l v_{si} u_{sj} = \frac{1}{l-1} \sum_{s=1}^l (\varphi_{si} - \bar{\varphi}_i)(u_{sj} - \bar{u}_j).
\end{aligned} \tag{3.9-4}$$

Здесь  $u_{si}$ ,  $v_{si}$ ,  $f_{si}$ ,  $\varphi_{si}$  и т. д. –  $i$ -я компонента векторов  $\vec{u}_s$ ,  $\vec{v}_s$ ,  $\vec{f}_s$ ,  $\vec{\varphi}_s$  и т. д. соответственно;  $\bar{v}_i = \bar{\varphi}_i - \bar{\varphi}$ ,  $\bar{u}_i = \bar{f}_i - \bar{f}$ , индекс  $s$  обозначает номер пары  $\vec{\varphi}_s, \vec{f}_s$ ;  $\mathbf{K}_{\varphi f}$  имеет смысл взаимоквариационной матрицы векторов  $\vec{\varphi}$  и  $\vec{f}$ .

Будем искать решающий оператор  $\mathbf{R}$  из условия минимизации среднеквадратических ошибок определения искомого вектора  $\vec{\varphi}$  (или  $\vec{v}$ ) с помощью соотношения (3.9-2). Запишем выражение для матрицы ошибок вектора  $\vec{\varphi}$

$$\tilde{\mathbf{D}} = E\left( (\vec{\varphi} - \vec{\varphi}) (\vec{\varphi} - \vec{\varphi})' \right), \tag{3.9-5}$$

где вектора  $\vec{\varphi}$  и  $\vec{\varphi}$  – истинное и восстановленное значения искомого вектора. Подставив в это выражение оценку решения (3.9-2), получим

$$\tilde{\mathbf{D}} = E\left( (\vec{\varphi} - \mathbf{R}\vec{f}) (\vec{\varphi} - \mathbf{R}\vec{f})' \right). \tag{3.9-6}$$

Раскрывая скобки, запишем матрицу ошибок в виде

$$\tilde{\mathbf{D}} = E(\mathbf{R}\vec{f}\vec{f}'\mathbf{R}') - E(\mathbf{R}\vec{f}\vec{\varphi}') - E(\vec{\varphi}\vec{f}'\mathbf{R}') + E(\vec{\varphi}\vec{\varphi}'). \tag{3.9-7}$$

Учитывая определения (3.9-4), получим

$$\tilde{\mathbf{D}} = \mathbf{R}\mathbf{K}_f\mathbf{R}' - \mathbf{R}\mathbf{K}_{f\varphi} - \mathbf{K}_{\varphi f}\mathbf{R}' + \mathbf{K}_\varphi. \quad (3.9-8)$$

В качестве количественной характеристики точности решения (точности восстановления  $\vec{\varphi}$ ), как и ранее, примем суммарное среднеквадратическое отклонение

$$\sigma_\varphi^2 = \text{Tr}\tilde{\mathbf{D}}. \quad (3.9-9)$$

Таким образом, задача нахождения решающего оператора  $\mathbf{R}$  (решающей матрицы) сводится к поиску такой матрицы, которая дает минимум функционала (3.9-9). Для решения этой задачи решим, как обычно, уравнение

$$\frac{d}{d\mathbf{R}} \text{Tr}\tilde{\mathbf{D}} = 0. \quad (3.9-10)$$

Подставляя выражение для матрицы ошибок (3.9-8) в уравнение (3.9-10) и учитывая, что  $\mathbf{K}_\varphi$  не зависит от искомого оператора  $\mathbf{R}$ , приходим к уравнению

$$\frac{d}{d\mathbf{R}} (\text{Tr}\mathbf{R}\mathbf{K}_f\mathbf{R}' - \text{Tr}\mathbf{R}\mathbf{K}_{f\varphi} - \text{Tr}\mathbf{K}_{\varphi f}\mathbf{R}') = 0, \quad (3.9-11)$$

где учтено свойство следа матрицы

$$\text{Tr}(\mathbf{A} + \mathbf{B}) = \text{Tr}\mathbf{A} + \text{Tr}\mathbf{B}$$

Воспользовавшись правилами вычисления производных функции матричного аргумента, которые даны в приложении, из уравнения (3.9-11) получим

$$2\mathbf{R}\mathbf{K}_f - 2\mathbf{K}_{\varphi f} = 0.$$

Здесь мы учли, что  $\mathbf{K}_f$  – симметричная матрица и то, что  $\mathbf{K}_{\varphi f} = \mathbf{K}'_{f\varphi}$ , что следует из построения взаимоквариационных матриц. Таким образом, решающий оператор имеет вид

$$\mathbf{R} = \mathbf{K}_{\varphi f}\mathbf{K}_f^{-1}. \quad (3.9-12)$$

Подставляя решающий оператор (3.9-12) в уравнение (3.9-2), получим оценку решения обратной задачи при регрессионном подходе

$$\vec{\varphi} = \mathbf{K}_{\varphi f}\mathbf{K}_f^{-1}\vec{f}. \quad (3.9-13)$$

Отметим ряд особенностей применения регрессионного подхода к решению обратных задач атмосферной оптики.

## Особенности регрессионного подхода

1) Расчет решающего оператора  $\mathbf{R}$  при наличии статистического ансамбля эмпирических (измеренных) данных  $\vec{\varphi}_s, \vec{f}_s$  не требует знания оператора прямой задачи (т. е. не требует формулировки модели переноса излучения для конкретной рассматриваемой обратной задачи) и знания количественных характеристик взаимодействия излучения со средой. Это бывает очень важно, если по каким-либо причинам построение оператора  $\mathbf{A}$  очень сложно или невозможно (например перенос излучения в сложном растительном покрове).

2) Успех применения регрессионного подхода в значительной степени определяется выбранным статистическим ансамблем пар векторов  $\vec{\varphi}_s, \vec{f}_s$  и «принадлежностью» искомых векторов  $\vec{\varphi}$  к этому ансамблю. В случае реализаций  $\vec{\varphi}$ , неадекватных априорному ансамблю, можно ожидать увеличения погрешностей восстановления. В качестве количественной меры этой адекватности можно использовать формулу плотности нормального распределения (3.7-8) для оценки вероятности реализации  $\vec{\varphi}$  с точки зрения ее принадлежности априорному ансамблю

$$P(\vec{\varphi}) = \frac{1}{(2\pi)^{m/2} |\mathbf{K}_\varphi|^{1/2}} \exp\left(-\frac{1}{2}(\vec{\varphi}, \mathbf{K}_\varphi^{-1} \vec{\varphi})\right). \quad (3.9-14)$$

Часто более удобно оценивать «адекватность» атмосферного состояния в пространстве измерений излучения  $\vec{f}$

$$P(\vec{f}) = \frac{1}{(2\pi)^{m/2} |\mathbf{K}_f|^{1/2}} \exp\left(-\frac{1}{2}(\vec{f}, \mathbf{K}_f^{-1} \vec{f})\right). \quad (3.9-15)$$

Поскольку формулы (3.9-14) и (3.9-15) представляют собой плотность нормального распределения, такая проверка адекватности применима лишь в случае гауссовского распределения априорного ансамбля. Тем не менее в большинстве случаев распределение атмосферных параметров близко к гауссовскому, и проверка принадлежности реализации априорному ансамблю описанным способом может использоваться.

На практике повышение точности регрессионного подхода достигается путем использования не одного (например глобального) ансамбля, а целого набора ансамблей векторов  $\vec{\varphi}_s, \vec{f}_s$  для различных

широтных зон, сезонов, типов подстилающей поверхности, условий облачности и т. д. В случае наземных дистанционных измерений целесообразно, естественно, использовать «локальную» статистику, например для различных сезонов. Выбор конкретного ансамбля можно осуществлять на основе расчета вероятности (3.9-15).

3) Важным достоинством регрессионного подхода к решению обратных задач атмосферной оптики является формальная простота его алгоритмической реализации и высокая вычислительная эффективность, что особенно важно при необходимости обрабатывать большие массивы данных измерений в оперативном режиме. Отметим, однако, что обращение матрицы  $\mathbf{K}_f$  в формуле (3.9-13) является нетривиальной задачей и требует специальных приемов, которые, например, достигаются учетом присутствия в измерениях  $\vec{f}_s$  ошибок измерений  $\vec{\varepsilon}_s$ . Регрессионный подход в тех или иных формах нашел достаточно широкое применение при решении обратных задач атмосферной оптики в оперативных системах обработки спутниковых измерений.

4) К недостаткам рассмотренного подхода следует отнести необходимость организации специальных независимых (например подспутниковых) измерений векторов  $\vec{\varphi}_s$ . В ряде случаев это может быть сопряжено с серьезными дополнительными финансовыми затратами. К счастью, например, в задаче термического зондирования атмосферы, а также в ряде других задач эти независимые измерения осуществляются постоянно по определенной программе на сети станций аэрологического зондирования. Это позволяет периодически вычислять (или уточнять) решающий оператор  $\mathbf{R}$  для различных широтных зон и сезонов. Разумеется, такие вычисления можно проводить для районов, хорошо охваченных традиционными измерениями.

### **Связь регрессионного подхода с другими методами решения некорректных обратных задач**

Интересно поставить вопрос о том, как согласуется регрессионный подход с другими методами решения обратных задач. Можно показать, что в случае линейной связи векторов  $\vec{\varphi}$  и  $\vec{f}$

полученная формула (3.9-13) эквивалентна оценке по методу статистической регуляризации. В самом деле, используя

$$\vec{f}^* = \mathbf{A}\vec{\varphi} + \vec{\varepsilon},$$

получим выражение для ковариационной матрицы  $\mathbf{K}_f$

$$\begin{aligned} \mathbf{K}_f &\equiv E(\vec{f}^* \vec{f}^{*'}) = E\left((\mathbf{A}\vec{\varphi} + \vec{\varepsilon})(\mathbf{A}\vec{\varphi} + \vec{\varepsilon})'\right) = \\ &= E(\mathbf{A}\vec{\varphi}\vec{\varphi}'\mathbf{A}') + E(\vec{\varepsilon}\vec{\varepsilon}'). \end{aligned} \quad (3.9-16)$$

При этом преобразовании учитывалось, что вектора  $\vec{\varphi}$  и  $\vec{\varepsilon}$  некоррелированы, т. е.

$$E(\vec{\varphi}\vec{\varepsilon}') = 0, \quad E(\vec{\varepsilon}\vec{\varphi}') = 0.$$

Так как

$$E(\mathbf{A}\vec{\varphi}\vec{\varphi}'\mathbf{A}') = \mathbf{A}E(\vec{\varphi}\vec{\varphi}')\mathbf{A}' = \mathbf{A}\mathbf{K}_\varphi\mathbf{A}' \quad \text{и} \quad E(\vec{\varepsilon}\vec{\varepsilon}') = \mathbf{\Sigma}, \quad (3.9-17)$$

то

$$\mathbf{K}_f = \mathbf{A}\mathbf{K}_\varphi\mathbf{A}' + \mathbf{\Sigma}. \quad (3.9-18)$$

Для матрицы  $\mathbf{K}_{\varphi f}$  можно записать

$$\mathbf{K}_{\varphi f} = E(\vec{\varphi} \vec{f}^{'}) = E\left(\vec{\varphi}(\mathbf{A}\vec{\varphi} + \vec{\varepsilon})'\right) = E(\vec{\varphi}\vec{\varphi}'\mathbf{A}') = \mathbf{K}_\varphi\mathbf{A}'. \quad (3.9-19)$$

Следовательно,

$$\mathbf{K}_{\varphi f} = \mathbf{K}_\varphi\mathbf{A}'. \quad (3.9-20)$$

Подставляя выражения (3.9-18) и (3.9-20) в соотношение (3.9-2), получим выражение для оценки решения

$$\vec{\tilde{\varphi}} = \mathbf{K}_\varphi\mathbf{A}'(\mathbf{A}\mathbf{K}_\varphi\mathbf{A}' + \mathbf{\Sigma})^{-1} \vec{f}^*. \quad (3.9-21)$$

С точностью до обозначений ( $\mathbf{D} = \mathbf{K}_\varphi$ ) выражение (3.9-21) совпадает с оценкой (3.7-18) метода статистической регуляризации, что говорит о взаимной связи различных подходов к решению обратных задач.

Наряду с рассмотренными выше достаточно обоснованными подходами, опирающимися на различные методы регуляризации, для решения обратных задач атмосферной оптики находят применение и более простые эмпирические методы. Рассмотрим один из них.

### 3.10. Итерационные методы решения обратных задач

Рассматриваемый класс методов проиллюстрируем на примере метода М. Шахина, предложенного для решения задачи термического зондирования атмосферы из космоса. Рассмотрим интегральное уравнение Фредгольма I рода

$$f(y) = \int_a^b K(x, y) \varphi(x) dx. \quad (3.10-1)$$

В реальной ситуации число измерений излучения (различных значений переменной  $y$ , для которых мы располагаем измерениями  $f^*(y)$ ) конечно, например, это могут быть измерения на различных частотах

$$f(y_i) = \int_a^b K(x, y_i) \varphi(x) dx, \quad i = 1, \dots, m. \quad (3.10-2)$$

Поясним идею метода, а затем дадим формальное его обоснование и описание. Вообще говоря, итерационные методы широко применяются для решения различных задач, и соответствующая математическая теория развита в достаточной степени (сжимающие отображения). Основная идея метода проста – если у нас имеется приближенный метод решения задачи, позволяющий улучшить имеющееся решение, воспользуемся им неоднократно, каждый раз улучшая предыдущее приближение к решению.

В частности, метод Шахина применим, если ядро интегрального уравнения  $K(x, y_i)$  при каждом  $i$  имеет выраженный локализованный максимум как функция переменной  $x$ . Предполагается также, что расположение этих максимумов  $x_i$  позволяет с помощью некоторой процедуры интерполяции восстановить функцию  $\varphi(x)$  на всем интервале  $[a, b]$  по ее значениям  $\varphi(x_i)$ . В этих предположениях каждое измерение  $f^*(y_i)$  связано со значением искомой функции  $\varphi$  в соответствующей точке  $x_i$  и позволяет оценить величину  $\varphi(x_i)$ . Поскольку такая оценка может быть лишь приближенной, то, получив

таким образом функцию  $\varphi_k(x)$ , можно рассчитать на основе соотношения (3.10-2) на  $k$ -м шаге итерационного процесса значения  $f_k(y_i)$  и, сравнив их с результатами измерений  $f^*(y_i)$ , получить новые оценки  $\varphi_{k+1}(x_i)$ . Этот итерационный процесс может быть продолжен, пока невязка не станет достаточно мала.

Перейдем к формальному рассмотрению метода. По теореме о среднем для уравнений (3.10-2) можно утверждать, что для каждого  $i$  существует такое число  $x_i \in [a, b]$ , что выполняется равенство

$$f(y_i) = \varphi(x_i) \int_a^b K(x, y_i) dx = \varphi(x_i) A_i, \quad i = 1, \dots, m. \quad (3.10-3)$$

Для двух функций  $\varphi_1(x)$  и  $\varphi_2(x)$ , можно записать соотношения

$$\begin{aligned} f_1(y_i) &= \varphi_1(x_i) A_i, \\ f_2(y_i) &= \varphi_2(x_i) A_i \end{aligned} \quad (3.10-4)$$

и составить следующую пропорцию:

$$\frac{f_1(y_i)}{f_2(y_i)} = \frac{\varphi_1(x_i)}{\varphi_2(x_i)}. \quad (3.10-5)$$

На основе этой пропорции М. Шахин и предложил простой алгоритм решения уравнения (3.10-1). Он состоит из следующих этапов. Пусть нам известно начальное приближение к искомому решению  $\varphi_0(x)$ . Вычислим по формуле (3.10-1) соответствующий ему вектор  $f_0(y_i)$ . Затем найдем следующее приближение к решению ( $\varphi_1(x_i)$ ) на основе соотношения, следующего из пропорции (3.10-5)

$$\varphi_1(x_i) = \varphi_0(x_i) \frac{f^*(y_i)}{f_0(y_i)}, \quad (3.10-6)$$

где  $f^*(y_i)$  – измеренные значения функции  $f$ . Далее процесс вычислений можно повторять. Рекуррентное соотношение соответствующего итерационного процесса можно записать в общей форме

$$\varphi_{k+1}(x_i) = \varphi_k(x_i) \frac{f^*(y_i)}{f_k(y_i)}. \quad (3.10-7)$$

В рамках метода Шахина могут использоваться различные критерии останковки итерационного процесса, например близость  $f_k(y_i)$  к измеренным величинам  $f^*(y_i)$ , т. е. малость невязки

$$|f_k(y_i) - f^*(y_i)| \leq \sigma, \quad (3.10-8)$$

где  $\sigma$  – погрешность измерений. Может быть также использован критерий сходимости в себе, т. е. малость изменения решения на шаге итерационного процесса.

Если ядра уравнений (3.10-1) имеют явно выраженные максимумы как функции переменной  $x$ , естественно выбирать  $x_i$  из условия

$$K(x_i, y_i) = \max_x K(x, y_i). \quad (3.10-9)$$

Достаточно тонким вопросом является необходимое для интегрирования в выражении (3.10-1) определение функции  $\varphi$  на всем интервале  $[a, b]$  по оценкам ее значений в конечном количестве точек  $(\varphi(x_1), \dots, \varphi(x_n))$ . Интересно отметить, что количество  $n$  точек и используемые методы интерполяции и экстраполяции фактически ограничивают гладкость получаемого решения и тем самым представляют собой неявно заданную априорную информацию о решении.

Практический успех применения описанного итерационного метода определяется видом ядра  $K(x, y_i)$  в выбранных для измерений значениях  $y_i$  и подбором метода интерполяции и критерия останковки итерационного процесса (3.10-7).

В ряде обратных задач атмосферной оптики с помощью метода Шахина удается получать достаточно хорошие решения. Предложены также и более общие итерационные методы, основанные на тех же идеях. Подобные методы сравнительно просты в реализации, но их применимость существенно зависит от вида ядер уравнения (3.10-2). С нашей точки зрения наиболее существенным их недостатком является неявное введение априорной информации, затрудняющее понимание того, в каком же классе функций ищется решение.

### 3.11. Влияние погрешностей задания оператора прямой задачи на решение обратных задач

Как мы упоминали ранее, на практике мы никогда не знаем оператор  $\mathbf{A}$  абсолютно точно. Причины этого многообразны [7, 10].

