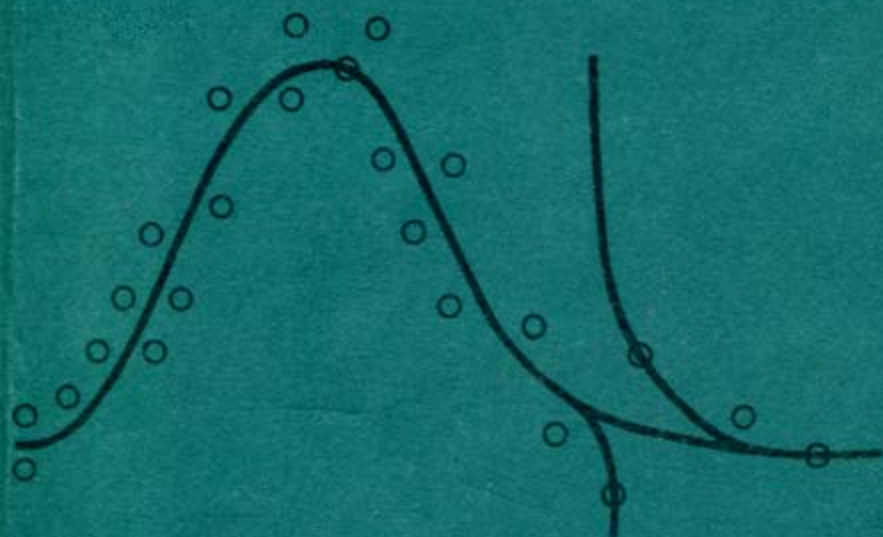


В.Н. Виноградов
Е.В. Гай
Н.С. Работнов

АНАЛИТИЧЕСКАЯ АППРОКСИМАЦИЯ ДАННЫХ в ядерной и нейтронной физике



ЭНЕРГОАТОМИЗДАТ

**В.Н. Виноградов
Е.В. Гай
Н.С. Работнов**

**АНАЛИТИЧЕСКАЯ
АППРОКСИМАЦИЯ
ДАНЫХ
в ядерной
и нейтронной
физике**



МОСКВА
ЭНЕРГОАТОМИЗДАТ
1987

Виноградов В.Н., Гай Е.В., Работнов Н.С. Аналитическая аппроксимация данных в ядерной и нейтронной физике. М.: Энергоатомиздат, 1987. 128 с.

Описан метод аналитической аппроксимации данных рациональными функциями одного переменного, ориентированной на теоретические и расчетные задачи нейтронной и ядерной физики. Детально представлены алгоритмы построения рациональных аппроксимант и вычисления погрешностей их параметров, проиллюстрированные многочисленными примерами их использования.

Для научных работников и инженеров, специализирующихся в областях ядерной, нейтронной, реакторной физики и прикладной математики, а также аспирантов и студентов старших курсов.

Табл. 15. Ил. 15. Библиогр. 73.

Рецензент Ф. Никитиу.

НАУЧНОЕ ИЗДАНИЕ

Владимир Николаевич Виноградов
Евгений Владимирович Гай
Николай Семенович Работнов

АНАЛИТИЧЕСКАЯ АППРОКСИМАЦИЯ ДАННЫХ В ЯДЕРНОЙ И НЕЙТРОННОЙ ФИЗИКЕ

Редактор *Е.В. Сатарова*
Художественные редакторы *А.Т. Кирьянов, Ю.В. Сазанская*
Технический редактор *Е.В. Пронь*
Корректор *С.Ю. Торокина*
Оператор *Т.В. Соколова*
ИБ № 1547

Набор выполнен в Энергоатомиздате на Композере ИБМ-82. Подписано в печать 20.01.87. Т-03539. Формат 60 x 88 1/16. Бумага офсетная №1. Печать офсетная. Усл. печ. л. 7,84. Усл. кр.-отт. 8,20. Уч.-изд. л. 8,39. Тираж 1900 экз. Заказ 1364. Цена 1 р 30 к.

Энергоатомиздат, 113114, Москва, М-114, Шлюзовая наб., 10
Московская типография № 9 Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли
109033, Москва, Ж-33, Волочаевская ул., 40

ОГЛАВЛЕНИЕ

Предисловие	3
Введение	5
Глава 1. Общие вопросы аналитической аппроксимации данных	7
1.1. Способы представления данных в нейтронно-физических расчетах	7
1.2. Роль аналитичности в задаче аппроксимации	8
1.3. Выбор системы аппроксимирующих функций при обработке экспериментальных результатов	10
1.4. Параметризация аппроксиманты опорными ординатами	12
Глава 2. Паде-аппроксимация, определение и основные свойства	14
2.1. Терминология и обозначения	14
2.2. Приближение Паде первого рода	16
2.3. Задача аналитического продолжения функции и сходимости рациональной аппроксиманты	18
2.4. Приближение Паде второго рода	20
2.5. Инвариантность приближений Паде. Преобразование Эйлера	22
Глава 3. Алгоритмы рациональной интерполяции	25
3.1. Детерминантное решение задачи интерполяции	25
3.2. Производные по опорным ординатам	30
3.3. Рекуррентные соотношения для рациональной интерполяции	33
3.4. Алгоритм построения двух аппроксимант с одинаковыми знаменателями	38
3.5. Тригонометрическая мероморфная интерполяция	41
3.6. Особые случаи рациональной интерполяции	44
Глава 4. Приближение Паде первого рода и его обобщения. Основные алгоритмы	48
4.1. Алгоритмы построения приближения Паде первого рода	48
4.2. Интерполяция с кратными узлами	52
4.3. Построение многоточечной аппроксиманты с учетом разложения по отрицательным степеням	54
4.4. Особые случаи приближения Паде первого рода	57
Глава 5. Рациональная аппроксимация экспериментальных зависимостей	61
5.1. Метод наименьших квадратов	61
5.2. Дискретная оптимизация паде-аппроксиманты (метод перебора)	64
5.3. Построение и анализ рациональной аппроксиманты. Шумовые полюсы	68
5.4. Примеры использования дискретной оптимизации для аппроксимации экспериментальных данных	71
5.5. Резонансный анализ на основе рациональной аппроксимации	78
Глава 6. Оценка погрешностей паде-аппроксиманты	82
6.1. Информационная и ковариационная матрицы при параметризации резонансными параметрами и опорными ординатами	82
6.2. Диагонализация информационной матрицы опорных ординат	86
6.3. Погрешности аппроксиманты	87
6.4. Статистически оптимальная рациональная интерполяция	92
6.5. Оценка количества информации, полученной при обработке и анализе экспериментальных данных	97
Глава 7. Некоторые приложения паде-аппроксимации	102
7.1. Расчет подгрупповых констант теории переноса нейтронов	102
7.2. Обратная задача теоретической спектроскопии	109
7.3. Экспоненциальный анализ	117
Список литературы	126

ПРЕДИСЛОВИЕ

Ядерная энергетика связана с фундаментальными научными исследованиями, пожалуй, сильнее и глубже, чем любая другая отрасль техники. Весьма важной областью этих исследований являются ядерная и нейтронная физика. Теория, эксперимент и технические приложения объединены здесь высоким уровнем использования математики — как ее общих результатов, так и прикладных, вычислительных методов. При этом в развитии последних, в приспособлении мощных математических инструментов к конкретным ядерно-физическим задачам сами физики традиционно активны.

Одним из заметных направлений такой деятельности в последнее время стало использование аппроксимации рациональными функциями (приближение Паде) в самых разнообразных задачах математической физики. В частности, использование этой аппроксимации оказалось весьма плодотворным и перспективным при обработке экспериментальных данных, однако в литературе оно отражено лишь журнальными публикациями с рассмотрением конкретных задач или краткими отступлениями в монографиях на другие темы. В то же время подробное освещение этого вопроса с целью более широкого внедрения аппроксимации рациональными функциями в практику обработки экспериментальных данных представляется нам необходимым и своевременным. Потребность в издании такого рода в какой-то мере должна удовлетворить настоящая книга.

Книги, в которых рассматриваются математические методы обработки экспериментальных данных, можно отнести к двум основным типам. Один из них — это монографии и учебники по теоретической статистике, в которых конкретные задачи или вообще не излагаются, или привлекаются лишь для иллюстрации. Книги второго типа — это практические руководства, в которых рассматривается в основном вычислительная сторона метода наименьших квадратов. Аналитические свойства функций, аппроксимирующих экспериментальные данные, в обоих случаях остаются в стороне от основного круга проблем.

