

УСПЕХИ ФИЗИЧЕСКИХ НАУК

519.24/.27

ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ ДЛЯ РЕШЕНИЯ НЕКОРРЕКТНЫХ ЗАДАЧ

В. Ф. Турчин, В. П. Козлов, М. С. Малевич

ВВЕДЕНИЕ

Термин «некорректные задачи» объединяет в настоящее время широкий круг разнообразных и разнородных проблем, значительная часть которых выходит за рамки данного обзора. Оставляя подробное выяснение математического смысла термина до гл. 1, заметим, что слова «некорректные (неправильно поставленные) задачи» не должны пониматься слишком буквально, хотя эти задачи, в отличие от «корректных», весьма чувствительны к правильной постановке и потому часто действительно оказываются поставленными неправильно. Одним из классов некорректных задач, правильной постановке которых немало способствует их ясное физическое содержание, являются так называемые некорректные обратные задачи математической физики. По сложившейся терминологии *) к п р я м ы м задачам математической физики относят задачи, ориентированные по ходу причинно-следственной связи, т. е. задачи разыскания следствий заданных причин: определение полей во времени и в пространстве при заданных источниках, вычисление реакций прибора по известному сигналу на входе и т. п.

Обратными задачами в этом понимании оказываются задачи, связанные с обращением причинно-следственной связи, т. е. задачи отыскания неизвестных причин известных следствий: определения характеристик источников поля по значениям поля в некоторых точках или областях пространства, восстановление входного сигнала по реакции на выходе прибора и т. п. Обратные задачи возникают обычно как задачи интерпретации тех или иных наблюдений.

Собственно говоря, любая задача измерения тех или иных характеристик физического объекта есть обратная задача в смысле данного определения. Однако имеется тенденция использовать термин «обратная задача» для обозначения достаточно сложных задач интерпретации, когда либо одновременно и не независимым образом измеряют много параметров физического объекта, либо число этих параметров вообще неопределенно велико (если состояние объекта описывается, например, некоторой функцией координат). Строго говоря, некорректными могут быть только задачи второго типа, и именно им будет уделено основное внимание; однако при большом числе параметров и в задачах первого типа могут наблюдаться

*) Эта терминология берет начало, по-видимому, от работы А. Н. Тихонова ¹.

характерные особенности некорректных задач, связанные, как мы увидим ниже, с информационной недоопределенностью. Способы решения в условиях такой информационной недоопределенности, по существу, одинаковы для задач обоих типов.

Хотя нельзя утверждать, что все «функциональные» обратные задачи заведомо некорректны в математическом смысле, в большинстве случаев они все-таки оказываются некорректными. Таковы обратная задача теории потенциала ², ³, связанная с интерпретацией гравиметрических наблюдений в геологоразведке, обратная задача теплопроводности ⁴, класс обратных задач теории переноса излучения ⁶⁻⁹, ¹²⁻¹⁸, ⁴⁶⁻⁵², возникающих как задачи «зондирования» сред по оптическим характеристикам выходящего излучения, класс «инструментальных» обратных задач, возникающих при попытке исключить влияние измерительного прибора в оптической и рентгеновской спектроскопии, инструментальной оптике и радиоастрономии ⁵, ¹⁰, ¹¹, ¹⁹⁻³⁰, ³⁴, ³⁵ и т. д. Уже этот далеко не полный перечень показывает, что некорректные задачи глубоко проникли в физику и, что гораздо важнее, дальнейшее развитие теории и техники современного физического эксперимента невозможно без ясного понимания математической природы соответствующих обратных задач.

Повышенный интерес к некорректным измерительным задачам не в последнюю очередь вызван развитием вычислительной техники, в частности, ЭВМ, сделавших доступной математическую обработку больших объемов числового материала. При этом часто наблюдается известная переоценка возможностей такой обработки и, как следствие, излишний оптимизм по отношению к измерительным методам, основанным на решении некорректных обратных задач *). С другой стороны, трудности решения обратных задач, иногда кажущиеся **), могут тормозить разработку и внедрение весьма полезных методов измерения. Поэтому изложение и некоторое обобщение имеющегося к настоящему времени опыта решения некорректных обратных задач представляются по крайней мере своевременными.

Важной чертой измерительных задач, которые, по существу, только и будут рассматриваться в настоящем обзоре, является вероятностная природа величин, наблюдаемых в реальном эксперименте. В простейших случаях говорят о «случайных помехах», «шумах» измерительной аппаратуры, искажающих «полезный сигнал»: в более сложных случаях сам этот сигнал является случайным, как например, при измерении интенсивности потока квантов.

Эта стохастичность является неотъемлемой чертой всякого реального эксперимента и, естественно, должна явным образом входить в формулировку обратной задачи. Статистические подходы и методы решения некорректных обратных задач являются поэтому прямым следствием указанной стохастичности эксперимента.

Целью настоящего обзора является освещение развитых за последнее время методов решения некорректных обратных задач, существенно опирающихся на математическую статистику и теорию информации. Прежде чем приступить к выполнению этой задачи, обсудим более подробно понятие некорректности и дадим беглый обзор обычных (не статистических) методов решения некорректных задач и трудностей, которые при этом возникают.

*) Это относится, по-видимому, к некоторым обратным задачам спутниковой метеорологии (см., например, ¹⁵).

**) Например, отсутствие аналитического решения интегрального уравнения вида (1).

I. НЕКОРРЕКТНЫЕ ОБРАТНЫЕ ЗАДАЧИ И ТРУДНОСТИ ИХ РЕШЕНИЯ

Некорректность обратной задачи

Типичная и чаще всего встречающаяся в математической физике обратная задача — это решение интегрального уравнения Фредгольма 1-го рода:

$$\int_a^b K(x, y) \varphi(x) dx = f(y), \quad c \leq y \leq d, \quad (1)$$

где $K(x, y)$ — ядро уравнения — определяет оператор $\hat{K}$ прямой задачи, переводящей неизвестную функцию $\varphi(x)$, описывающую состояние объекта измерения, в некоторую другую функцию $f(y)$, доступную непосредственной регистрации и потому считающуюся известной. Иногда уравнение (1) принимает специальный вид «свертки»:

$$\int_{-\infty}^{+\infty} K(y-x) \varphi(x) dx = f(y), \quad (2)$$

если ядро зависит только от разности аргументов. Мы будем рассматривать как уравнение (1), так и уравнение (2), последнее — из-за простоты относящихся к нему результатов.

Решение уравнения (1) может не существовать вообще или существовать не для всякой правой части $f(y)$. Например *), если

$$K(y, x) = y + x, \quad (3)$$

правая часть $f(y)$ должна быть линейной функцией y . Если же решение и существует, оно не обязано быть единственным. В условиях приведенного примера, если $f(y) = f_0 + f_1 y$, любая функция $\varphi(x)$, удовлетворяющая двум условиям,

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

является решением уравнения (1) с ядром (3).

В задачах физического происхождения мы уверены обычно в существовании функции $\varphi(x)$, стоящей под интегралом в левой части уравнения (1). Единственность также часто может быть гарантирована **).

Однако даже при условии существования и единственности решения уравнение (1) (или (2)) может обладать одной специфической особенностью, которая и делает задачу некорректной. Эта особенность — «сглаживающее» действие ядра. Поясним это на примере.

Пусть ядро $K(y, x)$ как функция второго аргумента есть непрерывная функция. Рассмотрим в качестве решений две функции $\varphi(x)$: произвольную функцию $\varphi_1(x)$ и $\varphi_2(x) = \varphi_1(x) + C \sin 2\pi n x / (b-a)$, где n — достаточно большое целое число. Без дальнейших вычислений ясно, что при любом сколь угодно большом значении C можно подобрать столь большое число n , что разность соответствующих правых частей $f_1 = \hat{K}\varphi_1$ и $f_2 = \hat{K}\varphi_2$ будет (при данном значении аргумента y) по абсолютной величине меньше любого наперед заданного (сколь угодно малого) числа ε , т. е. ядро рассмотренного типа может «сгладить» даже очень интенсивную, но высоко-

*) Пример заимствован из работы ³⁸.

**) Как, например, для обратной задачи теории потенциала ³⁹, обратной задачи спектроскопии и инструментальной оптики ⁴⁰ и ряда других задач.

частотную составляющую искомой функции до весьма малого уровня (и тем меньшего, чем выше ее частота). Будь правая часть $f(y)$ нам известна точно, в этом не было бы большой беды — ведь единственность решения обеспечена. Однако наличие помех, сопровождающих регистрацию функции $f(y)$, катастрофическим образом меняет дело. Действительно, пусть по условиям эксперимента мы можем контролировать близость зарегистрированной функции $f^*(y)$ к точному образу $f(y) = K\hat{\varphi}$ неизвестной функции $\varphi(x)$ лишь с точностью до ошибки ϵ :

$$\max_{c \leq y \leq d} |f^*(y) - f(y)| \leq \epsilon. \quad (4)$$

Легко понять, что если ядро уравнения (1) обладает описанным выше сглаживающим действием, то всегда можно найти две функции $\varphi_1(x)$ и $\varphi_2(x) = \varphi_1(x) + C \sin 2\pi nx / (b - a)$, образы которых $f_1 = K\hat{\varphi}_1$ и $f_2 = K\hat{\varphi}_2$ одновременно удовлетворяют неравенству (4) с одной и той же функцией $f^*(y)$. Таким образом, уравнению (1) с взятой из эксперимента правой частью $f^*(y)$ удовлетворяют с точностью до ошибки ϵ по крайней мере две различные функции, а точнее, бесчисленная совокупность функций, среди которых имеются экземпляры, сколь угодно сильно различающиеся между собой. Понятно поэтому, что, взявшись решать уравнение (1) с приближенной правой частью $f^*(y)$, мы не имеем никаких шансов получить истинное решение $\varphi(x)$, характеризующее действительное состояние объекта исследования, и почти наверное придем к некоторому ложному решению, содержащему неопределенно большие, быстро осциллирующие компоненты.

Фактически к описанной ситуации и сводится некорректность задачи (1). Точное определение ^{39, 41, 42, 56} позволяет освободиться лишь от специального вида условия (4), допуская более общие меры уклонения.

Оценка различия между функциями по величине (4), свойственная классическому анализу, неудобна для приложений, так как регистрируемая функция $f^*(y)$ является обычно реализацией некоторого случайного процесса и подчинена лишь вероятностным ограничениям, так что указать ее максимальное уклонение от точного значения можно либо «с большим запасом», либо нельзя совсем. Поэтому более адекватным представляется статистический подход к определению корректности ⁴³, сводящийся к изучению статистических свойств решения задачи (1) в зависимости от характеристик случайного процесса, реализациями которого являются регистрируемые в эксперименте функции $f^*(y)$.

Особенности статистического подхода удобно продемонстрировать на примере из инструментальной спектроскопии — редукции спектра, записанного на приборе с конечным разрешением, к «идеальному прибору» (с бесконечно высоким разрешением) ²². Соответствующая обратная задача формулируется в виде уравнения (2) *, дополненного предположением о том, что регистрируемая функция $f^*(y)$ есть сумма точной правой части $f(y)$ и стационарного случайного процесса $\delta(y)$ (помехи) с нулевым средним и корреляционной функцией $\Delta(\eta) = \langle \delta(y) \delta(y + \eta) \rangle$. Формальное решение уравнения (2) легко получается с помощью преобразования Фурье:

$$\varphi(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ipx} \frac{\tilde{f}(p)}{\tilde{K}(p)} dp, \quad (5)$$

где знаком $\sim$ отмечено преобразование Фурье от соответствующей функ-

*) Ядро уравнения (2) в этом случае носит специальное название «аппаратная функция».

ции. Найдем дисперсию функции $\varphi(x)$, если вместо f в правую часть формулы (5) подставить $f^* = f + \delta$. Как показано Раутианом²², соответствующее выражение имеет вид

$$D(\varphi) = \langle \varphi^2 \rangle - \langle \varphi \rangle^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{G(p)}{|\tilde{K}(p)|^2} dp, \quad (6)$$

где $G(p) = \tilde{\Delta}(p)$ — спектр мощности процесса $\delta(y)$. Заметим теперь, что преобразование Фурье $\tilde{K}(p)$ аппаратной функции (передаточная функция) обладает тем свойством, что

$$|\tilde{K}(p)| \rightarrow 0 \text{ при } |p| \rightarrow \infty. \quad (7)$$

Поэтому для того, чтобы дисперсия решения задачи (2) оставалась конечной, спектр мощности шумов $G(p)$ должен достаточно быстро убывать при $|p| \rightarrow \infty$. Это накладывает жесткие ограничения на класс процессов $\delta(y)$, допустимых в качестве помех. На практике эти условия никогда не выполняются, так как помеха всегда содержит компоненту «белого шума», и, следовательно, при $|p| \rightarrow \infty$ спектр $G(p)$ стремится к конечному пределу. Поэтому дисперсия решения по формуле (6) получается бесконечной, а подстановка в формулу (5) вместо $\tilde{f}(p)$ найденной из эксперимента функции $\tilde{f}^*(p)$ невозможна. Источник этой трудности очевиден: высокочастотные компоненты $f^*(y)$, которые возникают из-за наличия помехи и которых не должно быть в истинной функции $f(y)$, делятся согласно формуле (5) на малые собственные числа $\tilde{K}(p)$ и дают в решении большие (а при $|p| \rightarrow \infty$ бесконечно большие) осцилляции.

Легко видеть, что условие (7) эквивалентно классической некорректности задачи (2). Полезно посмотреть, как должно было бы обстоять дело, чтобы задача (2) была корректной. Для этого условие (7) должно быть заменено на противоположное: $|\tilde{K}(p)| > \varepsilon > 0$ при $|p| \rightarrow \infty$, и конечность дисперсии $D(\varphi)$ обеспечена для гораздо более широкого класса процессов $\delta(y)$: достаточно, чтобы дисперсия самого процесса $\delta(y)$ была ограничена. В этом и состоит согласно статистическому подходу⁴³ различие между некорректными и корректными задачами: решение корректной задачи оказывается статистически устойчивым по отношению к более широкому классу случайных процессов, чем решение некорректной задачи. Этим, в частности, объясняется тот факт, что методы решения корректных задач могли разрабатываться без специального привлечения статистики помех — помехи всегда могли быть учтены на конечном этапе расчета как возмущения точного решения*). Для некорректных задач такой способ не годится; дело в том, что реальные помехи, как правило, не являются «допустимыми» для них, и дисперсия точного решения оказывается на самом деле бесконечной. т. е. формально точное решение типа (5) попросту не имеет смысла.

А л г е б р а и з а ц и я н е к о р р е к т н о й з а д а ч и и о б у с л о в л е н н о с т ь

При численном решении интегрального уравнения типа (1) его всегда тем или иным способом сводят к системе линейных алгебраических уравнений:

$$\sum_{i=1}^n K_{ji} \varphi_i = f_j, \quad j = 1, 2, \dots, n, \quad (8)$$

*) Примером корректной задачи является решение уравнения Фредгольма 2-го рода, отличающегося от (1) наличием в левой части неизвестной функции $\varphi(y)$ в виде дополнительного слагаемого.

где φ_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.

Может показаться, что сведение уравнения (1) к системе алгебраических уравнений (8) исчерпывает проблему его решения — достаточно только выбрать порядок приближения n настолько большим, чтобы получить любую требуемую точность. Однако, помня о некорректности исходной задачи (1), следует ожидать, что на этом пути возникнет какое-то препятствие. Таким препятствием является плохая обусловленность системы уравнений (8), т. е. чрезвычайно сильная зависимость решения от вариаций свободного члена, а также от ошибок коэффициентов (8) и от ошибок счета.

Как показано Фаддеевым^{44, 45}, обусловленность системы (8) тесно связана с набором собственных чисел μ_K матрицы K^*K (K — матрица системы (8), K^* — транспонированная матрица).

Обусловленность (устойчивость) падает с ростом отношения $\mu_{\max}/\mu_{\min}$. При этом решение «разбалтывается» в направлениях собственных векторов ψ_K матрицы K^*K , соответствующих малым собственным числам μ_K . Более точно, чувствительность проекции вектора решения φ на направление ψ_K к вариациям компонент вектора f и элементов матрицы K пропорциональна μ_K^{-1} . Поэтому, в частности, требования к точности вычислений быстро растут по мере роста отношения $\mu_{\max}/\mu_{\min}$ даже при точно известном векторе f .

Алгебраизация некорректной задачи всегда (при достаточно большом порядке приближения n) дает плохо обусловленную систему уравнений, ибо можно показать, что если исходная задача некорректна, то $\mu_{\max}/\mu_{\min} \rightarrow \infty$, когда $n \rightarrow \infty$. Следовательно, выбирая порядок приближения достаточно большим, можно сделать систему уравнений (8) сколь угодно плохо обусловленной.

П р и б л и ж е н н о е р е ш е н и е н е к о р р е к т н ы х о б р а т н ы х з а д а ч

Подходы к решению обратных задач менялись в зависимости от того, как проявлялась и насколько была понятной их «некорректная» природа. В этом разделе мы перечислим некоторые из этих подходов, сохраняя по возможности хронологическую последовательность.