1) В общем случае подавляющее большинство обратных задач атмосферной оптики являются нелинейными (см. раздел 1.4). В некоторых случаях нелинейность проявляется в незначительной степени и для ускорения или упрощения расчетов решается близкая к исходной линейная задача. При этом (раздел 1.5) линейный оператор  $\mathbf{A}$  вычисляется для некоторого «среднего» (начального) состояния атмосферы. Это означает, что соотношение

$$\vec{f}^* = \hat{\mathbf{A}}\vec{\varphi} + \vec{\varepsilon}$$

можно трактовать как соотношение, в котором оператор  $\mathbf{A}$  задан с погрешностью  $\Delta\mathbf{A}$ , т. е.

$$\vec{f}^* = (\mathbf{A} + \Delta\mathbf{A})\vec{\varphi} + \vec{\varepsilon}, \quad (3.11-1)$$

причем эта погрешность существенным образом зависит от конкретного значения вектора  $\vec{\varphi}$ .

2) Конкретные вычисления оператора  $\mathbf{A}$  основываются, как мы уже неоднократно отмечали, на той или иной физико-математической модели переноса излучения. При формулировке этой модели неизбежно использование тех или иных приближений. Эти приближения, естественно, должны приводить к ошибкам в задании оператора  $\mathbf{A}$ .

3) Обратные задачи атмосферной оптики в большинстве случаев носят комплексный характер. Это означает, что на измеряемые функционалы излучения влияет не только одна искомая функция, но и ряд других, «второстепенных» функций, влияние которых хоть и мало, но не всегда пренебрежимо мало. Это их влияние также может приводить к ошибкам задания оператора  $\mathbf{A}$ , так как нам приходится использовать в качестве этих «второстепенных» функций их приближения.

4) В оператор прямой задачи  $\mathbf{A}$  входят количественные характеристики взаимодействия излучения со средой – коэффициенты поглощения, рассеяния и т. д. Эти характеристики известны нам из

предварительных исследований, например лабораторных измерений или расчетов, которые также никогда не бывают абсолютно точными. Так, например, расчеты параметров тонкой структуры полос поглощения: положений, интенсивностей, полуширин, формы контура спектральных линий поглощения – всегда используют те или иные приближения. Не всегда достаточно хорошо известны и функциональные зависимости этих параметров взаимодействия излучения со средой от параметров физического состояния этой среды (давления, температуры и т. д.). Таким образом, все это приводит к погрешностям вычисления оператора прямой задачи.

5) При численной реализации алгоритмов неизбежен переход от точных аналитических выражений к их дискретным аналогам (например, замена интеграла квадратурами). Теоретически погрешность такой дискретизации может быть сделана сколь угодно малой, но на практике оказывается обычно ограниченной требованиями к скорости вычислений, объему используемой памяти вычислительного устройства и т. д. Кроме того, не следует забывать о конечной точности представления чисел и выполнения арифметических операций в вычислительных системах, которые в ряде случаев могут приводить к существенным погрешностям в результатах вычислений.

Интересно отметить, что важность всех этих вроде бы очевидных обстоятельств стала понятна только после приобретения реального опыта решения обратных задач атмосферной оптики, в частности, решения задачи термического зондирования атмосферы из космоса. Так, в конце 60-х – начале 70-х годов несмотря на создание специализированной уникальной спутниковой аппаратуры для измерений уходящего теплового излучения в 15 мкм полосе  $\text{CO}_2$  первые опыты по ТЗА продемонстрировали относительно низкую точность восстановленных вертикальных профилей температуры. Как показал проведенный позднее анализ, основной причиной этого были именно ошибки задания прямого оператора задачи  $\mathbf{A}$ .

Ошибки задания оператора  $\mathbf{A}$  можно подразделить на *устраняемые* и *неустраняемые*. К первым из них можно отнести погрешности  $\Delta \mathbf{A}$ , обусловленные современным уровнем наших знаний об особенностях переноса излучения в реальной системе «атмосфера–подстилающая поверхность» и количественных характеристиках взаимодействия излучения со средой. Проводя дополнительные эксперименты и теоретические исследования, мы

можем уменьшать уровень соответствующих ошибок оператора прямой задачи.

Второй тип ошибок оператора  $\mathbf{A}$  обусловлен вариациями состояния атмосферы, и этот тип ошибок всегда будет сопутствовать решению обратных задач атмосферной оптики. Их минимизация возможна при комплексном подходе к решению обратных задач и использованию методов решения нелинейных обратных задач одновременно относительно как «основных», так и «второстепенных» функций. На самом деле этот тип ошибок также вызван нелинейностью задачи. Задача, линейная относительно одного из атмосферных параметров (например, температуры), оказывается нелинейной при учете влияния на излучение состава атмосферы.

Анализ влияния ошибок задания прямого оператора задачи на точность решения обратной задачи начнем с соотношения (3.11-1), которое можно записать в виде

$$\mathbf{A}\vec{\varphi} = \vec{f}^* + \Delta\mathbf{A}\vec{\varphi} + \vec{\varepsilon}. \quad (3.11-2)$$

Таким образом, наличие ошибок в операторе  $\mathbf{A}$  можно трактовать как появление дополнительных ошибок («эффективного шума») при решении задачи для точно заданного оператора  $\mathbf{A}$ . Характеристики этого эффективного шума  $\Delta\mathbf{A}\vec{\varphi}$  определяются свойствами ошибок  $\Delta\mathbf{A}$  и зависят от искомого решения  $\vec{\varphi}$ .

Рассмотрим общий случай нелинейной обратной задачи, которую согласно [18] запишем в виде

$$y = F(\vec{x}, \vec{b}) + \vec{\varepsilon}, \quad (3.11-3)$$

где  $\vec{b}$  – вектор  $n_b$  параметров модели, например параметров калибровки и т. д. Как мы уже отмечали, эти параметры никогда не известны точно. Как показывает соотношение (3.11-2), это является причиной появления случайных и систематических отличий между измеренными и рассчитанными величинами  $y$ . Линеаризуем соотношение (3.11-3) вблизи некоторого среднего состояния атмосферы и неизвестных параметров  $(\vec{x}, \vec{b})$ :

$$y = F(\vec{x}, \vec{b}) + \frac{\partial F}{\partial \vec{x}}(\vec{x} - \vec{x}) + \frac{\partial F}{\partial \vec{b}}(\vec{b} - \vec{b}) + \vec{\varepsilon}. \quad (3.11-4)$$

Вводя обозначения

$$\mathbf{K}_x = \frac{\partial F}{\partial \bar{x}} \text{ и } \mathbf{K}_b = \frac{\partial F}{\partial \bar{b}},$$

получим

$$\delta y = y - F(\bar{x}, \bar{b}) = \mathbf{K}_x \delta \bar{x} + \mathbf{K}_b \delta \bar{b} + \bar{\varepsilon}. \quad (3.11-5)$$

Если мы знаем или можем задать ковариационную матрицу  $\mathbf{D}_b$ , описывающую вариации параметров модели  $\bar{b}$  (или их ошибки), то можно вычислить ковариационную матрицу «эффективного шума»  $\mathbf{K}_b \mathbf{D}_b \mathbf{K}_b'$ .

Различные формулы для оценок решения уравнения (3.11-2) и матриц ошибок были получены В. П. Козловым и А. А. Яковлевым в рамках метода разложения решения по базису собственных векторов информационного оператора  $\mathbf{K}_\varphi \mathbf{A}' \Sigma^{-1} \mathbf{A}$ . Не приводя достаточно громоздкие выкладки и соотношения, отметим, что наличие возмущений оператора  $\Delta \mathbf{A}$  приводит к появлению дополнительной систематической ошибки в решении. Величина этой ошибки зависит от характера (свойств) возмущений  $\Delta \mathbf{A}$ . Анализ показывает, что большие, но осциллирующие возмущения ядра приводят к незначительным дополнительным систематическим погрешностям решения. Наиболее «опасны» плавные возмущения прямого оператора задачи. Расчеты реальных возмущений, вызываемых различными физическими причинами, показывают, что такие плавные возмущения реализуются во многих обратных задачах атмосферной оптики.

Таким образом, погрешности задания оператора  $\mathbf{A}$  могут существенно снижать точность дистанционного зондирования атмосферы. Это говорит о важности экспериментальных и теоретических исследований особенностей переноса радиации в атмосфере и количественных характеристик взаимодействия излучения со средой для создания (формулировки) адекватных физико-математических моделей при решении обратных задач.

Приведем простой пример, иллюстрирующий влияние ошибок оператора задачи на точность получаемых решений. Рассмотрим задачу определения общего содержания  $U$  поглощающего газа при измерениях поглощения солнечного излучения. (см. раздел 1.4, формула (1.4-8)). В простейшем случае монохроматических измерений однородной среды справедлив закон ББЛ

$$I_v(s) = I_v(0) \exp(-k_v U), \quad (3.11-6)$$

где  $I_v(s)$  и  $I_v(0)$  – измеряемая и падающая (первоначальная) интенсивность излучения,  $k_v$  – коэффициент ослабления. После несложных преобразований получим связь между относительными погрешностями

$$\frac{\Delta f_v}{f_v} = \frac{\Delta k_v}{k_v} + \frac{\Delta U}{U}, \quad (3.11-7)$$

где  $f_v = \frac{I_v(0)}{I_v(s)}$ . В случае абсолютно точных измерений выполняется соотношение

$$\left| \frac{\Delta k_v}{k_v} \right| = \left| \frac{\Delta U}{U} \right|. \quad (3.11-8)$$

Таким образом, относительная погрешность задания коэффициента ослабления в  $n\%$  приведет при решении обратных задач к дополнительной погрешности в  $n\%$  при определении содержания поглощающего газа  $U$ .

### **3.12. Оценка пространственной разрешающей способности дистанционных измерений**

До сих пор при рассмотрении обратных задач атмосферной оптики нас интересовали точностные характеристики оценок решения. Другой важной характеристикой дистанционных измерений атмосферных параметров является пространственное, прежде всего вертикальное, разрешение.

Понятие разрешения в физике впервые появилось в спектроскопии при рассмотрении задачи о возможности «разделения» в спектрах поглощения или излучения отдельных спектральных линий, расположенных достаточно близко друг от друга, конкретным спектральным прибором. Классическое определение спектрального разрешения Релея трактует его как наименьшее расстояние (в длинах волн, частотах) между двумя монохроматическими линиями с одинаковыми интенсивностями, при котором они уверенно разрешаются. Слово «разрешаются» означает, что изображение

одионочной спектральной линии можно отличить от изображения двух спектральных линий. Количественный критерий «уверенного» разрешения был сформулирован Релеем, как снижение поглощения (или излучения) между линиями на 20 % по сравнению с максимальными значениями в центрах линий в спектре поглощения (или излучения), зарегистрированного спектральным прибором. Разумеется, в качестве критерия разрешения может быть взята и другая величина, согласованная с погрешностью измерений и формой аппаратной функции конкретного спектрального прибора.

Напомним, что рассматриваемая задача описывается соотношением

$$f^*(\nu') = \int_{-\infty}^{\infty} f(\nu) A(\nu - \nu') d\nu, \quad (3.12-1)$$

где  $f(\nu)$  – монохроматический спектр,  $f^*(\nu')$  – измеренный спектр,  $A(\nu - \nu')$  – аппаратная функция спектрального прибора. Решение интегрального уравнения свертки (3.12-1) было рассмотрено в главе 2.

Рассматриваемые обратные задачи атмосферной оптики можно формально свести к уравнению, близкому к уравнению (3.12-1). Запишем исходное интегральное уравнение с учетом конечного числа измерений излучения

$$f_i = \int A_i(x) \varphi(x) dx, \quad i = 1, \dots, n. \quad (3.12-2)$$

Будем искать оценку решения  $\tilde{\varphi}(x)$  в виде линейной комбинации данных измерений  $f_i$  (про ошибки измерений пока «забудем»):

$$\tilde{\varphi}(x) = \sum_{i=1}^n h_i(x) f_i. \quad (3.12-3)$$

Подставив измерения (3.12-2) в уравнение (3.12-3), получим

$$\tilde{\varphi}(x) = \int H(x, x') \varphi(x') dx', \quad (3.12-4)$$

где

$$H(x, x') = \sum_{i=1}^n h_i(x) A_i(x'). \quad (3.12-5)$$

Из равенства (3.12-4) следует, что оценка  $\tilde{\varphi}(x)$  представляет собой результат усреднения истинного решения  $\varphi(x)$  с ядром  $H(x, x')$ .

Поэтому степень усреднения оценкой  $\tilde{\varphi}(x)$  истинного профиля  $\varphi(x)$  определяется характером функции  $H(x, x')$ , которая в свою очередь зависит от ядра уравнения (3.12-2) (т. е. оператора прямой задачи) и алгоритма решения обратной задачи, задаваемого соотношением (3.12-3).

В идеале, если  $H(x, x') = \delta(x' - x_{oi})$  (здесь  $\delta$  – дельта-функция), мы будем получать оценку решения  $\tilde{\varphi}(x_{oi}) = \varphi(x_{oi})$ . На практике этого добиться нельзя, но можно оптимизировать алгоритм решения обратной задачи (коэффициенты  $h_i(x)$  в соотношении (3.12-3)) для получения минимального усреднения.

Сравнение соотношений (3.12-1) и (3.12-4) показывает, что в обратных задачах атмосферной оптики осредняющее ядро  $H(x, x')$  играет роль «аппаратной функции» и может служить мерой пространственного (вертикального) разрешения косвенных измерений. В литературе имеются примеры использования «спектроскопического» подхода к анализу вертикального разрешения в задачах термического зондирования атмосферы.

Представляет интерес подход к построению решающего алгоритма, предложенный Г. Бэкусом и Ф. Гилбертом (см., например, [17]). В его основу положена минимизация пространственного разрешения решения обратной задачи.

Количественно «толщину» осредняющих ядер  $H(x, x')$  Гильберт и Бэкус предложили характеризовать величиной ширины

$$S(x) = N \int (x' - x)^2 H^2(x, x') dx', \quad (3.12-6)$$

( $N$  – нормирующий множитель) физический смысл которой можно пояснить на простом примере. Пусть осредняющее ядро  $H(x, x')$  представляет собой прямоугольник шириной  $l$ , центрированной в точке  $y$ , и ядро  $H(x, x')$  нормировано на единицу

$$\int H(x, x') dx' = 1. \quad (3.12-7)$$

Тогда ширина такого ядра  $H(x, x')$  в соответствии с определением (3.12-6) будет

$$S(x) = N \int_{x-l/2}^{x+l/2} (x - x')^2 \frac{1}{l^2} dx'. \quad (3.12-8)$$

Вычисление интеграла (3.12-8) дает

$$S(x) = \frac{N \times l}{12}.$$

Приняв нормирующий множитель  $N = 12$ , получим, что ширина осредняющего ядра, имеющего форму прямоугольника с размером сторон  $l$ ,  $S(x) = l$ , что оправдывает выбор формулы (3.12-6) в качестве количественной характеристики пространственного разрешения дистанционных измерений. Минимизируя величину  $S(x)$ , можно для каждого  $x$  найти коэффициенты  $h_i(x)$ , т. е. построить решающий оператор задачи, позволяющий получить наилучшее пространственное разрешение. Для этого, подставив разложение (3.12-5) и значение  $N = 12$  в определение ширины ядра (3.12-6), получим

$$\begin{aligned} S(x) &= 12 \int (x' - x)^2 \left( \sum_{i=1}^n h_i(x) A_i(x') \right)^2 dx' = \\ &= 12 \int (x' - x)^2 \left( \sum_{i=1}^n h_i(x) A_i(x') \right) \left( \sum_{j=1}^n h_j(x) A_j(x') \right) dx' = \\ &= 12 \sum_{i,j=1}^n h_i(x) h_j(x) \int (x' - x)^2 A_i(x') A_j(x') dx' = 12 \bar{h}'(x) \mathbf{K} \bar{h}(x), \end{aligned} \quad (3.12-9)$$

где  $\bar{h}(x)$  – вектор, элементами которого являются коэффициенты  $h_i(x)$ ,  $i = 1, \dots, n$ ,  $\mathbf{K}$  – матрица, элементы которой задаются соотношением

$$k_{i,j} = \int (x' - x)^2 A_i(x') A_j(x') dx'. \quad (3.12-10)$$

Из очевидных соображений должно выполняться условие нормировки (3.12-7). Подставив в него разложение (3.12-5), получим

$$\sum_{i=1}^n h_i(x) \int A_i(x') dx' = \bar{h}'(x) \bar{a} = 1, \quad (3.12-11)$$

где компоненты вектора  $a_i = \int A_i(x') dx'$ . Таким образом, получаем задачу на условный экстремум

$$\bar{h}(x) = \arg \min \bar{h}'(x) \mathbf{K} \bar{h}(x), \quad (3.12-12)$$

при условии  $\bar{h}'(x)\bar{a} = 1$ .

Решая задачу (3.12-12) методом неопределенного множителя Лагранжа, получим для любого  $x$  вектор коэффициентов  $h_i(x)$

$$\bar{h}(x) = \frac{\mathbf{K}^{-1}\bar{a}}{\bar{a}'\mathbf{K}^{-1}\bar{a}} \quad (3.12-13)$$

и, таким образом, построим решающий оператор задачи (3.12-2).

Величину дисперсии определения искомой функции, обусловленную ошибками измерений с ковариационной матрицей ошибок  $\Sigma$ , можно записать в следующей форме:

$$\sigma_\varphi^2(x) = \bar{h}'(x)\Sigma\bar{h}(x). \quad (3.12-14)$$

Если коэффициенты  $h_i(x)$  выбраны на основе только решения задачи минимизации (3.12-12) без учета соотношения (3.12-14), то величина погрешности восстановления  $\sigma_\varphi^2(x)$  может быть слишком большой, т. е. высокое вертикальное разрешение косвенного метода будет сопровождаться низкой точностью восстановления искомого параметра. Добиться одновременно минимизации  $S(x)$  и  $\sigma_\varphi^2(x)$  невозможно, так как уменьшение  $S(x)$  ведет к увеличению  $\sigma_\varphi^2(x)$  и наоборот.