В отличие от книг этих двух типов, в предлагаемом издании основное внимание уделяется вопросам, связанным с использованием именно аналитических свойств функций при обработке экспериментальных данных. Сведения из математической статистики привлекаются лишь в пределах необходимого минимума. Особое место занимает построение ал-

горитмов, позволяющих эффективно использовать для обработки экспериментальных данных следующие по сложности после полиномов класс аналитических функций — рациональные функции. Приводятся примеры использования этих алгоритмов, реализованных на ЭВМ, в задачах ядерной и нейтронной физики. Основные методические результаты представляют собой дальнейшее развитие математического аппарата приближения Паде. Мы надеемся, что эта книга окажется полезной для специалистов в области обработки данных, а также для всех, кто использует приближение Паде высоких порядков в теоретической и расчетной работе. Ее можно использовать и для первого знакомства с этим вопросом.

Настоящая монография является итогом коллективной работы, начатой в 1973 г. по инициативе одного из авторов (Е.В. Гая), которым получены основные методические результаты, изложенные в гл. 3, 4. Остальные главы написаны совместно.

Авторы выражают искреннюю признательность С.А. Бадикову, в соавторстве с которым написаны работы, положенные в основу изложения материала двух последних глав, а также своим многочисленным коллегам, сотрудникам Физико-энергетического института, за полезные обсуждения. С особой благодарностью мы вспоминаем ныне покойного Л.Н. Усачева, чье внимание и поддержка много значили для нас.

В.Н. Виноградов, Е.В. Гай, Н.С. Работнов

ВВЕДЕНИЕ

Для физического расчета ядерно-энергетических установок требуются весьма большой объем вычислений и множество входных данных. Существенную часть последних составляют функции одного переменного, описывающие энергетическую зависимость сечений разнообразных ядерных реакций под действием нейтронов. Их измерение — одна из важнейших задач экспериментальной ядерной физики, а сбор, хранение, оценка этих данных и представление их в виде, наиболее удобном для использования, стали самостоятельной и немаловажной научной дисциплиной.

Потребность в удобном аналитическом представлении одномерных функциональных зависимостей постоянно возникает на всех этапах исследований — от построения теоретических моделей до обработки результатов эксперимента. При этом следует различать две основные задачи.

1. Аппроксимация известных функций, когда необходимо сложную функцию, значения которой могут быть вычислены сколь угодно точно, но с большими затратами труда, заменить простой, легко вычисляемой и близкой к исходной по некоторым критериям. В этой задаче сведения об исходной функции могут быть заданы в различных формах: в виде ее значений в некоторой области изменения аргумента, коэффициентов ряда Тейлора (не обязательно в области сходимости этого ряда), обобщенных моментов с некоторыми весами, результатов применения различного рода интегральных преобразований, характеристик особых точек и экстремумов и т.д. Функция "наилучшего приближения" (внутри класса аппроксимирующих функций, выбранного из соображений удобства дальнейшего использования) может определяться самыми разнообразными требованиями: минимизацией максимального или среднего квадратического отклонения от исходной функции, заданием асимптотик, моделированием особенностей функций и т.д.

2. Аппроксимация дискретной одномерной зависимости, полученной в результате эксперимента. В этом случае значения аппроксимируемой функции известны лишь при некоторых значениях аргумента и с некоторой погрешностью, т.е. содержат в качестве аддитивной добавки случайную величину, о статистическом распределении которой имеется ограниченная или полная информация. В качестве критерия выбора аппроксиманты обычно принимают требование статистической оптимальности типа критерия χ^2 .

В обоих случаях результирующая аппроксиманта становится впоследствии объектом различных математических операций, законность использования которых часто решающим образом зависит от близости ее аналитических свойств к аналитическим свойствам истинной функции, прежде всего от сходства их особенностей в комплексной плоскости. С этой точки зрения весьма популярная в силу своей простоты и разработанности методов полиномиальная аппроксимация обладает серьезным недостатком — полным отсутствием у аппроксиманты особых точек, кроме бесконечно удаленной.

В настоящем издании основное внимание уделяется приближению Паде — аппроксимации рациональными функциями, у которых имеются полюсы, моделирующие особенности приближаемой функции. В последнее время этот метод находит все более широкое применение в самых разных задачах математической физики (см., например, [1–7]). При обработке данных его использование сдерживалось нелинейностью соответствующей задачи метода наименьших квадратов и своеобразной формой неустойчивости — появлением так называемых шумовых дублетов (см. § 5.3). В работах авторов этой книги [8 – 10] были предложены способы обхода указанных трудностей и создан приемлемый метод паде-аппроксимации экспериментальных зависимостей. Вместе с тем в необходимых случаях рассматриваются и другие способы аппроксимации.

ГЛАВА 1. ОБЩИЕ ВОПРОСЫ АНАЛИТИЧЕСКОЙ АППРОКСИМАЦИИ ДАННЫХ

1.1. СПОСОБЫ ПРЕДСТАВЛЕНИЯ ДАННЫХ В НЕЙТРОННО-ФИЗИЧЕСКИХ РАСЧЕТАХ

Первичной исходной информацией для большинства нейтронно-физических расчетов являются данные о зависимости сечений различных ядерных реакций, вызываемых нейтронами, от энергии нейтронов E . Если речь идет об интегральных сечениях, то соответствующие данные описываются функциями, иногда весьма сложными, от этой единственной переменной. Дифференциальные сечения зависят еще от углов вылета и энергий продуктов реакции. В настоящей книге речь пойдет почти исключительно об аппроксимации функциональных зависимостей, хотя излагаемые в ней результаты можно, разумеется, использовать и при работе с дифференциальными сечениями, фиксируя в них все переменные, кроме одной.

Результатом ядерно-физического эксперимента по измерению сечения $\sigma(E)$ всегда служит набор дискретных значений $\sigma(E_i)$, соответствующих некоторому набору значений энергии. Эта информация может быть весьма детальной, т.е. насчитывать многие тысячи пар значений энергия — сечение (так называемое поточечное задание нейтронных сечений). Однако для расчетов обычно требуется знать сечение при произвольном значении энергии, отсюда возникает необходимость использования разного рода интерполяционных формул, которые позволяют вычислять $\sigma(E)$ между первичными точками. Сами экспериментальные значения содержат неконтролируемую случайную погрешность и поэтому не всегда удобны в качестве узлов интерполяционных формул. Вместо них при поточечном задании все чаще используются оцененные данные, представляющие собой результат применения некоторых процедур сглаживания первичной информации.

Такая информация близка к полной в области низких энергий нейтронов, соответствующих разрешенным резонансам. В этой области можно описать зависимость $\sigma(E)$ функцией многих параметров. Вид этой функции определяется физическими соображениями, а значения параметров — подгонкой под экспериментальные данные. С повышением энергии экспериментальные, а значит, и оцененные значения $\sigma(E)$ становятся результатом усреднения по все более широким энергетическим интервалам с некоторым, не всегда хорошо известным весом, и детальные сведения о резонансной структуре сечений теряются. Для этих областей приводятся вместе с поточечным заданием или вместо него средние значения резонансных параметров и характеристики их статистических

распределений, т.е. их флюктуации относительно средних значений. Подробная информация об этих способах задания ядерных данных содержится в описаниях многочисленных библиотек на магнитных носителях информации, из которых наиболее употребительны американская ENDF/B (версии IV и V [11, 12]), английская UKNDL [13], японская JENDL [14].

Идеальным было бы непосредственное использование в реакторно-физических расчетах всей имеющейся поточечной информации о нейтронных сечениях вместе с интерполяционными формулами, однако ресурсы современных ЭВМ такой возможности пока не дают. Даже там, где имеется детальная поточечная информация, ее приходится заменять менее детальной, но зато существенно меньшей по объему. Самый распространенный метод такого сокращения входной информации – получение групповых констант, которые и используются непосредственно в реакторно-физических расчетах. Методы их получения и применения изложены в [15, 16]. Простейший из этих методов – усреднение известных сечений по энергетическим интервалам с фиксированными для данной системы констант границами. Число таких интервалов колеблется от немногих десятков до многих сотен.

В практических расчетах требуются, кроме того, функционалы сечений, более сложные, чем их средние значения в каких-то интервалах, например пропускание (см. § 7.3), факторы резонансной самоэкранировки (см. § 7.1) и т.д. Они, в свою очередь, являются функциями некоторых параметров: пропускание зависит от толщины образца, факторы самоэкранировки – от так называемого сечения разбавления и т.д.

На всех этапах перевода информации о нейтронных сечениях из одной формы в другую аппроксимация экспериментальных, оцененных или усредненных данных играет важную роль.