Отметим прежде всего, что некорректные обратные задачи можно в известном смысле «решать», и весьма успешно, не только не применяя специально разработанные для этого в самое последнее время математические методы, но часто вообще не подозревая о некорректности и связанных с нею проблемах. Речь идет об извлечении физически достоверной информации о тех или иных характеристиках точного решения φ (или, в физических терминах, состоянии объекта исследования) без отыскания самого решения; при этом форма точного решения либо известна (с точностью до определяемых параметров), либо несущественна. Приведем пример. Наше уверенность в справедливости одной из важнейших физических теорий — квантовой механики — основана в значительной степени на интерпретации данных экспериментальной спектроскопии атомов и молекул, т. е., по существу, на решении обратной задачи (2). Но следует ли из математической некорректности задачи (2) некорректность квантовой механики как физической теории? Хотя ответ нетрудно предугадать, проследить соответствующую аргументацию*), может быть, полезно. Дело

*) Эта аргументация не исчерпывает, разумеется, проблемы обоснования квантовой механики как таковой.

в том, что экспериментальным основанием классической квантовой механики послужила совокупность данных о положениях и интенсивностях спектральных линий в эмиссионных спектрах атомов и простых молекул. Но, как хорошо известно (в общем случае это следует из вида уравнения (2)), интенсивность и з о л и р о в а н н о й спектральной линии может быть найдена по приборному спектру f независимо от вида истинного спектра Φ (формы линии) и даже от ядра уравнения (2) (о ядре нужно знать только нормирующий множитель $\tilde{K}(0)$ — чувствительность прибора). То же, при дополнительном предположении о симметрии формы линии и ядра $\hat{K}$, относится и к определению положения спектральной линии (частоты перехода). При этом, как показывает анализ (см., например, ²²), обе указанные характеристики определяются с конечной ошибкой при естественных предположениях о характере шума, так что некорректность задачи (2) не отражается на достоверности результатов, если соблюдаются известные предосторожности, известные из практики работы со спектральными приборами. Другие, далеко не столь тривиальные примеры подобной ситуации, доставляет спектроскопия светорассеивающих веществ. Здесь требуются зачастую немалые аналитические усилия ^{51, 52}, чтобы выделить ту схему эксперимента, в которой результаты непосредственных измерений, с одной стороны, позволяют определить оптические параметры исследуемого вещества, а с другой стороны, хотя бы приближенно, не зависят от точного вида (часто неизвестного) структуры светорассеивающей среды, определяющей совместно с измерительным прибором ядро, вообще говоря, некорректной обратной задачи, сформулированной относительно спектра поглощения исследуемого вещества *).

Прямая возможность обойти трудности, вызванные некорректностью задачи, состоит иногда в подходящей параметризации решения Φ , основанной на физической специфике конкретной задачи. Например, в разобранном выше примере изолированной спектральной линии мы можем знать (если наблюдается спектр вещества в газовой фазе) форму контура — допустим, доплеровского или лоренцовского, в зависимости от условий эксперимента. Тогда из функционального уравнения (2) можно получить алгебраическое уравнение для единственного неизвестного параметра — ширины линии **), решение которого может быть получено, например, методом наименьших квадратов. В общем случае подобные соображения приводят к системе связанных уравнений, которая решается обычным путем (если только не оказывается плохо обусловленной). Процедуры такого рода часто используются при обработке экспериментальных данных.

Если говорить о более общих методах решения, то, по существу, первая попытка решения некорректной задачи (2) была предпринята еще лордом Рэлеем в 1871 г., предложившим итерационный метод коррекции щелевых искажений в спектроскопии. С тех пор подобные методы предлагались неоднократно (см. по этому поводу обзор ²³). Как было показано Раутианом ²², все они эквивалентны некоторым итерационным методам решения уравнения (2). Дело в том, что в спектроскопии ядро уравнения (2) имеет характер очень узкого импульса, т. е. соответствующий оператор $\hat{K}$, будучи применен к достаточно гладкой функции, мало отличается от единичного оператора $\hat{I}$ (при подходящей нормировке). Поэтому для решения (2) можно попытаться использовать формальное

*) Вопрос о допустимой неопределенности ядра уравнения (1) представляет важную, но еще не решенную проблему ^{51, 52}. Все существующие методы решения уравнения (1) требуют точного знания ядра (по поводу чувствительности решения к ошибкам ядра см. ниже).

**) Интенсивность и положение, как указано выше, определяются независимо.

разложение обратного оператора задачи в ряд Неймана:

$$\hat{K}^{-1} = \sum_{n=0}^{\infty} (\hat{I} - \hat{K})^n, \quad (8a)$$

т. е. находить поправки к нулевому приближению $\varphi_0 = f$ путем последовательного применения оператора $\hat{I} - \hat{K}$. Начальные поправки действительно оказываются относительно малыми, если начинать с достаточно гладких функций f , но в дальнейшем быстро растут, обнаруживая все более быстрые осцилляции, причем этот рост обусловлен присутствующими в правой части (2) помехами *). Поэтому при применении итерационных алгоритмов исследователь вынужден был сам решать, когда следует прервать процесс итераций, руководствуясь теми или иными соображениями о реальном или шумовом происхождении возникающих после каждой новой итерации деталей решения.

Другой общий метод решения задачи типа (1)–(2) состоит попросту в решении соответствующей алгебраической системы (8). Если алгебраизация проведена удачно, то иногда можно получить приемлемое решение, ограничившись небольшим порядком приближения n , при котором неустойчивость еще не проявляется. Чаще всего алгебраизацию выполняют, разлагая $\varphi(x)$ по системе ортогональных функций (желательно, имеющих физический смысл) и оставляя первые n членов разложения. Этот метод, наряду с методом итераций, получил широкое распространение на практике **) в то время, когда в силу технической трудности вычислений и нельзя было использовать высокие порядки приближения, так что сам факт некорректности задачи типа (1) часто мог оставаться неизвестным авторам вычислений. Однако по мере развития вычислительной техники, когда стало возможным иметь дело с системами большой размерности (или итерациями высокого порядка), некорректность стала проявляться практически в виде «разбалтывания» решения по мере роста порядка приближения. Первоначальный подход в этом случае состоял в том, чтобы регулировать степень приближения в зависимости от характера получающегося решения, останавливаясь на том приближении, когда «разбалтывание» (оцениваемое из сравнения решения с ожидаемым видом неизвестной функции) еще не слишком велико. Удобно рассмотреть этот подход на примере уравнения (2), где имеется простая возможность регулировать степень приближения, пользуясь формальным решением (5).

Ограничиваясь восстановлением спектра сигнала $\tilde{\varphi}(p)$ в конечной полосе частот $|p| \leq p_0$, перепишем (5) в виде

$$\varphi^*(x) = \int_{-p_0}^{+p_0} e^{ipx} \frac{\tilde{\varphi}(p)}{\hat{K}(p)} dp, \quad (5a)$$

дающем приближенное решение уравнения (2). Дисперсия приближенного решения φ^* определяется соответственно «урезанным» интегралом (6) и остается конечной, если p_0 не слишком велика. Изменяя p_0 , можно менять соотношение между «точностью приближения» искомой функции и величиной ошибки; при этом повышение «точности» (за счет увеличения p_0) ведет автоматически к росту ошибки ²².

*) Ряд (8a) сходится, если только правая часть f совпадает с точным образом $\hat{K}\varphi$ истинного решения.

**) Нет никакой возможности, да и необходимости, перечислять относящиеся сюда работы.

Для выбора частоты среза p_0 привлекались различные соображения: исключение нулей передаточной функции $\tilde{K}(p)$ ²¹, ожидаемый вид иско- мой функции^{22, 34}, вид зарегистрированной функции^{25 *}).

Имеется еще один, несколько более общий способ регулирования степени приближения при использовании решения (5). Этот способ состоит в домножении подынтегрального выражения в (5) на некоторую, вообще говоря, произвольную функцию $\tilde{g}(p)$, достаточно быстро убывающую при $|p| \rightarrow \infty$, чтобы соответствующий интеграл (6) оставался конечным:

$$\varphi^*(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \tilde{g}(p) \frac{\tilde{f}(p)}{\tilde{K}(p)} e^{ipx} dp. \quad (56)$$

Заметим, что (56) переходит в (5а) при специальном выборе фактора $\tilde{g}$. Нетрудно понять, что использование (56) эквивалентно разысканию «сгла- женной 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 «заглаживания» можно регулировать, меняя параметры функ- ции $\tilde{g}(p)$ и даже сам вид этой функции. Однако такая крайняя свобода воздействия на решение уравнения (2) уже должна настораживать: ясно, что при субъективном выборе «сглаживающего фактора» и вообще способа приближенного решения чрезвычайно легко исказить результат. Что это так, показывают примеры неоднозначной интерпретации радиоастрономи- ческих наблюдений, приведенные в работе²¹.

Стремление избавиться от произвольных факторов и разработать общие методы решения некорректных задач, наряду с математическими исследованиями природы некорректности, привели в последнее десятиле- тие к разработке новых подходов и методов решения некорректных задач.

Фундаментальное значение для этих подходов имеют следующие соображения. Некорректную задачу можно рассматривать как эффектив- но недоопределенную. Действительно, при классическом понимании некорректности в качестве решения задачи (1) может фигурировать любая функция φ , удовлетворяющая условию

$$r_e(f^*, \hat{K}\varphi) \leq \varepsilon, \quad (9)$$

где $r_e(f^*, f)$ — мера уклонения регистрируемой функции f^* от точной правой части f , зависящая от условий эксперимента. Поскольку среди этих функций имеются и «плохие», возникает задача доопределения зада- чи (9) с тем, чтобы получить единственное решение, по возможности близкое к истинному.

Доопределить задачу (9) можно, вообще говоря, различными спосо- бами. Однако любой способ доопределения должен основываться на тех или иных представлениях о характере искомого решения, иначе говоря, на априорной информации о решении. Различные способы решения некорректных задач, включая рассматриваемые ниже статистические методы, отличаются, явно или неявно, видом используемой априорной информации.

) В работе²⁵ описана программа численного преобразования Фурье, в которой из интеграла (5) исключались те интервалы частот, где спектр реализации $f^(p)$ казался обусловленным одним шумом ($|f^*(p)| < kG(p)$, k — коэффициент запаса).

Первой работой данного направления явилась работа Филлипса⁵³, в которой из множества функций φ , удовлетворяющих условию (9), предлагалось выбрать «самую гладкую» функцию, а точнее, функцию, минимизирующую норму производной

$$\int_a^b \left(\frac{d\varphi}{dx} \right)^2 dx = \min. \quad (10)$$

Обычно удается показать, что искомый минимум достигается на границе области, определяемой неравенством (9); следовательно, мы можем заменить (9) на равенство

$$r_e(f^*, \hat{K}\varphi) = e. \quad (9a)$$

Возникает задача на условный экстремум. Решая ее по методу Лагранжа, получаем уравнение

$$r_e(f^*, \hat{K}\varphi) + \alpha \int_a^b \left(\frac{d\varphi}{dx} \right)^2 dx = \min, \quad (11)$$

где α — неопределенный множитель, который определяется из (9a). Для евклидовой метрики условие (11) приводит после алгебраизации к системе линейных уравнений. Применяя этот метод на практике, Филлипс отметил, что при определении параметра α из (9a) восстановленная функция получается слишком заглаженной. Однако можно подобрать такое значение α (в каждом конкретном случае), что в результате решение (11) получится гораздо более близкое к истинному решению. Таким образом, в методе Филлипса параметр α остался фактически неопределенным.

Для алгебраизированной системы (8) несколько более общая форма условия (10) была рассмотрена Туоми⁵⁴:

$$\Omega(\varphi) = \sum_{i,j}^n H_{ij} \varphi_i \varphi_j = \min, \quad (8a)$$

где матрица H положительно определена. В остальном подход Туоми совпадает с подходом Филлипса. Условие (11) принимает вид

$$K^* K \varphi + \alpha H \varphi = K^* f, \quad (11a)$$

множитель Лагранжа α определяется из соотношения

$$|K\varphi - f| = e.$$

Другой способ доопределения задачи (9) (см. библиографию в работе⁴²; там же исследованы относящиеся сюда математические вопросы) является в некотором смысле дополнительным к только что описанному. Мы рассмотрим его на примере уравнения (8). Квазирешением (терминология из⁴²) системы (8) назовем вектор φ , удовлетворяющий двум условиям:

$$\Omega(\varphi) = \sum_{i,j} H_{ij} \varphi_i \varphi_j \leq C, \quad (12)$$

$$|K\varphi - f| = \min, \quad (13)$$

где матрица H , как и выше, положительно определена. Если известно, что минимум достигается на границе области, определяемой неравенством (12), задача (12)–(13) сводится к уравнению (11a), но множитель Лагранжа будет определяться из соотношения (12), в котором нужно оставить знак равенства.

Условие (12) выделяет некоторую ограниченную замкнутую область в качестве множества допустимых решений и эквивалентно наличию априорной информации о принадлежности искомого решения данному замкнутому множеству. В работе ⁵⁵ использованы более слабые ограничения, вытекающие из физического смысла решения задачи (1) относительно плотности распределения частиц по размерам (см. ⁴⁶), а именно, неотрицательность плотности. Условие (12) заменяется в этом случае на

$$\sum_{i=1}^n H_{ij} \varphi_i \geq 0 \quad \text{при всех } j, \quad (12a)$$

и задача решается методами линейного программирования.

А. Н. Тихонов ⁵⁶⁻⁵⁸ ввел понятие регуляризации решения некорректной задачи. Под этим понимается построение семейства корректных задач, зависящего от параметра регуляризации α , которое обладает тем свойством, что при $\alpha \rightarrow 0$ и при одновременном стремлении к нулю погрешности правой части решение корректной задачи стремится к истинному решению исходной некорректной задачи. Уравнение (11) было постулировано и изучено Тихоновым независимо от Филлипса как регуляризованное уравнение с параметром регуляризации α . При практическом применении метод Тихонова совпадает с методом Филлипса с той, быть может, разницей, что неопределенность параметра α естественно вытекает из самой идеи регуляризации, в то время как при подходе Филлипса она является неприятностью, до известной степени дискредитирующей идею метода (об этом см. ниже). Впрочем, следует отметить, что в ряде работ ⁵⁹⁻⁶¹, развивающих идеи Тихонова, предлагались алгоритмы определения α , однако их вряд ли можно признать обоснованными, так как остается открытым вопрос об источнике или хотя бы точном виде априорной информации.

Как уже отмечалось, описанные выше способы решения некорректных задач исходят из предположения, что эксперимент позволяет контролировать точную верхнюю границу ошибки $|f^*(y) - f(y)|$. Последовательное 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, оно не отвечает природе реального эксперимента и не позволяет аккуратно оценить ошибку восстановления. Поэтому представляется более естественным рассматривать проблему решения некорректной задачи с учетом статистического характера ошибок эксперимента и с использованием других типов априорной информации, в том числе статистической.

II. ИНФОРМАЦИЯ, ПОЛУЧАЕМАЯ ИЗ ЭКСПЕРИМЕНТА, И РЕШЕНИЕ ОБРАТНОЙ ЗАДАЧИ

Разрешающая способность и информационная метрика

Анализ «информирующих возможностей» эксперимента естественно начать с простейшей задачи, где эти возможности вообще проявляются, а именно, с задачи различения двух близких состояний объекта измерения. Постановка подобной задачи в инструментальной оптике и спектроскопии с целью оценки предельных возможностей оптических приборов — проблема разрешающей способности — восходит к Рэлею (обзор дальнейших работ см. в ²²). В основе подхода Рэлея лежит идея сравнения (с помощью

данного оптического или спектрального прибора) двух стандартных «объектов»; например, в случае спектроскопа — спектрального дублета и одиночной линии. При статистическом подходе этому соответствует формулировка задачи в терминах теории статистических решений^{31 *}).