В связи с этим необходимо идти на определенный компромисс. Одним из вариантов такого компромисса является минимизация линейной комбинации  $S(x)$  и  $\sigma_\varphi^2(x)$ , заданной, например, в виде

$$Q(x) = qS(x) + (1-q)\gamma\sigma_\varphi^2(x), \quad (3.12-15)$$

где коэффициент  $\gamma$  обеспечивает согласование размерностей обоих слагаемых. Изменение параметра  $q$  позволяет строить оценки решения  $\tilde{\varphi}(x)$ , либо минимизируя ошибку восстановления ( $q = 0$ ), либо максимизируя вертикальное разрешение ( $q = 1$ ). Выбирая промежуточные значения параметра  $q$ , мы задаем определенный компромисс между пространственным разрешением и точностью восстановления.

Другими вариантами может быть добавление к задаче (3.12-12) еще одного ограничения, например задание допустимой погрешности решения обратной задачи уравнением

$$\sigma_{\varphi}^2(x) = \sigma_0^2(x)$$

или задание допустимой величины пространственного разрешения  $S(x)$ , заведомо большей, чем найденная из (2.1-12), с минимизацией погрешности измерений  $\sigma_{\varphi}^2(x)$ .

### 3.13. О решении нелинейных обратных задач атмосферной оптики

Как мы уже неоднократно упоминали ранее, подавляющее большинство обратных задач атмосферной оптики является нелинейным. Поэтому рассматриваемое нами до сих пор соотношение

$$f^*(y) = \int_a^b A(x, y) \varphi(x) dx + \varepsilon(y) \quad (3.13-1)$$

выполняется лишь с ограниченной точностью в окрестности  $\varphi = 0$ . Кроме того, при численной реализации мы чаще всего решаем конечномерный аналог нашей непрерывной (функциональной) задачи. Соответственно, вместо функционального уравнения (3.13-1) рассмотрим записанное в общей форме соотношение

$$\vec{f}^* = \mathbf{A}(\vec{\varphi}) + \vec{\varepsilon}. \quad (3.13-2)$$

Здесь  $\mathbf{A}$  обозначает нелинейный оператор, сопоставляющий векторам пространства  $\mathbf{R}^n$ , описывающим состояние атмосферы, вектора пространства  $\mathbf{R}^m$ , представляющие собой измеряемые характеристики поля излучения.

Заметим, что если выражение (3.13-1) подразумевает, что функция  $\varphi$  представляет собой профиль только одного из атмосферных параметров (например, температуры или концентрации озона), то вектор  $\vec{\varphi}$  в выражении (3.13-2) может быть составлен из нескольких векторов, каждый из которых представляет высотный профиль одного из атмосферных параметров. На практике обычно всегда таких параметров несколько – полосы поглощения различных газов, как правило, перекрываются, в широких спектральных областях наблюдается влияние аэрозольного рассеяния и поглощения излучения, и т. д. Поэтому простота непосредственного использования формы записи (3.13-2) для формулировки комплексной обратной

задачи относительно одновременного определения профилей нескольких атмосферных параметров важна для постановки реальных задач.

Многие из методов решения обратных задач, рассмотренных нами выше в линейном приближении, могут быть переформулированы для нелинейной постановки задачи. Так, например, соотношение, лежащее в основе МНК (раздел 3.2), может быть записано как

$$\tilde{\varphi} = \arg \min_{\tilde{\varphi} \in \mathbf{R}^n} \left\| \mathbf{A}(\tilde{\varphi}) - \tilde{f}^* \right\|_{\Sigma^{-1}}^2. \quad (3.13-3)$$

Методы численного решения экстремальных задач, подобных (3.13-3), широко известны. В их основе лежит метод наискорейшего спуска, дополненный приемами, позволяющими избегать различных затруднений. В частности, для поиска оценки решения  $\tilde{\varphi}$  может быть использован, например, известный метод Марквардта–Левенберга.

Сходным образом формулируются в нелинейной постановке и другие методы, основанные на минимизации невязки, например, методы регуляризации (3.6)

$$\tilde{\varphi} = \arg \min_{\tilde{\varphi} \in \mathbf{R}^n} \left( \left\| \mathbf{A}(\tilde{\varphi}) - \tilde{f}^* \right\|_{\Sigma^{-1}}^2 + \alpha \left\| \tilde{\varphi} \right\|_{\mathbf{H}}^2 \right) \quad (3.13-4)$$

или метод статистической регуляризации (3.7)

$$\tilde{\varphi} = \arg \min_{\tilde{\varphi} \in \mathbf{R}^n} \left( \left\| \mathbf{A}(\tilde{\varphi}) - \tilde{f}^* \right\|_{\Sigma^{-1}}^2 + \left\| \tilde{\varphi} - \bar{\varphi} \right\|_{\mathbf{D}^{-1}}^2 \right). \quad (3.13-5)$$

Отметим, что для задачи (3.13-5), как и для некоторых других случаев, помимо общих алгоритмов минимизации разработаны специализированные методы решения, например [12, 21, 22].

Решение задачи (3.13-5) может быть получено в результате итерационного процесса, основанного на линеаризации исходного уравнения вблизи очередного приближения к решению и на получении следующего приближения по формулам метода статистической регуляризации для линейного случая

$$\tilde{\varphi}_{k+1} = \arg \min_{\tilde{\varphi} \in \mathbf{R}^n} \left( \left\| \mathbf{A}(\tilde{\varphi}_k) + \mathbf{A}_k (\tilde{\varphi} - \tilde{\varphi}_k) - \tilde{f}^* \right\|_{\Sigma^{-1}}^2 + \left\| \tilde{\varphi} - \bar{\varphi} \right\|_{\mathbf{D}^{-1}}^2 \right), \quad (3.13-6)$$

где  $\mathbf{A}_k$  – матрица частных производных оператора  $\mathbf{A}$  в точке предыдущего приближения к решению  $\tilde{\varphi}_k$ . Отсюда можно получить ряд эквивалентных рекуррентных формул, например:

$$\tilde{\varphi}_{k+1} = \bar{\varphi} + (\mathbf{A}'_k \boldsymbol{\Sigma}^{-1} \mathbf{A}_k + \mathbf{D}^{-1})^{-1} \mathbf{A}'_k \boldsymbol{\Sigma}^{-1} \times \\ \times (\tilde{f}^* - \mathbf{A}(\tilde{\varphi}_k) - \mathbf{A}_k(\bar{\varphi} - \tilde{\varphi}_k)), \quad (3.13-7a)$$

$$\tilde{\varphi}_{k+1} = \bar{\varphi} + \mathbf{DA}'(\mathbf{A}_k \mathbf{DA}' + \boldsymbol{\Sigma})^{-1} (\tilde{f}^* - \mathbf{A}(\tilde{\varphi}_k) - \mathbf{A}_k(\bar{\varphi} - \tilde{\varphi}_k)). \quad (3.13-7б)$$

Эти две формулы алгебраически эквивалентны, но в зависимости от конкретной задачи одна из них оказывается более удобной, чем другая. При космических измерениях прозрачности по Солнцу на касательных трассах (см. раздел 1.4) количество измерений  $m$  часто во много раз превосходит количество определяемых атмосферных параметров  $n$ , и формула (3.13-7a), в которой размерность обращаемой матрицы равна  $n$ , удобнее для численной реализации. В наземных измерениях уходящего излучения атмосферы ситуация обычно обратная ( $n \gg m$ ) и удобнее формула (3.13-7б).

Необходимо отметить, что в нелинейном случае вопрос единственности минимума в выражениях (3.13-3)–(3.13-6) в отличие от линейной задачи остается открытым и требует отдельного исследования в конкретных случаях. Для этого может использоваться моделирование по схеме так называемого замкнутого численного эксперимента.

Численное моделирование эксперимента является одним из обязательных этапов создания систем интерпретации дистанционных измерений и позволяет провести детальное исследование возможностей конкретного эксперимента и отработать детали алгоритмов интерпретации. Оно включает три основных этапа:

- 1) расчет полей излучения на основе заданных значений, определяемых в эксперименте атмосферных параметров;
- 2) моделирование на основе результатов п. 1) измеряемых прибором величин с учетом функций пространственного и спектрального осреднения, шумов и погрешностей калибровки аппаратуры и т. д.;
- 3) решение обратной задачи с использованием результатов п. 2) вместо результатов измерений реального эксперимента.

Сопоставление полученных в результате профилей атмосферных параметров с положенными в основу моделирования «точными» профилями позволяет наглядно оценить погрешности моделируемого метода, его высотное разрешение и другие характеристики.

На этом мы заканчиваем рассмотрение современных методов решения обратных задач атмосферной оптики.

## Глава 4

### Информативность косвенных измерений и оптимизация эксперимента

#### 4.1. Введение

Полученные в предыдущем разделе различные формулы для оценок решения  $\tilde{\varphi}$  и ковариационных матриц  $\tilde{\mathbf{D}}$  ошибок восстановления обратных задач атмосферной оптики демонстрируют их зависимость от оператора прямой задачи  $\mathbf{A}$ . При заданной физико-математической модели переноса излучения структура матрицы  $\mathbf{A}$  будет зависеть от условий измерений излучения, т. е. конкретных характеристик используемого прибора. К этим характеристикам следует, например, отнести число спектральных каналов измерений, их положение и ширину, вид аппаратной функции в каждом канале и т. д. Для касательной геометрии измерений (на горизонт планеты) важными параметрами измерений являются диапазон и число прицельных высот  $h_0$ , для которых осуществляются измерения, угловое разрешение прибора. Естественно поставить вопрос – каковы должны быть условия измерений, чтобы добиться максимальной точности восстановления искомой функции. Решение этой задачи возможно при использовании специального математического аппарата оптимального планирования эксперимента.

Качество дистанционного зондирования зависит от погрешностей измерений. Последняя определяется чувствительностью используемого приемного элемента, временем измерений и т. д. Таким образом, при оптимизации условий измерений нужно учитывать и временные параметры планируемых экспериментов.

Современные подходы к оптимизации условий измерений требуют прежде всего выбора численной характеристики качества измерений, величину которой мы и будем максимизировать, улучшая

измерения. Наиболее универсальной из таких характеристик является количество информации, содержащейся в измеренных величинах. Таким образом, связанной с обсуждаемой проблемой оптимизации является проблема оценки информативности измерений. Вопрос может быть сформулирован следующим образом: какую (в количественном выражении) информацию об искомой функции  $\varphi(x)$  мы можем получить с помощью косвенных измерений, т. е. из измерений  $f(y)$ ?

Рассмотрение задач оптимизации эксперимента имеет важное практическое значение, так как, как правило, приборы для дистанционного зондирования представляют собой уникальные и дорогостоящие устройства. При их изготовлении необходимо четко сформулировать техническое задание, характеризующее основные свойства аппаратуры: число каналов измерений, их положение, спектральные ширины и т. д. Стоимость прибора в существенной степени зависит от его характеристик, и в реальной ситуации необходимо найти компромисс между информативностью прибора и его стоимостью. Поэтому критерием оптимальности может быть выбрана, например, некоторым образом взвешенная сумма информативности измерений и оценки удешевления аппаратуры или ее разработки.

#### **4.2. Количественные меры информативности косвенных измерений**

Предварительный анализ информационных характеристик косвенных измерений был начат нами в главе 2. Напомним, что там рассматривался простой пример задачи распознавания одного из двух возможных состояний объекта исследования – функций  $\varphi_1(x)$  и  $\varphi_2(x)$  – и вводилось понятие информационного расстояния  $\rho$  в пространстве состояний изучаемого объекта. Информационное расстояние характеризует содержащуюся в результатах эксперимента информацию о состояниях  $\varphi_1(x)$  и  $\varphi_2(x)$  с точки зрения их различимости.

Другие информационные характеристики косвенных измерений были введены в рамках статистических подходов к решению обратных задач на основе сравнений априорной и апостериорной

ковариационных матриц искомой функции (вектора). Действительно, если отсутствуют какие-либо измерения вектора  $\vec{f}$ , то наилучшей оценкой вектора  $\vec{\varphi}$  в рассматриваемом случае является средний вектор  $\bar{\vec{\varphi}}$ , а величину его возможного разброса (вариаций) характеризует либо непосредственно априорная ковариационная матрица  $\mathbf{D}$ , либо функционалы от этой матрицы. При наличии измерений вектора  $\vec{f}$  по формулам, приведенным в главе 3, можно получить оценку решения  $\vec{\tilde{\varphi}}$  и ее апостериорную матрицу ошибок  $\tilde{\mathbf{D}}$ . Тогда судить об информативности измерений можно на основе сравнения априорной матрицы  $\mathbf{D}$  и апостериорной  $\tilde{\mathbf{D}}$ .

Вествотер и Стренд [23] предложили использовать в качестве информационных характеристик косвенных измерений следующие величины:

$$R = Tr(\mathbf{D} - \tilde{\mathbf{D}}), \quad (4.2-1)$$

$$F = \frac{Tr(\mathbf{D} - \tilde{\mathbf{D}})}{Tr \mathbf{D}}, \quad (4.2-2)$$

$$U = \left( \frac{Tr \tilde{\mathbf{D}}}{n} \right)^{1/2}. \quad (4.2-3)$$

Здесь величина  $R$  характеризует общее абсолютное уменьшение неопределенности в знании вектора  $\vec{\varphi}$ , величина  $F$  – относительное уменьшение неопределенности, а  $U$  – среднеквадратическую погрешность определения одной компоненты искомого вектора  $\vec{\varphi}$ . Максимальное количество информации для данных условий измерений (эксперимента) может быть получено из измерений с нулевой ошибкой, и ему соответствует выражение

$$R = Tr(\mathbf{D} - \mathbf{DA}'(\mathbf{ADA}')^{-1}\mathbf{AD}). \quad (4.2-4)$$

Использование выражения (4.2-4) на практике затруднительно в силу необходимости обращения плохо обусловленной матрицы  $\mathbf{ADA}'$ . Довольно часто в работах по анализу информативности косвенных измерений используют величину  $F$  или близкие к выражению (4.2-2) определения. Так, в идеальной ситуации, никогда недостижимой,  $F=1$ . Этот случай соответствует  $\tilde{\mathbf{D}}=0$ , т. е. абсолютно точному

решению обратной задачи. Наоборот,  $F = 0$  соответствует отсутствию информации об искомом векторе в измерениях излучения ( $\mathbf{D} = \tilde{\mathbf{D}}$ ).

Для характеристики пространственных информационных возможностей дистанционных измерений часто используются функции

$$\Phi(x) = \left( \frac{\tilde{\sigma}_\varphi^2(x)}{\sigma_\varphi^2(x)} \right)^{1/2} \quad (4.2-5)$$

или

$$\Psi(x) = \left( 1 - \frac{\tilde{\sigma}_\varphi^2(x)}{\sigma_\varphi^2(x)} \right)^{1/2}, \quad (4.2-6)$$

где  $\sigma_\varphi^2(x)$  и  $\tilde{\sigma}_\varphi^2(x)$  – диагональные элементы матриц  $\mathbf{D}$  и  $\tilde{\mathbf{D}}$ , характеризующие дисперсии (априорные и апостериорные) искомого вектора  $\vec{\varphi}$ , т. е. естественные вариации и погрешности определения вектора  $\vec{\varphi}$  соответственно.

В случае идеальных, абсолютно точных косвенных измерений  $\tilde{\sigma}_\varphi^2 = 0$ , что соответствует  $\Phi(x) = 0$  и  $\Psi(x) = 1$ . Отсутствие информации о векторе  $\vec{\varphi}$  в дистанционных измерениях соответствует  $\Phi(x) = 1$  и  $\Psi(x) = 0$  ( $\sigma^2(x) = \tilde{\sigma}^2(x)$ ).

Для сравнения информативности различных экспериментов по дистанционному зондированию можно использовать также сами матрицы ошибок  $\tilde{\mathbf{D}}$ , их определители  $|\tilde{\mathbf{D}}|$ , максимальные дисперсии отдельных компонент  $\max_i \sigma_i^2$  и т. д.

В качестве естественной меры информативности применяется также понятие энтропии как меры упорядоченности исследуемой системы (объекта). Если система может находиться в одном из  $n$  состояний с вероятностью  $P_j$ ,  $j = 1, \dots, n$ , то энтропия такой системы равна

$$I = - \sum_{j=1}^n P_j \log P_j. \quad (4.2-7)$$

Энтропия максимальна, если  $P_1 = P_2 = \dots = P_n = 1/n$ , т. е. все состояния системы равновероятны. С ростом вероятности нахождения

системы в одном из состояний энтропия убывает, а упорядоченность системы возрастает (см. Приложение).

Отметим, что основание логарифма в различных выражениях для энтропии и количества информации (ниже) не имеет существенного значения. Переход от одного основания к другому просто приводит к изменению масштаба величины количества информации. Наиболее часто в качестве оснований логарифма используют числа 2, 10 или  $e$  (основание натурального логарифма). При этом единицы информации называются бит, дит, и нат соответственно. Так, например,  $\Delta I = 1$  бит означает информацию о результате опыта, имеющего два равновероятного исхода. Поэтому мы не указываем в формулах основание логарифма.

Предположим, что проведен эксперимент, в результате которого получен вектор значений  $\vec{f}^*$ . Пусть  $P_0(\vec{\varphi})$  – априорная плотность распределения вектора неизвестных параметров  $\vec{\varphi} = (\varphi_1, \dots, \varphi_n)$ . Обозначим  $P(\vec{\varphi} / \vec{f}^*)$  апостериорную плотность вероятности  $\vec{\varphi}$  (условная плотность вероятности). Используем в качестве меры полученной информации о векторе  $\vec{\varphi}$  в процессе эксперимента приращение (отрицательное, т. е. уменьшение) энтропии  $\Delta I$ . Согласно формуле (4.2-7), получим

$$\begin{aligned} \Delta I(\vec{\varphi}) = & \sum_{j=1}^n \int P(\varphi_j / \vec{f}^*) \log P(\varphi_j / \vec{f}^*) d\varphi_j - \\ & - \sum_{j=1}^n \int P_0(\varphi_j) \log P_0(\varphi_j) d\varphi_j \end{aligned} \quad (4.2-8)$$

$$\text{при условии нормировки } \sum_{j=1}^n \int P_0(\varphi_j) d\varphi_j = 1.$$

Величина  $\Delta I$  называется мерой количества информации (по Шеннону), приобретенной в эксперименте. Заметим, что бессмысленно говорить о количестве информации, извлекаемой из измерений, безотносительно к априори имеющейся информации. Иными словами, корректно использовать выражение «приращение информации, получаемое из измерений» и обозначения вида  $\Delta I$ . Тем не менее мы будем иногда употреблять выражения, подобные «количество информации, извлекаемой из измерений», всегда подразумевая ее приращение, и использовать обозначение  $I$ .