Обработка результатов современного эксперимента, как правило, многоступенчатый и сложный процесс. Он обычно включает: внесение поправок на геометрию опыта и эффективность детектирующих устройств; вычитание фона и эффектов, связанных с наличием примесей в образцах; учет конечного энергетического разрешения аппаратуры и эффекта Доплера; окончательный расчет значений измеряемых величин с учетом их возможных корреляций. Только после этого можно приступить к анализу с целью получения физической информации. Этот анализ, вообще говоря, однозначен лишь в предположении, что физические закономерности описываются аналитическими функциями.

1.2. РОЛЬ АНАЛИТИЧНОСТИ В ЗАДАЧЕ АППРОКСИМАЦИИ

Термин "аналитический" в зависимости от контекста может иметь различный смысл. Так, первое употребление этого слова, с которым мы встречаемся еще в школе, это "аналитическое задание функции" в отличие от графического или табличного. Речь идет о том, что, определив

простейшие функции, более сложные функциональные зависимости сводят к некоторым комбинациям простейших, при этом свойства полученной функции (гладкость, дифференцируемость и др.) требуют дополнительного рассмотрения.

Аналитическое в этом смысле описание экспериментальной зависимости осуществляется, в частности, с помощью полиномов, рациональных функций, сплайнов (т.е. сшиванием различных полиномов в различных областях изменения аргумента), но наличие аналитической формулы при одних значениях аргумента в такой постановке еще не ограничивает выбор формулы при других его значениях, для этого требуется дополнительная информация или дополнительные критерии.

Гораздо уже понятие аналитической функции [17, 18] комплексного аргумента $f(z) = u(x, y) + iv(x, y)$, где x и y — действительная и мнимая части аргумента; u и v — действительная и мнимая части функции. Есть два эквивалентных определения такой функции: по Вейерштрассу — это функция, задаваемая своим разложением в ряд Тейлора в области его сходимости и продолжениями этого ряда в других областях; по Коши—Риману (Д'Аламберу—Эйлера) — в точках аналитичности должны выполняться условия $du/dx = dv/dy$ и $dv/dx = -du/dy$, соответствующие существованию в этих точках производной по комплексной переменной $z = x + iy$, которая не зависит от направления дифференцирования в комплексной плоскости. Первое из этих определений удобно для изучения свойств конкретных аналитических функций, второе — для проверки аналитичности функции.

Хотя в определении речь идет об аналитичности в точке, одно из самых существенных для физических применений свойств аналитических функций заключается в следующем: задание аналитической функции на некотором отрезке полностью и однозначно определяет эту функцию или ее ветвь во всей комплексной плоскости. Представление о единстве и детерминированности физических законов лучше всего согласуется именно с требованием аналитичности используемых функций. Один из аргументов в пользу такого требования состоит в том, что дифференциальные уравнения, описывающие физические явления, это уравнения с аналитическими коэффициентами, а решениями таких уравнений являются аналитические функции. Известно, что квантовая механика совсем не требует такого сужения класса используемых функций (см. [19]), так что это, скорее, эмпирический закон, чем математическая теорема.

При обработке экспериментальных данных с ростом числа и точности измерений ситуация приближается к следующей: известны значения функции на участке действительной оси; требуется параметризовать эту зависимость и, что самое главное, выявить физические закономерности, найти параметры физических моделей, которые, как только что упоминалось, описываются аналитическими функциями. Подробнее об этом см. в § 5.5.

1.3. ВЫБОР СИСТЕМЫ АППРОКСИМИРУЮЩИХ ФУНКЦИЙ ПРИ ОБРАБОТКЕ ЭКСПЕРИМЕНТАЛЬНЫХ РЕЗУЛЬТАТОВ

Аппроксимация аналитическими функциями позволяет определить параметры физических моделей. В свою очередь, во многих случаях знание физической природы изучаемого явления непосредственно определяет выбор достаточно простого класса аппроксимирующих функций. Приведем примеры такого рода, не останавливаясь пока на смысле входящих в формулы параметров.

1. Угловые распределения продуктов ядерной реакции, проходящей через уровень составного ядра с определенным спином J , описываются полиномами от $\cos \theta$ (θ — угол между направлениями пучка налетающих частиц и вылета продуктов реакции), причем максимальная степень полинома $N_{\max} = 2J$.

2. Энергетическая зависимость сечений ядерных реакций под действием нейтронов в резонансной области энергий определяется положением полюсов соответствующей S -матрицы, а потому хорошо аппроксимируется полюсным разложением, т.е. рациональной функцией от энергии E , умноженной на $E^{-1/2}$:

$$\sigma(E) = \frac{1}{\sqrt{E}} \frac{P_N(E)}{Q_M(E)} = \frac{1}{\sqrt{E}} \sum_{i=1}^l \frac{\alpha_i(E - E_i) + \beta_i}{\gamma_i^2 + (E - \epsilon_i)^2}, \quad (1.1)$$

где $P_N(E)$ и $Q_M(E)$ — полиномы степеней N и M соответственно.

3. Радиоактивный распад ядер происходит по экспоненциальному закону, и зависимость от времени активности образцов, содержащих различные нуклиды, представляется суммой экспонент:

$$A(t) = \sum_{i=1}^l a_i e^{-p_i t}. \quad (1.2)$$

4. Функции, описывающие периодические процессы в электрических цепях, разлагаются в ряды Фурье:

$$f(t) = \sum_{i=1}^l a_i \sin(\omega_i t + \delta_i). \quad (1.3)$$

5. Обобщением случаев 3 и 4 является представление в виде суммы экспонент с комплексными параметрами, когда в (1.2) параметры a_i и p_i могут принимать комплексные значения, что позволяет наряду с периодическими описывать и аperiodические процессы, включая затухающие колебания.

6. Функции времени, которые обладают спектром Фурье, сосредоточенным в ограниченном интервале частот, т.е. такие, что

$$f(t) = \frac{1}{2\pi} \int_{-2\pi W}^{2\pi W} e^{i\omega t} F(\omega) d\omega, \quad (1.4)$$

в соответствии с теоремой Котельникова [20] разлагаются по так называемым функциям отсчетов:

$$f(t) = \sum_{n=-\infty}^{\infty} f\left(\frac{n}{2W}\right) \frac{\sin(2\pi Wt - n\pi)}{2\pi Wt - n\pi}, \quad (1.5)$$

т.е. полностью определяются своими значениями в эквидистантной последовательности точек. Это разложение широко используется в теории информации и теории линий связи.

7. Если статистическое распределение некоторой физической величины представляет собой суперпозицию нормальных распределений с различными средними значениями и дисперсиями, то такое распределение разлагается в сумму гауссианов

$$f(z) = \sum_{i=1}^l a_i e^{-\frac{(z - \mu_i)^2}{2\sigma_i^2}}. \quad (1.6)$$

В пп. 1–7 перечислены, пожалуй, все важнейшие используемые на практике классы аппроксимирующих функций. Теория и методы этих аппроксимаций детальнее всего разработаны для полиномов и разложений Фурье (тригонометрических полиномов), которым посвящена весьма обширная литература.

Аналитическая аппроксимация одномерной зависимости, заданной дискретной последовательностью пар значений аргумента и функции, может преследовать различные цели. Об одной мы уже упоминали: параметры аппроксиманты могут иметь непосредственный физический смысл. Так, ϵ_i и $2\gamma_i$ в (1.1) — энергии ядерных уровней и их полные ширины; p_i в (1.2) связаны с периодом полураспада отдельных нуклидов T_i соотношением $T_i = \ln 2/p_i$.

В других случаях аналитическое выражение нужно просто для того, чтобы иметь возможность вычислить значение измеряемой функции при произвольном значении аргумента.

Немаловажной представляется возможность уточнения результатов эксперимента с помощью аналитической аппроксимации. Поясним это общеизвестным простым примером. Если значение некоторой физической величины измеряется N_{ex} раз и погрешность каждого измерения является нормально распределенной случайной величиной со стандартным отклонением Δf , то погрешность среднего значения \bar{f} по N_{ex} измерениям равна $\Delta f/\sqrt{N_{\text{ex}}}$ — 1, т.е. служит более надежной оценкой величины f , чем любое из измеренных значений. Такой результат обработки эксперимента можно рассматривать как результат аппроксимации $f(i)$ простейшей аналитической функцией — горизонтальной прямой $f(i) = \bar{f} = \text{const}$. В более сложных случаях расчет погрешности аппроксиманты не так прост, но в общем возможность уточнения сохраняется, если число параметров аппроксиманты меньше числа точек, в которых осуществляются измерения.