Пусть $\varphi_1(x)$ и $\varphi_2(x)$ — две функции, описывающие фиксированные состояния объекта измерения, и последний находится в одном из этих состояний. В результате эксперимента над объектом мы получаем наблюдаемую функцию $f^*(y) = f(y) + \delta(y)$, где $f(y)$ — функция, связанная с истинным состоянием $\varphi(x)$ уравнением (1), а $\delta(y)$ — случайная функция (шум). Следуя³¹, будем считать $\delta(y)$ нормальным стационарным случайным процессом с нулевым средним и спектром мощности $G(p)$. Задача состоит в том, чтобы, наблюдая $f^*(y)$, решить, в каком из состояний φ_1 или φ_2 , объект измерения находится в действительности. Будем характеризовать различимость состояний максимальной вероятностью правильных решений P_r при условии, что сравниваемые состояния априори равновероятны и используется оптимальная решающая процедура. Величина P_r принимает значения между $1/2$ и 1: нижний предел отвечает абсолютно неразличимым состояниям, верхний — абсолютно различимым. Можно показать, что P_r зависит от данных задачи (ядра $K(y, x)$, функций φ_1 и φ_2 и статистики помех) только через величину ρ , которая и характеризует непосредственно степень различимости состояний:

$$P_r = 1/2 + 1/2\Phi(\rho/2), \quad (14)$$

$$\rho^2 \equiv \|\varphi_1 - \varphi_2\|^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{|\hat{K}(\varphi_1 - \varphi_2)|^2}{G(p)} dp, \quad (15)$$

где $\Phi(x)$ — интеграл вероятности, а знак $\sim$ означает преобразование Фурье от выражения, стоящего под знаком модуля. Выражение (15) получается в предположении, что вблизи краев интервала наблюдения (c, d) функция $\hat{K}\varphi$ обращается в нуль при любом φ и интервал корреляции шума много меньше длины интервала (c, d) ³². Величина ρ , определяемая по формуле (15) как мера различимости состояний φ_1 и φ_2 с помощью «прибора», описываемого оператором $\hat{K}$ и спектром мощности шума $G(p)$, обладает весьма важными свойствами. Прежде всего, случай $\rho(\varphi_1 - \varphi_2) = 0$ соответствует полной неразличимости (эквивалентности) состояний в данном эксперименте. Второй предельный случай $\rho = \infty$ соответствовал бы «бесконечно большой» (абсолютной) различимости, чего на практике обычно не бывает. Удобно принять $\rho = 1$ в качестве пороговой величины, определяющей предел различимости³³). Величину ρ естественно назвать информационным расстоянием в пространстве состояний изучаемого объекта. Действительно, можно показать, что ρ обладает всеми свойствами геометрического расстояния — неотрицательностью, симметричностью относительно сравниваемых состояний и справедливостью «неравенства треугольника»: $\rho(\varphi_1 - \varphi_2) \leq \rho(\varphi_1 - \varphi) + \rho(\varphi - \varphi_2)$. Расстояние ρ характеризует содержащуюся в результатах эксперимента информацию о состояниях φ_1 и φ_2 с точки зрения их различимости³⁴).

*) В теории оптических приборов аналогичный подход был независимо использован в работах³² и³³.

**) Подставляя $\rho = 1$ в (14), получим, что вероятность правильной идентификации состояния φ в этом случае всего на 13,7% превышает вероятность угадывания (50%) и быстро убывает при дальнейшем уменьшении ρ .

***) О количественном подходе к информационной емкости множества в пространстве сигналов см. в⁶².

Неопределенность решения обратной задачи

Как следует из результатов предыдущего раздела, для каждого «истинного решения» φ_0 задачи (1), характеризующего действительное состояние объекта измерения, существует бесконечно много функций φ , соответствующих состояниям, не отличимым от действительного по результатам реального (статистического) эксперимента, а именно таких, которые удовлетворяют неравенству

$$\|\varphi - \varphi_0\| \leq 1. \quad (16)$$

Другими словами, функции φ заполняют некоторую сферу единичного радиуса в функциональном пространстве решений задачи (1), которую естественно назвать «сферой неопределенности». Действительно, любая из функций φ , удовлетворяющая неравенству (16), может быть принята в качестве решения обратной задачи (1), так как эксперимент не дает информации, достаточной, чтобы отличить эту функцию от любой другой, также удовлетворяющей неравенству (16) *). По смыслу метрики (15) сфера (16) характеризует максимальную точность, которая может быть вообще достигнута при решении задачи (1). Поэтому, в частности, всякая попытка дальше «уточнить» решение чисто математическими средствами, без привлечения дополнительной информации, аналогична попытке создания «информационного вечного двигателя», производящего информацию из ничего.

Однако практическая ценность метода решения уравнения (1), приводящего к произвольной функции из сферы (16), была бы невелика: если только исходная задача некорректна, указанная сфера содержит, главным образом, функции φ , которым не отвечают никакие физические состояния измеряемого объекта. Для выделения «физической части» сферы неопределенности необходимо указать какие-то признаки, отличающие «физические» решения от «нефизических» или, как говорят обычно, нужна априорная информация о возможных состояниях объекта измерения или, что то же, о функциях, допустимых в качестве решения уравнения (1). Оставляя более подробное обсуждение вопросов использования априорной информации при решении обратных задач до следующего параграфа, рассмотрим случай отсутствия всякой априорной информации³⁷.

Ясно, конечно, что решать в этом случае каким-либо методом непосредственно уравнение (1) совершенно не имеет смысла. Однако иногда представляет интерес не столько само решение уравнения (1), сколько значения тех или иных функционалов от этого решения, например, значение полной интенсивности спектральной линии безотносительно к форме ее спектрального контура. Способ анализа таких задач с использованием метрики (15) был предложен в работе³⁷ применительно к классу линейных функционалов от искомого решения.

Ограничимся, следуя³⁷, случаем разностного ядра (2), для которого все результаты записываются особенно просто. Формула (15) переписется в этом случае в виде

$$\rho^2 = \|\varphi_1 - \varphi_2\|^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{|\tilde{K}(p)|^2 |\tilde{\varphi}_1(p) - \tilde{\varphi}_2(p)|^2}{G(p)} dp, \quad (15a)$$

*) Строго говоря, чтобы это утверждение было совсем точным, необходимо в правой части (16) заменить 1 на $1/2$, что принципиально не меняет дела: нам удобнее для дальнейшего сохранить единицу.

где, как и выше, $\tilde{K}(p)$ есть преобразование Фурье от ядра $K(x)$. Оценим предельную точность определения линейного функционала

$$A = A_\varphi = \int_{-\infty}^{+\infty} a(x) \varphi(x) dx, \quad (17)$$

где $a(x)$ — весовая функция, определяющая функционал. Поместим сферу неопределенности в начало координат, положив в условии (16) $\varphi_0(x) \equiv 0$. Подставляя функции $\varphi(x)$, содержащиеся в этой сфере, в формулу (17), мы будем получать значения величины A , которые, по определению сферы неопределенности, практически следует отождествить со значением $A_{\varphi_0} = 0$, соответствующим центру сферы. Другими словами, если мы обозначим через $\|A\|$ максимальное значение A на сфере (16), то все значения A , удовлетворяющие неравенству $|A| \leq \|A\|$, не могут быть с помощью данного прибора отличны от нуля, т. е. $\|A\|$ представляет ошибку определения величины A по данным эксперимента*). Так как расстояние (15а) зависит только от разности сравниваемых функций, сфера (16) при изменении своего центра смещается как целое. Отсюда (и из линейности функционала A) следует независимость ошибки от истинного спектра.

Далее, можно показать, что величина $\|A\|$ может быть вычислена по формуле

$$\|A\|^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{|\tilde{a}(p)|^2 G(p)}{|\tilde{K}(p)|^2} dp, \quad (18)$$

где $\tilde{a}(p)$ — преобразование Фурье от весовой функции функционала. Величина $\|A\|^2$ фактически представляет дисперсию ошибки при вычислении функционала A , обусловленной шумами эксперимента.

Следует обратить внимание на то, что правая часть (18) остается конечной только для достаточно узкого класса функционалов. Если для какого-нибудь функционала интеграл (18) расходится, это означает невозможность определения этого функционала в данном эксперименте без дополнительных сведений о неизвестной функции $\varphi(x)$. Структура выражения (10) показывает, что класс допустимых функционалов ($\|A\| < \infty$) состоит из функционалов с весовыми функциями $a(x)$, существенно более гладкими, чем ядро уравнения (2), если только спектр мощности шумов $G(p)$ не убывает патологически быстро при $|p| \rightarrow \infty$. В частности, заведомо не являясь допустимыми функционалами значения функции φ в заданных точках.

III. ЖЕСТКИЕ АПРИОРНЫЕ ОГРАНИЧЕНИЯ

Для большей наглядности мы будем рассматривать алгебраизированный вариант обратной задачи (8), предполагая, что порядок n достаточно велик, чтобы обеспечить требуемую точность приближения.

Наложить жесткие априорные ограничения на искомую функцию $\varphi(x)$ (а в алгебраизованном варианте — на вектор φ в n -мерном пространстве R^n) — это значит выделить в пространстве R^n некую область D и искать решения лишь в этой области, объявив все решения вне области D бессмысленными, «нефизическими». Легко видеть, что вопрос об информационной обусловленности задачи (8) сводится к вопросу о взаимном расположении области D и сферы неопределенности (16) или, что то же самое, к вопросу о геометрии и размерах области D в метрике (15).

*) Характеристики прибора и помех входят через определение метрики (15а).

Систему (8) следует считать информационно недоопределенной, если сфера (16) с центром в любой точке $\varphi_0 \in D$ содержит точки, не принадлежащие области D : Действительно, в тех направлениях, в которых сфера (16) «вылезает» из области D , эксперимент не дает возможности уточнить значение параметра состояния по сравнению с априорным диапазоном. Более того, формальное решение системы (8), например, по способу наименьших квадратов, даст погрешность координаты в этих направлениях, во много раз превосходящую этот диапазон. Это означает, в сущности, что экспериментальные данные не содержат информации о соответствующем параметре состояния. Наличие таких «не обеспеченных» информацией параметров среди неизвестных системы (8) и составляет смысл данного выше определения. Понятие информационно недоопределенной системы может быть более удобно выражено в терминах геометрических характеристик области D в метрике (8). А именно, назовем область D вырожденной в одном или нескольких направлениях, если максимальный диаметр области D в этих направлениях не превосходит единицы. Тогда, очевидно: 1) вырождение области D эквивалентно недоопределенности системы (8), 2) система (8) (соответствующий эксперимент) не позволяет определить параметры, соответствующие направлениям вырождения.

Геометрическая формулировка особенно ясно показывает, что свойство недоопределенности не зависит от выбора базиса при алгебраизации исходной задачи (1) *), т. е. от выбора координат в пространстве R^n . Это свойство определяется только самим экспериментом и областью D .

Вернемся теперь к задаче определения «физического» решения системы (8). Ясно, что для этого необходимо заранее исключить вырожденные направления, так как именно они и являются источником «нефизичности». Пусть R' есть максимальное подпространство пространства решений R^n , не содержащее направлений вырождения области D , — разрешенное подпространство. Тогда решением плохо обусловленной системы ($r < n$) разумно считать проекцию неизвестного вектора φ на подпространство R' .

Для случая, когда область D задана своими центральными моментами второго порядка — тензором инерции C ,

$$C^{ik} = \langle (\varphi^i - \langle \varphi^i \rangle) (\varphi^k - \langle \varphi^k \rangle) \rangle,$$

в работе ⁶³ предложен метод последовательного «исчерпывания» главных направлений, приводящий к фактическому построению разрешенного подпространства задачи. В вычислительном отношении метод сводится к проблеме собственных значений оператора GC :

$$\sum_{i,k=1}^n g_{ik} C^{il} u^k = \lambda u^l, \quad (19)$$

где g_{ik} — элементы информационной матрицы Фишера ⁶⁴. Пусть $\lambda_1 > \lambda_2 > \dots > \lambda_n$ — собственные числа задачи (19), а $u_1, u_2, \dots, u_n$ — соответствующие им собственные векторы. Можно показать, что каждое собственное число λ_k численно равно отношению среднего квадрата (дисперсии) априори допустимой вариации компоненты вектора φ в направлении u_k к дисперсии ошибки определения этой компоненты из системы (8). Ясно, что при $\lambda_k < 1$ соответствующее направление u_k следует считать вырожденным. Таким образом, разрешенное подпространство R' должно содержать только направления u_k , для которых $\lambda_k > 1$, и все их линейные комбинации. Другими словами, первые собственные векторы $u_1, \dots, u_r$ образуют базис разрешенного подпространства R' ; при этом

*) При условии, что базис позволяет описать все точки области D , т. е. имеет достаточную размерность.

размерность этого подпространства равна числу собственных чисел λ_k , превышающих единицу. Искомое решение задачи (1) выражается при этом в виде

$$\varphi^*(x) = \varphi_0(x) + \sum_{k=1}^r \varphi'^k \eta'_k(x), \quad (20)$$

где базис $\{\eta'_k\}$ строится из базиса $\{\eta_i\}$, использованного при получении системы (8), с помощью собственных векторов:

$$\eta'_k(x) = \sum_{i=1}^n u_k^i \eta_i(x),$$

а $\varphi_0(x) = \sum_{i=1}^n \langle \varphi^i \rangle \eta_i(x)$ есть решение, соответствующее центру области D .

Явное решение задачи дается формулами, выражающими коэффициенты разложения (20) через измеренные значения

$$\varphi'^k = \sum_{i=1}^n \sum_{j=1}^m f_j k_{ji} u_k^i / \sum_{i=1}^n \sum_{j=1}^m k_{ji} k_{ji} u_k^i u_k^i. \quad (21)$$

Решение, получаемое по формулам (20), (21), удовлетворяет, насколько это возможно, требованиям, которые естественно предъявить к решению некорректной задачи при жестких априорных ограничениях. Операция перехода $f \rightarrow \varphi$ «почти» сохраняет информацию о различии любых векторов из R^n , так как если сравниваемые векторы лежат в R^r , то расстояние ρ между ними не уменьшается; если же они имеют составляющие в дополнительном подпространстве R^{n-r} , то уменьшение имеет порядок $\sqrt{\lambda_{r+1}} < 1$, т. е. отождествляются состояния, практически не отличимые в данном эксперименте. Кроме того, решение φ^* , будучи случайным, почти наверное принадлежит «физической» области D , так как его математическое ожидание, взятое по распределению помех, совпадает с истинным решением, принадлежащим D по определению, а помехи, действующие в невырожденных направлениях, не могут вывести решение за пределы, существенно превышающие априорный диапазон.

IV. ВЕРОЯТНОСТНЫЕ ОГРАНИЧЕНИЯ И МЕТОД СТАТИСТИЧЕСКОЙ РЕГУЛЯРИЗАЦИИ

Метод статистической регуляризации был развит в работах ⁶⁵⁻⁶⁸. Он состоит в том, что априорная информация об искомой функции вносится путем задания некоторого распределения вероятностей. Это приводит к замене точного решения уравнения на некоторое приближенное, «регуляризованное» решение. Существуют различные способы задания априорного распределения вероятностей, поэтому существуют и различные варианты метода статистической регуляризации, которые будут кратко описаны ниже. Но сначала мы изложим принципиальные положения, на которых основан этот метод. Хотя по сравнению с введением жестких ограничений на искомую функцию введение ограничений вероятностным образом требует в ряде случаев большей априорной информации, оно тем не менее обладает следующими преимуществами.

Во-первых, задача решения некорректного уравнения выступает в приложениях как задача обработки экспериментальных данных, для которой введение вероятностных понятий является неизбежным, ибо ошибка правой части носит случайный характер и может быть охарактеризована лишь вероятностным образом. Поэтому вероятностный способ задания априорной информации приводит к использованию единообраз-

ного аппарата и является более естественным. Это особенно ясно видно, когда встает вопрос об ошибке восстановления.

Во-вторых, вероятностный способ позволяет более полно использовать предыдущий опыт, учитывая его в априорном распределении.

В-третьих, когда такого опыта нет, вероятностный способ позволяет сформулировать весьма слабые предположения об искомой функции, принципиально невыразимые на языке теории множеств (принадлежность к «слоистому ансамблю», о котором см. ниже).

Как и в предыдущем разделе, мы предположим, что наше уравнение уже алгебраизировано, т. е. сведено к системе уравнений

$$\sum_{i=1}^n k_{ji} \varphi_i = f_j, \quad j = 1, 2, \dots, m. \quad (22)$$

В отличие от уравнений (8), мы допускаем, что число уравнений m не равно числу неизвестных n . Наличие или отсутствие точного решения уравнения (22) никакой роли играть не будет. Величины φ_i могут быть любыми линейными функционалами от $\varphi(x)$, однако для метода статистической регуляризации всего естественнее выбирать в качестве φ_i значения $\varphi(x)$ в опорных точках, и это будет подразумеваться в дальнейшем для удобства изложения. Векторы φ и $\mathbf{f}$ в пространствах R^n и R^m соответственно мы часто будем называть просто функциями φ и f .

Чтобы изложить сущность метода статистической регуляризации, необходимо сформулировать нашу задачу как задачу математической статистики и ввести соответствующие понятия.

Стратегии и оценки

Пусть измеряются m величин f_j , которые, если бы не ошибки измерения, были связаны с n неизвестными величинами φ_i соотношениями (22). Благодаря ошибкам измерения величины f_j , полученные в результате измерений, отличаются от их идеальных значений, даваемых уравнениями (22). Чтобы не вводить новых обозначений, мы будем отныне через f_j обозначать не идеальные, а реальные значения величины f_j . Тогда уравнения (22) будут соблюдаться лишь приближенно, а точные уравнения будут иметь вид

$$\sum_{i=1}^n k_{ji} \varphi_i + \delta_j = f_j, \quad j = 1, 2, \dots, m,$$

где δ_j — ошибки измерения величин f_j , образующие случайный вектор δ в пространстве R^m . Обозначим через $P_\delta(\delta)$ плотность вероятности этого вектора. Очевидно, величины f_j являются также случайными и зависят как от истинных значений φ_i , так и от ошибок δ_j . Условная плотность вероятности вектора $\mathbf{f}$ при заданном векторе φ есть

$$P(\mathbf{f}|\varphi) = P_\delta(\mathbf{f} - \hat{K}\varphi). \quad (23)$$

(Заметим, что функция $P(\mathbf{f}|\varphi)$ $m+n$ переменных f_j и φ_i полностью характеризует экспериментальное устройство, т. е. как связь между φ и $\mathbf{f}$, так и статистические свойства ошибок измерения.) Теперь мы можем сформулировать задачу следующим образом. Неизвестные величины φ_i характеризуют состояние природы. Некий случайный процесс дает нам величины f_j , причем условная плотность вероятности $P(\mathbf{f}|\varphi)$ известна. Что можно по величинам f_j сказать о величинах φ_i ? Ответ на подобного рода вопросы и есть основная задача математической статистики.