Поскольку информация, извлекаемая из эксперимента, зависит не только от условий эксперимента, но и от случайных реализаций вектора  $\vec{f}^*$ , рассматривают также средние величины информации

$$\bar{\Delta I}(\vec{\varphi}) = \sum_{j=1}^n \int P(\vec{f}^*) \Delta I(\varphi_j) d\varphi_j^* , \quad (4.2-9)$$

$$\text{где } P(\vec{f}^*) = \sum_{j=1}^n \int P(\vec{f}^* / \varphi_j) P_0(\varphi_j) d\varphi_j \quad (4.2-10)$$

– априорная плотность распределения вектора  $\vec{f}^*$ . Можно показать, что наряду с соотношением (4.2-9) справедливо также представление

$$\bar{\Delta I}(\vec{\varphi}) = \sum_{j=1}^n \int P(\vec{f}^* / \varphi_j) P_0(\varphi_j) \log \frac{P(\vec{f}^* / \varphi_j)}{P(\vec{f}^*)} d\vec{f}^* d\varphi_j . \quad (4.2-11)$$

Анализ информативности дистанционных измерений оказывается наиболее эффективным в случае, когда как искомые параметры, так и измеряемые компоненты являются случайными величинами, распределенными по нормальному закону. Выражения для информативности становятся достаточно простыми и удобными для непосредственной интерпретации.

Будем считать, что априорное распределение вектора неизвестных  $\vec{\varphi}$  нормально и его плотность

$$P_0(\vec{\varphi}) = \frac{|\mathbf{K}_\varphi|^{-1/2}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}(\mathbf{K}_\varphi^{-1}\vec{\varphi}, \vec{\varphi})\right), \quad (4.2-12)$$

где  $\mathbf{K}_\varphi$  – априорная ковариационная матрица  $\vec{\varphi}$ . В силу нормальности распределения (4.2-12) и рассматриваемой задачи

$$\mathbf{A}\vec{\varphi} = \vec{f}, \quad \vec{f}^* = \vec{f} + \vec{\varepsilon}, \quad (4.2-13)$$

вектор  $\vec{f}$  является также случайным и распределенным по нормальному закону с ковариационной матрицей

$$\mathbf{K}_f = \mathbf{A}\mathbf{K}_\varphi\mathbf{A}' . \quad (4.2-14)$$

Пусть ошибки измерений некоррелированы, тогда ковариационная матрица вектора  $\vec{\varepsilon}$  определяется равенством

$$\mathbf{K}_\varepsilon = \sigma_\varepsilon^2 \mathbf{I}_m , \quad (4.2-15)$$

где  $\mathbf{I}_m$  – единичная матрица. В таком случае для ковариационной матрицы  $\tilde{\mathbf{K}}_f$  случайного вектора  $\vec{f}^*$  справедливо соотношение

$$\tilde{\mathbf{K}}_f = \mathbf{K}_f + \mathbf{K}_\varepsilon. \quad (4.2-16)$$

Для количества информации (в смысле Шеннона)  $I(\vec{\varphi}, \vec{f}^*)$  о случайном векторе  $\vec{\varphi}$ , содержащемся в  $m$  измеренных величинах вектора  $\vec{f}^* = \vec{f} + \vec{\varepsilon}$ , можно записать следующее выражение :

$$I(\vec{\varphi}, \vec{f}^*) = \frac{1}{2} \left( \log(|\mathbf{K}_\varphi| |\tilde{\mathbf{K}}_f|) - \log|\tilde{\mathbf{K}}_{\varphi f}| \right), \quad (4.2-17)$$

где  $\tilde{\mathbf{K}}_{\varphi f}$  – ковариационная матрица совместного распределения пары векторов  $\vec{\varphi}$  и  $\vec{f}^*$ . Для матрицы  $\tilde{\mathbf{K}}_{\varphi f}$  можно использовать клеточное представление

$$\tilde{\mathbf{K}}_{\varphi f} = \begin{vmatrix} \mathbf{K}_\varphi & \mathbf{K}_\varphi \mathbf{A}' \\ \mathbf{A} \mathbf{K}_\varphi & \tilde{\mathbf{K}}_f \end{vmatrix}. \quad (4.2-18)$$

Очевидно, что при  $\mathbf{A} \equiv 0$ , т. е. в случае отсутствия измерений,

$$|\tilde{\mathbf{K}}_{\varphi f}| = |\mathbf{K}_\varphi| |\tilde{\mathbf{K}}_f|. \quad (4.2-19)$$

Поэтому первое слагаемое в правой части равенства (4.2-17) определяет априорные сведения о векторах  $\vec{\varphi}$  и  $\vec{f}^*$ . Второе слагаемое характеризует апостериорную информацию об этих векторах. Таким образом, выражение (4.2-17), как и исходное определение (4.2-8), позволяет описать количественным образом приращение информации, получаемое в результате измерений (эксперимента).

После ряда преобразований формула (4.2-17) может быть записана в более простом виде

$$I(\varphi, \vec{f}^*) = \frac{1}{2} \log \frac{|\tilde{\mathbf{K}}_f|}{(\sigma_\varepsilon^2)^m}. \quad (4.2-20)$$

Величина  $|\tilde{\mathbf{K}}_f|$  представляет собой обобщенную дисперсию «сигнала»  $\vec{f}^*$  (т. е. измеряемой характеристики). В знаменателе выражения (4.2-20) стоит обобщенная дисперсия вектора ошибок измерений  $\vec{\varepsilon}$ . Поэтому в данном случае количество информации (4.2-20) имеет простую наглядную интерпретацию: количество

информации определяется логарифмом обобщенного отношения сигнал/шум. Величина  $I(\vec{\varphi}, \vec{f}^*)$  по соотношению (4.2-20) характеризует приобретенную информацию в пространстве измеряемых величин. Можно получить выражение для количества информации  $I$  и в пространстве искомых векторов  $\vec{\varphi}$

$$I(\vec{\varphi}, \vec{f}^*) = \frac{1}{2} \log \frac{|\mathbf{K}_\varphi|}{|\tilde{\mathbf{K}}_\varphi|}, \quad (4.2-21)$$

где  $\tilde{\mathbf{K}}_\varphi$  – апостериорная ковариационная матрица ошибок искомого вектора  $\vec{\varphi}$ , характеризующая его остаточную неопределенность. Для апостериорной матрицы  $\tilde{\mathbf{K}}_\varphi = \tilde{\mathbf{D}}$  (введенное в главе 3 обозначение) можно, например, использовать полученное ранее выражение

$$\tilde{\mathbf{K}}_\varphi = \tilde{\mathbf{D}} = \mathbf{K}_\varphi - \mathbf{K}_\varphi \mathbf{A}' \tilde{\mathbf{K}}_f \mathbf{A} \mathbf{K}_\varphi. \quad (4.2-22)$$

Таким образом, количество информации  $I(\vec{\varphi}, \vec{f}^*)$  согласно выражению (4.2-21), можно интерпретировать как логарифм отношения неопределенностей в знании искомого вектора  $\vec{\varphi}$  до и после эксперимента.

Выражение для количества информации по Шеннону в виде (4.2-21) можно преобразовать следующим образом. Запишем апостериорную ковариационную матрицу  $\tilde{\mathbf{D}}$  (или что то же  $\tilde{\mathbf{K}}_\varphi$ ) в виде (формула (3.7-26))

$$\tilde{\mathbf{D}} = (\mathbf{A}' \boldsymbol{\Sigma}^{-1} \mathbf{A} + \mathbf{D}^{-1})^{-1} = (\mathbf{F} + \mathbf{D}^{-1})^{-1}, \quad (4.2-23)$$

где  $\mathbf{F}$  – матрица Фишера. Тогда

$$\Delta I(\vec{\varphi}, \vec{f}^*) = \log \frac{|\mathbf{D}|}{|(\mathbf{F} + \mathbf{D}^{-1})^{-1}|}. \quad (4.2-24)$$

Используя преобразование, следующее из свойств определителей:

$$\frac{|\mathbf{D}|}{|(\mathbf{F} + \mathbf{D}^{-1})^{-1}|} = |\mathbf{D}| |(\mathbf{F} + \mathbf{D}^{-1})| = |\mathbf{D}\mathbf{F} + \mathbf{I}|, \quad (4.2-25)$$

получим

$$\Delta I(\vec{\varphi}, \vec{f}^*) = \log |\mathbf{DF} + \mathbf{I}|. \quad (4.2-26)$$

Матрица  $\mathbf{DF}$ :

$$\mathbf{DF} = \mathbf{DA}'\Sigma^{-1}\mathbf{A} = \mathbf{C}, \quad (4.2-27)$$

называется информационной матрицей Козлова. Напомним, что для любой матрицы  $\mathbf{B}$  размерности  $n$

$$\det \mathbf{B} = |\mathbf{B}| = \prod_{i=1}^n \lambda_i, \quad (4.2-28)$$

где  $\lambda_i$  – собственные значения матрицы  $\mathbf{B}$ . Подставив равенство (4.2-28) в формулу (4.2-26) для количества информации по Шеннону, можно записать еще одно выражение для количества информации

$$\Delta I(\vec{\varphi}, \vec{f}^*) = \log \prod_{i=1}^n (1 + \lambda_i), \quad (4.2-29)$$

где  $\lambda_i$  – собственные значения (числа) информационной матрицы Козлова. Наконец, воспользовавшись известным свойством логарифмов, получим для количества информации

$$\Delta I(\vec{\varphi}, \vec{f}^*) = \sum_{i=1}^n \log(1 + \lambda_i). \quad (4.2-30)$$

Таким образом, количественные характеристики информативности косвенных измерений по Шеннону полностью определяется структурой собственных чисел матрицы  $\mathbf{C}$ .

В.П. Козлов предложил также использовать в качестве информационной характеристики *информационный объем*

$$V^2 = \prod_{i=1}^r \lambda_i, \quad (4.2-31)$$

где  $r = \max\{i : \lambda_i \geq 1\}$ , т. е. при расчете информационного объема  $V$  учитываются лишь те собственные числа  $\lambda_i$ , для которых выполняется неравенство  $\lambda_i \geq 1$  (предполагается, что собственные числа упорядочены по убыванию). В определении (4.2-31) используются только собственные числа  $\lambda_i \geq 1$ , поскольку с собственными числами, меньшими единицы, практически не связано при однократной интерпретации (однократных измерениях) никакой полезной информации. Каждое собственное число  $\lambda_i$  информационной матрицы

С может быть представлено как отношение априорной дисперсии некоторого параметра к дисперсии ошибки определения этого параметра в рассматриваемом эксперименте. Таким образом, понятно, что при  $\lambda_i < 1$  эксперимент не позволяет уточнить априорную неопределенность соответствующего параметра. Сравнивая выражения для количества информации по Шеннону (соотношения (4.2-29) и (4.2-30)) с объемом информации по Козлову (соотношение (4.2-31)), видим, что в расчет  $I(\bar{\varphi}, \bar{f}^*)$  входят все собственные числа  $\lambda_i, i = 1, \dots, n$ , матрицы **C** и, таким образом, вклад в полное количество информации вносят и малоэффективные с точки зрения возможности их уточнения в данном эксперименте компоненты  $\lambda_i < 1 (i = r + 1, \dots, n)$ . Разумеется, подход Шеннона также верен в том смысле, что эти компоненты несут определенную информацию о решении и накопление и использование этой информации возможно при *повторных* измерениях. Подход Козлова эту информацию не учитывает, поскольку она фактически не является полезной при дистанционных оптических измерениях атмосферных параметров, как правило, не воспроизводимых в точности.

Можно получить более удобное для использования выражение для информационного объема  $V$ . Для этого заметим сначала, что матрица **C**, определенная соотношением (4.2-27), действует в пространстве искомым векторов  $\bar{\varphi}$ . Рассмотрим теперь матрицу **C**<sup>\*</sup>, действующую в пространстве векторов  $\bar{f}$ , т. е. в пространстве измерений

$$\mathbf{C}^* = \mathbf{A} \mathbf{D} \mathbf{A}' \boldsymbol{\Sigma}^{-1}. \quad (4.2-32)$$

Можно показать, что ненулевые собственные числа матриц **C** и **C**<sup>\*</sup> совпадают, а соответствующие им собственные вектора связаны между собой соотношением

$$\bar{\psi}^*_i = \mathbf{A} \bar{\psi}_i, \quad i = 1, \dots, \min(m, n). \quad (4.2-33)$$

Действительно, пусть  $\bar{\psi}_i$  – собственные вектора матрицы **C**. Тогда по определению

$$\mathbf{D} \mathbf{A}' \boldsymbol{\Sigma}^{-1} \mathbf{A} \bar{\psi}_i = \lambda_i \bar{\psi}_i, \quad i = 1, \dots, n. \quad (4.2-34)$$

Умножим выражение (4.2-34) слева на матрицу **A**, тогда

$$\mathbf{A} \mathbf{D} \mathbf{A}' \boldsymbol{\Sigma}^{-1} \mathbf{A} \bar{\psi}_i = \lambda_i \mathbf{A} \bar{\psi}_i. \quad (4.2-35)$$

Если обозначить  $\vec{\psi}^*_i = \mathbf{A}\vec{\psi}_i$ , то получим

$$\mathbf{ADA}'\Sigma^{-1}\vec{\psi}^*_i = \lambda_i\vec{\psi}^*_i, \quad (4.2-36)$$

что и доказывает сделанное утверждение.

Рассмотрим случай равноточных измерений вектора  $\vec{f}^*$ , т. е.  $\Sigma = \sigma^2\mathbf{I}$ . При этом матрица  $\mathbf{C}^*$  оказывается симметричной:

$$\mathbf{C}^* = \sigma^{-2}\mathbf{ADA}'. \quad (4.2-37)$$

Выражение для матрицы  $\mathbf{C}^*$  можно также записать в виде

$$\mathbf{C}^* = \frac{\mathbf{K}_f}{\sigma^2}, \quad (4.2-38)$$

где  $\mathbf{K}_f = \mathbf{ADA}'$  – ковариационная матрица измеряемого вектора (сигнала). Таким образом, анализ информативности косвенных измерений может быть достаточно легко осуществлен на основе ковариационных матриц  $\mathbf{K}_f$ , которые можно построить на основе экспериментальных измерений или рассчитать даже для нелинейной прямой задачи.

Этот подход к оценке информативности измерений может оказаться очень удобен для исследования нелинейных задач. Весь наш предыдущий анализ в этом разделе проводился только для линейного случая, но в нелинейном случае мы можем, построив тем или иным способом ковариационную матрицу  $\mathbf{K}_f$ , выбирать схему измерения, максимизируя количество информации, содержащейся в измерениях.

Особенную ценность может также представлять анализ эмпирических ковариационных матриц, полученных на основе обработки больших массивов измерений излучения. Предположим, что минимальное собственное число матрицы  $\mathbf{C}^*$  не меньше единицы:

$$\lambda_{\min}(\mathbf{C}^*) \geq 1. \quad (4.2-39)$$

Эксперимент, удовлетворяющий условию (4.2-39), естественно назвать *информационно-обусловленным*, так как он позволяет эффективно оценивать  $m$  параметров искомого решения по  $m$  наблюдениям. Для информационно-обусловленного эксперимента выражение для информационного объема принимает особенно простой вид

$$V^2 = \det \mathbf{C}^* .$$

Наконец, отметим, что информационный объем Козлова имеет простой физический смысл – он характеризует число состояний изучаемого объекта (различных профилей температуры, влажности и т. д.) различимых (выделяемых) с помощью конкретного метода измерений. Например, анализ ковариационных матриц температуры и излучения показывает, что прямые радиозондовые измерения вертикального профиля температуры в диапазоне давлений 1–1000 мб со случайной ошибкой измерений 0,5 К позволяют различить  $2,2 \cdot 10^{22}$  (23 независимых компоненты) состояний термического режима земной атмосферы. С другой стороны, косвенные измерения вертикальных профилей температуры с помощью использования спектральных измерений уходящего микроволнового излучения (1500 измерений в спектральной области 1–200 ГГц) позволяют выделить  $0,81 \cdot 10^{15}$  (14 независимых компонент) состояний при ошибке измерений радиояркостной температуры излучения в 0,5 К.

#### **4.3. Оптимизация условий измерений**

Важное практическое значение, особенно при спутниковом зондировании атмосферы и подстилающей поверхности, имеет выбор оптимальных характеристик измерительного устройства (прибора). Это обусловлено следующими важнейшими обстоятельствами:

1) Для эффективного использования результатов дистанционного зондирования атмосферы и подстилающей поверхности при решении различных научных и прикладных задач необходимо получать информацию о параметрах физического состояния среды с гарантированными точностью, вертикальным и горизонтальным разрешением, требуемым пространственным охватом и периодичностью измерений. Разнообразные требования к упомянутым выше характеристикам сформулированы в настоящее время специалистами в области метеорологии, климатологии, океанологии и т. д.

2) Создание специализированной аппаратуры и проведение измерений (особенно из космоса) требует значительных финансовых затрат.

Ясно, что эти обстоятельства взаимно противоречивы, и в каждом конкретном случае необходимо найти компромиссное решение. Особенно важное значение имеет этот компромисс при создании оперативных спутниковых систем наблюдений, т. е. систем длительного функционирования, поставляющих информацию для анализа и прогноза погоды, контроля состояния климатической системы и параметров окружающей среды. При оптимизации условий измерений, например, уходящего излучения системы атмосфера–подстилающая поверхность, требуется в общем случае оптимизация следующих характеристик:

а) параметров орбиты спутника (высота, наклонение и т. д.), которые определяют в значительной степени пространственный охват наблюдений, их периодичность и время измерений;

б) характера углового сканирования спутникового прибора и его пространственного (углового) разрешения, которые определяют пространственные области наблюдений, периодичность, горизонтальное разрешение;

в) времени (продолжительности) единичного измерения и полного пространственного скана;

г) спектральных областей измерений, числа спектральных каналов, их спектрального разрешения. (То же можно сказать и об угловых характеристиках прибора при касательном зондировании атмосферы.)