Аналитическая аппроксимация позволяет также сжать информацию, т.е. сократить объем числовых массивов, подлежащих хранению, что весьма существенно в случае нейтронных данных, поскольку объем соответствующих машинных библиотек уже сейчас составляет многие сотни мегабайт.

1.4. ПАРАМЕТРИЗАЦИЯ АППРОКСИМАНТЫ ОПОРНЫМИ ОРДИНАТАМИ

Даже ограничившись определенным классом аппроксимирующих функций, параметры аппроксиманты в большинстве случаев все еще можно выбирать разными способами. Рассмотрим возможные варианты такого выбора на примере полиномиальной аппроксиманты. Полином $P_N(z)$ степени N можно задать:

1) значениями коэффициентов a_n при степенях аргумента:

$$P_N(z) = \sum_{n=0}^N a_n z^n; \quad (1.7)$$

2) значениями корней полинома z_n и коэффициентом a_N при старшей степени:

$$P_N(z) = a_N \prod_{n=1}^N (z - z_n); \quad (1.8)$$

3) значениями самого полинома как функции z при $N+1$ значениях аргумента z_ν :

$$P_N(z) = \sum_{\nu=0}^N P_\nu \prod_{\mu \neq \nu} \frac{(z - z_\mu)}{(z_\nu - z_\mu)}, \quad (1.9)$$

где $P_\nu \equiv P_N(z_\nu)$. (Выражение (1.9) известно как интерполяционная формула Лагранжа [21]).

Эти варианты эквивалентны в том смысле, что полное число параметров во всех случаях одинаково и равно $N+1$, и разные их наборы взаимно однозначно выражаются друг через друга. Однако в различных задачах можно по разным соображениям предпочесть тот или иной конкретный вариант. Так, разложения (1.7) и (1.9) являются линейными функциями параметров a_n и P_ν соответственно, а (1.8) нелинейно по z_n . В то же время, имея значения z_n , можно весьма просто вычислять a_n и P_ν , а обратная процедура не имеет, в общем случае, алгебраического решения и должна выполняться численными методами, которые быстро усложняются с ростом N .

Представления (1.7) и (1.8) единственны, в то время как, варьируя абсциссы z_ν , заданный полином можно представить в виде (1.9) бесконечным числом способов. Эта гибкость интерполяционной формулы

Лагранжа широко используется в теории полиномиальной аппроксимации. Выбор z_ν (назовем их опорными абсциссами, а P_ν — опорными ординатами) можно подчинять самым различным требованиям оптимальности. Таким способом получают многие широко употребляющиеся аппроксимационные, квадратурные и прочие формулы [21].

Параметризация опорными ординатами возможна и, как будет показано ниже, полезна и для других классов аппроксимирующих функций. Так, при разложении (1.5) по функциям отсчетов также используются в качестве параметров опорные ординаты. Заметим, что выбор абсцисс в этом случае подчинен общему требованию — они должны образовывать эквидистантные последовательности, шаг которых определяется границей фурье-спектра.

В общем случае, если задано L -параметрическое семейство функций одной переменной $f(z; p_1, p_2, \dots, p_L)$, то, выбирая L , вообще говоря, произвольных значений аргумента z , можно записать систему уравнений

$$f(z_\nu; p_1, p_2, \dots, p_L) = f_\nu, \quad \nu = 1, 2, \dots, L. \quad (1.10)$$

Если система (1.10) имеет в некоторой области значений f_ν решение, однозначно определяющее вектор $\{p_\nu\}$ как функцию вектора $\{f_\nu\}$, то f_ν можно рассматривать как новые параметры. Если удастся получить для $f(z)$ как функции этих аргументов явное выражение или достаточно простой алгоритм вычисления, то величинами f_ν можно пользоваться, как любыми другими параметрами аппроксиманты — находить для них оптимальные значения, брать по ним производные, вычислять ковариационные матрицы этих параметров и т.д. При этом отмеченная выше для полиномов гибкость этой параметризации сохраняется в общем случае и оказывается очень полезной.

Опорные ординаты обладают следующими очевидными достоинствами: имеют все одинаковую размерность (когда речь идет об аппроксимации функциональной зависимости между физическими величинами), значение функции при фиксированном аргументе является их однородной функцией первого порядка, если

$$f(z; \lambda f_1, \lambda f_2, \dots, \lambda f_L) = \lambda f(z; f_1, f_2, \dots, f_L), \quad (1.11)$$

т.е. умножение всех опорных ординат на одно и то же число эквивалентно умножению всей функции на то же число (обратное свойство, разумеется, тривиально и выполняется всегда).

Для всех рассматриваемых классов аппроксимирующих функций свойство (1.11) выполняется. Легко, однако, привести примеры, когда оно нарушается. Это, в частности, относится ко всем функциям с "неподвижной точкой", которые при любом выборе параметров имеют постоянное значение при некотором значении аргумента, например $\exp(a_1 z + a_2 z^2 + \dots + a_n z^n)$ при $z = 0$.

Зависимость функции от опорных ординат также и симметрична (разумеется, если переставлять вместе с ординатами и значения опорных абсцисс).

Однородность функциональной зависимости от опорных ординат (1.11) представляет собой весьма важное свойство аппроксиманты. По теореме Эйлера об однородных функциях первого порядка справедливо следующее их представление:

$$f(z; f_1, f_2, \dots, f_L) = \sum_{\nu=1}^L f_{\nu} \frac{\partial f(z)}{\partial f_{\nu}}. \quad (1.12)$$

Это соотношение, вообще говоря, не линейно по f_{ν} , поскольку производные также могут зависеть от f_{ν} .

Производные $\partial f(z)/\partial f_{\nu}$ естественно называть коэффициентами чувствительности функции к значениям опорных ординат. Эти коэффициенты сами являются функциями переменного z , но при фиксированных абсциссах z_{ν} обязательно зависят от f_{ν} . Так, в случае полиномиальной аппроксимации дифференцирование выражения (1.9) по P_{ν} дает полиномы, зависящие при заданном N только от выбора z_{ν} :

$$\frac{\partial P_N(z)}{\partial P_{\nu}} = \prod_{\mu \neq \nu} \frac{(z - z_{\mu})}{(z_{\nu} - z_{\mu})}. \quad (1.13)$$

Следует отметить также общее свойство коэффициентов чувствительности, вытекающее непосредственно из определения опорных ординат как независимых параметров:

$$\left. \frac{\partial f(z)}{\partial f_{\nu}} \right|_{z=z_{\mu}} = \delta_{\mu\nu}. \quad (1.14)$$

ГЛАВА 2. ПАДЕ-АППРОКСИМАЦИЯ, ОПРЕДЕЛЕНИЕ И ОСНОВНЫЕ СВОЙСТВА

2.1. ТЕРМИНОЛОГИЯ И ОБОЗНАЧЕНИЯ

Функцию, являющуюся отношением двух полиномов, одни авторы (см., например, [18, 22]) называют дробно-рациональной функцией, другие (см., например, [23, 24]) – просто рациональной функцией. В свою очередь, функции, не имеющие в конечной области других особенностей, кроме полюсов, в одних книгах называются рациональными [17, 22], а в других – мероморфными [23, 24], что в переводе означает "дробные". Везде в дальнейшем под рациональной функцией будем понимать отношение полиномов.

Некоторые сложности представляет также употребление терминов "интерполяция", "аппроксимация" и их производных. Под интерполяцией всегда подразумеваем проведение кривой по точкам, т.е. построение

ние функции, принимающей при некоторых значениях аргумента заданные значения, а под аппроксимацией — любые способы построения приближения к функции, в том числе и интерполяцию. Такой же смысл имеет термин "аппроксиманта" — приближающая функция любого типа.

В дальнейшем будем использовать следующие обозначения:

$$f^{[N, M]}(z) = \frac{P_N(z)}{Q_M(z)} = \sum_{n=0}^N p_N^n z^n / \sum_{m=0}^M q_M^m z^m \quad (2.1)$$

— рациональная функция, в числителе которой полином $P_N(z)$ степени N , а в знаменателе — полином $Q_M(z)$ степени M ; $L = N + M + 1$ — ранг приближения. Иногда под рангом приближения будем подразумевать не только полное число его параметров L , но и его структуру $[N, M]$, т.е. степени числителя и знаменателя. Поскольку при одной и той же степени числителя могут быть приближения с различными степенями знаменателей и наоборот, там, где это необходимо, будем использовать обозначение $f^{[N, M]}(z) = P_{N, M}(z)/Q_{N, M}(z)$, сохраняя прежний смысл индексов N и M . В некоторых случаях нет необходимости указывать степени числителя и знаменателя, поскольку они однозначно определяются начальным приближением, цепочкой рекуррентного построения и рангом приближения L : тогда используется обозначение $f^L(z) = P_{(L)}(z)/Q_{(L)}(z)$. Все рекуррентные соотношения одинаковы для числителей и знаменателей рациональных приближений; поэтому в них вместо $P_{N, M}(z)$ и $Q_{N, M}(z)$ употребляется единое обозначение $R_{N, M}(z)$. Коэффициенты полиномов $P_N(z)$ и $Q_M(z)$ определены равенством (2.1). Там, где это необходимо, будут использоваться также обозначения $p_{N, M}^n$, $q_{N, M}^m$, $p_{(L)}^n$ и $q_{(L)}^m$. Коэффициенты рекуррентных соотношений зависят от цепочки построения и степеней числителя и знаменателя входящих в них приближений; ниже употребляются одни и те же буквы α и β для обозначения разных коэффициентов в различных соотношениях. Из контекста всегда ясно, к какому соотношению и к какому набору индексов относятся приводимые формулы для α и β .