Вопрос «Что можно сказать о φ_i ?» недостаточно точен. Когда мы пытаемся его уточнить, то в первую очередь напрашивается такая формулировка: сколь вероятны те или иные значения величины φ_i ? Иначе говоря, какова плотность вероятности $P(\varphi | f)$ вектора φ при условии, что в результате измерений получен вектор f ?

Чтобы ответить на этот вопрос, необходимо ввести какую-то априорную плотность вероятности $P(\varphi)$, т. е. предположить, что опыт по измерению f является одним из серии подобных опытов, которые производятся при различных состояниях природы φ , выбираемых случайно в соответствии с плотностью вероятности $P(\varphi)$. Тогда апостериорная вероятность определится по формуле Байеса:

$$P(\varphi | f) = \frac{P(\varphi) P(f | \varphi)}{\int P(\varphi) P(f | \varphi) d\varphi}. \quad (24)$$

Без введения априорной вероятности понятие апостериорной вероятности $P(\varphi | f)$ теряет всякий смысл.

Если мы не требуем столь детальной информации о функции как апостериорное распределение вероятности для нее, то мы не обязаны вводить и априорного распределения вероятности. Однако («не в дверь, так в окно») априорные распределения появляются и при самой общей постановке вопроса, при которой математическая статистика рассматривается как теория принятия решений в условиях неполного знания состояния природы. Для чего нам вообще нужна какая бы то ни было информация? Очевидно, для того, чтобы принимать решения о выполнении тех или иных действий. Если бы мы знали истинное состояние природы φ , мы приняли бы какое-то решение. Но мы не знаем φ , а знаем лишь f . Математическая статистика должна дать нам рекомендации, как действовать в таких случаях; она должна указать стратегию принятия решения при неизвестном φ , но известном f . Чтобы избежать обсуждения специфических для каждой конкретной проблемы особенностей, статистик может дать такую рекомендацию: «Вот вам вектор φ^0 , и действуйте так, как если бы $\varphi = \varphi^0$ ». Такое значение φ^0 называется оценкой φ , а алгоритм вычисления φ^0 по f и его дальнейшего использования вместо истинного φ называется оценочной стратегией.

Применяя одну и ту же стратегию, мы в зависимости от случая получаем лучший или худший результат, т. е. оценка φ^0 в большей или меньшей степени приближается к истинному φ . Естественно поставить вопрос об отыскании такой стратегии, которая в среднем даст наилучший результат. Однако для того чтобы определить понятие «в среднем», необходимо ввести априорное распределение вероятностей для неизвестного $P(\varphi)$. Поэтому наилучшая в среднем стратегия (такая стратегия называется байесовой) зависит от априорного распределения вероятностей $P(\varphi)$. Чтобы вычислить байесову оценку любой величины, в частности, оценку φ^0 вектора φ , надо по формуле Байеса (24) вычислить апостериорное распределение $P(\varphi | f)$ и усреднить данную величину по этому распределению. В частности,

$$\varphi_i^0 = \int \varphi_i P(\varphi | f) d\varphi, \quad i = 1, 2, \dots, n. \quad (25)$$

Помимо байесовых стратегий, в математической статистике рассматриваются еще так называемые минимаксные стратегии, которые минимизируют «вред» от замены истинных величин на их оценки в наименее благоприятном случае. Для физических приложений эти стратегии менее естественны. Кроме того, байесовы стратегии в совокупности обладают своего рода «полнотой». Назовем разумной такую стратегию, для кото-

рой не существует другой стратегии, которая является лучшей всегда, т. е. независимо от исхода каких бы то ни было случайных событий. Можно доказать, что каждая разумная стратегия является байесовой стратегией, соответствующей некоторому априорному распределению вероятности. Следовательно, изучая байесовы стратегии, мы изучаем все разумные стратегии.

Из формулы Байеса (24) мы видим, что апостериорная вероятность есть произведение априорной вероятности $P(\varphi)$ и условной вероятности $P(f|\varphi)$ (знаменатель не зависит от φ и вводится для нормировки). Поэтому апостериорная информация о φ есть сумма информации априорной и извлеченной из опыта. Если один из сомножителей гораздо информативнее другого, то апостериорная вероятность близка к первому сомножителю и почти не зависит от второго. Этот случай встречается очень часто, причем информативным фактором является, разумеется, условная вероятность $P(f|\varphi)$, иначе незачем было бы ставить эксперимент. Пусть, например, $P(f|\varphi)$ как функция φ имеет резкий максимум в некоторой области пространства V_0 и быстро стремится к нулю за пределами этой области, а $P(\varphi)$ слабо меняется в этой области. Тогда от конкретного вида $P(\varphi)$ апостериорная вероятность $P(\varphi|f)$ практически не зависит, поэтому мы можем положить $P(\varphi) = \text{const}$ в любой области V_1 , включающей V_0 .

Итак, каждая разумная стратегия принятия решений в условиях неопределенности является байесовой стратегией, соответствующей некоторому априорному распределению вероятностей. Когда мы не имеем никакой априорной информации, мы принимаем решения, которые кажутся нам в этой ситуации естественными и, возможно, наилучшими. В действительности они являются наилучшими для мира, в котором все мыслимые возможности (описанные определенным способом) являются равновероятными. Мы увидим ниже, что обычное, прямое решение уравнений (22) без применения методов регуляризации вместе с обычной оценкой ошибки по методам классической теории ошибок есть байесова стратегия для априорной плотности вероятности $P(\varphi) = \text{const}$. 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.

Прежде чем переходить к рассмотрению различных способов внесения априорной информации, мы конкретизируем функцию плотности вероятности ошибок. Сделаем обычное и, как правило, вполне оправданное допущение, что ошибки δ_j при разных j независимы и распределены по нормальному закону с математическим ожиданием, равным нулю. Среднеквадратичную ошибку измерения величины δ_j обозначим через s_j . Тогда условная вероятность (23) примет вид

$$P(f|\varphi) = \prod_{j=1}^m (2\pi s_j^2)^{-1/2} \exp \left\{ -\frac{1}{2s_j^2} \left[f_j - \sum_{i=1}^n k_{ji}\varphi_i \right]^2 \right\}. \quad (26)$$

Это выражение можно значительно упростить, если ввести пропорциональные величинам f_j величины g_j , измеряемые с равной абсолютной точностью,

$$g_j = \frac{s}{s_j} f_j,$$

и соответствующую матрицу L с матричными элементами

$$L_{ji} = \frac{s}{s_j} k_{ji}.$$

Величину s (ошибку измерения g_j) определим так, чтобы преобразование $f \rightarrow g$ было унитарным:

$$s^m = \prod_{j=1}^m s_j.$$

Таким образом, s есть среднее геометрическое всех ошибок измерения f_j . Уравнения (22) переходят в

$$L\varphi = g, \quad (27)$$

однако мы по-прежнему будем рассматривать в качестве основных физических переменных величины f_j , а к вектору g будем относиться, как к вспомогательной конструкции.

Условную вероятность (26) мы можем теперь записать в виде

$$P(f|\varphi) = (2\pi s^2)^{-m/2} \exp \left\{ -\frac{1}{2s^2} |g - L\varphi|^2 \right\}. \quad (28)$$

Решение без регуляризации

Положим $P(\varphi) = 1/V$ в весьма большом объеме V (который мы в дальнейшем устремим к бесконечности, если это окажется возможным) и $P(\varphi) = 0$ вне этого объема. Тогда в объеме V апостериорная вероятность $P(\varphi|f)$ будет с точностью до константы совпадать с $P(f|\varphi)$, даваемым выражением (28).

С каждым распределением вероятности можно связать бесконечное множество, называемое статистическим ансамблем, элементы которого получаются путем независимых случайных выборок, подчиненных данному распределению. Использование понятия статистического ансамбля часто способствует краткости и наглядности языка, апеллируя к привычным теоретико-множественным представлениям. Апостериорное распределение $P(\varphi|f)$ для случая, когда $P(\varphi) = \text{const}$, определяет ансамбль векторов (функций) φ , который мы будем называть полным ансамблем нерегуляризованных решений и обозначать через $P_\Pi(\varphi)$ (мы идентифицируем ансамбли, указывая соответствующие плотности вероятности φ). Итак,

$$P_\Pi(\varphi) = c_1 \exp \left\{ -\frac{1}{2s^2} |g - L\varphi|^2 \right\}. \quad (29)$$

Компоненты φ_i^0 нерегуляризованного решения суть средние от φ_i по этому ансамблю, их погрешности — корни из дисперсий φ_i .

Раскрывая квадрат модуля вектора в (29) и включая в константу множитель, не зависящий от φ , получаем

$$P_\Pi(\varphi) = c_2 \exp \left\{ -\frac{1}{2s^2} [(\varphi, L^*L\varphi) - 2(L^*g, \varphi)] \right\}. \quad (30)$$

Ансамбль $P_\Pi(\varphi)$ можно представлять себе как множество точек в пространстве R^n , плотность которых пропорциональна $P_\Pi(\varphi)$. Чтобы уяснить характер этого множества, введем в пространстве R^n новые координатные оси, а именно, ортонормированную систему собственных векторов ψ^k матрицы L^*L . Так как L^*L — симметричная, неотрицательно определенная матрица, все векторы ψ^k — действительные, и им соответствуют неотрицательные собственные числа, которые мы обозначим через λ_k^2 .

Обозначим через $\tilde{\varphi}_k$ проекцию φ на ψ^k . Величины $\tilde{\varphi}_k$ можно принять в качестве новых координат вектора φ . В новых переменных плотность

распределения полного ансамбля решений имеет вид

$$P_{\Pi}(\varphi) = c_2 \exp \left\{ - \sum_{k=1}^n \frac{1}{2s^2} (\lambda_k^2 \tilde{\varphi}_k^2 - 2\tilde{h}_k \tilde{\varphi}_k) \right\}, \quad (30a)$$

где $\tilde{h}_k$ — проекция на ψ^k вектора L^*g .

Таким образом, те величины $\tilde{\varphi}_k$, для которых $\lambda_k \neq 0$, являются статистически независимыми и распределены по нормальному закону с математическим ожиданием $\tilde{h}_k/\lambda_k^2$ и дисперсией s^2/λ_k^2 . Следовательно, область, в которой в основном сосредоточены точки ансамбля $P_{\Pi}(\varphi)$, имеет в направлении ψ^k размеры порядка s/λ_k . Те величины $\tilde{\varphi}_k$, для которых $\lambda_k = 0$, фактически не входят в выражение (30a) (легко показать, что если $\lambda_k^2 = 0$, то и $\tilde{h}_k = 0$). При движении в направлении вектора ψ^k функция $P_{\Pi}(\varphi)$ не меняется. Из этого уже видно, что если хотя бы одно из собственных значений λ_k обращается в нуль, мы не можем положить $V \rightarrow \infty$, ибо интеграл $\int P_{\Pi}(\varphi) d\varphi$ в этом случае не существует. Информация о $\tilde{\varphi}_k$ в нашем уравнении полностью отсутствует, и мы обязаны наложить на $\tilde{\varphi}_k$ какие-то априорные ограничения.

Если все λ_k отличны от нуля (а для этого необходимо и достаточно, чтобы $\det(L^*L)$ был отличен от нуля), то можно считать, что $V < \infty$. Бейесова стратегия для этого случая сводится к методу наименьших квадратов. Бейесова оценка φ^0 находится из уравнения

$$L^*L\varphi^0 = L^*g,$$

которое для случая $m=n$ совпадает с исходным уравнением (27), а ошибка восстановления σ_i в точке i может быть найдена по формуле

$$\sigma_i^2 = s^2 ((L^*L)^{-1})_{ii}, \quad i = 1, 2, \dots, n. \quad (31)$$

Для характеристики ошибки решения φ^0 в целом мы можем вычислить среднеквадратичную по всем i ошибку σ :

$$\frac{1}{n} \sum_{i=1}^n \sigma_i^2 = \frac{s^2}{n} \text{Sp}((L^*L)^{-1}) = \frac{s^2}{n} \sum_{k=1}^n \frac{1}{\lambda_k^2}$$

(мы воспользовались инвариантностью следа матрицы при преобразовании координат).

Таким образом, сильное возрастание ошибки нерегуляризованного решения имеет место тогда, когда хотя бы некоторые собственные значения λ_k матрицы L^*L весьма малы, и происходит оно вследствие неопределенности соответствующих ψ -компонент.

Если матрица K имеет хотя бы пару строк, близких друг к другу с точностью до общего множителя, или вообще ее строки почти линейно зависимы, т. е., говоря качественно, «похожи друг на друга», то этим же свойством будет обладать и матрица L , а среди собственных значений матрицы L^*L появятся малые числа. Тогда нерегуляризованное решение оказывается бесполезным, так как оно содержит ошибку, во много раз большую, чем сама восстановленная функция. Именно такие матрицы — с почти линейно зависимыми строками — и получаются при алгебраизации некорректной задачи. Впрочем, это может случиться и при корректном исходном уравнении.

Р е ш е н и е в а н с а м б л е, з а д а н н о м к о н е ч н о й в ы б о р к о й

Допустим, что априорный ансамбль возможных решений охарактеризован путем задания N его векторов Φ^v , которые, следовательно, можно считать полученными случайной выборкой, подчиненной априорному распределению вероятности $P(\Phi)$. Этот случай может возникнуть, например, при систематическом прямом измерении величин Φ_i в некоторых стабильных условиях с целью накопления статистического материала, который может быть использован для надежного косвенного определения Φ_i по измеренным в тех же условиях величинам f_j , когда прямые измерения отсутствуют.

Если выборка достаточно представительна, то регуляризованное решение и его ошибку можно получить как байесовское решение и ошибку, при вычислении которых усреднение по априорному распределению заменяется (приближенно) на усреднение по выборке.

Пусть задана некая функция $F(\Phi)$ величин Φ_i . Среднее от этой функции по апостериорному распределению (24) можно представить в виде

$$\langle F(\Phi) \rangle = \int F(\Phi) P(\Phi | f) d\Phi = \frac{\langle F(\Phi) P(f | \Phi) \rangle_{\text{ap}}}{\langle P(f | \Phi) \rangle_{\text{ap}}},$$

где $\langle G(\Phi) \rangle_{\text{ap}}$ означает среднее от функции $G(\Phi)$ по априорному распределению $P(\Phi)$.

Заменим среднее по ансамблю на среднее по конечной выборке:

$$\langle G(\Phi) \rangle_{\text{ap}} \approx \frac{1}{N} \sum_{v=1}^N G(\Phi^v).$$

Тогда для $\langle F(\Phi) \rangle$ получаем

$$F(\Phi) = \sum_{v=1}^N w_v F(\Phi^v) / \sum_{v=1}^N w_v. \quad (32)$$

где

$$w_v = \exp \left\{ - \sum_{j=1}^m \frac{1}{2s_j^2} \left[f_j - \sum_{i=1}^n k_{ji} \Phi_i^v \right]^2 \right\} \quad (33)$$

— зависящие от f весовые множители для различных векторов Φ^v , пропорциональные условным вероятностям $P(f | \Phi^v)$.

Положив $F(\Phi) = \Phi_i$ и $F(\Phi) = (\Phi_i - \langle \Phi_i \rangle)^2$, получим по формулам (32) и (33) регуляризованное решение и его ошибку.

Р е ш е н и е в а н с а м б л е, з а д а н н о м к о р р е л я ц и о н н о й м а т р и ц е й

Положим, что математическое ожидание вектора Φ по априорному ансамблю $P(\Phi)$ обращается в нуль. Это предположение будет подразумеваться также всюду в дальнейшем. (Ясно, что оно не ограничивает общности рассмотрения, так как всегда можно сделать параллельный перенос.) Допустим также, что тем или иным способом нам стала известна корреляционная матрица априорного ансамбля:

$$C_{ij} = \langle \Phi_i \Phi_j \rangle_{\text{ap}} = \int \Phi_i \Phi_j P(\Phi) d\Phi. \quad (34)$$

Чтобы при определении априорной вероятности $P(\Phi)$ внести минимум информации кроме той, которая содержится в соотношениях (34), мы из

всех функций $P(\varphi)$, удовлетворяющих условию (34), выберем ту, которая содержит минимум информации о φ , т. е. минимизирует функционал

$$I[P(\varphi)] = \int \ln P(\varphi) P(\varphi) d\varphi, \quad (35)$$

дающий (с точностью до константы) количественную меру этой информации.

Можно показать, что такая плотность распределения имеет вид

$$P(\varphi) = c_3 \exp \left\{ -\frac{1}{2} (\varphi, C^{-1}\varphi) \right\}. \quad (36)$$

Корреляционная матрица C — симметричная положительно определенная матрица. Следовательно, существует ортонормированная система собственных векторов χ^l , $l = 1, \dots, n$, матрицы C с положительными собственными числами, которые мы обозначим через γ_l^2 . (Случай, когда какие-либо $\gamma_l = 0$, может быть рассмотрен как предельный.) Обозначим через φ_l проекцию вектора φ на ось χ^l . В координатах φ_l плотность распределения $P(\varphi)$ примет вид

$$P(\varphi) = c_3 \exp \left\{ -\sum_{l=1}^n (\varphi_l^2 / 2\gamma_l^2) \right\}.$$

Следовательно, компоненты φ_l разложения φ по собственным векторам корреляционной матрицы статистически независимы и распределены нормально со среднеквадратичным значением γ_l .