При этом важно напомнить, что случайные погрешности измерений, в существенной степени влияющие на точностные характеристики искомых параметров, зависят от приведенной чувствительности приемника, качества передачи телеметрической информации, а также времени измерений, углового и спектрального разрешения измерений. Например, для тепловых приемников среднеквадратический уровень случайных шумов измерений  $\sigma_\varepsilon$  зависит от времени измерений  $t$  и спектрального разрешения  $\Delta\nu$  следующим образом:

$$\sigma_\varepsilon \sim \frac{1}{\Delta\nu\sqrt{t}}. \quad (4.3-1)$$

Для других типов приемников эта зависимость может иметь несколько другой вид.

Проблема комплексной оптимизации условий измерений, т. е. оптимизации одновременно всей совокупности указанных выше

характеристик в силу своей сложности и громоздкости до настоящего времени не решалась. Как правило, к настоящему времени рассматривались частные проблемы оптимизации:

- – выбор оптимальной высоты спутника и характеристик его орбиты;
- – выбор оптимального горизонтального разрешения и характера пространственного сканирования;
- – определение оптимальных спектральных характеристик приборов.

Следует отметить, что результаты оптимизации, т. е. оптимальные условия измерений, в существенной степени зависят от цели измерений. Так, например, в задаче температурного зондирования это – погрешности измерений, область измерений по высоте, вертикальное разрешение, определение температуры или геопотенциала и т. д.

При оптимизации условий измерений необходимо использовать тот или иной количественный критерий. Так, при использовании вероятностной априорной информации наряду с оценкой искомого решения в главе 3 были получены формулы для остаточной ковариационной матрицы  $\tilde{\mathbf{D}}$  (матрицы ошибок), исчерпывающе описывающей точностные характеристики решения. Из этих же формул наглядно видно, что эти оценки зависят от оператора прямой задачи  $\mathbf{A}$ . При заданной физико-математической модели переноса излучения этот оператор зависит от условий измерений, например, числа, положения и ширины спектральных интервалов измерений излучения, времени (продолжительности) измерения. Это открывает возможности для постановки задач оптимального планирования эксперимента, т. е. выбора таких условий проведения измерений и, таким образом, задания определенного вида оператора  $\mathbf{A}$ , которые позволяют получить оценки искомого решения с желательными или определенными экстремальными свойствами.

Наиболее очевидным из экстремальных свойств искомой оценки решения является требование минимальности средней (суммарной) дисперсии получаемой оценки. Если обозначить  $\vec{\beta}$  вектор независимых параметров, определяющих условия измерений (его принято называть *планом эксперимента* или просто *планом*), то задачу

на построение оптимального плана можно сформулировать следующим образом:

$$\tilde{\beta} = \arg \min_{\beta} Tr(\tilde{\mathbf{D}}(\tilde{\beta})). \quad (4.3-2)$$

План эксперимента  $\tilde{\beta}$ , отвечающий условию (4.3-2), носит название А-оптимального плана, а запись  $\tilde{\mathbf{D}}(\tilde{\beta})$  отражает тот факт, что матрица ошибок  $\tilde{\mathbf{D}}$  зависит от выбранного плана эксперимента.

Другим достаточно распространенным критерием оптимальности плана является требование минимальности определителя матрицы  $\tilde{\mathbf{D}}$  (D-оптимальный план):

$$\tilde{\beta} = \arg \min_{\beta} |\tilde{\mathbf{D}}(\tilde{\beta})|. \quad (4.3-3)$$

В этом случае требование экстремальности означает, что полученные оценки искомых параметров обладают эллипсоидом рассеяния с минимальным объемом.

При построении оптимальных планов эксперимента могут использоваться и другие функционалы от остаточной ковариационной матрицы. В общем случае план  $\beta$  называют  $\Psi$ -оптимальным, если выполняются следующее экстремальное свойство:

$$\tilde{\beta} = \arg \min_{\beta} \Psi(\tilde{\mathbf{D}}(\tilde{\beta})), \quad (4.3-4)$$

где  $\Psi$  – некоторый функционал от матрицы  $\tilde{\mathbf{D}}$ .

Выбор критериев оптимальности при построении планов может отличаться не только видом функционала  $\Psi$ , но и использованием вместо матрицы  $\tilde{\mathbf{D}}$  некоторой другой матрицы, например, информационной матрицы Козлова. Как мы показали в предыдущем разделе, матрица Козлова  $\mathbf{C}$  тесно связана с характеристиками информативности эксперимента и количеством информации по Шеннону (как, собственно, и матрица  $\tilde{\mathbf{D}}$ ). Использование матрицы  $\mathbf{C}$  вместо матрицы  $\tilde{\mathbf{D}}$  позволяет строить оптимальные планы эксперимента, обладающие экстремальными свойствами относительно информационных характеристик эксперимента.

Рассмотрим далее ряд методов оптимизации условий измерений на примере оптимизации спектральных характеристик приборов.

При физико-математической формулировке обратных задач атмосферной оптики можно получить линейное соотношение, связывающее вариации искомого профиля и вариации монохроматической интенсивности излучения в интегральной форме:

$$f(y) = \int_a^b A(x, y) \varphi(x) dx, \quad (4.3-5)$$

где под  $y$  будем понимать частоту (или длину волны) излучения. Однако любой измерительный прибор измеряет не интенсивность излучения, а поток излучения (т. е. энергию) в конечных спектральных интервалах  $[v_{1i}, v_{2i}]$

$$f_i = \int_{v_{1i}}^{v_{2i}} \rho_i(y) f(y) dy, \quad (4.3-6)$$

где  $\rho_i(y)$  – аппаратная функция  $i$ -го канала измерений, или, с учетом уравнения (4.3-5),

$$f_i = \int_{v_{1i}}^{v_{2i}} \rho_i(y) \int_a^b A(x, y) \varphi(x) dx dy. \quad (4.3-7)$$

(Аналогичные рассуждения можно провести и для конечного времени измерений и конечной угловой апертуры прибора.) Переходя в соотношении (4.3-7) от интеграла к конечной сумме, например, с помощью той или иной квадратурной формулы, получим

$$f_i = \sum_{j=1}^m \tilde{a}_i(x_j) \omega_j \varphi_j, \quad i = 1, \dots, n, \quad (4.3-8)$$

где

$$\tilde{a}_i(x_j) = \int_{v_{1i}}^{v_{2i}} \rho_i(y) A(x_j, y) dy, \quad (4.3-9)$$

а  $\omega_j$  – квадратурные коэффициенты. Обозначив  $a_{i,j} = \tilde{a}_i(x_j) \omega_j$ , приходим к уже использованной нами ранее матричной форме записи задачи (разумеется, дискретизация может быть произведена и другими способами)

$$\vec{f} = \mathbf{A} \vec{\varphi}. \quad (4.3-10)$$

По существу здесь отличие, например, от уравнения (4.3-6) состоит в том, что вместо набора отсчетных точек  $\{y_i\}_{i=1,\dots,n}$  в нашем рассмотрении оказывается набор весовых функций  $\{\rho_i(y)\}_{i=1,\dots,n}$ , характеризующих спектральную чувствительность прибора в различных каналах измерений (в литературе  $\rho_i(y)$  часто также называют аппаратными функциями). При использовании специальных приемов нормировки на величины  $\rho_i(y)$  накладываются разные условия, например:

$$0 \leq \rho_i(y) \leq 1. \quad (4.3-11)$$

Отметим, что это условие может не выполняться для отдельных типов спектральных приборов, для которых  $\rho_i(y)$  может принимать отрицательные значения. Иногда на весовые функции  $\rho_i(y)$  накладываются и более жесткие условия.

Таким образом, общая постановка задачи оптимального планирования спектральных характеристик прибора может быть сформулирована следующим образом.

Требуется выбрать число  $n$  и систему весовых функций  $\{\rho_i(y)\}_{i=1,\dots,n}$  в заданной спектральной области  $[y_1, y_2]$ , подчиненных условиям типа (4.3-11), чтобы максимизировать некий информационный функционал  $\Phi$  (или минимизировать некий функционал от матрицы  $\tilde{\mathbf{D}}$ ).

В методе оптимизации Козлова в качестве меры информативности эксперимента используется величина информационного объема  $V$ , определяемого с помощью собственных чисел  $\lambda_i$  информационной матрицы  $\mathbf{C} = \mathbf{DF}$  ( $\mathbf{F}$  – матрица Фишера) или (что то же самое) ковариационной матрицы «сигнала» (см. предыдущий раздел). Напомним, что при расчете информационного объема используются  $\lambda_i \geq 1$ , а выражение для информационного объема  $V$  при равнооточных измерениях можно записать в виде

$$V^2 = \det \mathbf{C}^* = \det (\sigma_\varepsilon^{-2} \mathbf{A} \mathbf{D} \mathbf{A}'). \quad (4.3-12)$$

### **Задача построения оптимального прибора**

Рассмотрим теперь одну частную задачу планирования эксперимента. Мы предполагаем, что спектральные каналы измерений,

не обязательно односвязные, могут быть «составлены» из элементарных спектральных интервалов. Нашей задачей является построение оптимальной спектральной схемы измерений, т. е. системы спектральных каналов прибора.

Положим, что способ дискретизации непрерывной задачи заключается в том, что неизвестные функции  $\varphi(x)$  представляются в виде линейной комбинации базисных функций  $\mu_k(x)$ ,  $k = 1, \dots, m$ , т. е.

$$\varphi(x) = \sum_{k=1}^m \xi_k \mu_k(x). \quad (4.3-13)$$

Поставим задачу оптимизации эксперимента относительно параметров  $\xi_k$ . Иначе говоря, требуется найти систему весовых (аппаратных) функций  $\{\rho_i(y)\}_{i=1, \dots, n}$ , максимизирующую информационный объем  $V$  или, что то же самое,  $\det \mathbf{C}^*$ .

Выражение (4.3-13) индуцирует в пространстве «сигнала» (измеряемых потоков излучения) разложение

$$f(y) = \sum_{k=1}^m \xi_k b_k(y), \quad (4.3-14)$$

где

$$b_k(y) = \int_a^b A(x, y) \mu_k(x) dx. \quad (4.3-15)$$

Тогда в представлении (4.3-13) матрица  $\mathbf{A}$  в уравнении (4.3-10) есть матрица с элементами вида

$$a_{i,k} = \int_{y_{1i}}^{y_{2i}} \rho_i(y) b_k(y) dy. \quad (4.3-16)$$

Введем также матрицу априорных ковариаций искомого вектора  $\vec{\xi}$ :

$$E(\vec{\xi}, \vec{\xi}') = \mathbf{D}. \quad (4.3-17)$$

По свойству определителей имеем

$$V^2 = \det \mathbf{C}^* = \sigma^{-2m} \det \mathbf{D} (\det \mathbf{A})^2 \quad (4.3-18)$$

(разумеется, раз мы говорим об определителе матрицы  $\mathbf{A}$ , она предполагается квадратной), и условие оптимальности по отношению к информационному объему  $V$  принимает вид

$$|\det \mathbf{A}| = \max. \quad (4.3-19)$$

Задача (4.3-19) есть в точности задача D-оптимального планирования регрессионного эксперимента в  $m$ -мерном пространстве распределений. Решение этой задачи, т. е. система оптимальных весовых функций  $\{\rho_i(y)\}_{i=1,\dots,n}$ , удовлетворяющих ограничениям типа (4.3-11) и условию оптимальности (4.3-19) такова, что  $\rho_i(y)$  принимают только два значения – ноль и единица, т. е. каждый спектральный канал измерений является объединением некоторого числа спектральных интервалов внутри основной области измерений  $[y_1, y_2]$  (можно также вводить дополнительные ограничения на число этих спектральных интервалов для упрощения оптимального плана эксперимента).

Существенно, что для зависимости погрешности измерений от суммарной ширины спектрального интервала  $\Delta\nu$  и времени  $t$  вида (4.3-1) В. П. Козловым предложен эффективный итерационный алгоритм, сходящийся за небольшое число шагов к решению задачи (4.3-19) и позволяющий фактически находить оптимальные системы весовых функций  $\{\rho_i(y)\}$ . Тем самым может считаться решенной и поставленная задача оптимизации условий эксперимента (в данном случае, спектральных). Для других видов зависимости погрешности измерений от суммарной ширины спектрального интервала подобные алгоритмы строгого поиска оптимума в настоящее время неизвестны, но существуют эмпирические алгоритмы, которые позволяют находить квазиоптимальное, т. е. близкое к оптимальному решение.

### **Оптимальный выбор положения спектральных каналов**

Рассмотрим другую задачу оптимизации. Многие современные спектральные приборы выполняют измерения в большом (до десятков тысяч) числе спектральных каналов. Использование всех измерений в достаточно сложных алгоритмах интерпретации требует чрезмерно больших вычислительных ресурсов, в первую очередь большого времени расчетов. С другой стороны, спутниковая аппаратура выполняет до нескольких тысяч сеансов измерений за сутки, а

прикладные задачи требуют быстрой обработки данных и получения результатов интерпретации иногда не позднее, чем через несколько часов после измерений. В этой ситуации возникает задача выбора небольшого числа измерений для использования в алгоритмах интерпретации.

Таким образом, другой подход к оптимизации измерений состоит в выборе отсчетных точек  $y_i$  при их общем фиксированном числе и фиксированном спектральном разрешении. Следует отметить, что с ростом числа отсчетных точек  $m$  (точек измерений функции  $f$ ) точность решения задачи (4.3-10) непрерывно растет. (Даже если мы будем дополнительно измерять в тех же спектральных интервалах, точность будет возрастать в связи с уменьшением уровня случайных ошибок измерений, как это происходит при многократных измерениях одной и той же величины). Однако в то же время с увеличением  $m$  увеличивается линейная зависимость между уравнениями системы (4.3-10) и уменьшается вклад каждой последующей отсчетной точки  $y_i$  в информативность эксперимента. Эти соображения, принимая во внимание «стоимость» дополнительных измерений, также должны ограничивать общее число отсчетных точек.

Для решения задач оптимизации в такой постановке используются, как правило, алгоритмы, основанные на переборе различных планов измерений.

### **Алгебраический подход к построению спектральной схемы измерений**

Первый, по существу, алгебраический подход к рассмотрению проблем оптимизации эксперимента был развит Туоми. Он основан на том, что система (4.3-10) при большом числе отсчетных точек  $y_i$  может быть «избыточной» вследствие линейной зависимости между уравнениями. Это означает, что существует линейная комбинация наблюдений

$$l = \sum_{i=1}^m l_i f_i = (\bar{l}, \bar{f}^*) \text{ при } \sum_{i=1}^m l_i^2 = 1, \quad (4.3-20)$$

которая с точностью до ошибок измерений равна нулю, т. е.

$$E(l^2) \leq \sigma^2, \quad (4.3-21)$$

где  $\sigma^2$  – дисперсия ошибок. Иными словами, линейная зависимость измерений означает, что некоторые измеренные величины могут быть получены как линейная комбинация остальных измерений. Поэтому часть измерений можно исключить, уменьшив тем самым число спектральных каналов прибора и не потеряв информации, содержащейся в исключаемых измерениях.

На эту идею и опирается подход Туоми к выбору отсчетных точек. Он предложил критерий выбора того наблюдения из входящих в линейную комбинацию, которое точнее всего может быть «предсказано» по остальным. Соответствующее уравнение просто вычеркивается из системы (4.3-10). Затем анализ повторяется с новой системой, содержащей уже  $m-1$  уравнений. Процесс прекращается, когда условие (4.3-21) окажется невыполненным.

Описанная процедура довольно трудоемка, а оптимальность результата неочевидна. Определенный выигрыш состоит в том, что предложенный подход позволяет получить «хорошую» (достаточно хорошо обусловленную) систему (4.3-10).

Основной недостаток такого подхода, как впервые указал В.П. Козлов, состоит в том, что он игнорирует важные особенности реальных спектрометрических измерений. Козлов, в частности, указал, что «избыточное» наблюдение не следует отбрасывать, а его необходимо «объединять» с другими «подобными» измерениями. Это позволяет получить значительный выигрыш в информативности измерений, в частности, в связи с тем, что случайные погрешности измерений могут (например для приемников теплового излучения) уменьшаться с увеличением ширины спектральных интервалов измерений. В результате также уменьшаются погрешности восстановления атмосферных параметров.

### **Алгоритмы выбора положения спектральных каналов**

При наличии статистической информации об искомом решении можно рассчитать матрицу ошибок косвенного метода  $\tilde{\mathbf{D}}$  и использовать при выборе числа измерений и их положения (например в спектре излучения) на основе анализа матрицы  $\tilde{\mathbf{D}}$ , характеризующей погрешности решения обратной задачи. В качестве критерия оптимальности можно взять функцию

$$\Psi = Tr \tilde{\mathbf{D}}.$$

Тогда условие оптимальности имеет вид

$$\Psi = Tr \tilde{\mathbf{D}} = \min, \quad (4.3-22)$$

где матрица  $\tilde{\mathbf{D}} = (\mathbf{A}'\Sigma^{-1}\mathbf{A} + \mathbf{D}^{-1})^{-1}$ . Условие (4.3-22) определяет выбор отсчетных точек  $\{y_i\}_{i=1,\dots,m}$  из некоторого интервала  $[c,d]$ , если зафиксировано число этих точек. Если при решении задачи оптимизации это число не зафиксировано, то необходимо проводить серию расчетов при разных  $m$ , а оптимальное число измерений  $m$  определять в результате анализа последовательности матриц  $\tilde{\mathbf{D}}$ . В связи с тем, что с ростом числа измерений функция  $\Psi$  монотонно уменьшается, необходимо найти некоторый компромисс между точностью и «стоимостью» эксперимента. Точные и однозначные условия этого компромисса сформулировать непросто.

Если число  $m$  необходимых измерений зафиксировано, то алгоритм поиска оптимального положения точек измерений  $y_i$  строится как перебор всех возможных сочетаний по  $m$  из полного набора  $M$  возможных измерений, распределенных таким образом во всей области измерений  $[c,d]$ . На каждом шаге перебора вычисляется функция  $\Psi$ , затем находится ее минимальное значение и, следовательно, оптимальное положение точек измерений. Общий объем вычислений пропорционален величине числа сочетаний из  $M$  по  $m$   $C_M^m$ :

$$C_M^m = \frac{M!}{m!(M-m)!}.$$

В табл. 4.3-1 показано число вариантов расчетов при различных  $m$  (числа измерений) и  $M$  (исходного числа возможных измерений).