Основы теории аппроксимации рациональными функциями заложены еще в работах Коши, Якоби, Фробениуса (см. [1]), а многие важные результаты получены в теории цепных дробей и методе обратных разностей. Французский математик Паде, по-видимому, их не знал и дал в своей диссертации первое систематическое изложение теории, которая в последние десятилетия интенсивно используется в различных физических и вычислительных задачах. Исторический обзор развития теории дан в статье [3, с.1].

Существуют различные обобщения паде-аппроксимации, определяемые, в частности, для операторных и матричных переменных, а также для функций многих переменных. Однако в этой книге мы не будем на них останавливаться.

2.2. ПРИБЛИЖЕНИЕ ПАДЕ ПЕРВОГО РОДА

Приближением Паде первого рода (Паде-1) к аналитической функции $f(z)$ называется рациональная функция* (2.1), для которой первые $L = N + M + 1$ членов разложения в ряд Тейлора по степеням z совпадают с соответствующими членами ряда Тейлора функции $f(z)$, или, что то же самое, функция, имеющая с $f(z)$ $L - 1$ одинаковую низшую производную при $z = 0$, т.е. такая, что

$$\left. \frac{d^k f^L(z)}{dz^k} \right|_{z=0} = \left. \frac{d^k f(z)}{dz^k} \right|_{z=0}, \quad k = 0, 1, \dots, L - 1. \quad (2.2)$$

Введем обозначение a_k для коэффициентов формального разложения функции $f(z)$ в степенной ряд:

$$f(z) = \sum_{k=0}^{\infty} \left. \frac{d^k f(z)}{dz^k} \right|_{z=0} \frac{z^k}{k!} = \sum_{k=0}^{\infty} a_k z^k. \quad (2.3)$$

В пределах круга сходимости ряда Тейлора функции $Q^{-1}(z)$ условия (2.2) эквивалентны следующей системе линейных уравнений для определения коэффициентов полиномов в числителе и знаменателе правой части (2.1):

$$\left. \begin{aligned} \sum_{m=0}^M q_M^m a_{n-m} &= p_N^n, \quad 0 \leq n \leq N; \\ \sum_{m=0}^n q_M^n a_{n-m} &= 0, \quad N < n < L - 1. \end{aligned} \right\} \quad (2.4)$$

Решение этой системы можно найти с помощью простых рекуррентных соотношений, не прибегая к численному обращению матриц.

В отличие от задачи приближенного описания известных функций в любой из решаемых нами (см. гл. 7) задач известны лишь числовые значения конечного числа коэффициентов a_k , поэтому вопрос о сходимости полученных приближений решается так, как это обычно делается при решении физических задач с помощью теории возмущений — просто сравнением последовательных приближений. При этом использование паде-аппроксимации часто приводит к результатам, которые могут показаться парадоксальными: ряд теории возмущений расходится, и его частичные суммы смысла не имеют, т.е. разложение (2.3) носит формальный характер, а построенные по нему приближения Паде сходятся.

Перечислим и проиллюстрируем примерами основные преимущества приближения Паде-1 перед частичными суммами ряда Тейлора, на основании которых это приближение строится.

*Такое приближение называют также аппроксимацией рациональными функциями со свободными (плавающими) полюсами в отличие от аппроксимации рациональными функциями с заданными полюсами [23].

1. Область сходимости приближения Паде-1 гораздо шире, чем у степенного ряда. Рассмотрим функцию $f(z) = \ln(1+z)$, для которой точка $z = -1$ является особой, а именно точкой ветвления. Ряд Тейлора любой функции сходится лишь в круге, радиус которого равен расстоянию от начала координат до ближайшей особой точки, поэтому ряд

$$\ln(1+z) = z - z^2/2 + z^3/3 - z^4/4 + z^5/5 + \dots \quad (2.5)$$

можно использовать для приближенного вычисления функции лишь при $|z| < 1$. Выпишем несколько первых приближений Паде-1 для этой функции

$$\left. \begin{aligned} f^{[1,0]}(z) &= z; & f^{[1,1]}(z) &= \frac{2z}{z+2}; & f^{[2,1]}(z) &= \frac{z(z+6)}{2(2z+3)}; \\ f^{[2,2]}(z) &= \frac{3z(z+2)}{z^2+6z+6}; & f^{[3,2]}(z) &= \frac{z(z^2+21z+30)}{3(3z^2+12z+10)}. \end{aligned} \right\} \quad (2.6)$$

Обозначим сумму первых λ членов ряда (2.5) S_λ , а значение построенного по ним приближения f^λ , где $\lambda = N+M$ ($a_0 \neq 0$). Сравнение этих величин друг с другом и с точными значениями приводится ниже:

z	$\ln(1+z)$	S_1	f^1	S_2	f^2
0,5	0,4055	0,5	0,5	0,375	0,400
2,0	1,0986	2,0	2,0	0	1
S_3	f^3	S_4	f^4	S_5	f^5
0,4166	0,4062	0,401	0,4050	0,40629	0,40550
2,6667	1,1429	-1,333	1,0909	5,0667	1,1014

Видно, что приближения (2.6) позволяют вычислять $\ln(1+z)$ и при $z = 2$, когда ряд (2.5) расходится.

2. В общей части областей сходимости ряда Тейлора и приближения Паде-1 последнее сходится быстрее. Данные п. 1 подтверждают это. Рассмотрим другой пример. Приближение Паде-1, построенное по трем первым членам ряда Тейлора, для функции $f(z) = \sin z = z - z^3/6 + z^5/120 + \dots$ имеет вид

$$f^{[3,2]}(z) = (60z - 7z^3)/(60 + 3z^2). \quad (2.7)$$

При постановке в (2.7) $z = \pi/4 \approx 0,7854$ получается $\sin(\pi/4) \approx 0,7071$ со всеми четырьмя верными знаками, в то время как первые три члена разложения в ряд дают $\sin(\pi/4) \approx 0,691$. Таким образом, используя периодичность синуса и соотношение $\sin(\pi/2 - z) = \sqrt{1 - \sin^2 z}$, с помощью приближения (2.7) можно построить полную таблицу четырехзначных значений этой функции. Благодаря быстрой сходимости приближение Паде стало в последнее время основным средством для расчета значений элементарных и специальных функций в вычислительной математике [25].

3. Паде-аппроксиманта моделирует полюсы приближаемых функций*. Рассмотрим последовательные паде-аппроксиманты для мероморфной функции $f(z) = \operatorname{tg} z$, имеющей полюсы при $z_n = \pi(2n + 1)/2$:

$$f^{[1,0]}(z) = z; \quad f^{[1,2]}(z) = \frac{3z}{3 - z^2}; \quad f^{[3,2]}(z) = \frac{z(15 - z^2)}{15 - 6z^2}.$$

Уже $f^{[3=2]}(z)$ дает для положения первого полюса $z = \sqrt{5/2} \approx 1,581$, что совпадает с $\pi/2$ в пределах погрешности 1%. С увеличением ранга приближения появляются новые полюсы, а предыдущие воспроизводят все более точно.

2.3. ЗАДАЧА АНАЛИТИЧЕСКОГО ПРОДОЛЖЕНИЯ ФУНКЦИИ И СХОДИМОСТЬ РАЦИОНАЛЬНОЙ АППРОКСИМАНТЫ

Приведенные выше примеры показывают, что использование приближения Паде позволяет выйти за пределы области сходимости ряда Тейлора. Остановимся на этом несколько подробнее.