Используя (36) и (28), получаем по формуле Бейеса (24) апостериорное распределение

$$P(\varphi | f) = c_4 \exp \left\{ -\frac{1}{2s^2} [(\varphi, (L^*L + s^2C^{-1})\varphi) - 2(L^*g, \varphi)] \right\}, \quad (37)$$

которое снова оказывается нормальным. Согласно (25) регуляризованное решение $\varphi^C = \langle \varphi \rangle$ есть решение уравнения

$$(L^*L + s^2C^{-1})\varphi^C = L^*g. \quad (38)$$

Средний квадрат ошибки восстановления есть

$$\sigma_i^2 = s^2 ((L^*L + s^2C^{-1})^{-1})_{ii}. \quad (39)$$

Метод решения обратной задачи по формулам (38) и (39) был предложен также в работах ^{83, 84}, исходя из близких соображений.

Статистический ансамбль с плотностью вероятности, равной произведению плотностей вероятности двух других ансамблей, можно назвать пересечением этих ансамблей. Такое пересечение имеет много общего с теоретико-множественным пересечением. Если перемножаемые ансамбли таковы, что описываются плотностью вероятности типа «все или ничего», т. е. принимающей постоянное значение в некоторой области пространства, обращаясь в нуль вне этой области, то их пересечение также есть ансамбль такого типа и занимаемая им область есть пересечение областей, занимаемых исходными ансамблями. В нашем случае ансамбли описываются нормальным распределением, которое теоретически продолжается до бесконечности, однако за пределами определенной области оно убывает так быстро, что ситуация почти такова, как при ансамблях указанного выше типа.

Апостериорный ансамбль (37) есть пересечение полного ансамбля решений (29) с априорным ансамблем (36); следовательно, в теоретико-множественной интерпретации он состоит из векторов, которые являются членами обоих ансамблей. Выберем определенный уровень значимости, например 0,99, и будем представлять ансамбль областью (телом)

в пространстве R^n , ограниченной поверхностью $P(\varphi) = \text{const}$, такой, что интеграл от $P(\varphi)$ по этой области равен 0,99. Вследствие нормальности распределений все области будут ограничены эллипсоидами (иногда вырождающимися). Главные оси эллипсоида, изображающего полный ансамбль решений, направлены по векторам ψ^k , главные оси эллипсоида, соответствующего априорному ансамблю, — по векторам χ^l . Вообще говоря, эти оси не совпадают.

Исследуем картину пересечения ансамблей при ошибке измерения s , стремящейся к нулю.

Если все собственные значения λ_k отличны от нуля, то при $s \rightarrow 0$ полный ансамбль решений стягивается в точку, соответствующую точному решению уравнения (8). Следовательно, в эту же точку стягивается и апостериорный ансамбль регуляризованных решений (независимо от вида $P(\varphi)$, если только $P(\varphi^a)$, где φ^a — точное решение, не обращается в нуль). Таким образом, регуляризованное решение в этом случае стремится к точному, а его ошибка стремится к нулю.

Если одно или несколько собственных значений λ_k равны нулю, то при $s \rightarrow 0$ полный ансамбль решений стремится к гиперплоскости, параллельной соответствующим векторам ψ^k . Пересечение этой гиперплоскости с эллипсоидом априорного ансамбля определяет некоторую вырожденную эллиптическую область в пространстве R^n , которая и характеризует апостериорный ансамбль. Центр этой области есть регуляризованное решение, ее размеры определяют ошибку восстановления, которая, таким образом, при $s \rightarrow 0$ стремится к конечному пределу, отличному от нуля.

Исследование пересечения ансамблей в пространстве R^n показывает, что информативность эксперимента зависит от относительной ориентации главных осей эллипсоидов полного ансамбля решений и априорного ансамбля. Если оси первого эллипсоида, соответствующие максимальным значениям λ_k , являются для второго эллипсоида «дальними» направлениями, т. е. такими, что расстояние от центра до границы в этих направлениях близко к максимальному, то эксперимент информативен. Если эти оси направлены по «ближним» направлениям априорного эллипсоида, эксперимент мало информативен. В частности, можно представить себе такой случай, когда при данном априорном распределении эксперимент вообще не дает никакой информации. Пусть γ_l обращается в нуль для некоторой группы индексов l ; соответствующие векторы χ^l определяют подпространство в пространстве R^n . И пусть матрица L^*L такова, что все ее собственные векторы ψ^k , для которых $\lambda_k \neq 0$, лежат в этом подпространстве. Легко видеть, что в этом случае апостериорное распределение совпадает с априорным. Из эксперимента можно определить лишь те проекции вектора φ , которые равны нулю априори.

Анализ регуляризованного решения весьма упрощается, если предположить, что матрицы L^*L и C могут быть одновременно приведены к диагональному виду (такой случай имеет место, например, при трансляционной симметричности задачи в целом). Обозначим, для определенности, общие собственные векторы через ψ^k и перенумеруем их так, чтобы λ_k не возрастали. Тогда из (38) и (39) находим в ψ -представлении

$$\tilde{\varphi}_k^C = \frac{h_k}{\lambda_k^2 + s^2/\gamma_k^2}, \quad (40)$$

$$\sigma_k^2 = \frac{s^2}{\lambda_k^2 + s^2/\gamma_k^2}. \quad (41)$$

Регуляризованное решение (40) мы можем следующим образом связать с точным решением φ (если оно существует):

$$\tilde{\varphi}_k^C = z_k \tilde{\varphi}_k, \quad (40a)$$

где

$$z_k = \frac{\lambda_k^2 \gamma_k^2}{\lambda_k^2 \gamma_k^2 + s^2} \quad (406)$$

— «обрезающий фактор», значение которого зависит от соотношения между $\lambda_k^2 \gamma_k^2$ — средним квадратом k -й компоненты в правой части и s^2 — средним квадратом ошибки измерения одной компоненты. Если $\lambda_k^2 \gamma_k^2 \gg s^2$, т. е. полезный сигнал много больше шума, то $z_k \approx 1$ и в регуляризованном решении k -я компонента практически такая же, как в нерегуляризованном (точном). Если $\lambda_k^2 \gamma_k^2 \ll s^2$ (сигнал много меньше шума), то $z_k \ll 1$ и k -я гармоника в решении, отражающая исключительно ошибки измерения, подавляется. Сравнивая (41) с (31), мы видим, что σ_k^2 для регуляризованного решения также находится в отношении z_k к σ_k^2 для нерегуляризованного решения. При $\lambda_k^2 \gamma_k^2 \gg s^2$ ошибка практически такая же, как и без регуляризации; при $\lambda_k^2 \gamma_k^2 \ll s^2$ из (41) находим $\sigma_k^2 \approx \gamma_k^2$; неопределенность k -й компоненты равна ее неопределенности в априорном ансамбле. Итак, роль регуляризации заключается в том, что ложную информацию, которую дает нерегуляризованное решение о k -компонентах с малым λ_k , она заменяет на информацию об этих компонентах, содержащуюся в априорном распределении.

Неравенство $\lambda_k^2 \gamma_k^2 \ll s^2$ является критерием, выполнение которого дает основания считать, что λ_k практически равно нулю и условная вероятность $P(\mathbf{f} | \Phi)$ от $\tilde{\varphi}_k$ не зависит. В спектре матрицы L^*L может оказаться большой разрыв, так что при некотором k_0 собственное значение $\lambda_{k_0}^2$ в сто или в тысячу раз больше, чем следующее значение $\lambda_{k_0+1}^2$. Тогда при некотором s возникнет такая ситуация, что для $k \leq k_0$ обрезающий фактор z_k можно считать равным единице, а для $k > k_0$ — нулю, и эта ситуация будет сохраняться при уменьшении s во много раз (пока $k_0 + 1$ -я компонента не станет информативной). При решении регуляризованного уравнения это приведет к тому странному на первый взгляд явлению, что при многократном уменьшении ошибки s не меняется ни регуляризованное решение, ни его погрешность.

Решение в ансамбле ограниченных или гладких функций

Допустим, нам известно, что искомая функция $\varphi(x)$ «более или менее гладкая». Уточнить эту информацию можно, введя какой-либо функционал, характеризующий степень гладкости функции, например, норму ее q -й производной (значение q мы не конкретизируем):

$$\Omega[\varphi(x)] = \int \left[\frac{d^q \varphi(x)}{dx^q} \right]^2 dx, \quad (42)$$

и указав ожидаемое примерное значение этого функционала:

$$\Omega[\varphi(x)] \approx \omega.$$

Точно так же можно использовать сведения об ограниченности функции $\varphi(x)$. Для этого достаточно в формуле (42) положить $q = 0$. Можно сформировать функционал Ω как линейную комбинацию норм нескольких производных. Во всех этих случаях после выполнения алгебраизации априорная информация предстанет в виде

$$(\varphi, \Omega\varphi) = \sum_{i,j=1}^n \varphi_i \Omega_{ij} \varphi_j \approx \omega,$$

где Ω — симметричная неотрицательно определенная матрица, являющаяся конечно-разностным эквивалентом соответствующего функционала.

Очевидно, мы должны ввести такое априорное распределение $P(\varphi)$, что

$$\int (\varphi, \Omega \varphi) P(\varphi) d\varphi = \omega. \quad (43)$$

Чтобы внести минимум произвола, мы, как и в предыдущем разделе, из всех $P(\varphi)$, удовлетворяющих условию (43), выберем $P(\varphi)$ с минимальной информацией (35). Такой функцией $P(\varphi)$ будет

$$P_\alpha(\varphi) = c_\alpha \exp \left\{ -\frac{\alpha}{2} (\varphi, \Omega \varphi) \right\}, \quad (44)$$

где

$$\alpha = n/\omega. \quad (45)$$

Таким образом, мы пришли к случаю, рассмотренному в предыдущем разделе, с корреляционной матрицей $C = (\alpha \Omega)^{-1}$. Все результаты, полученные выше, сохраняют свою силу. В частности, регуляризованное решение φ^α подчиняется уравнению (38):

$$(L^*L + s^2 \alpha \Omega) \varphi^\alpha = L^*g. \quad (46)$$

Это уравнение, как уже указывалось, впервые было выведено Филлипсом⁵³. Независимо от Филлипса, оно было постулировано Тихоновым⁵⁶ как корректное уравнение, приближенно представляющее некорректное уравнение (27).

Решая регуляризованное уравнение (46), следует иметь в виду, что матрица Ω может иметь собственные числа, равные нулю. Если, например, Ω описывает норму первой производной, то для вектора φ , изображающего постоянную функцию, т. е. такого, что все его компоненты φ_i одинаковы, $(\varphi, \Omega \varphi) = 0$. Для матрицы Ω , дающей норму второй производной, существуют два линейно независимых вектора, обладающих этим свойством, и т. д. Это значит, что соответствующее априорное распределение уходит вдоль этих векторов в бесконечность с ненулевым значением («бесконечно дальние» направления). (Распределение (44) надо в этом случае понимать как определенное в некотором большом объеме V .) Если бесконечно дальние направления априорного ансамбля являются также бесконечно дальними или просто очень дальними направлениями полного ансамбля решений, то регуляризованного решения, соответственно, не существует или оно обладает очень большой ошибкой. В этом случае необходимо изменить Ω (например, включить в функционал норму самой функции), чтобы ограничить плотность вероятности $P(\varphi)$ по бесконечно дальним направлениям. Однако, как правило, в этом нет необходимости, ибо функционал Ω , изображающий норму производной, используется тогда, когда для регуляризации достаточно информации о гладкости функции, и следовательно, относительно сложения с постоянной, линейной и т. п. функциями решение достаточно устойчиво. На языке ансамблей это означает, что бесконечно дальние направления априорного ансамбля являются близкими направлениями полного ансамбля решений. Поэтому мы можем считать, что априорное распределение (44) отлично от нуля лишь в некотором объеме V , а после умножения на полный ансамбль решений можем устремить V к бесконечности.

Решение в наиболее узком допустимом ансамбле

Вспомним снова о функциях плотности вероятности типа «все или ничего». Если и априорный ансамбль, и полный ансамбль решений описываются плотностью вероятности такого типа, то может случиться, что соответствующие им области в пространстве R^n не имеют общих точек,

и следовательно, ансамбля, являющегося пересечением этих двух ансамблей, не существует. Такая ситуация указывает на противоречие между информацией, извлекаемой из опыта, и априорной информацией, и мы должны констатировать, что в данном априорном ансамбле решения проблемы (т. е. решения уравнения (8) с указанной точностью) не существует.

Если ансамбли описываются нормальными распределениями, то их пересечение существует всегда, однако интуитивно ясно, что если центральные области ансамблей, где сосредоточено, скажем, 99% или 99,9% всех точек, не пересекаются, а пересечение происходит за счет хвостов нормального распределения, то такое пересечение является чем-то вроде побочного эффекта и не имеет физического смысла. Поэтому необходим критерий, позволяющий отделить те случаи, когда апостериорный ансамбль действительно является решением проблемы, от тех случаев, когда он представляет «игру хвостов» нормальных распределений. Этот критерий (см. ⁶⁵) естественно основать на оценке той погрешности, с которой решения, входящие в апостериорный ансамбль, удовлетворяют уравнению (8) или (27) (последнее предпочтительнее, так как в полном ансамбле решений погрешность g_i не зависит от i). Выражение

$$\frac{1}{n} (g - L\varphi)^2$$

дает средний (по номеру компоненты i) квадрат погрешности, с которой выполняется уравнение (27). Усредняя его по некоторому ансамблю, мы получим средний квадрат погрешности для всех функций (векторов) ансамбля. Очевидно, что если для усреднения мы возьмем полный ансамбль решений (29), то получим величину s^2 по ее определению:

$$s^2 = \left\langle \frac{1}{n} (g - L\varphi)^2 \right\rangle_{\Pi}.$$

Если же мы будем усреднять по апостериорному ансамблю, то получим, вообще говоря, другую величину, которую обозначим через s'^2 :

$$s'^2 = \left\langle \frac{1}{n} (g - L\varphi)^2 \right\rangle_{\text{апост}} = \frac{1}{n} \int (g - L\varphi)^2 P(\varphi | f) d\varphi. \quad (47)$$

Теперь критерий, о котором говорилось выше, мы можем сформулировать в виде неравенства $s'^2 \leq s^2$. Оно имеет следующий смысл. Апостериорный ансамбль мы получаем путем отбора из полного ансамбля решений тех функций φ , которые принадлежат к априорному ансамблю. Если в результате этого отбора среднеквадратичное отклонение $L\varphi$ от g не увеличилось, т. е. $s'^2 \leq s^2$, то это значит, что принадлежность к априорному ансамблю не мешает функции φ быть решением уравнения (27). Апостериорный ансамбль в этом случае удовлетворяет предъявляемым к нему требованиям и является наиболее полным решением проблемы. Если же $s'^2 > s^2$, то это значит, что принадлежность функции φ к априорному ансамблю несовместима (статистически) с соблюдением уравнения (27) с точностью до s . Апостериорный ансамбль в этом случае состоит из функций φ , которые (в среднем) не удовлетворяют с нужной точностью исходному уравнению; следовательно, решения задачи в данном априорном ансамбле не существует.

Если на основании предыдущего опыта нам хорошо известна корреляционная матрица C , то вряд ли есть необходимость подвергать решение, полученное согласно (38), проверке критерием $s'^2 \leq s^2$. Действительно, в этом случае мы не сомневаемся, что истинная функция φ принадлежит к выбранному нами априорному ансамблю, а если это так, то $s'^2 \approx s^2$ (см. ⁶⁵). (Отметим, что неравенство $s'^2 \leq s^2$ надо понимать в соответствии с духом статистики как $s'^2 < s^2$ или $s'^2 \approx s^2$.) То же относится и к

решению в ансамбле гладких (или ограниченных) функций, когда величина ω , а следовательно, и α , нам достаточно надежно известна. Но часто бывает, что надежных данных о величине ω нет. Тогда можно поставить задачу об отыскании наибольшего α , при котором еще существует решение (см. ⁶⁶). Рассмотрим семейство априорных ансамблей (44) с произвольным α и назовем допустимым такой ансамбль, что $s'^2 \leq s^2$. При $\alpha = 0$ ансамбль (44) покрывает все функции φ , приписывая им равные вероятности. С возрастанием α ансамбль сужается (здесь нет необходимости давать строгое определение этого понятия, а интуитивно оно очевидно), оставляя в своем составе все более и более гладкие функции, и при $\alpha = \infty$ вырождается в ансамбль, содержащий лишь функцию, тождественно равную нулю.

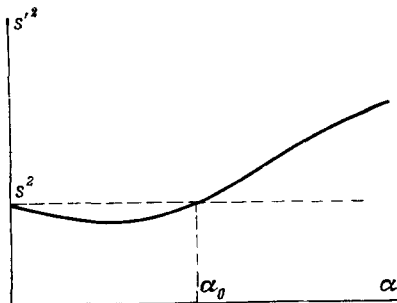


Рис. 1. Зависимость s'^2 от α .