Таблица 4.3-1

Число вариантов расчетов

$m$	$M = 20$	$M = 100$
1	20	100
2	190	4 950
3	1 140	16 170
4	4 845	3 921 225
5	15 504	75 287 520
...	...	...
10	184 756	17 310 309 456 440

Если учесть, что в современных обратных задачах возникают проблемы оптимизации при  $m \approx 20-100$  и  $M = 10^4 - 10^5$ , то объем необходимого числа вариантов расчетов достигает астрономических величин, делающих невозможным полный перебор всех планов эксперимента даже при использовании современных ЭВМ. Число вариантов полного перебора возможных планов в методе Козлова, описанном нами ранее, еще во много раз больше.

Рассмотренные задачи оптимизации относятся к классу экстремальных задач, методы решения которых рассматриваются в бурно развивающейся в последние десятилетия соответствующей области математики. Несмотря на ряд успешных результатов, полученных для определенных классов экстремальных задач, точное их решение во многих случаях требует полного перебора всех возможных планов. Как уже говорилось выше, В. П. Козлову удалось построить эффективный алгоритм [4], позволяющий избежать полного перебора и при небольших вычислительных затратах найти оптимальный план для сформулированной выше постановки задачи. В большинстве других случаев (например для зависимости погрешностей измерения от ширины спектрального интервала, отличающейся от зависимости (4.3-1)) подобные алгоритмы отсутствуют, и вместо поиска оптимального решения приходится использовать различные алгоритмы частичного перебора, позволяющие в большинстве случаев построить достаточно близкий к оптимальному план эксперимента, называемый обычно квазиоптимальным. К подобным алгоритмам относится, например, предложенный О. М. Покровским алгоритм решения задачи выбора положения спектральных каналов (постановку задачи см. выше).

### **Алгоритм решения задачи об оптимальном выборе спектральных каналов**

Этот алгоритм основывается на использовании формулы (4.2-20) для количества информации

$$\Delta I(x, \tilde{y}) = \frac{1}{2} \log \frac{|\tilde{\mathbf{K}}_f|}{(\sigma^2)^m}. \quad (4.3-23)$$

Прежде всего выясним геометрический смысл определителя  $|\tilde{\mathbf{K}}_f|$ .

Пусть  $a_i = (a_{i1}, a_{i2}, \dots, a_{in})$ ,  $i = 1, \dots, l$  – строка матрицы  $\mathbf{A}$  операторного уравнения (4.3-10), соответствующая значению  $y_i$ . Введем в рассмотрение симметричную матрицу

$$\Phi = \begin{vmatrix} K_\varphi & 0 \\ 0 & I_\varepsilon \end{vmatrix},$$

$(l+n)$ -мерные вектора  $\tilde{a}_i = (a_{i1}, \dots, a_{in}, 0, \dots, 0, \sigma, 0, \dots, 0)$ ,  $i = 1, \dots, l$ ,  $\sigma$  находится на позиции  $n+i$ , и скалярное произведение в пространстве этих векторов  $(\tilde{a}, \tilde{a}) = (\Phi \tilde{a}, \tilde{a})$ . Как следует из теории матриц,  $|\tilde{\mathbf{K}}_f|$  – определитель Грама векторов  $\tilde{a}_i$  в смысле введенного скалярного произведения

$$\Gamma(\tilde{a}_1, \dots, \tilde{a}_n, \Phi) = \left| \{(\tilde{a}_i, \tilde{a}_j)\}_{i,j=1}^l \right|. \quad (4.3-24)$$

Следовательно,  $|\tilde{\mathbf{K}}_f|$  представляет собой величину объема параллелепипеда, построенного на векторах  $\tilde{a}_1, \dots, \tilde{a}_l$  в метрике  $(\tilde{a}, \tilde{a})$ .

Если план эксперимента допускает  $m$  измерений, то задача абсолютной максимизации количества информации (4.3-23) может быть решена путем полного перебора сочетаний из  $l$  значений отсчетных точек  $y_i$  по  $m$  и вычисления  $C_l^m$  определителей вида (4.3-24).

В рассматриваемых нами проблемах оптимизации условий эксперимента в обратных задачах атмосферной оптики число  $l$  может быть велико, поэтому полный перебор всех допустимых планов невозможен. В связи с этим Покровский предложил эмпирический алгоритм последовательного определения квазиоптимальной совокупности значений  $y_i$  на заданной сетке  $\{y_i\}_{i=1}^l$  ( $l \geq m$ ). Идея алгоритма проста и заключается в том, что каналы выбираются последовательно, причем на каждом шаге добавляется канал, дающий наибольший прирост количества информации по Шеннону.

Предположим, что уже выбрано  $m$  значений  $y_{i_k}$ ,  $k = 1, \dots, m$ . Рассмотрим приращение количества информации  $\delta I_{m+1}$  при использовании дополнительного измерения при  $y_{i_{m+1}}$ .

Используя выражение (4.3-24), имеем

$$\delta I_{m+1} = \Delta I_{m+1} - \Delta I_m = \frac{1}{2} \log \frac{\Gamma(\tilde{a}_{i_1}, \tilde{a}_{i_2}, \dots, \tilde{a}_{i_m}, \tilde{a}_{i_{m+1}}, \Phi)}{\Gamma(\tilde{a}_{i_1}, \tilde{a}_{i_2}, \dots, \tilde{a}_{i_m}, \Phi)}. \quad (4.3-25)$$

Величина отношения определителей в (4.3-25) представляет собой квадрат расстояния вектора  $\tilde{a}_{i_{m+1}}$  до гиперплоскости, построенной на совокупности векторов  $\tilde{a}_{i_1}, \tilde{a}_{i_2}, \dots, \tilde{a}_{i_m}$ , соответствующих уже выбранным значениям  $y_{i_k}, k=1, \dots, m$ . Тогда из геометрических соображений можно получить равенство

$$\begin{aligned} & \frac{\Gamma(\tilde{a}_{i_1}, \tilde{a}_{i_2}, \dots, \tilde{a}_{i_m}, \tilde{a}_{i_{m+1}}, \Phi)}{\Gamma(\tilde{a}_{i_1}, \tilde{a}_{i_2}, \dots, \tilde{a}_{i_m}, \Phi)} = \\ & = \sigma^2 + \frac{\Gamma(\tilde{a}_{i_1}, \tilde{a}_{i_2}, \dots, \tilde{a}_{i_m}, \hat{a}_{i_{m+1}}, \Phi)}{\Gamma(\tilde{a}_{i_1}, \tilde{a}_{i_2}, \dots, \tilde{a}_{i_m}, \Phi)}, \end{aligned} \quad (4.3-26)$$

в котором вектор  $\hat{a}_{i_{m+1}} = (a_{i_{m+1}1}, \dots, a_{i_{m+1}n}, 0, \dots, 0)$  в отличие от вектора  $\tilde{a}_{i_{m+1}}$  уже не несет в себе «шума». Правая часть соотношения (4.3-26) содержит две составные части приращения информации: «шум»  $\sigma^2$  и полезную информацию, вносимую новым вектором  $\tilde{a}_{i_{m+1}}$ .

Указанное разделение информации составляет основу для анализа информационного содержания эксперимента и определяет алгоритм последовательного оптимального выбора точек измерений  $y_{i_{m+1}}$ . Значение  $y_{i_{m+1}}$  находят из условия максимизации определителя  $\Gamma(\hat{a}_{i_{m+1}}, \Phi)$ . При этом требуется вычисление  $l$  таких определителей. Далее оптимальный выбор  $y_{i_{m+1}}$  при условии, что  $y_{i_k}, k=1, \dots, m$ , уже определены, сводится к нахождению значения  $y_{i_{m+1}}$ , которому

соответствует вектор  $\tilde{a}_{i_{m+1}}$ , максимизирующий  $\Gamma(\tilde{a}, \Phi)$ , т. е. приращение полезной информации.

Вопрос о необходимом числе измерений  $m$  (числе точек  $y_{i_k}$ ,  $k=1, \dots, m$ ) в методике Покровского решается на основе анализа значений  $\Delta I_m$ ,  $\delta I_{m+1}$  и величины отношения «сигнал/шум»

$$\frac{\Gamma(\tilde{a}_{i_1}, \tilde{a}_{i_2}, \dots, \tilde{a}_{i_m}, \tilde{a}_{i_{m+1}}, \Phi)}{\Gamma(\tilde{a}_{i_1}, \tilde{a}_{i_2}, \dots, \tilde{a}_{i_m}, \Phi) \sigma^2}.$$

Большинство рассмотренных (а также ряд других) методик оптимизации условий измерений, отличаясь по исходным предпосылкам, характеру используемой априорной информации об искомом решении в рассматриваемой обратной задаче, выбору критерия оптимизации, тем не менее укладываются в одну общую схему [5]. Моделируется некий «полный» радиационный эксперимент, включающий измерения, например, уходящего излучения в большом числе спектральных каналов, ширины и положения которых зафиксированы. Отбор небольшого (по сравнению с первоначальной совокупностью) числа каналов диктуется фактически не максимальной информативностью возможного эксперимента, а фактом убывающей «полезности» дополнительных каналов измерений. Такая «оптимальная» система может только приближаться по информативности к «полному» эксперименту, но не может его превзойти.

Как мы уже отмечали ранее, В. П. Козлов впервые указал на принципиальные недостатки рассмотренной схемы в применении ее к разработке спектральной аппаратуры, так как она не принимает во внимание природу рассматриваемого класса экспериментов. В рассматриваемых экспериментах мы имеем дело с измерениями спектрального распределения энергии излучения. Получение большого числа каналов измерений требует увеличения спектрального разрешения (уменьшения ширин спектральных интервалов) и, как следствие, приводит к уменьшению энергии полезного сигнала в каждом канале. Поэтому «полный» радиационный эксперимент, сравнение с информацией которого (или соответствующей

дисперсионной матрицей) осуществляется при оптимизации, может не являться наиболее информативным (или наиболее точным).

Поэтому хотя уравнения соответствующей линейной системы, соответствующие различным спектральным интервалам измерений, и являются практически линейно зависимыми, полезно не отбрасывать эти «лишние» уравнения (и соответствующие измерения), как это делается, например, в методике Туоми, а объединять энергию излучения в «зависимых» спектральных интервалах в один канал измерений. Именно это и делается в методике оптимизации условий измерений Козлова. Эта методика успешно использовалась, в частности, при рассмотрении ряда обратных задач атмосферной оптики ИК области спектра.

#### ***4.4. Пример исследования информативности и построения оптимальной спектральной схемы измерений***

Проиллюстрируем использование аппарата оптимального планирования дистанционного эксперимента на примере оптимизации измерений в надир собственного излучения системы атмосфера–подстилающая поверхность в микроволновой (50–71 ГГц) области спектра [13].

При выборе величин предельного спектрального разрешения и погрешности измерений использовались параметры прибора AMSU, разработанного для космических измерений в указанном диапазоне спектра. Предельное спектральное разрешение прибора составляет 3 МГц при ошибке измерений в этом канале 2,5 К.

В соответствии с этими параметрами вся спектральная область была разбита на 7001 спектральный интервал шириной 3 МГц. Гипотетические измерения во всех этих каналах одновременно представляют собой так называемый полный радиационный эксперимент и при заданном спектральном разрешении предоставляют исчерпывающую информацию о наблюдаемой во всем диапазоне спектра яркостной температуре уходящего излучения.

Расчеты информационного содержания полного радиационного эксперимента позволили оценить потенциальные возможности зондирования в исследуемой спектральной области и определить

необходимое число измерительных каналов для извлечения всей полезной информации, содержащейся в поле радиации. В табл. 4.4-1 приведены первые 20 собственных значений  $\lambda_i$  ковариационной матрицы уходящего микроволнового излучения.

Таблица 4.4-1

Первые собственные значения глобальной ковариационной матрицы яркостных температур уходящего микроволнового излучения  $T_{я}$  в спектральной области 50–71 ГГц ( $K^2$ )

$i$	$\lambda_i$	$i$	$\lambda_i$	$i$	$\lambda_i$	$i$	$\lambda_i$
1	325984	6	2251,7	11	8,9115	16	0,0423
2	233572	7	976,92	12	2,9575	17	0,0118
3	42661	8	290,23	13	0,9998	18	0,0036
4	8859,1	9	88,662	14	0,4192	19	0,0017
5	5484,9	10	32,168	15	0,1433	20	0,0005

Анализ собственных значений, приведенных в табл. 4.4-1, позволяет определить по критерию  $\lambda_i > \sigma^2$  число независимых информационных компонент излучения при различных погрешностях его измерений и оценить потенциальные возможности температурного зондирования, задавшись произвольной точностью этих измерений.

При уровне погрешности измерений 2,5 К (радиометр AMSU) в уходящем микроволновом излучении содержится 11 независимых компонент. Следовательно, при указанной погрешности измерений можно считать информационно-обусловленным прибор, выполняющий измерения в 11 спектральных каналах, поэтому определение оптимальных условий измерений было проведено для этого числа каналов.

В рассматриваемом диапазоне спектра величина среднеквадратической случайной ошибки измерений  $\sigma_{T_{я}}$  обратно пропорциональна корню квадратному из полной спектральной ширины канала, поэтому в данном случае не может быть использован разработанный В.П. Козловым для ИК диапазона метод поиска оптимального плана. Был разработан алгоритм частичного перебора, позволяющий построить квазиоптимальный план эксперимента.

Приведем основные соотношения, на которые опирается этот алгоритм.

Каждый канал оптимального плана измерений задается как подмножество набора элементарных частотных интервалов  $\Delta f_j, j=1, \dots, n$ , шириной 3 МГц, покрывающих в совокупности исследуемую спектральную область. При этом оптимальному плану соответствует бинарная матрица  $\{p_{i,j}\}_{i=1, \dots, m, j=1, \dots, n}$ , каждая строка которой описывает один канал спектрального прибора. Ее элементы  $p_{i,j}$  принимают значение 1, если  $j$ -й спектральный интервал прозрачен в  $i$ -м канале моделируемого прибора, и 0 в противном случае.

Оптимальным планом измерений мы считаем спектральную схему, максимизирующую информационный объем  $V$  измерительной системы.

В этих предположениях задача сводится к нахождению максимума квадрата определителя

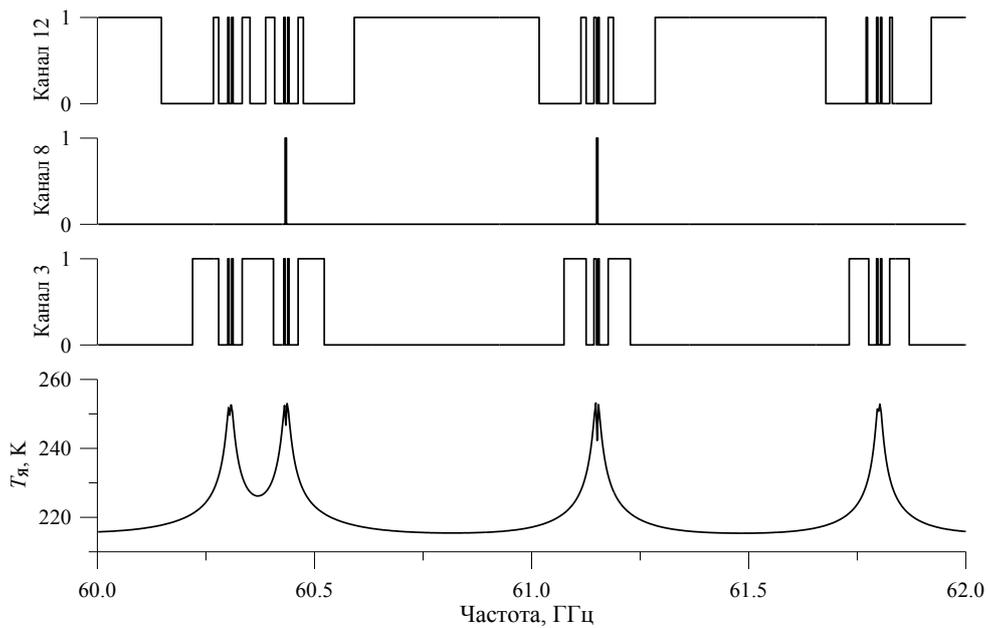
$$\left| \begin{array}{c} \left\{ \frac{\sum_{l=1}^n p_{i,l} x_{j,l}}{\sum_{l=1}^n p_{i,l}} \right\}_{i,j=1, \dots, m} \right|^2, \quad (4.4-1)$$

где  $x_{j,l}$  – матрица, составленная из собственных векторов ковариационной матрицы яркостной температуры уходящего микроволнового излучения, помноженных на корни квадратные из соответствующих собственных чисел. Эти собственные вектора соответствуют независимым компонентам информации, содержащейся в излучении, и одновременно независимым компонентам профиля температуры, которые мы можем восстановить из всего объема измерений в рассматриваемой спектральной области.

Целевая функция (4.4-1) отличается весами  $1/\sum_{l=1}^n p_{i,l}$  от целевой функции, рассмотренной В.П. Козловым для ИК области спектра, что потребовало создания специального алгоритма поиска экстремума.

Полученные оптимальные планы имеют сложную спектральную структуру. Отдельный канал измерений состоит из многих спектральных интервалов различной ширины, не связанных между

собой и расположенных в различных частях всей рассмотренной спектральной области. Для иллюстрации на рис. 4.4-1. приведен небольшой участок трех каналов оптимального плана, рассчитанного для  $N = 11$ . (В нижней части рис. 4.4-1 приведен спектр яркостной температуры уходящего излучения.)



**Рис. 4.4-1.** Фрагмент оптимального плана и спектр яркостной температуры уходящего излучения в микроволновой области спектра.

Ординаты 1 и 0 соответствуют наличию или отсутствию приема излучения в соответствующем элементарном спектральном интервале. Из приведенного рисунка видно, что конфигурация оптимального плана коррелирует со спектральным поведением спектра уходящего излучения. В частности, в оптимальных каналах суммируется излучение спектральных интервалов с близкими оптическими характеристиками.

В табл. 4.4-2 приведены величины информационного объема для различных спектральных схем измерений уходящего излучения.