Принципиальную возможность выхода за пределы области сходимости ряда Тейлора дает его аналитическое продолжение. Для этого надо вычислить с помощью ряда значения функции и ее производных в точке z_0 , лежащей внутри области сходимости, и перейти к ряду Тейлора относительно точки z_0 по степеням $t = z - z_0$, т.е. фактически сделать замену переменной $z = t + z_0$ в исходном ряду. Поясним это на примере функции $1/(1+z)$. Особой точкой этой функции является полюс при $z = -1$, поэтому ряд Тейлора для нее $\sum_{n=0}^{\infty} (-1)^n z^n$ (геометрический

ряд) сходится лишь при $|z| < 1$. Переходя к разложению относительно $z_0 = 3/4$, получаем

$$\sum_{n=0}^{\infty} (-1)^n z^n = \sum_{n=0}^{\infty} (-1)^n (t + 3/4)^n = \frac{4}{7} \sum_{n=0}^{\infty} (-1)^n (4t/7)^n.$$

Последний ряд сходится при $|t| < 7/4$, следовательно, с его помощью можно, например, вычислять $1/(1+z)$ и при $z = 2$, поскольку тогда $t = 5/4$. Таким образом, коэффициенты ряда Тейлора в совокупности несут гораздо большую информацию о значениях функции, чем его частичные суммы. Надо только уметь ее извлекать и один из способов сделать это заключается в построении паде-аппроксимант.

Приближение Паде-1 к сумме геометрической прогрессии уже по первым двум членам дает $P_0(z)/Q_1(z) = (1+z)^{-1}$, т.е. точно восстанавливает исходную функцию. В общем случае приближение Паде, конечно, функцию точно не восстанавливает, однако оно позволяет аппроксими-

* Рациональные приближения, моделирующие более сложные особенности функций, рассмотрены в [26].

мировать суммы рядов не только в пределах их кругов сходимости, но и вне их, причем полученные значения отнюдь не стремятся во внешней области к частичным суммам исходных рядов.

Такой подход к суммированию расходящихся рядов был развит еще Эйлером, который не рассматривал сумму ряда как предел последовательности его частичных сумм, а считал, что "сумма некоторого бесконечного ряда есть конечное выражение, из разложения которого возникает этот ряд" (цит. по книге Фихтенгольц Г.Н. Основы математического анализа М.: Физматгиз, 1960. Т. 2. С. 106).

В течение длительного времени этот подход отвергался, и примеры рядов, которые таким образом суммировал Эйлер, во многих учебниках еще недавно приводились в качестве иллюстрации того, как неправильно обращались с бесконечными рядами великие математики XVIII в. Развитая в XX в. теория суммирования расходящихся рядов фактически основана на идеях Эйлера. Рациональные функции — это как раз пример таких "конечных выражений", разложения которых в ряд Тейлора вне круга сходимости дают расходящиеся ряды.

Как уже упоминалось, при решении практических задач сходимость последовательных приближений Паде определяется просто их вычислением, однако даже если они сходятся, то еще не известно, сходятся ли они именно к той функции, по коэффициентам ряда Тейлора которой они построены. В последние два десятилетия теории сходимости таких приближений посвящено множество работ. Обзор состояния теории можно найти, например, в [1]. Суммируем вкратце основные результаты, полученные в этом направлении.

Прежде всего, при достаточно большом ранге приближение Паде точно восстанавливает рациональные функции, в том числе и полиномы. Многочисленные примеры показывают, что с ростом ранга приближения улучшается описание функций в точках их аналитичности, а расположение полюсов аппроксиманты несет информацию об особых точках функции. Например, разрез на участке действительной оси является той областью, в которой скапливаются полюсы аппроксиманты. Пояснить это можно следующим образом: приближение Паде ряда с действительными коэффициентами по построению может иметь на действительной оси только действительные значения и поэтому может моделировать различные значения аппроксимирующей функции при подходе к одной и той же точке разреза с разных сторон только полюсами. Ясно, что такие рассуждения отнюдь не служат строгим доказательством.

В строгой теории сходимости приближений Паде можно выделить три основных направления.

Первое из них дает доказательства сходимости частных последовательностей приближений при некоторых ограничениях на функцию. Простейшая из теорем этого типа утверждает, что если в круге радиусом ρ находится M полюсов (с учетом их кратности) функции $f(z)$ и нет других особенностей, то последовательность приближений $f^{[N, M]}(z)$

при $N \rightarrow \infty$ сходится к $f(z)$ в этом круге. В более сложных теоремах при некоторых ограничениях на коэффициенты ряда Тейлора функции $f(z)$ доказывалась сходимость во всей плоскости.

Следующая совокупность теорем получена в том случае, когда рассматриваются подпоследовательности приближений, распределение полюсов которых удовлетворяет некоторым условиям, и доказывалась сходимость таких подпоследовательностей.

Наконец, третий подход к проблеме сходимости состоит в доказательстве того, что если полюсы аппроксиманты нарушают сходимость, то область нарушения в некотором смысле мала.

Наиболее полезно, по-видимому, так называемое утверждение Бейкера (Baker), у которого нет строгого доказательства, но нет и опровергающих контрпримеров: если функция регулярна при $|z| \leq 1$, кроме полюсов и точки $z = 1$, тогда, по крайней мере, последовательность $P_N(z)/Q_N(z)$ при $N \rightarrow \infty$ сходится к этой функции внутри круга $|z| \leq 1$.

2.4. ПРИБЛИЖЕНИЕ ПАДЕ ВТОРОГО РОДА

Приближением Паде второго рода (Паде-2) для функции $f(z)$ называется рациональная функция $f^{[N, M]}(z)$ вида (2.1) такая, что ее значения при L значениях аргумента совпадают со значениями $f(z)$ в тех же точках:

$$f^{[N, M]}(z_i) = f(z_i), \quad i = 1, 2, \dots, L, \quad L = N + M + 1. \quad (2.8)$$

Равенства (2.8) приводят к следующей системе линейных уравнений для определения коэффициентов полиномов в числителе и знаменателе аппроксиманты:

$$\sum_{n=0}^N p_n z_i^n - f(z_i) \sum_{m=0}^M q_m z_i^m = 0. \quad (2.9)$$

Методы решения этой системы с помощью рекуррентных соотношений будут описаны в гл. 3. Построение приближения Паде-2 будем также называть рациональной интерполяцией. Из определения приближения Паде-2 следует, что для него весьма естественна параметризация опорными ординатами, описанная в § 1.4.

Для построения приближения Паде-1 необходимо уметь вычислять в одной точке производные аппроксимирующей функции высокого порядка. При табличном задании функции этот метод практически не применим, особенно если опорные значения имеют заметную погрешность. Именно поточечно, таблично заданные функции являются объектом аппроксимации приближением Паде-2. Как будет показано в гл. 5, рациональную интерполяцию можно эффективно использовать для аппроксимации функций в тех случаях, когда ранг аппроксиманты меньше числа

точек, в которых задана исходная функция, и выбор наилучшего приближения производится с помощью метода наименьших квадратов или другими методами оптимизации.

Приближение Паде-2 во многих случаях не только обеспечивает получение интерполяционных формул, гораздо более эффективных, чем, скажем, формулы полиномиальной интерполяции Лагранжа, но и позволяет экстраполировать значения функции, заданной на ограниченном интервале, за его пределы, чего практически никогда не позволяет сделать полиномиальное приближение. Что особенно нетривиально, приближение Паде-2 дает возможность, например, довольно точно устанавливать положение действительных полюсов интерполируемой функции, лежащих за пределами интервала интерполяции.

Примеры такого рода будут приведены в гл. 5.

Одна из областей применимости приближения Паде-2 — изучение в различных задачах математической физики аналитических свойств функций, которые в силу физической природы рассматриваемого явления известны лишь при целочисленных значениях аргумента. Таковы, например, зависимости амплитуд от орбитального момента в квантовой теории (см. [5]).

Оригинальный способ использования приближения Паде-2 для ускорения сходимости числовых рядов изложен в [5, 6]. Суть его ясна из следующего примера. Рассмотрим ряд

$$S = \sum_{n=1}^{\infty} 1/n^2$$

и обозначим S_N его частичные суммы до $n = N$ включительно. Рассмотрим S_N как функцию переменной $z = N^{-1}$, т.е. $S_N = f(z)$. Тогда точная полная сумма $S = S_{\infty} = f(0)$. Если взять всего три первые частичные суммы $S_1 = 1$, $S_2 = 1,25$, $S_3 = 1,361$, построить по ним приближение Паде-2 для $f(z)$ и вычислить его значение при $z = 0$, то получится $f^{[1,1]}(0) = 1,65$ при точном числовом значении $1,645 \dots$ Для обеспечения такой точности непосредственным суммированием пришлось бы взять около 20 членов исходного ряда — весьма значительное ускорение сходимости.