В работе ⁶⁶ исследована зависимость s'^2 от α для трансляционно-симметричного случая и показано, что она имеет вид, изображенный на рис. 1. Точка $\alpha = \alpha_0$ характеризует наиболее узкий допустимый ансамбль гладких функций. Она может быть найдена путем численного расчета кривой $s'^2(\alpha)$.

Итак, если параметр α , характеризующий степень гладкости искомой функции, неизвестен, то можно определить значение α , соответствующее

наиболее «гладкому» допустимому ансамблю, и найти решение в этом ансамбле. Таким образом, и найденное решение должно быть в каком-то смысле наиболее гладким, поэтому интересно сравнить этот подход с подходом Филлипса ⁵³, который также заключается в нахождении наиболее гладкого решения, приближенно удовлетворяющего исходному уравнению. В соответствии с подходом Филлипса параметр α должен был бы определяться из уравнения

$$\frac{1}{n} (g - L\varphi^\alpha)^2 = s^2, \quad (48)$$

где φ^α — регуляризованное решение. Однако, как уже говорилось, функция, полученная при таком алгоритме определения α , оказывается слишком заглаженной, и лучшие результаты получаются, если брать α , которые приводят к меньшей, чем s^2 , правой части в уравнении (48). При статистическом подходе параметр α определяется из условия

$$s'^2 = \frac{1}{n} (g - L\varphi^\alpha)^2 + \frac{1}{n} \langle (L(\varphi - \varphi^\alpha))^2 \rangle = s^2. \quad (49)$$

Уравнение (49) отличается от (48) тем, что к левой части прибавлена величина, зависящая от дисперсии по апостериорному ансамблю, что соответствует уравнению (48) с уменьшенной правой частью; поэтому, пользуясь условием (49), мы получаем менее заглаженное решение. В работе ⁶⁶ показано, что этот алгоритм определения параметра α приводит к весьма удовлетворительным результатам.

Решение в слоистом ансамбле

В работе ⁶⁷ предложен другой метод регуляризации, дающий не самое гладкое (в каком бы то ни было смысле) решение, а решение наиболее вероятной гладкости.

Пусть нам известно, что искомая функция является в какой-то степени гладкой и могла бы быть получена в результате выборки из какого-то

ансамбля гладких функций $P_\alpha(\varphi)$; однако параметр α , характеризующий этот ансамбль и, следовательно, гладкость искомой функции, нам не известен. Значит ли это, что регуляризация невозможна и лучшее, что мы можем сделать, — это использовать нерегуляризованное решение?

Мы видели, что нерегуляризованное решение соответствует предположению $P(\varphi) = \text{const}$. Но неопределенность величины α отнюдь не влечет предположения $P(\varphi) = \text{const}$, ибо это предположение означает полное отсутствие корреляций между значениями φ_i . В то же время, если φ изображает непрерывную физическую функцию, то между значениями φ_i с близкими i должна существовать корреляция, хотя степень коррелированности (зависящая от параметра α) нам не известна и может быть произвольной. Такое предположение относительно φ можно описать априорной вероятностью

$$P(\varphi) = \int_0^\infty P(\alpha) P_\alpha(\varphi) d\alpha, \quad (50)$$

где $P(\alpha)$ — априорная вероятность α , которую мы можем считать принимающей постоянное значение в любой сколь угодно большой области положительных α .

Ансамбль с плотностью вероятности (50) можно назвать «слоистым», понимая под слоями ансамбля гладких функций с различными α . Слоистый ансамбль (50) может быть использован для выражения априорной информации о φ в тех случаях, когда известна физическая природа φ , но неизвестен параметр, характеризующий основной физический фактор. Пусть, например, φ представляет зависимость координаты от времени при движении точечной массы m в одномерном поле силы F . Следовательно, функция $\varphi(x)$ (x — время) есть решение уравнения Ньютона, и вторая производная $\varphi''(x)$ имеет порядок величины F/m . Мы можем ничего не знать о величинах m и F в нашей физической системе и тем не менее утверждать, что φ подчиняется априорному распределению (50), где $P_\alpha(\varphi)$ дается формулой (44), а матрица Ω представляет норму второй производной функции $\varphi(x)$ (в формуле (42) $q = 2$). Этой информации оказывается достаточно для нахождения регуляризованного решения и его ошибки.

Используя (50) в качестве априорного распределения, получаем по формуле Бейеса (24) апостериорное распределение

$$P(\varphi | \mathbf{f}) = \frac{\int P(\alpha) P_\alpha(\varphi) P(\mathbf{f} | \varphi) d\alpha}{\int P(\alpha) P_\alpha(\varphi) P(\mathbf{f} | \varphi) d\alpha d\varphi}. \quad (51)$$

Бейесова оценка любой функции $F(\varphi)$ есть среднее от $F(\varphi)$ по этому распределению:

$$\langle F(\varphi) \rangle = \int P(\varphi | \mathbf{f}) F(\varphi) d\varphi.$$

Подставим перед собой задачу выразить $\langle F(\varphi) \rangle$ через интеграл по α от бейесовой оценки той же функции в случае определенного α :

$$\langle F(\varphi) \rangle_\alpha = \int P(\varphi | \mathbf{f}, \alpha) F(\varphi) d\varphi, \quad (52)$$

где через $P(\varphi | \mathbf{f}, \alpha)$ мы обозначили апостериорное распределение при априорном распределении $P_\alpha(\varphi)$ (оно получается по формуле (37) заменой C^{-1} на $\alpha\Omega$).

Вводя распределения

$$P(f|\alpha) = \int P(f|\varphi) P_\alpha(\varphi) d\varphi, \quad (53)$$

$$P(\alpha|f) = \frac{P(\alpha) P(f|\alpha)}{\int P(\alpha) P(f|\alpha) d\alpha}, \quad (54)$$

находим

$$\langle F(\varphi) \rangle = \int \langle F(\varphi) \rangle_\alpha P(\alpha|f) d\alpha. \quad (55)$$

Таким образом, $\langle F(\varphi) \rangle$ получается путем дополнительного усреднения $\langle F(\varphi) \rangle_\alpha$ по α с весовой функцией $P(\alpha|f)$, т. е. апостериорной вероятностью параметра гладкости α при известном результате измерений f .

Используя предыдущие результаты для $\langle F(\varphi) \rangle_\alpha$, можно вычислить $\langle F(\varphi) \rangle$ путем численного интегрирования по α в (55). Однако если измерение f дает достаточно большое количество информации о параметре α , то функция $P(\alpha|f)$ будет иметь резкий максимум при некотором значении α_0 . Тогда вместо усреднения по α можно воспользоваться формулами для определенного α , положив (апостериори) $\alpha = \alpha_0$. Значение α_0 может быть найдено путем численного анализа кривой $P(\alpha|f)$.

Величина α_0 есть, по существу, статистическая оценка параметра α по методу наибольшего правдоподобия. Границы между различными «слоями» гладких функций, входящими в ансамбль (50), не являются строго определенными вследствие вероятностной постановки всей задачи. Однако если задана функция φ , то мы можем с помощью методов математической статистики оценить значение α , характеризующее тот слой, из которого с наибольшей вероятностью была взята данная функция. Когда мы измеряем функцию f для определения функции φ , эта последняя нам, разумеется, не известна. Но нам известна вероятностная характеристика процесса, порождающего f при заданной φ . Таким образом, для определения α мы получаем следующую двухступенчатую схему:

$$\alpha \xrightarrow{P_\alpha(\varphi)} \varphi \xrightarrow{P(f|\varphi)} f$$

из которой находим условную вероятность f при заданном α , выражаемую формулой (53). Если эта функция известна, то оценка α по заданной f есть классическая задача математической статистики.

Возникает вопрос: в каких случаях можно заменять усреднение по α на решение при наиболее вероятном α ? Чтобы ответить на этот вопрос, надо вычислить ширину по α пика в функции $P(\alpha|f)$, иначе говоря, погрешность $\Delta\alpha$ определения параметра α . Выражение для погрешности $\Delta\alpha$ имеет весьма ясный физический смысл в случае, когда матрицы Ω и L могут быть одновременно приведены к диагональному виду. Для этого случая получаем (см. ⁶⁸⁾

$$\Delta\alpha/\alpha_0 = \sqrt{2/n_{\text{eff}}},$$

где

$$n_{\text{eff}} = \sum_{k=1}^n z_k^2,$$

а z_k — обрезающий фактор (406). Величина n_{eff} есть эффективное число независимых компонент искомой функции, определяемых из эксперимента. Те компоненты, для которых $z_k \ll 1$, не содержат информации об α и не вносят вклада в уменьшение $\Delta\alpha$.

Алгоритм определения наиболее вероятного α и изложенный в предыдущем разделе алгоритм определения α для наиболее узкого допустимого ансамбля приводят, вообще говоря, к разным значениям α_0 . Однако в ⁶⁵ показано, что α_0 , соответствующее наиболее узкому допустимому ансамблю, также может рассматриваться как состоятельная оценка α . Поэтому в тех случаях, когда $\Delta\alpha$ не слишком велико, указанные два метода приводят к близким результатам, которые асимптотически совпадают при $\Delta\alpha \rightarrow 0$.

Некоторые результаты и проблемы

В работах ⁶⁶⁻⁶⁸ были описаны программы для вычислительных машин, реализующие метод статистической регуляризации, и результаты некоторых математических экспериментов, проведенных с целью проверки этого метода. Эксперименты проводились следующим образом. Выбиралось некоторое ядро уравнения (матрица) K и «истинная» функция (вектор) ϕ . Затем определялась «истинная» правая часть уравнения как результат умножения K на ϕ . К вектору $K\phi$ прибавлялся вектор случайной ошибки с независимыми нормально распределенными компонентами. Полученный таким образом вектор f , имитирующий экспериментальный результат, использовался для восстановления вектора ϕ с помощью указанных программ. Результат восстановления сравнивался с истинным вектором ϕ с учетом теоретически полученной ошибки восстановления (вектор σ).

В различных программах использовались методы наиболее узкого допустимого ансамбля, наиболее вероятного α и усреднения по α . Исследовалась также функция $P(\alpha | f)$. Во всех экспериментах было найдено, что с точностью до ошибки, которую указывает теория, восстановленная функция совпадала с истинной. В качестве иллюстрации приведем один пример из работы ⁶⁷ (рис. 2). Ядро уравнения имело разностный вид — соответствующая функция разрешения показана на рисунке. «Экспериментальная» правая часть f показана вместе с ошибкой измерения. Крестики дают результат восстановления при выборе наиболее вероятнейшего α (режим 2), гистограмма — при усреднении по α (режим 3).

Применение к некоторым задачам геофизики методов решения некорректных задач в ансамбле, заданном корреляционной матрицей, описывается в разделе V настоящей статьи.

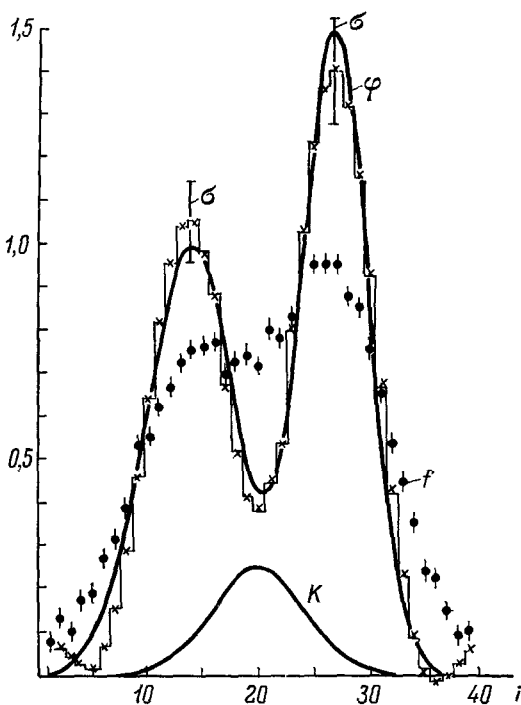


Рис. 2. Восстановление функции $\phi(i)$ по результату ее свертки с кривой разрешения $K(i)$.

σ — ошибка восстановления.

Использование априорного распределения вероятностей в тех случаях, когда нам известно только, что искомая функция гладкая, а действительные ее статистические характеристики не известны, ставит серьезную и общую проблему — проверку того, что найденное в конечном счете регуляризованное решение, действительно, может быть с достаточно большой вероятностью получено путем выборки из априорного ансамбля. Для этого могут использоваться обычные статистические критерии правдоподобия гипотезы.

Такую проверку желательно было бы делать как при заданном заранее α , так и при определении α апостериори. В тех случаях, когда пересечение априорного ансамбля и полного ансамбля решений происходит на «хвостах» гауссианов, критерий правдоподобия даст, очевидно, отрицательный результат. В примерах, которые решались в работах ⁶⁶⁻⁶⁸, можно было почти не сомневаться, что критерий правдоподобия должен дать положительный результат, так как n_{eff} было невелико и подгонкой одного параметра α можно достичь достаточной степени правдоподобия. Однако в более сложных случаях, когда эксперимент содержит много информации о Φ , это может оказаться не так. Как быть в таких случаях?

Ответ напрашивается сам собой: вводить два или больше параметров в априорный ансамбль и определять их апостериори, как мы делали это с одним параметром α . Таким образом, априорная плотность вероятности будет иметь вид «многомерно-слоистого» ансамбля:

$$P(\Phi) = \int P(\alpha_1, \dots, \alpha_v) P_{\alpha_1, \dots, \alpha_v}(\Phi) d\alpha_1 \dots d\alpha_v.$$

Вместо того чтобы сразу определять наименее вероятный набор параметров $\alpha_1, \dots, \alpha_v$, можно предложить следующую процедуру. Сначала придадим всем параметрам α_i , кроме одного, какие-то определенные начальные значения (выбираемые, скажем, на основании прошлого опыта). Находим наименее вероятное значение этого параметра и определяем степень правдоподобия. Если она удовлетворительна, то ищем решение в соответствующем ансамбле. Если она неудовлетворительна, увеличиваем число неизвестных параметров на единицу и т. д. Число параметров, которые в конечном счете окажутся оцениваемыми апостериори, зависит от информативности эксперимента относительно Φ .

Такая постановка вопроса порождает следующую проблему: каков должен быть набор параметров $\alpha_1, \dots, \alpha_v$ и в каком порядке их следует оценивать? В качестве второго параметра естественно предложить порядок производной q в формуле (42). А какое взять начальное значение q ?

Чтобы приблизиться к решению такого рода проблем, необходимы фундаментальные исследования статистической структуры функций, встречающихся в различных классах физических задач. Эти исследования явятся пополнением и уточнением априорной информации, необходимой для адекватного решения некорректных, а по существу, недоопределенных задач физики.

V. ОБРАТНЫЕ ЗАДАЧИ ОПТИЧЕСКОГО ЗОНДИРОВАНИЯ АТМОСФЕРЫ

Рассмотренный в разделе IV метод статистической регуляризации был испытан на задачах определения вертикальных распределений (профилей) температуры и влажности атмосферы по измерениям спектров собственного излучения Земли. Эти задачи в настоящее время весьма актуальны в связи с использованием искусственных спутников Земли для зондирования атмосферы оптическими методами, т. е. для определе-

ния температуры, влажности и других параметров нижних слоев атмосферы (0—50 км) в любой точке земного шара. Учитывая, что в задачах физики атмосферы пространственные и временные вариации температуры и влажности должны быть известны с довольно высокой степенью точности (1—2° по температуре при диапазоне изменений 200—320° К, 0,1—0,2 г/кг по удельной влажности при диапазоне изменений 0,5—20 г/кг), обеспечение надежных решений указанных обратных задач становится практической необходимостью.

Так как, во-первых, эти вариации носят случайный характер, т. е. профили температуры $T(\zeta)$ и влажности $q(\zeta)$ являются случайными функциями высоты *), а во-вторых, для земной атмосферы имеются богатые данные о статистических характеристиках вертикальной структуры полей $T(\zeta)$ и $q(\zeta)$, то вполне естественно применение к решению этих задач метода статистической регуляризации.

Определение вертикального профиля температуры атмосферы

Идея определения вертикального профиля температуры атмосферы по измерениям собственного излучения Земли в достаточно узких спектральных интервалах полосы поглощения CO_2 вблизи 15 мкм со спутников была предложена в работе Каплана ⁶. В основе этой идеи лежит тот физический факт, что излучение в разных участках избранной полосы генерируется в различных слоях атмосферы и, следовательно, определяется температурой этих слоев. Соответствие это достаточно однозначно, поскольку относительная концентрация CO_2 постоянна и хорошо известна до очень больших высот, а поглощением водяного пара и других атмосферных субстанций можно пренебречь **).