*Таблица 4.4-2*

Величины информационных объемов измерений  
для различных спектральных схем

Спектральная схема измерений	Информационный объем
AMSU	$0,25 \cdot 10^9$
Оптимальный план	$0,19 \cdot 10^{14}$
Полный радиационный эксперимент	$0,67 \cdot 10^{14}$

Из табл. 4.4-2 можно видеть, что оптимальные схемы измерений характеризуются величинами информационного объема, превышающими на пять порядков соответствующие значения для радиометра AMSU. Расчеты погрешностей определения профиля температуры показали, что преимущества оптимальных систем микроволнового температурного зондирования атмосферы достигают 0,2–0,6 К в тропосфере и 1,2–1,5 К в стратосфере по сравнению с аппаратурой AMSU.

Таким образом, приведенный пример наглядно демонстрирует преимущества оптимальной спектральной схемы.

**П р и л о ж е н и е**

**Краткий обзор необходимых  
сведений из области  
математики**

***Небольшое предисловие***

Это Приложение ни в коем случае не претендует на полноту изложения рассматриваемых разделов математики и служит лишь для того, чтобы напомнить читателю основные понятия и факты, необходимые для понимания материала. Почти все собранные здесь сведения приведены по известному справочнику [8]. За редким исключением доказательства теорем отсутствуют, мы предполагаем, что при необходимости читатель сможет обратиться к соответствующим учебникам.

Несколько слов о терминологии и обозначениях. Термин «оператор» подразумевает отображение из одного функционального или векторного пространства в другое. Для обозначения операторов мы используем заглавные полужирные буквы (например **A**). Символы такого же вида используются для обозначения матриц и множеств. Термин «функционал» обозначает отображение из функционального или векторного пространства в множество вещественных чисел. Для обозначения функционалов, функций и чисел используются обычные латинские или греческие буквы (например  $\Psi, x$ ).

В части пособия, относящейся к функциональному анализу, используются следующие обозначения. Запись «функция  $f$ » обозначает функцию, т. е. отображение из множества вещественных чисел в множество вещественных чисел, в отличие от записи  $f(x)$ , которая обозначает число, являющееся значением функции  $f$  от значения аргумента  $x$  (тоже числа).

Наряду с этими обозначениями встречаются и традиционные для классического математического анализа, например, «функция  $f(x)$ »

обозначает не значение функции в точке  $x$ , а функцию как отображение. Какие обозначения имеются в виду в каждом конкретном случае, понятно из контекста.

### **П.1. Понятия расстояния и метрического пространства. Пространства $L_2$ и $C_{[a,b]}$**

Пусть заданы линейное векторное пространство  $\mathbf{F}$  и вещественная функция  $\rho_F(f_1, f_2)$ ,  $f_1, f_2 \in \mathbf{F}$ , удовлетворяющая следующим условиям:

- 1)  $\rho_F(f_1, f_2) > 0$ ;  $\rho_F(f_1, f_2) = 0$  только при  $f_1 = f_2$ ;
- 2)  $\rho_F(f_1, f_2) = \rho_F(f_2, f_1)$ ;
- 3)  $\rho_F(f_1, f_2) < \rho_F(f_1, f_3) + \rho_F(f_2, f_3)$  для любого  $f_3 \in \mathbf{F}$  (неравенство треугольника).

Тогда функция  $\rho_F$  называется расстоянием, а пространство  $\mathbf{F}$  метрическим пространством. Также говорят, что  $\mathbf{F}$  наделено метрикой  $\rho_F$ , или метризовано.

Мы будем рассматривать метрические пространства  $L_2$  – пространство функций, интегрируемых с квадратом на интервале  $[a, b]$  и  $C_{[a,b]}$  – пространство непрерывных на  $[a, b]$  функций. Метрика в этих пространствах определяется следующими расстояниями  $\rho_F$ :

$$\rho_{L_2}(f_1, f_2) = \sqrt{\int_a^b (f_1(y) - f_2(y))^2 dy}; \quad (\text{П.1-1})$$

$$\rho_C(f_1, f_2) = \max_{y \in [a, b]} |f_1(y) - f_2(y)|. \quad (\text{П.1-2})$$

### **П.2. Метод множителей Лагранжа (условный экстремум)**

Рассмотрим применение метода для поиска экстремума функции многих переменных с ограничениями.

Чтобы найти максимумы и минимумы действительной функции  $f(x_1, x_2, \dots, x_n)$

переменных  $x_1, x_2, \dots, x_n$ , подчиненных достаточно гладкому *дополнительному условию* в виде уравнения связи

$$\varphi(x_1, x_2, \dots, x_n) = 0, \quad (\text{П.2-1})$$

применяют следующий прием: для каждого из значений вещественной переменной  $\lambda$  ищется экстремум функции

$$\Phi(x_1, x_2, \dots, x_n) \equiv f(x_1, x_2, \dots, x_n) + \lambda \varphi(x_1, x_2, \dots, x_n). \quad (\text{П.2-2})$$

Полученные в результате параметрические зависимости  $x_1(\lambda), x_2(\lambda), \dots, x_n(\lambda)$  (или связывающие их уравнения, также параметрически зависящие от  $\lambda$ ) совместно с уравнением связи (П.2-1) позволяют разрешить его относительно *множителя Лагранжа*  $\lambda$  и найти решение исходной экстремальной задачи.

Обычно для практического использования метод формулируется в терминах необходимого условия экстремума

$$\frac{\partial \Phi}{\partial x_i} = 0, \quad i = 1, \dots, n, \quad (\text{П.2-2a})$$

что технически удобно, поскольку позволяет говорить о совместном решении системы  $n$  уравнений (П.2-2a) и уравнения связи (П.2-1) относительно  $n+1$  неизвестных ( $x_i, i = 1, \dots, n$  и  $\lambda$ ).

Доказательство проведем от противного. Допустим, что существуют  $x'_1, x'_2, \dots, x'_n$  такие, что  $f(x'_1, x'_2, \dots, x'_n) < f(x_1, x_2, \dots, x_n)$  и относительно штрихованных переменных выполняется условие (П.2-1). Но тогда поскольку условие (П.2-1) выполняется относительно обоих наборов переменных, то и  $\Phi(x'_1, x'_2, \dots, x'_n) < \Phi(x_1, x_2, \dots, x_n)$ , что противоречит построению набора  $x_1, x_2, \dots, x_n$ . Это противоречие доказывает применимость метода. Разумеется, к формулировке (П.2-2a) следует отнести все оговорки, связанные с необходимым условием экстремума.

Метод неопределенных множителей Лагранжа допускает разнообразные расширения, в частности, он может быть сформулирован для решения интересующей нас вариационной экстремальной задачи. В этом случае ищется функция  $\varphi$ , минимизирующая функционал  $F(\varphi)$  при условии

$$R(\varphi) = 0, \quad (\text{П.2-3})$$

где  $R$  также функционал от функции  $\varphi$ . Для поиска решения рассматривается задача минимизации функционала

$$\Phi(\varphi, \lambda) = \lambda R(\varphi) + F(\varphi), \quad (\text{П.2-4})$$

решение которой  $\varphi_\lambda$  зависит от множителя Лагранжа  $\lambda$  как от параметра. Тогда множитель Лагранжа может быть найден из совместного рассмотрения этой зависимости и условия (П.2-3). Доказательство аналогично предыдущему случаю.

Для наших целей важно отметить, что минимизация (П.2-4) производится при фиксированном  $\lambda$ , и реализующая минимум функция  $\varphi_\lambda$  зависит от множителя Лагранжа параметрически.

### **П.3. Основные формулировки и факты теории случайных процессов**

Мы приводим информацию, относящуюся только к единственному вещественному случайному процессу.

**Случайные процессы.** Случайный процесс есть случайная функция  $x(t)$  от независимой переменной  $t$ . Каждое испытание дает определенную функцию  $X(t)$ , которая называется реализацией процесса или выборочной функцией. Случайный процесс можно рассматривать либо как совокупность реализаций процесса  $X(t)$ , либо как совокупность случайных величин, зависящих от параметра  $t$ .

**Средние по множеству наблюдений. Корреляционные функции.** Для случайного процесса  $x(t)$  особый интерес представляют средние  $\mathbf{M}x(t) = \xi(t)$ ,  $\mathbf{M}x^2(t)$  и автокорреляционная функция

$$R_{xx}(t_1, t_2) = \mathbf{M}(x(t_1)x(t_2)). \quad (\text{П.3-1})$$

(Символы  $\mathbf{M}$  и  $\mathbf{D}$  в теории случайных процессов традиционно обозначают математическое ожидание и дисперсию соответственно.)

Средние и автокорреляционная функция выражают важнейшие свойства случайного процесса и часто представляют все, что известно о данном процессе. Заметим, что

$$\mathbf{M}x^2(t) = R_{xx}(t, t), \quad \mathbf{D}x(t) = R_{xx}(t, t) - (\xi(t))^2. \quad (\text{П.3-2})$$

**Стационарные случайные процессы.** Случайный процесс называется стационарным, если все его конечномерные распределения

вероятностей инвариантны относительно сдвига по параметру  $t$ . Отсюда следует, что средние значения

$$\mathbf{M}x(t) = \mathbf{M}x(0) = \mathbf{M}x = \xi, \quad \mathbf{M}x^2(t) = \mathbf{M}x^2(0) = \mathbf{M}x^2 \quad (\text{П.3-3})$$

постоянны, а автокорреляционная функция зависит только от запаздывания  $\tau = t_2 - t_1$ , причем

$$R_{xx}(\tau) = R_{xx}(-\tau).$$

**Спектральная плотность по множеству наблюдений.** Для стационарного случайного процесса  $x(t)$  спектральная плотность  $\Phi_{xx}(\omega)$  по множеству наблюдений (реализаций) определяется соотношением Хинчина–Винера

$$\Phi_{xx}(\omega) = \int_{-\infty}^{\infty} R_{xx}(\tau) e^{-i\omega\tau} d\tau. \quad (\text{П.3-4})$$

При соответствующих условиях имеет место формула обращения

$$R_{xx}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{xx}(\omega) e^{i\omega\tau} d\omega. \quad (\text{П.3-5})$$

Преобразование Фурье (П.3-4) вводится обычно для упрощения соотношений между корреляционными функциями входного и выходного сигналов в линейной стационарной системе. Существование преобразования (П.3-4) требует кроме существования  $\mathbf{M}x^2$  еще и достаточно быстрого стремления к нулю величины  $R_{xx}(\tau)$  при  $\tau \rightarrow \infty$ .

Для действительного процесса

$$\Phi_{xx}(\omega) = 2 \int_0^{\infty} R_{xx}(\tau) \cos(\omega\tau) d\tau, \quad (\text{П.3-6})$$

$$R_{xx}(\tau) = \frac{1}{\pi} \int_0^{\infty} \Phi_{xx}(\omega) \cos(\omega\tau) d\omega. \quad (\text{П.3-7})$$

Для действительного случайного процесса  $x(t)$  подстановка  $\tau = 0$  в соотношение (П.3-7) дает

$$\mathbf{M}|x|^2 = R_{xx}(0) = \frac{1}{\pi} \int_0^{\infty} \Phi_{xx}(\omega) d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{xx}(\omega) d\omega. \quad (\text{П.3-8})$$

$\Phi_{xx}(\omega)$  можно толковать как спектральное разложение средней «мощности»  $\mathbf{M}|x|^2$  по частотам  $\omega$  в  $(0, \infty)$  или  $(-\infty, \infty)$ . Второй интеграл дает разложение по всем положительным и отрицательным частотам со спектральной плотностью  $\Phi_{xx}(\omega)$  (размерность  $\Phi_{xx}(\omega)$  есть (единицы  $x$ )<sup>2</sup>/герц), первый интеграл – по неотрицательным («действительным») частотам с «односторонней» спектральной плотностью  $2\Phi_{xx}(\omega)$ .

#### **П.4. Вектора, матрицы и конечномерные линейные операторы**

Заметим, что в силу изоморфизма матриц над полем вещественных чисел и конечномерных линейных операторов имеет место аналогия между теоремами, относящимися к этим пространствам. Также некоторые понятия и теоремы могут быть обобщены и на бесконечномерные векторные пространства (пространства функций) (см. предисловие).

#### **Скалярное произведение**

Скалярное произведение векторов  $\vec{a}, \vec{b}$  в  $m$ -мерном евклидовом пространстве определяется формулой

$$(\vec{a}, \vec{b}) = \sum_{k=1}^m a_k b_k, \quad (\text{П.4-1})$$

где индекс  $k$  обозначает  $k$ -ю компоненту вектора. Если  $\vec{a} = \vec{b}$ , то из (П.4-1) получаем

$$(\vec{a}, \vec{a}) = \|\vec{a}\|^2 = \sum_{k=1}^m a_k^2, \quad (\text{П.4-2})$$

т. е. норма вектора может быть определена через его скалярный квадрат. Таким образом, скалярное произведение векторов задает длину вектора, следовательно, расстояние и метрику пространства.

Скалярное произведение ортогональных векторов равно нулю и обратно: если скалярное произведение ненулевых векторов равно нулю, то эти вектора ортогональны.

Скалярное произведение вектора на вектор единичной длины равно проекции первого вектора на ось, определяемую вторым.

### Энергетическая норма

Наряду с конечномерным метрическим пространством, наделенным евклидовой метрикой, определяемым скалярным произведением, можно рассмотреть и другие метрики.

Пусть  $S$  – симметричный положительно определенный оператор (соответствующая ему матрица) в пространстве  $\mathbf{R}^n$ . Тогда скалярное произведение  $(\vec{x}, S\vec{y})$  задает норму вектора (и тем самым расстояние)  $\|\vec{x} - \vec{y}\|_S^2 = \rho_S(\vec{x}, \vec{y})^2 = (\vec{x} - \vec{y}, S(\vec{x} - \vec{y}))$ , называемую иногда энергетической нормой.

### Собственные вектора и собственные числа матрицы

Если для матрицы  $A$  выполняется соотношение

$$A\vec{\psi} = \lambda\vec{\psi}, \quad (\text{П.4-3})$$

где  $\vec{\psi}$  – некоторый вектор и  $\lambda$  – вещественное число, то число  $\lambda$  называется собственным числом матрицы  $A$ , а вектор  $\vec{\psi}$  называется собственным вектором матрицы  $A$ , соответствующим ее собственному числу  $\lambda$ . Уравнение (П.4-3) означает, что действие матрицы на ее собственный вектор сводится только к его растяжению (сжатию при  $|\lambda| < 1$ ) в  $\lambda$  раз. Заметим, что собственные вектора определены с точностью до множителя, т. е. если  $\vec{\psi}$  – собственный вектор, соответствующий собственному числу  $\lambda$ , то для любого вещественного числа  $a$  и вектор  $a\vec{\psi}$  также является собственным вектором, соответствующим тому же собственному числу.

В рамках нашего курса мы обычно рассматриваем матрицы, собственные числа которых неотрицательны. Мы всегда будем предполагать, что собственные числа упорядочены по убыванию, а собственные вектора нормированы так, что имеют единичную длину.

## Некоторые свойства собственных чисел и собственных векторов

Симметричная матрица  $\mathbf{A}$  неотрицательно определена тогда и только тогда, когда существует матрица  $\mathbf{B}$ , такая, что  $\mathbf{A} = \mathbf{B}'\mathbf{B}$ .

Все собственные числа неотрицательно определенной симметричной матрицы неотрицательны.

Собственные вектора нормальной матрицы (нормального оператора) попарно ортогональны. (Матрица  $\mathbf{A}$  называется нормальной, если  $\mathbf{A}'\mathbf{A} = \mathbf{A}\mathbf{A}'$ .)

## Матрица в базисе из собственных векторов

В различных приложениях часто используется рассмотрение матрицы (линейного оператора) в базисе из ее собственных векторов. Особенно удобно такое рассмотрение в случае ортонормальности этого базиса. В этом случае матрица становится диагональной, причем на диагонали ее стоят собственные числа. Унитарная матрица преобразования координат при этом составлена из собственных векторов исходной матрицы.

## Норма матрицы

Существуют различные определения нормы матрицы. В наших приложениях мы используем следующее: по определению норма матрицы  $\|\mathbf{A}\|_{\mathbf{R}^m}$

$$\|\mathbf{A}\|_{\mathbf{R}^m} = \max_{\vec{\varphi} \neq 0} \frac{\|\mathbf{A}\vec{\varphi}\|_{\mathbf{R}^m}}{\|\vec{\varphi}\|_{\mathbf{R}^m}}, \quad (\text{П.4-4})$$

где  $\|\vec{\varphi}\|_{\mathbf{R}^m}$  – евклидова норма вектора  $\vec{\varphi}$ .

Очевидное свойство нормы матрицы:  $\|\mathbf{A}\vec{\varphi}\|_{\mathbf{R}^m} \leq \|\mathbf{A}\|_{\mathbf{R}^m} \|\vec{\varphi}\|_{\mathbf{R}^m}$ .

Норма матрицы  $\mathbf{A}$  равна максимальной абсолютной величине ее собственного числа. Докажем это. Воспользуемся определением нормы матрицы (П.4-4), заменив вектор  $\vec{\varphi}$  его разложением по базису из собственных векторов  $\{\psi_i\}_{i=1,m}$  матрицы  $\mathbf{A}$ :

$$\begin{aligned} \|\mathbf{A}\| &= \max_{\vec{\varphi} \neq 0} \frac{\left\| \mathbf{A} \sum_{k=1}^m a_{\varphi k} \vec{\psi}_k \right\|}{\left\| \sum_{k=1}^m a_{\varphi k} \vec{\psi}_k \right\|} = \max_{a_{\varphi k}} \frac{\left\| \sum_{k=1}^m a_{\varphi k} \lambda_k \vec{\psi}_k \right\|}{\left\| \sum_{k=1}^m a_{\varphi k} \vec{\psi}_k \right\|} \leq \\ &\leq |\lambda_{\max}| \max_{\vec{\varphi}} \frac{\left\| \sum_{k=1}^m a_{\varphi k} \vec{\psi}_k \right\|}{\left\| \sum_{k=1}^m a_{\varphi k} \vec{\psi}_k \right\|} = |\lambda_{\max}|. \end{aligned}$$

С другой стороны,  $\|\mathbf{A}\| = \max_{\vec{\varphi} \neq 0} \frac{\|\mathbf{A}\vec{\varphi}\|}{\|\vec{\varphi}\|} \geq \frac{\|\mathbf{A}\vec{\psi}_{\max}\|}{\|\vec{\psi}_{\max}\|} = |\lambda_{\max}|$ . Таким образом, действительно,  $\|\mathbf{A}\| = \lambda_{\max}$ .