В том же обзоре [5] описан итерационный метод решения трансцендентных и алгебраических уравнений вида $f(z) = 0$ с помощью построения рациональной интерполяции $f^{1,1}(z) = (Az + B)/(Cz + D)$, проходящей через точки $f(z_1)$, $f(z_{l+1})$, $f(z_{l+2})$. Следующая итерация состоит в построении такого же приближения с заменой $z_l \rightarrow z_{l+3} = -A/B$ и т.д. Этот метод обеспечивает кубическую сходимость — если точки z_l , z_{l+1} , z_{l+2} отличаются от истинного корня на Δz_l , Δz_{l+1} , Δz_{l+2} , то для следующего приближения $\Delta z_{l+3} \sim \Delta z_l \Delta z_{l+1} \Delta z_{l+2}$. Таким образом, уменьшение погрешности от 0,1 до 10^{-9} достигается за два шага, при этом с одинаковым успехом ищутся как действительные, так и комплексные корни уравнений.

Очень важно для наших целей использование приближения Паде-2 для аппроксимации сложных экспериментальных зависимостей — резонансных кривых с числом параметров до нескольких десятков.

2.5. ИНВАРИАНТНОСТЬ ПРИБЛИЖЕНИЙ ПАДЕ. ПРЕОБРАЗОВАНИЯ ЭЙЛЕРА

Рассмотрим теперь свойства инвариантности приближений Паде. Вопрос ставится так: какие преобразования можно делать с функцией $f(z)$ и переменной z , чтобы соответственно преобразованное приближение Паде было приближением того же ранга $[N, M]$ для преобразованной функции относительно преобразованной переменной?

Прежде всего, заметим, что в отличие от полиномиальных приближений рациональные нельзя складывать — если сумма полиномов $P_N(z)$ и $P'_N(z)$ степени N есть полином $P''_N(z)$ той же степени, то сумма $P''_N(z)/Q_M(z)$ рациональных функций $P_N(z)/Q_M(z)$ и $P'_N(z)/Q'_M(z)$ уже не есть рациональная функция того же ранга, если $Q_M(z) \neq Q'_M(z)$. Следовательно, рациональное приближение заданного ранга к сумме функций не есть сумма приближений того же ранга к каждому из слагаемых. Такая неаддитивность, конечно, к числу достоинств рациональной аппроксимации не относится, однако она не представляет принципиальной трудности, поскольку вся необходимая информация об аппроксиманте суммы содержится в аппроксимантах слагаемых. В случае приближения Паде-1 можно, например, разложить слагаемые в ряды, сложить ряды и по суммарному ряду построить рациональное приближение требуемого ранга $[N, M]$. Для приближения Паде-2, даже не рассматривая алгебраического решения этой проблемы, по сумме рациональных функций, используя машинные алгоритмы, строят рациональное приближение заданного ранга везде, где это необходимо (см. § 7.2).

При рациональной интерполяции очевидна инвариантность приближений относительно линейного преобразования (сдвига и растяжения) независимой переменной — если приближение $P_N(z)/Q_M(z)$ проходит через L точек $f(z_i)$, то $T_N(t)/S_M(t) = P_N(at + \beta)/Q_M(at + \beta)$ проходит через L точек $\varphi(t_i)$, где $\varphi(t) = f(z(t))$. Такая инвариантность используется при поиске корней полиномов числителя и знаменателя приближения Паде-2. Обычно это удобнее делать, отображая сначала исходный интервал изменения аргумента на $[-1, 1]$ или $[0, 1]$ с помощью линейного преобразования.

При более общем дробно-линейном преобразовании $z = (at + \beta)/(\gamma t + \delta)$ получим

$$P_N(z)/Q_M(z) = (\gamma t + \delta)^M T_N(t)/(\gamma t + \delta)^N S_M(t).$$

Построенная таким образом рациональная функция аргумента t проходит через все L точек $\varphi(t_i)$, где $\varphi(t) = f[(at + \beta)/(\gamma t + \delta)]$, но будет функцией того же ранга, что и $f^{[N, M]}(z)$, лишь в том случае, когда

$N = M$, т.е. только диагональные приближения инвариантны относительно дробно-линейных преобразований независимой переменной. Очевидно, что такой же вывод получится при дробно-линейном преобразовании функции $\varphi(z) = [\alpha f(x) + \beta]/[\gamma f(z) + \delta]$, т.е. если подвергнуть такому же преобразованию диагональное рациональное приближение $P_N(z)/Q_N(z)$, получим диагональное приближение $T_N(z)/S_N(z)$, которое проходит через L точек $\varphi(z_i)$. Отметим, что при интерполяции любые преобразования, которые делаются с исходной и интерполирующей функциями, дают решение задачи интерполяции с тем же числом узлов, но рациональное приближение считается инвариантным относительно преобразования лишь в том случае, если при этом не меняется его ранг $[N, M]$.

Задача аппроксимации функции по конечному числу членов ее ряда Тейлора в этом отношении принципиально отличается от задачи интерполяции. Как и в § 2.2, рассмотрим аналитическое продолжение ряда Тейлора функции $f(z)$, т.е. перейдем от разложения функции вблизи $z = 0$ к разложению вблизи $z = z_0$. Обозначив $t = z - z_0$, получим:

$$f(z) = \sum_i a_i z^i = \sum_k b_k t^k, \quad b_k = \sum_{i=k}^{\infty} z_0^{i-k} a_i \frac{i!}{k!(i-k)!}.$$

Следовательно, при таком преобразовании любое b_k зависит от всех a_i , $i \geq k$, тогда как при построении аппроксиманты используется лишь конечное число a_i . Отсюда следует, что построить аппроксиманту аналитического продолжения ряда Тейлора функции, имея лишь аппроксиманту исходного ряда, принципиально невозможно — необходимо привлекать информацию об отброшенных членах ряда Тейлора. Поэтому и приближение Паде-1, в частности, не может быть инвариантным относительно преобразования сдвига независимой переменной $z = t + z_0$.

Однако существует важный класс преобразований, а именно преобразования Эйлера, при которых конечное число членов разложения по t выражается через такое же число членов разложения по z :

$$\left. \begin{aligned} z &= at/(1 + \beta t); \quad f(z) = \sum_i a_i z^i; \\ \varphi(t) &= f(z(t)) = \sum_k c_k t^k; \\ c_0 &= a_0, \\ c_k &= \sum_{n=1}^k (-1)^{k-n} \beta^{k-n} a_n \alpha^n \frac{(k-1)!}{(n-1)!(k-n)!}. \end{aligned} \right\} \quad (2.10)$$

По определению приближения Паде-1 $P_N(z)/Q_M(z) = f(z) + O(z^L)$, где $O(z^L)$ при $z \rightarrow 0$ убывает, как z^L . Поскольку $O(z^L)$ при подстановке $z = at/(1 + \beta t)$ переходит в $O(t^L)$, рациональная функция ар-

гумента t

$$(1 + \beta t)^M T_N(t) / (1 + \beta t)^N S_M(t) = P_N[z(t)] / Q_M[z(t)]$$

будет аппроксимировать ряд Тейлора функции $\varphi(t)$ с точностью до $O(t^L)$. Отсюда следует, что диагональные приближения инвариантны относительно преобразований Эйлера. Эти преобразования, как и аналитическое продолжение ряда Тейлора, позволяют выйти за пределы круга сходимости последнего. Для пояснения снова рассмотрим функцию $f(z) = \ln(1+z)$ и преобразование переменной $z = t/(1-t)$, которое приводит к новой функции $\varphi(t) = -\ln(1-t)$. Исходная функция аналитична во всей плоскости комплексной переменной z , кроме разреза на действительной оси $[-\infty, 1]$, а функция $\varphi(t)$ аналитична во всей плоскости, кроме разреза $[1, \infty]$. Ряды Тейлора каждой из функций $f(z)$ и $\varphi(t)$:

$$f(z) = z - z^2/2 + z^3/3 - z^4/4 + \dots;$$

$$\varphi(t) = -(t + t^2/2 + t^3/3 + t^4/4 + \dots)$$

сходятся в круге единичного радиуса, но кругу $|t| < 1$ в плоскости z соответствует область $\operatorname{Re} z > -1/2$, и преобразованный ряд позволяет вычислять $\ln(1+z)$, в частности, при любом действительном $z > -1/2$. Переходить от суммы конечного числа членов ряда по z к такому же числу членов ряда по t можно, не зная всего ряда, а лишь используя формулу (2.10) для связи коэффициентов c_k и a_k .