Зависимость между спектральной интенсивностью излучения I_ν частоты ν , измеряемой со спутника, и температурой атмосферы $T(\zeta)$ описывается решением уравнения переноса тепловой радиации. Для наиболее простого случая абсолютно черной поверхности Земли и при измерениях I_ν в направлении местной вертикали эта зависимость будет иметь вид

$$I_\nu[T(\zeta), q(\zeta)] = B_\nu[T(1)] P_\nu[w(\zeta)] - \int_0^1 B_\nu[T(\zeta)] \frac{\partial P_\nu[w(\zeta)]}{\partial \zeta} d\zeta. \quad (56)$$

Здесь $B_\nu[T(\zeta)]$ — функция Планка, а $P_\nu[w(\zeta)]$ — функция пропускания атмосферы, характеризующая ослабление излучения некоторого источника столбом атмосферы (0, ζ) единичного сечения, содержащим

эффективную массу поглощающего газа $w(\zeta) = \int_0^\zeta t^n q(t) dt$ (n — кон-

станта, учитывающая изменение полуширины линий поглощения газа в неоднородной по высоте атмосфере). Функция P_ν в (56) берется с учетом спектральной чувствительности аппаратуры, а частота ν соответствует, например, середине интервала спектрального разрешения (иногда под индексом ν мы будем понимать просто номер участка спектра, для которого проводилось измерение I_ν).

*) В качестве переменной, характеризующей высоту, мы здесь, как это принято, используем величину ζ — давление атмосферы на данной высоте.

**) В действительности поглощение водяного пара и аэрозоля может внести заметный вклад в излучение на крыльях полосы CO_2 . Кроме того, определенную роль могут играть вариации концентрации CO_2 , существующие в атмосфере.

Соотношение (56), рассматриваемое как интегральное уравнение относительно $T(\xi)$, позволяет определить вертикальный профиль температуры, если известны интенсивность излучения I_ν как функция ν и функция пропускания P_ν , являющаяся ядром уравнения. Кажущаяся простота уравнения (56) послужила причиной того, что при первых попытках его решения в работах Уорка и Ямамото ^{69, 70} было использовано формальное приведение (56) к системам алгебраических уравнений, получаемых либо путем аппроксимации интеграла конечной суммой ⁷⁰, либо путем разложения искомой функции в ряды полиномов ⁶⁹. Как и следовало ожидать, в результате решения таких систем без принятия определенных мер предосторожности могли получаться величины, сколь угодно далекие от истинного решения и даже физически бессмысленные (см., например, ⁷¹), причем уточнение аппроксимаций, дающее увеличение порядка алгебраических систем, ведет к росту неустойчивости решения. Здесь проявляется некорректность обратной задачи оптического зондирования, которая является прямым следствием физического механизма, определяющего перенос собственного излучения в стратифицированной атмосфере. Благодаря этому механизму происходит сглаживание вариаций излучения в отдельных слоях атмосферы, причем степень сглаживания характеризуется ядром уравнения (56). По этой причине некоторые детали искомых функций, типа упомянутых выше резких изменений градиента температуры, либо вообще полностью заглаживаются и никак не проявляются в измеряемой интенсивности I_ν , либо дают вклад на уровне случайных ошибок измерений I_ν или ошибок расчетов. При обращении уравнений вида (56) эти ошибки усиливаются, что и приводит к неустойчивости решения.

Уравнение (56) нелинейно относительно функции $T(\xi)$, однако оно может быть с достаточной для практических целей точностью линеаризовано. Используя отношение $\varepsilon_\nu(T) = B_\nu(T)/B_0(T)$, где $B_0(T)$ — функция Планка для одного из N используемых участков полосы CO_2 , и пренебрегая зависимостью ε_ν от T (что допустимо только для очень узкого спектрального интервала полосы), Ямамото ⁶⁹ получил уравнение

$$\frac{I_\nu}{\varepsilon_\nu} = B_0[T(1)] P_\nu[w(1)] - \int_0^1 B_0[T(\xi)] \frac{\partial P_\nu[w(\xi)]}{\partial \xi} d\xi, \quad (57)$$

линейное относительно $B_0(T)$.

Для решения уравнения (57) в работах ^{69, 72} использованы разложения искомой функции по полиномам Лагранжа и Чебышева ⁶⁹ или по тригонометрическим функциям ⁷². В работах ¹⁶⁻¹⁸, а в последнее время и в работах ^{74, 75}, для этой цели применялись естественные ортогональные функции (векторы), являющиеся собственными функциями (векторами) корреляционной функции (матрицы), которые получаются путем статистической обработки материалов аэрологического зондирования атмосферы. Следует при этом отметить, что естественные ортогональные функции и обеспечивают оптимальную аппроксимацию любого случайного профиля из рассматриваемого ансамбля, т. е. позволяют параметризовать решение уравнения (56) минимальным числом коэффициентов разложения (по сравнению с любым другим ортогональным базисом). Однако это не дает автоматической гарантии решения соответствующей обратной задачи с той точностью, которая обеспечена ошибками измерений I_ν . Дело в том, что при формальном применении таких разложений не всегда согласовано число учитываемых членов ряда с ядром уравнения (56) (такое согласование было проведено, например, в упомянутых выше работах ^{62, 63}).

Задача определения вертикальных профилей температуры решалась в работах ^{76, 77} методом Тихонова ⁵⁶⁻⁵⁸.

Линеаризацию уравнения (56), удобную для применения статистических методов регуляризации, можно осуществить, представляя $T(\zeta)$ в виде суммы

$$T(\zeta) = \bar{T}(\zeta) + T'(\zeta), \quad (58)$$

где $\bar{T}(\zeta)$ — среднее вертикальное распределение температуры для данного ансамбля, а $T'(\zeta)$ — отклонение от $\bar{T}$ ($T' \ll \bar{T}$). Подставляя (58) в уравнение (56) и учитывая лишь линейные члены в разложении $B_v(T) = B_v(\bar{T}) + \frac{\partial B_v(\bar{T})}{\partial T} T' + \dots$, получим линейное уравнение относительно $T'(\zeta)$:

$$f_v = \frac{\partial B_v[\bar{T}(1)]}{\partial T} \bar{P}_v(1) - \int_0^1 \frac{\partial B_v[\bar{T}(\zeta)]}{\partial T} T'(\zeta) \frac{\partial \bar{P}_v(\zeta)}{\partial \zeta} d\zeta, \quad (59)$$

где $f_v = I_v - \bar{I}_v$, $\bar{I}_v = I_v[\bar{T}(\zeta)]$, $\bar{P}_v(\zeta) = P_v[w(\zeta)]$.

Для определения вертикальных профилей температуры с применением статистического метода регуляризации к уравнению (59) в работе ⁷⁸ использовались аэростатные измерения собственного излучения I_v , полученные с помощью многоканального спектрометра в пяти участках полосы CO_2 15 $\mu\text{м}$ (677,5; 691; 703 и 709 см^{-1}) и в «окне прозрачности» 899 см^{-1} ⁷⁵ (разрешение $\sim 5 \text{ см}^{-1}$), а также спектры излучения, измеренные с помощью фурье-спектрометра с разрешением порядка 2 см^{-1} ⁷⁹. Измерения ^{75, 79} проводились над одними и теми же районами [Палестайн (Техас) и Су-Фоллс (Южн. Дакота)] в летний сезон. Ошибки измерений, по указаниям авторов ^{75, 79}, не превосходили 1%.

Одновременно с I_v непосредственно измерялись вертикальные профили температуры и влажности, что дало возможность, с одной стороны, контролировать методы восстановления $T(\zeta)$, а с другой, учесть в ядре уравнения (59) реальное пропускание атмосферы. В качестве ядра уравнения (59) использовались функции пропускания, рассчитанные в работе ⁷⁵ для районов измерений I_v .

Определение вариации вертикальных профилей $T(\zeta)$ проводилось в работе ⁷⁸ по замкнутой и незамкнутой схемам. В замкнутой схеме (обозначим ее A_1) по известной реализации профиля $T(\zeta)$ и среднему профилю $\bar{T}(\zeta)$ с помощью (56) рассчитывались I_v и $\bar{I}_v$. Затем к I_v добавлялась случайная ошибка со среднеквадратичным отклонением 1%, после чего с применением формул (38), (39) к уравнению (59) восстанавливалась $T(\zeta)$. Сравнения восстановленных по схеме A_1 профилей $T(\zeta)$ с исходными приведены на рис. 3.

При оценке этих результатов следует иметь в виду, что об эффективности самого метода статистической регуляризации нужно судить лишь по «замкнутой схеме». При использовании «незамкнутой схемы» могут играть роль физические факторы, не учтенные (или неправильно учтенные) в исходном уравнении. В самом деле, среднеквадратичная ошибка σ_1 восстановления $T(\zeta)$ по схеме A_1 не превосходит 3° на самом неблагоприятном для метода уровне тропопавзы ($\zeta = 0,25 - 0,3 \text{ атм}$; $z = 10 - 12 \text{ км}$). Действительные же ошибки восстановления $T(\zeta)$ путем использования формул (38), (39) для непосредственных измерений I_v (незамкнутая схема A_2) оказались несколько большими, как это следует из рис. 3. Дело в том, что фактическая погрешность исходной информации f_v при данном ядре уравнения и принятом методе расчета в большинстве случаев превосходит ошибку измерений I_v в несколько раз.

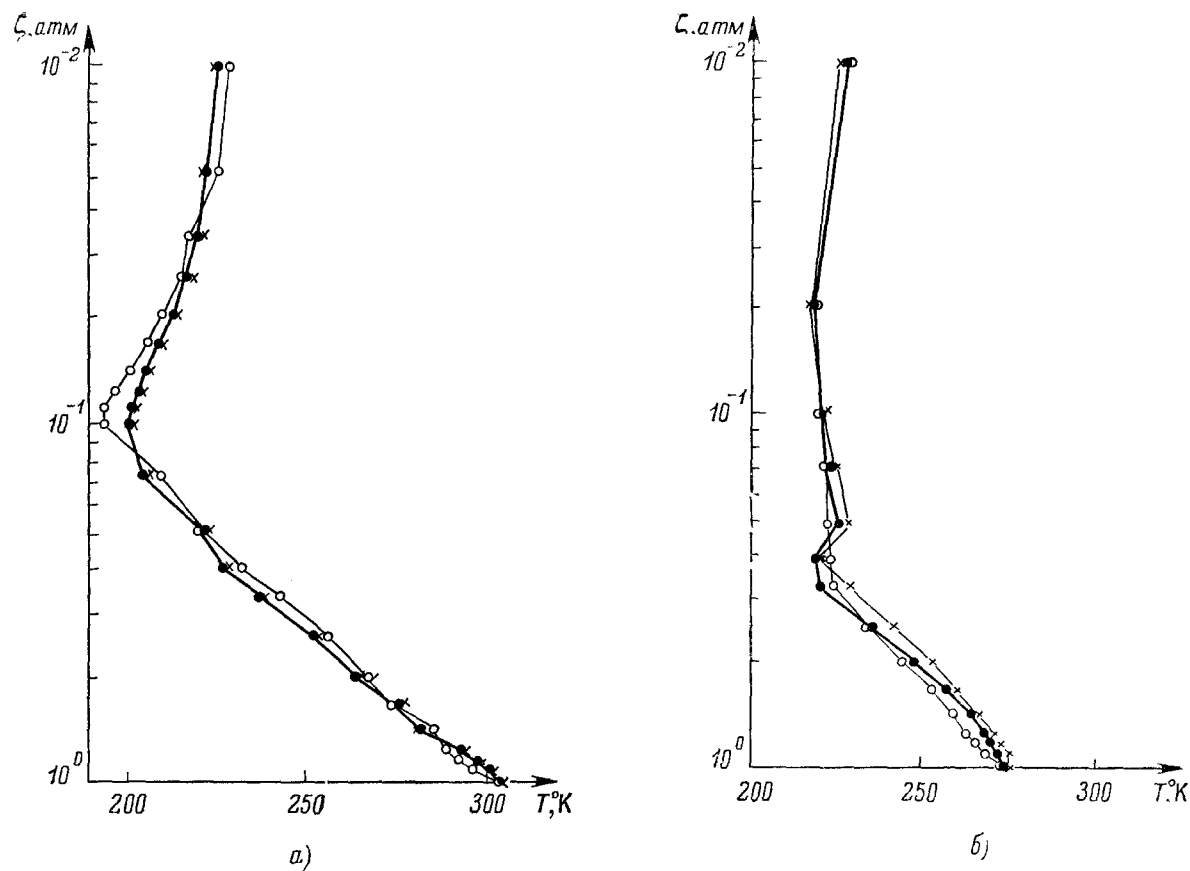


Рис. 3. Примеры восстановления вертикальных профилей температуры методом статистической регуляризации. ●, × — восстановление по замкнутой и незамкнутой схемам (соответственно) по аэростатным измерениям I_v в полосе CO_2 вблизи 15 мкм⁷²: а) над Палестайном (Техас), б) над Су-Фоллс (Южн. Дакота); ○ — истинные профили.

Вопрос о влиянии ошибок ядра уравнения (59) на ошибки определения профиля температуры или другого атмосферного параметра имеет крайне важное значение при практическом применении любого метода к задачам оптического зондирования атмосферы. В реальных условиях функция пропускания P_v в полосе CO_2 15 $\mu\text{м}$ зависит от ряда факторов, для определения которых также необходимо решать обратные задачи на основе некоторой дополнительной информации. Важнейшими из них является поглощение теплового излучения водным аэрозолем, в том числе каплями воды облаков, и водяным паром, крылья полос которого перекрываются с полосой CO_2 15 $\mu\text{м}$.

Для учета аэрозольного поглощения пока нельзя предложить сколько-нибудь обоснованного метода, поскольку оптические свойства аэрозоля и облаков в инфракрасной области спектра еще мало изучены. Что касается учета водяного пара, то для этой цели необходимо решать обратную задачу по определению вертикального профиля его концентрации, используя данные об излучении Земли в полосах паров H_2O .

О п р е д е л е н и е в е р т и к а л ь н о г о п р о ф и л я в л а ж н о с т и а т м о с ф е р ы

Вертикальный профиль относительной концентрации водяного пара $q(\xi)$ или эффективной массы влажности $w(\xi)$ может быть определен из уравнения (56), если измерено излучение Земли в какой-либо полосе паров H_2O , например, в полосе 6,3 $\mu\text{м}$. В этом случае ядром уравнения (56) служит функция $B_v[T(\xi)]$, т. е. для определения $q(\xi)$ или $w(\xi)$ необходимо иметь профиль температуры атмосферы. Следовательно, целесообразно ставить комплексную задачу определения $T(\xi)$ и $q(\xi)$ по одновременным измерениям I_v в полосах CO_2 и H_2O . Тогда профиль $T(\xi)$, определенный по излучению Земли в полосе CO_2 вблизи 15 $\mu\text{м}$ без учета поглощения паров H_2O и рассматриваемый как первое приближение, можно использовать для определения $q(\xi)$ по измерениям в полосе H_2O вблизи 6,3 $\mu\text{м}$, а затем, при необходимости, уточнить оба решения, учитывая перекрывание полос CO_2 и паров H_2O .

Физический принцип, на котором основана возможность определения профиля влажности, аналогичен принципу определения профилей температуры. Излучение атмосферы в центральной части полосы несет информацию о малых концентрациях водяного пара в верхней тропосфере или стратосфере. В областях меньшего поглощения и на крыле полосы I_v дает информацию о влажности в более низких слоях атмосферы. Веса, с которыми эта информация входит в I_v , определяются вертикальными профилями температуры и характером зависимости функции пропускания от влажности.

Задача определения $q(\xi)$ является некорректной задачей, несколько более сложной по сравнению с задачей определения $T(\xi)$. Усложнение связано прежде всего с существенно нелинейной зависимостью функции пропускания от $q(\xi)$. Кроме того, P_v непосредственно зависит от $w(\xi)$ и, следовательно, относительно $q(\xi)$ имеет место двойное сглаживание по высоте. Это повышает чувствительность $q(\xi)$ к случайным ошибкам измерений и расчетов, что и было проиллюстрировано в работе¹⁸, где профиль $q(\xi)$ определялся путем разложения отклонений $q'(\xi) = q(\xi) - \bar{q}(\xi)$ по собственным векторам корреляционной матрицы некоторого ансамбля профилей влажности ($\bar{q}(\xi)$ — средний профиль по этому ансамблю). Однако этот подход, так же как предложенный в работах^{80, 81}

способ представления профилей $q(\zeta)$ в виде парабол $q_0 \zeta^*$ с неизвестными параметрами q_0 и κ , не обеспечивает надежного решения некорректной

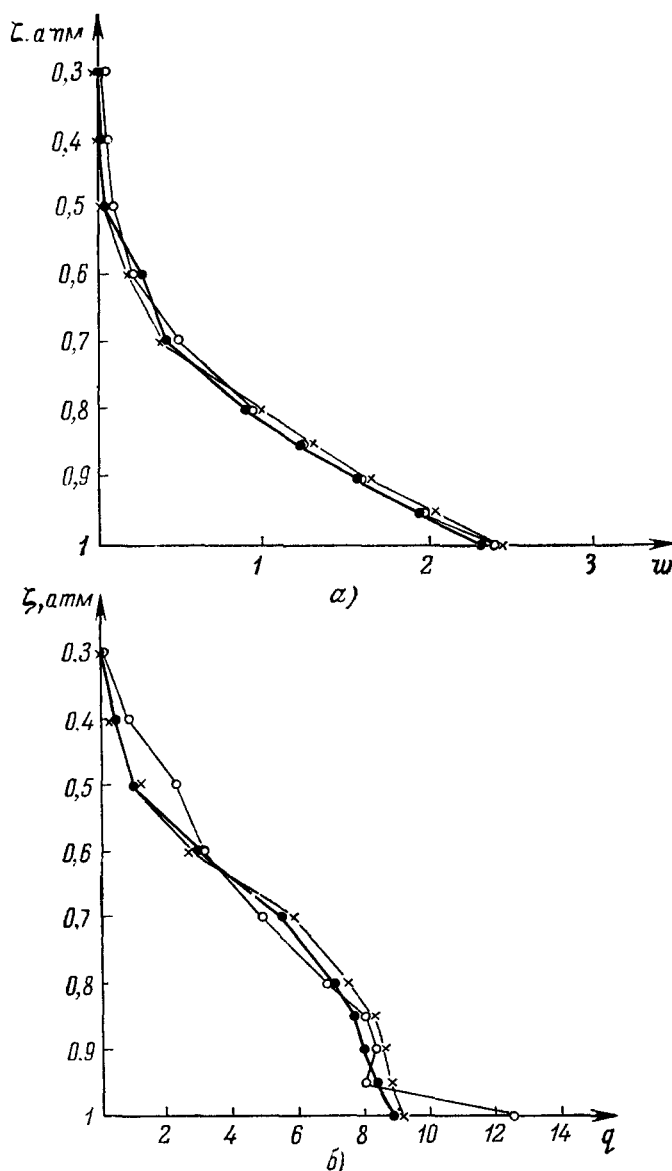


Рис. 4. Примеры ⁸² восстановления вертикальных профилей эффективной массы $w(\zeta)$ (а) и концентрации водяного пара $q(\zeta)$ (б) по аэростатным измерениям I_v в полосе 6,3 мкм ⁷⁹. Обозначения те же, что и на рис. 3.