### Обратная матрица

Если для матрицы  $\mathbf{A}$  существует такая матрица  $\mathbf{B}$ , что для любого вектора  $\vec{\varphi} \in R^m$  верно  $\mathbf{B}(\mathbf{A}\vec{\varphi}) = \vec{\varphi}$ , то матрица  $\mathbf{B}$  называется обратной матрице  $\mathbf{A}$  и обозначается  $\mathbf{A}^{-1}$ .

Собственные числа обратной матрицы представляют собой обратные величины собственных чисел прямой матрицы, а соответствующие собственные вектора матриц совпадают.

Это утверждение легко доказывается. Пусть  $\mathbf{A}\vec{\psi}_j = \lambda_j \vec{\psi}_j$ . Умножим это равенство слева на  $\mathbf{A}^{-1}$ . Тогда  $\mathbf{A}^{-1}\mathbf{A}\vec{\psi}_j = \mathbf{A}^{-1}\lambda_j \vec{\psi}_j$ , и так как  $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ , то  $\vec{\psi}_j = \lambda_j \mathbf{A}^{-1}\vec{\psi}_j$  или  $\mathbf{A}^{-1}\vec{\psi}_j = \frac{1}{\lambda_j} \vec{\psi}_j$ . Последнее равенство также показывает, что обратная матрица имеет те же собственные вектора  $\vec{\psi}_j$ .

### Системы линейных уравнений и определитель матрицы

*Определитель* (детерминант)  $\det \mathbf{A}$ ,  $|\mathbf{A}|$  квадратной матрицы  $\mathbf{A}$

$$\mathbf{A} = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} \quad (\text{П.4-5})$$

есть сумма  $n!$  слагаемых вида  $(-1)^r a_{1k_1} a_{2k_2} \dots a_{nk_n}$ , каждое из которых соответствует одному из  $n!$  различных упорядоченных множеств  $k_1, k_2, \dots, k_n$  полученных  $r$  попарными перестановками элементов множества  $1, 2, \dots, n$ .

#### Некоторые свойства определителей

$$|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|; |c\mathbf{A}| = c^n |\mathbf{A}|; |\mathbf{A}| = \prod_{i=1}^n \lambda_i.$$

Здесь  $\mathbf{A}$  и  $\mathbf{B}$  – матрицы,  $c$  – число,  $\{\lambda_i\}_{i=1,n}$  – собственные числа матрицы  $\mathbf{A}$ .

**Системы линейных уравнений. Правило Крамера.** Система  $n$  линейных уравнений с  $n$  неизвестными

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2, \\ \dots & \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= b_n, \end{aligned}$$

$$\text{или } \sum_{k=1,n} a_{ik}x_k = b_i, \quad i = 1, \dots, n, \text{ или } \mathbf{A}\vec{x} = \vec{b}$$

в случае, если ее определитель  $D = |\mathbf{A}|$  не равен нулю, имеет единственное решение

$$x_k = \frac{D_k}{D}, \quad k = 1, \dots, n, \quad (\text{П.4-6})$$

где  $D_k$  – определитель матрицы, полученной заменой элементов  $k$ -го столбца матрицы  $\mathbf{A}$   $a_{1k}, a_{2k}, \dots, a_{nk}$  на столбец свободных членов  $b_1, b_2, \dots, b_n$ .

**Системы линейных уравнений: общая теория.** Система  $m$  линейных уравнений с  $n$  неизвестными

$$\sum_{k=1, n} a_{ik} x_k = b_i, \quad i = 1, \dots, m \quad (\text{П.4-7})$$

имеет решение в том и только в том случае, если матрицы

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \text{ и } \begin{pmatrix} a_{11} & \dots & a_{1n} & b_1 \\ a_{21} & \dots & a_{2n} & b_2 \\ \dots & \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} & b_m \end{pmatrix} \quad (\text{П.4-8})$$

(матрица системы и расширенная матрица системы) имеют один и тот же ранг  $r$ . В противном случае система несовместна (уравнения системы несовместны).

Решение единственно, если  $r = m = n$ . Если обе матрицы (П.4-8) имеют ранг  $r \leq m$ , то уравнения (П.4-7) линейно независимы при  $r = m$  и линейно зависимы при  $r < m$ . В случае  $r < m$  некоторые  $m - r$  уравнений можно выразить в виде линейных комбинаций остальных  $r$  уравнений (независимых), и им удовлетворяют решения этих  $r$  уравнений. Линейно независимые уравнения определяют некоторые  $r \leq m$  неизвестных как линейные функции остальных  $n - r$  неизвестных, остающихся произвольными.

## П.5. Сведения из теории вероятностей

**Вводные замечания.** Вероятностями называются значения некоторой действительной функции, определенной на классе идеализированных событий, которые представляют собой результаты испытания (опыта или наблюдения). Вероятности вводятся посредством определенных аксиом, абстрагируемых из основных свойств статистических относительных частот. Практически понятие вероятности проявляется в том, что обычно относительная частота случайного события в каждой последовательности независимых повторных испытаний приближается к соответствующей вероятности.

*Случайная величина* есть любая переменная  $x$ , значения которой  $x = X$  образуют множество элементарных событий или, другими словами, обозначают точки в пространстве выборок. Соответствующее распределение вероятностей называется распределением случайной величины  $x$ .

Распределение случайной величины  $x$  задается ее функцией распределения

$$\Phi_x(X) \equiv \Phi(X) \equiv P\{x < X\}, \quad (\text{П.5-1})$$

где  $P\{x < X\}$  – вероятность того, что значение случайной величины окажется меньше числа  $X$ .

Действительная случайная величина  $x$  называется непрерывной, если ее функция распределения  $\Phi_x(X) \equiv \Phi(X)$  непрерывна и имеет производную – плотность распределения вероятностей величины  $x$ :

$$\varphi_x(X) \equiv \varphi(X) \equiv \lim_{\Delta X \rightarrow 0} \frac{P\{X \leq x < X + \Delta X\}}{\Delta X} = \frac{d\Phi}{dX}. \quad (\text{П.5-2})$$

Отметим формулы

$$P\{x < X\} \equiv \Phi(X) = \int_{-\infty}^X \varphi(x) dx, \quad (\text{П.5-3})$$

$$P\{a \leq x < b\} = \Phi(b) - \Phi(a) = \int_a^b \varphi(x) dx, \quad (\text{П.5-4})$$

$$\int_{-\infty}^{\infty} \varphi(x) dx = \Phi(\infty) = 1. \quad (\text{П.5-5})$$

*Условная вероятность*  $P\{E/E_1\}$  события  $E$  при условии осуществления события  $E_1$  определяется аксиомой: вероятность совмещения событий  $E \cap E_1$  равна  $P\{E \cap E_1\} = P\{E_1\}P\{E/E_1\}$  (правило умножения вероятностей).

*Математическое ожидание* (среднее значение) функции  $y(x)$  от непрерывной случайной величины  $x$  есть

$$E(y(x)) = \mathbf{M} y(x) = \int_{-\infty}^{\infty} y(x) \varphi(x) dx. \quad (\text{П.5-6})$$

В частности, математическое ожидание (среднее значение, центр распределения)  $\mathbf{M}x = \xi$  и дисперсия  $\mathbf{D}x = \sigma^2$  случайной величины  $x$  определяются по формулам

$$\mathbf{M}x = \xi = \int_{-\infty}^{\infty} x\varphi(x)dx, \quad (\text{П.5-7})$$

$$\mathbf{D}x = \sigma^2 = \mathbf{M}(x - \xi)^2 = \int_{-\infty}^{\infty} (x - \xi)^2 \varphi(x)dx.$$

Моментом порядка  $r$  случайной величины  $x$  относительно числа  $X$  называется математическое ожидание  $\mathbf{M}(x - X)^r$ .

### Многомерные распределения вероятностей

Если случайное событие описывается упорядоченным набором действительных чисел  $(X_1, X_2, \dots, X_n)$ , то этот набор представляет собой значение  $n$ -мерной случайной величины  $\vec{x} = (x_1, x_2, \dots, x_n)$ . Можно говорить о системе случайных величин или о  $n$ -мерном случайном векторе.

Распределение многомерной случайной величины  $\vec{x} = (x_1, x_2, \dots, x_n)$  задается функцией совместного распределения

$$\begin{aligned} \Phi_x(X_1, X_2, \dots, X_n) &\equiv (X_1, X_2, \dots, X_n) \equiv \\ &\equiv P\{x_1 < X_1, x_2 < X_2, \dots, x_n < X_n\}. \end{aligned} \quad (\text{П.5-8})$$

Остановимся на двумерном случае. Распределение системы двух случайных величин  $x_1, x_2$  определяет условное распределение величины  $x_1$  при  $x_2 = X_2$  и условное распределение величины  $x_2$  при  $x_1 = X_1$ .

Условные распределения величин  $x_1$  и  $x_2$  описываются условными плотностями распределений

$$\varphi_{1/2}(X_1 / X_2) = \frac{\varphi(X_1, X_2)}{\varphi_2(X_2)}, \quad \varphi_{2/1}(X_2 / X_1) = \frac{\varphi(X_2, X_1)}{\varphi_1(X_1)}. \quad (\text{П.5-9})$$

Отметим формулу Байеса для двух случайных величин

$$\varphi_{1/2}(X_1 / X_2) = \frac{\varphi_1(X_1)\varphi_{2/1}(X_2 / X_1)}{\int_{-\infty}^{\infty} \varphi_1(x_1)\varphi_{2/1}(X_2 / x_1)dx_1}. \quad (\text{П.5-10})$$

Вернемся к  $n$ -мерному случаю;  $n$  математических ожиданий  $\mathbf{M}x_1 = \xi_1, \dots, \mathbf{M}x_n = \xi_n$  определяют точку  $(\xi_1, \xi_2, \dots, \xi_n)$ , которая

называется центром распределения вероятностей. Величины  $\mathbf{M}(x_1 - X_1)^{r_1} (x_2 - X_2)^{r_2} \dots (x_n - X_n)^{r_n}$  называются моментами порядка  $r_1 + r_2 + \dots + r_n$  относительно точки  $(X_1, X_2, \dots, X_n)$ . Центральные моменты второго порядка представляют особый интерес и имеют специальные названия и обозначения:

$$\lambda_{ik} = \lambda_{ki} = \mathbf{M}(x_i - \xi_i)(x_k - \xi_k) = \begin{cases} \mathbf{D} x_i = \sigma_i^2 & \text{при } i = k \text{ (дисперсия)}, \\ \text{cov} \{x_i, x_k\} & \text{при } i \neq k \text{ (ковариация)}. \end{cases}$$

Эти моменты определяют матрицу моментов  $\{\lambda_{ik}\}$ , называемую также ковариационной матрицей. Ее определитель называется обобщенной дисперсией  $n$ -мерного распределения. Коэффициенты корреляции

$$\rho_{ik} = \rho(x_i, x_k) = \frac{\lambda_{ik}}{\sqrt{\lambda_{ii} \lambda_{kk}}} = \mathbf{M} \frac{x_i - \xi_i}{\sigma_i} \frac{x_k - \xi_k}{\sigma_k}, \quad i, k = 1, \dots, n, \quad (\text{П.5-11})$$

определяют корреляционную матрицу  $n$ -мерного распределения. Ковариационная и корреляционная матрицы вещественны и симметричны.

Энтропия  $H(x)$  распределения вероятностей для одномерной случайной величины  $x$  определяется формулой

$$H(x) = \mathbf{M} \log_2 \frac{1}{\varphi(x)} = - \int_{-\infty}^{\infty} \varphi(x) \log_2 \varphi(x) dx. \quad (\text{П.5-12})$$

## Интеграл вероятности

Интегралом вероятности называется функция неотрицательного вещественного аргумента

$$\Phi(U) = \frac{1}{\sqrt{2\pi}} \int_0^U \exp(-t^2/2) dt = \frac{1}{2} \text{erf} \left( \frac{U}{\sqrt{2}} \right) = \Phi_u(U) - \frac{1}{2}. \quad (\text{П.5-13})$$

Заметим, что  $\Phi$  монотонно возрастает на своей области определения и  $\lim_{U \rightarrow \infty} \Phi(U) = 0,5$ . Здесь  $\Phi_u(U)$  – нормальная функция распределения, erf – функция ошибок.

## **П.6. Дифференцирование функции векторного и матричного аргумента**

Приведем правила дифференцирования нескольких видов функций матричного и векторного аргумента, которые используются в пособии:

1) Дифференцирование обобщенного скалярного произведения по векторному аргументу

$$\frac{d}{d\vec{\varphi}}(\vec{\varphi}, \mathbf{B}\vec{f}) = \mathbf{B}\vec{f}, \quad (\text{П.6-1})$$

$$\frac{d}{d\vec{\varphi}}(\vec{\varphi}, \mathbf{C}\vec{\varphi}) = \mathbf{C}\vec{\varphi} + \mathbf{C}'\vec{\varphi}. \quad (\text{П.6-2})$$

В случае симметричной матрицы  $\mathbf{C}$ , т. е. при  $\mathbf{C} = \mathbf{C}'$ , можно также записать

$$\frac{d}{d\vec{\varphi}}(\vec{\varphi}, \mathbf{C}\vec{\varphi}) = 2\mathbf{C}\vec{\varphi}.$$

2) Дифференцирование следа матрицы

$$\frac{d}{d\vec{x}} \text{Tr}(\vec{x}\mathbf{B}) = \mathbf{B}', \quad (\text{П.6-3})$$

$$\frac{d}{d\vec{x}} \text{Tr}(\mathbf{B}\vec{x}') = \mathbf{B}, \quad (\text{П.6-4})$$

$$\frac{d}{d\vec{x}} \text{Tr}(\vec{x}\mathbf{B}\vec{x}') = \vec{x}\mathbf{B}' + \vec{x}\mathbf{B}. \quad (\text{П.6-5})$$

Если  $\mathbf{B}$  – симметричная матрица ( $\mathbf{B} = \mathbf{B}'$ ), то последнее равенство записывается в виде

$$\frac{d}{d\vec{x}} \text{Tr}(\vec{x}\mathbf{B}\vec{x}') = 2\vec{x}\mathbf{B}. \quad (\text{П.6-6})$$

Доказательство этих соотношений не представляет сложности – достаточно записать матричные соотношения в координатной форме.

## Л и т е р а т у р а

1. *Бирюлина М.С.* Разработка и усовершенствование спутниковых методов определения вертикальных профилей и общего содержания атмосферного озона: Дис. канд. физ.-мат. наук /Ленингр. гос. ун-т. Л., 1984, 134 С.
2. *Зуев В.Е., Комаров В.С.* Статистические модели температуры и газовых компонент атмосферы. Л., 1986.
3. *Киселев В.Н., Кузнецов А.Д., Розанов В.В., Тимофеев Ю.М.* Математическое обеспечение автоматизированной обработки данных аэрологических наблюдений, выполненных с помощью зарубежных измерительных систем. Л., 1989.
4. *Козлов В.П.* Об одной задаче оптимального планирования статистического эксперимента // Теория вероятностей и ее применения. 1974. Т. 19. №1. С. 226–230.
5. *Козлов В. П.* Математические вопросы обращения радиационных данных // Инверсия Абеля и ее обобщения. Новосибирск, 1978. С. 68–95.
6. *Козлов В.П.* Избранные труды по теории планирования эксперимента и обратным задачам оптического зондирования. СПб., 2000.
7. *Кондратьев К.Я., Тимофеев Ю.М.* Термическое зондирование атмосферы со спутников. Л., 1970.

8. *Корн Г., Корн Т.* Справочник по математике для научных работников и инженеров. М., 1974.
9. *Линник Ю.В.* Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений. М., 1962.
10. *Малкевич М.С.* Оптические исследования атмосферы со спутников. М., 1973.
11. *Обухов А.М.* О статистически ортогональных разложениях эмпирических функций // Изв. АН СССР. Сер. Геофиз. 1959. № 3.
12. *Покровский О.М., Тимофеев Ю.М.* Общий статистический подход к решению обратных задач атмосферной оптики // Метеорология и гидрология. 1972. №1. С. 52–59.
13. *Поляков А.В., Мелас В.Б., Тимофеев Ю.М., Шульгина Е.М.* Оптимизация спектральных условий измерений в задаче микроволнового температурного зондирования атмосферы со спутников // Исследование Земли из Космоса. 1998. №1. С. 21–28.
14. *Тимофеев Ю.М., Васильев А.В.* Теоретические основы атмосферной оптики. СПб., 2000.
15. *Тихонов А.Н., Арсенин В.Я.* Методы решения некорректных задач. М., 1983.
16. *Турчин В.Ф., Козлов В.П., Малкевич М.С.* Использование методов математической статистики для решения некорректных задач // Успехи физических наук. 1970. Т. 102, вып. 3. С. 345–386.
17. *Cram L.E.* Inversion problem in radiative transfer theory: The Backus-Gilbert formalism // J.Q.S.R.T. 1978. Vol. 20. P. 305–315.

18. *Houghton J.T., Taylor F.W., Rodgers C.D.* Remote sounding of atmospheres. (Cambridge Planetary Science series-) Cambridge, 1984.
19. *Menke W.* Geophysical Data Analysis: Discrete Inverse Theory. New York, 1989.
20. *Rodgers C.D.* Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation // *Rev. Geophys. and Space Phys.* 1976. Vol. 14. P. 603–624.
21. *Rodgers C.D.* Characterization and error analysis of profiles retrieved from remote sounding measurements // *J. Geoph. Res.* 1990. Vol. 95. N D5. P. 5587–5595.
22. *Rodgers C.D.* Some theoretical aspects of remote sounding in the Earth's atmosphere // *J.Q.S.R.T.* 1971. Vol. 11. P. 767–777.
23. *Strand O.N., Westwater E.R.* Minimum RMS estimation of the numerical solution of a Fredholm integral equation of the first kind// *SIAM J. Numer. Anal.* 1968. Vol. 5. P. 287–295.
24. *Twomey S.* Introduction to the mathematics of inversion in remote sensing and indirect measurements. Oxford, 1977.
- 25.