задачи. Особенно большие ошибки получаются в области сильных изменений искомой функции (например, в пограничном слое атмосферы).

Применение метода статистической регуляризации к уравнению

$$f_v = B_v[T(1)] \frac{\partial P_v[\bar{w}(1)]}{\partial w} w'(1) - \int_0^1 B_v[T(\zeta)] \frac{\partial}{\partial \zeta} \left\{ \frac{\partial P_v[\bar{w}(\zeta)]}{\partial w} w'(\zeta) \right\} d\zeta, \quad (60)$$

полученному путем линеаризации (56) относительно $w'(\zeta) = w(\zeta) - \bar{w}(\zeta)$, где

$$\bar{w}(\zeta) = w_0 \int_0^{\zeta} t^n \bar{q}(t) dt, \quad w'(\zeta) = w_0 \int_0^{\zeta} t^n q'(t) dt,$$

$$f_v = I_v - \bar{I}_v, \quad \bar{I}_v = I_v [T(\zeta), \bar{q}(\zeta)],$$

дает возможность непосредственно определить вертикальный профиль $w(\zeta)$ ⁸². В качестве априорной информации используется корреляционная матрица, построенная по вариациям профилей $w'(\zeta)$. Профили же $q(\zeta)$ вычисляются по восстановленным таким образом $w(\zeta)$ путем дифференцирования по ζ . По-прежнему восстановление $w(\zeta)$ и $q(\zeta)$ осуществляется по замкнутой (A_1) и незамкнутой (A_2) схемам. Схема A_1 , позволяющая проверить эффективность метода обращения, состоит в том, что по известным профилям $T(\zeta)$, $q(\zeta)$, $\bar{q}(\zeta)$ вычисляются величины I_v и $\bar{I}_v$. Затем, в предположении, что $T(\zeta)$ и $q(\zeta)$ известны, профили $w(\zeta)$ и $q(\zeta)$ восстанавливаются по вычисленным I_v , к которым добавляется случайная ошибка с заданной дисперсией. В схеме A_2 профили $q(\zeta)$ и $w(\zeta)$ восстанавливаются непосредственно по измеренным величинам I_v , ошибка измерений которых характеризуется заданной дисперсией.

Результаты определения $w(\zeta)$ и $q(\zeta)$ методом статистической регуляризации, полученные в ⁸² по измерениям I_v ⁷⁹, приведены на рис. 4. Нетрудно видеть (см. рис. 4,а), что восстановленный по замкнутой схеме профиль $\bar{w}(\zeta)$ практически совпадает с исходным. В незамкнутой схеме совпадение $\bar{w}(\zeta)$ и $w(\zeta)$ также вполне удовлетворительное, хотя погрешность восстановления $w(\zeta)$ в приземном слое может достигать 10%. Однако и эти ошибки оказываются существенными, как только мы переходим к определениям профилей $q(\zeta)$ путем дифференцирования $w(\zeta)$. Из рис. 4, б следует, что в приземном слое не удается приблизиться к исходному распределению $q(\zeta)$, обладающему здесь большим градиентом, независимо от того, какая используется схема — замкнутая или незамкнутая.

Можно, однако, применять метод статистической регуляризации для профиля $q(\zeta)$, представленного в виде разложения по собственным векторам. Тогда непосредственно определяемыми величинами будут коэффициенты разложения q_k . Необходимая для определения q_k априорная информация будет представляться в диагональной матрице собственных чисел корреляционной матрицы профилей $q(\zeta)$. Следует при этом указать, что для удовлетворительного восстановления профиля $q(\zeta)$ в пограничном слое воздуха и в других областях сильных вариаций градиента $q(\zeta)$ требуется более детальная по ζ априорная информация, что в свою очередь требует проведения аэрологического зондирования атмосферы на большем числе уровней.

ЗАКЛЮЧЕНИЕ

Понятие о некорректно поставленных задачах возникло в математической физике, поэтому попытки решения таких задач первоначально были ограничены рамками аналитического подхода. Однако в приложениях некорректные задачи возникают обычно как задачи обработки экспериментальных данных, в связи с чем к ним все чаще стали подходить с позиций математической статистики. При таком подходе стало ясно, что некорректные задачи отличаются от корректных тем, что для получения удовлетворительного решения они требуют нетривиальной дополнительной информации об искомой функции, априорной по отношению к тому опыту, на основании которого составляется решаемое уравнение.

Априорная информация может быть более или менее детальной и иметь своим источником общие соображения, вытекающие из физической сущности задачи (например, гладкость искомой функции), или конкретные данные опытов.

Явное введение априорной информации позволяет строго определить ошибку решения. В отличие от чисто аналитического подхода, который дает лишь оценку ошибки сверху (обычно сильно завышенную), статистический подход дает среднеквадратичную ошибку, необходимую физику для интерпретации результатов. Он дает также возможность оценить информативность эксперимента, т. е. количество и структуру реальной информации, независимо от процедуры решения. На этих оценках фактически основывается оптимальное планирование эксперимента, включая выбор схемы измерений и оптимизацию измерительной аппаратуры.

В заключение отметим, что статистические методы решения некорректных задач продемонстрировали свою эффективность в ряде практически важных приложений и в настоящее время получают все более широкое распространение. Дальнейший прогресс статистических методов решения некорректных задач состоит главным образом в возрастании умения ввести недостающую априорную информацию в постановку задачи и в разработке математической техники решения, когда задача уже поставлена и ядра уравнений, описывающих преобразование искомым величин в измеряемые, известны с достаточной степенью точности. В тех случаях, когда ядра уравнений отягощены существенными ошибками (а к таким случаям относятся многие задачи обработки экспериментальных данных), возникает важная проблема обеспечения репрезентативности используемых параметров задачи. Эта проблема, однако, присуща всем методам решения некорректных задач.

Авторы считают приятным долгом выразить благодарность Г. В. Розенбергу за внимание к работе и полезные обсуждения.

Институт физики атмосферы

АН СССР

Институт прикладной математики

АН СССР

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

1. А. Н. Тихонов, ДАН СССР 39, 195 (1943).
2. П. С. Новиков, ДАН СССР 18, 165 (1938).
3. В. С. Bullard, R. J. В. Соорер, Proc. Roy. Soc. (London) A194, 332 (1948), 332.
4. F. John, Ann. mat. pura e appl. 4, 129 (1955).
5. Б. С. Цыбаков, В. П. Яковлев, Изв. вузов (радиофизика) 1, № 5—6, 98 (1958).
6. L. D. Kaplan, JOSA 49, 1004 (1959).
7. J. A. Curcio, JOSA 51, 545 (1961).
8. М. В. Масленников, Ж. вычисл. матем. и матем. физ. 2, 1044 (1962).
9. S. Twomey, G. T. Severynse, J. Atm. Sci. 20, 329 (1963).
10. А. Нора, Optik 19, 357 (1963).
11. А. Нора, Optik 19, 409 (1963).
12. М. P. Freeman, S. Katz, JOSA 53, 1172 (1963).
13. P. Elder, T. Jerric, J. M. Birkenland, Appl. Opt. 4, 589 (1965).
14. R. N. Tourin, B. Krawow, Appl. Opt. 4, 237 (1965).
15. M. T. Chahine, J. Opt. Soc. Amer. 58, 1871 (1968).
16. M. S. Malkovich, Some problems on interpretation of radiation measurements from satellites. Proc. XV Astron. Congress, v. II, Warszawa, 1965, стр. 123.
17. М. С. Малкевич, В. И. Татарский, Косм. исследования 3, 444 (1965).
18. М. С. Малкевич, В. И. Татарский, в сб. «Исслед. косм. пространства», М., «Наука», 1965, стр. 104.

19. Н. Д. Н ю б е р г, ДАН СССР 4, 278 (1934).
20. R. N. B r a c e w e l l, J. A. R o b e r t s, Austr. J. Phys. 7, 615 (1954).
21. R. N. B r a c e w e l l, Proc. IRE 46, 106 (1958).
22. С. Г. Р а у т и а н, УФН 66, 475 (1958).
23. H. J. L o n d o n, W. L. M i r a n k e r, J. Math. Analys and Applicat. 2, 97 (1963).
24. J. C o m e s, V. N a z e l, J. Phys. et le Radium 22, 359 (1961).
25. J. S. R o l l e t t, L. A. H i g g s, Proc. Phys. Soc. 79, 87 (1962).
26. А. Ф. Б о н д а р е в, Оптика и спектроскопия 12, 510 (1962).
27. Б. Н. Г р е ч у ш н и к о в, Оптика и спектроскопия 12, 135 (1962).
28. А. А. Д м и т р и е в, «Косм. исследования» 1, 221 (1963).
29. A. G i r a r d, Communication presented on symposium on molecular structure and spectroscopy. Ohio State Univ., Colombus, 1963, Preprint.
30. А. С. К а г а н, В. М. С н о в и д о в, ЖЭТФ 34, 759 (1964).
31. В. П. К о з л о в, Оптика и спектроскопия 16, 301 (1964).
32. J. L. H a r r i s, JOSA 54, 606 (1964).
33. C. W. H e l s t r o m, The detection and resolution of optical signals, IEEE Trans. on Infrared Theory JT-10, 1964, стр. 275.
34. Г. Г. П е т р а ш, Труды ФИАН СССР 27, 3 (1964).
35. Р. Г. А к о н д ж а н о в, Э. Б. В а й н ш т е й н, Оптика и спектроскопия 18, 495 (1965).
36. Н. С. Ш е с т о в, Выделение оптических сигналов на фоне случайных помех, М., «Сов. радио», 1967.
37. В. П. К о з л о в, К вопросу об оптимальной редукции в теории спектральных приборов. Труды XVI Всесоюз. совещ. по спектроскопии (Москва, 1965), М., «Наука», 1969, стр. 73.
38. G. F o x, E. T. G o o d w i n, Phil. Trans. Roy. Soc. (London) A245 501 (1953).
39. М. М. Л а в р е н т ь е в, О некоторых некорректных задачах математической физики, Новосибирск, Изд-во СО АН СССР, 1962.
40. Л. А. Х а л ф и н, О теоретико-информационном подходе к теории спектральных приборов. Труды Всесоюз. совещ. по теории вероятностей и математической статистике (Ереван), Изд-во АН Арм. ССР, 1960, стр. 187.
41. J. H a d a m a r d, Sur les problemes aux derivees partielles et leur significations physiques, Bull. Univ. Princeton 13, 82 (1902).
42. В. К. И в а н о в, О некорректно поставленных задачах. Матем. сб., нов. сер. 61 (103), 211 (1963).
43. Л. А. Х а л ф и н, В. Н. С у д а к о в, ДАН СССР 157, 1058 (1964).
44. Д. К. Ф а д д е е в, Труды МИАН СССР 53, 387 (1959).
45. Д. К. Ф а д д е е в, В. Н. Ф а д д е е в а, Ж. вычисл. матем. и матем. физ. 1, 412 (1961).
46. К. С. Ш и ф р и н, А. Я. П е р е л ь м а н, Оптика и спектроскопия 15, 533, 667, 803 (1963).
47. К. С. Ш и ф р и н, А. Я. П е р е л ь м а н, ДАН СССР 158, 578 (1964).
48. К. С. Ш и ф р и н, А. Я. П е р е л ь м а н, Изв. АН СССР, сер. «Физ. атм. и океана» 1, 964 (1965).
49. К. С. Ш и ф р и н, И. Б. К о л м а к о в, Изв. АН СССР, сер. «Физ. атм. и океана» 2, 851 (1966).
50. К. С. Ш и ф р и н, Изв. АН СССР, сер. «Физ. атм. и океана» 2, 928 (1966).
51. Г. В. Р о з е н б е р г, УФН 91, 569 (1967).
52. Г. В. Р о з е н б е р г, УФН 95, 159 (1968).
53. D. L. P h i l l i p s, J. Associat. Comput. Machin. 9, 84 (1962).
54. S. T w o m e y, J. Associat. Comput. Machin. 10, 97 (1963).
55. Ф. Ф р и д р и х, в сб. «Методы вычислений», вып. 4, Изд-во Ленингр. ун-та, 1967, стр. 102.
56. А. Н. Т и х о н о в, ДАН СССР, 151, 501 (1963).
57. А. Н. Т и х о н о в, ДАН СССР, 153, 49 (1963).
58. А. Н. Т и х о н о в, В. Б. Г л а с к о, Ж. вычисл. матем. и матем. физ. 4, 564 (1964).
59. В. Я. А р с е н и н, В. К. И в а н о в, Ж. вычисл. матем. и матем. физ. 8, 310 (1968).
60. В. Я. А р с е н и н, В. В. И в а н о в, ДАН СССР 182, 9 (1968).
61. В. К. И в а н о в, Ж. вычисл. матем. и матем. физ. 6, 1089 (1966).
62. В. П. К о з л о в, ДАН СССР 166, 779 (1966).
63. В. П. К о з л о в, Изв. АН СССР, сер. «Физ. атм. и океана» 2, 137 (1966).
64. С. К у л ь б а к, Теория информации и статистика, М., «Наука», 1967.
65. В. Ф. Т у р ч и н, Ж. вычисл. матем. и матем. физ. 7, 1270 (1967).
66. В. Ф. Т у р ч и н, Ж. вычисл. матем. и матем. физ. 8, 230 (1968).
67. В. Ф. Т у р ч и н, В. З. Н о з и к, Изв. АН СССР, сер. «Физ. атм. и океана» 5, 29 (1969).

68. В. Ф. Турчин, В. З. Нозик, Препринт ФЭИ 138, ФЭИ, 1969.
 69. G. Yamamoto, J. Meteorol. 18, 581 (1961).
 70. D. Q. Wark, J. Geophys. Res. 66, 77 (1961).
 71. М. С. Малкевич, Косм. исследования 2, 246 (1964).
 72. D. T. Hilleary, D. Q. Wark, D. G. James, Nature 205, 489 (1965).
 73. М. С. Малкевич, В. Р. Козлов, И. А. Горчакова, Tellus 21, 389 (1969).
 74. D. Q. Wark, H. E. Fleming, Monthly Weath. Rev. 94, 351 (1966).
 75. D. Q. Wark, F. Saiedy, D. Q. James, Monthly Weath. Rev. 95, 468 (1967).
 76. В. Б. Гласко, Ю. М. Тимофеев, Изв. АН СССР, сер. «Физ. атм. и океана» 4, 303 (1968).
 77. О. М. Покровский, Ю. М. Тимофеев, Изв. АН СССР, сер. «Физ. атм. и океана» 5, 1082 (1969).
 78. В. Ф. Турчин, М. С. Малкевич, И. А. Горчакова, Изв. АН СССР, сер. «Физ. атм. и океана» 5, 449 (1969).
 79. L. W. Chaney, L. T. Loh, M. T. Surh, A Fourier transform spectrometer for measurement of atmospheric thermal radiation. Techn. Rep. Univ. of Michigan, 05863-12-T, May 1967.
 80. W. L. Smith, Monthly Weath. Rev. 95, 363 (1967).
 81. W. L. Smith, Monthly Weath. Rev. 96, 387 (1968).
 82. И. А. Горчакова, М. С. Малкевич, В. Ф. Турчин, Изв. АН СССР, сер. «Физ. атм. и океана» 6, 565 (1970).
 83. O. N. Strand, E. R. Westwater, J. Associat Comput. Machin. 15, 100 (1968).
 84. E. R. Westwater, O. N. Strand, J. Atm. Sci. 25, 750 (1968).
-