Итак, даже в случае полиномиальной аппроксимации преобразование Эйлера с соответствующими параметрами позволяет вычислять аппроксиманту функции вне круга сходимости ряда Тейлора, переходя от

$$P_L(z) = \sum_{n=0}^L a_n z^n \text{ к аппроксиманте } F_L(t) = \sum_{n=0}^L c_n [t(z)]^n = H_L(z) /$$

$/(a - \beta z)^L$. Эта аппроксиманта — уже не полиномиальная функция z , а рациональная с равными степенями полиномов числителя и знаменателя. Диагональное приближение Паде, которое описывает то же число членов ряда Тейлора функции $f(z)$, будет при $L - 1 = 2N$ иметь вид: $P_N(z)/Q_N(z)$ со степенями полиномов числителя и знаменателя, вдвое меньшими, чем у $H_L(z)/(a - \beta z)^L$. Как уже говорилось, преобразование Эйлера такого приближения приводит к аппроксиманте $T_N(t)/S_N(t)$, которая равна в точке $t = t(z)$ просто $P_N(z)/Q_N(z)$. Поэтому, если при некотором t последовательные приближения $T_N(t)/S_N(t)$ при $N \rightarrow \infty$ сходятся к функции $\varphi(t) \equiv f[z(t)]$, то приближения $P_N(z)/Q_N(z)$ сходятся к $f(z)$ при любом выборе параметров преобразований Эйлера. Эти рассуждения помогают понять, почему с помощью приближения Паде можно суммировать расходящиеся ряды — преобразованиями Эйлера их можно свести к сходящимся, а приближение Паде ин-

вариантно относительно преобразований Эйлера. Кроме того, становится понятнее, почему область сходимости приближения Паде гораздо шире области сходимости степенного ряда.

Рассмотрим теперь дробно-линейное преобразование не аргумента, а функции $\varphi(z)$ и применим его к диагональному приближению Паде-1 $f^{[N, N]}(z)$ для функции $f(z)$:

$$\varphi^{[N, N]}(z) = [\alpha P_N(z) + \beta Q_N(z)] / [\gamma P_N(z) + \delta Q_N(z)].$$

Инвариантность приближения Паде-2 относительно такого преобразования очевидна. В случае приближения Паде-1, используя равенство $f(z) = f^{[N, N]}(z) + O(z^L)$, нетрудно показать, что $\varphi(z) = \varphi^{[N, N]}(z) + O(z^L)$, т.е. диагональное приближение преобразованной функции есть преобразование диагонального приближения к исходной функции. Это преобразование, в частности, позволяет связать аппроксиманты функций $f(z)$ и $[f(z)]^{-1}$: выбирая $\alpha = \delta = 0$, $\beta = \gamma = 1$, получаем, что функции $[f(z)]^{-1}$ соответствует аппроксиманта $[f^{[N, N]}(z)]^{-1} = Q_N(z)/P_N(z)$.

ГЛАВА 3. АЛГОРИТМЫ РАЦИОНАЛЬНОЙ ИНТЕРПОЛЯЦИИ

3.1. ДЕТЕРМИНАНТНОЕ РЕШЕНИЕ ЗАДАЧИ ИНТЕРПОЛЯЦИИ

Первым этапом обработки и анализа экспериментальных данных является построение гладкой кривой, которая служит оценкой истинной физической зависимости. Широкое использование на этом этапе рациональных приближений становится возможным благодаря методу перебора, который позволяет получать оценки параметров аппроксиманты, не прибегая к решению системы нелинейных уравнений или использованию более сложных методов оптимизации. Идея метода перебора состоит в том, что среди множества интерполирующих кривых ранга L , построенных по различным наборам L экспериментальных точек из общего числа N_{ex} , при $N_{\text{ex}} \gg L$ будут и такие кривые, которые достаточно близки к статистически оптимальной. Подробнее этот метод изложен в гл. 5.

Метод перебора становится технически возможным лишь благодаря использованию соответствующих алгоритмов рекуррентного построения рациональной интерполяции. Эти алгоритмы аналогично формуле Ньютона [21] в задаче полиномиальной интерполяции позволяют увеличивать на единицу число узлов интерполяции, т.е. ранг приближения, при использовании лишь двух предыдущих приближений. При этом от добавляемой точки зависят лишь коэффициенты рекуррентного соотношения. Получение алгоритмов такого рода, несимметричных относительно последовательности выбора узлов, но удобных для машинного построения большого числа вариантов интерполяции и вычисления производных, является основной целью настоящей главы.

Задача интерполяции рациональными функциями, т.е. приближение Паде-2, имеет при заданных степенях полиномов числителя и знаменателя аппроксиманты N и M и заданном числе узлов интерполяции $L = N + M + 1$ не больше одного решения. Действительно, если отношения полиномов $P_N(z)/Q_M(z)$ и $P'_N(z)/Q'_M(z)$ принимают значения $f(z_i)$ в L точках z_i , то полином $P_N(z) Q'_M(z) - P'_N(z) Q_M(z)$ степени $N + M$ будет равен нулю в этих $N + M + 1$ точках и поэтому равен нулю тождественно. Следовательно, полиномы $P_N(z)$, $Q_M(z)$ и $P'_N(z)$, $Q'_M(z)$ могут различаться лишь общим множителем a , т.е. $P_N(z) = aP'_N(z)$, $Q_M(z) = aQ'_M(z)$, который будем называть нормировочным. При обработке резонансных кривых этот множитель удобно выбирать так, чтобы коэффициент при старшей степени знаменателя был равен единице. Случаи, когда построение рациональной интерполяции с заданными N и M невозможно, будут рассмотрены в § 3.6.

Детерминантное решение этой задачи было получено Якоби. Проще всего к нему прийти, присоединяя к системе (2.9) еще одно уравнение, которое следует из определения интерполирующей функции:

$$\sum_{n=0}^N p_N^n z^n - f^{[N,M]}(z) \sum_{m=0}^M q_M^m z^m = 0.$$

Получающаяся при этом система $L + 1$ линейных уравнений относительно $L + 1$ неизвестных p_N^n и q_M^m имеет нетривиальное решение, если ее детерминант равен нулю. Используя простейшие свойства детерминантов, последнюю строку этого детерминанта можно представить в виде суммы двух строк:

$$\begin{aligned} & (z^N, z^{N-1}, \dots, 1; -f^{[N,M]}(z), -zf^{[N,M]}(z), \dots, -z^M f^{[N,M]}(z)) = \\ & = (z^N, z^{N-1}, \dots, 1; 0, 0, \dots, 0) - (0, 0, \dots, 0; f^{[N,M]}(z), \\ & zf^{[N,M]}(z), \dots, z^M f^{[N,M]}(z)), \end{aligned}$$

затем перейти к сумме детерминантов и вынести в одном из них общий множитель $f^{[N,M]}(z)$. В результате получится следующее выражение для $f^{[N,M]}(z)$:

$$f^{[N,M]}(z) = - \begin{vmatrix} z_1^N & z_1^{N-1} & \dots & 1 & f_1 & z_1 f_1 & \dots & z_1^M f_1 \\ z_2^N & z_2^{N-1} & \dots & 1 & f_2 & z_2 f_2 & \dots & z_2^M f_2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ z_L^N & z_L^{N-1} & \dots & 1 & f_L & z_L f_L & \dots & z_L^M f_L \\ z^N & z^{N-1} & \dots & 1 & 0 & 0 & \dots & 0 \end{vmatrix}$$

$$\begin{vmatrix}
 z_1^N & z_1^{N-1} & \dots & 1 & f_1 & z_1 & f_1 & \dots & z_1^M & f_1 \\
 z_2^N & z_2^{N-1} & \dots & 1 & f_2 & z_2 & f_2 & \dots & z_2^M & f_2 \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 z_L^N & z_L^{N-1} & \dots & 1 & f_L & z_L & f_L & \dots & z_L^M & f_L \\
 0 & 0 & \dots & 0 & 1 & z & \dots & \dots & z^M &
 \end{vmatrix}
 \equiv - \frac{\begin{vmatrix} C_{N,M}^L \\ \tilde{z}, 0 \end{vmatrix}}{\begin{vmatrix} C_{N,M}^L \\ 0, \tilde{z} \end{vmatrix}}.$$

(3.1)

Здесь $f_i \equiv f(z_i)$; $C_{N,M}^L$ — прямоугольная матрица, структура которой ясна из (3.1).

В дальнейшем мы будем использовать следующее обозначение: $C_{N(k)M(l)}^{L(i)}$ — матрица, которая получается из $C_{N,M}^L$ отбрасыванием i -й строки и столбцов с z_n^k и $f_n z_n^l$. Если отбрасываются последние строки или крайние столбцы, будем писать просто $C_{N_1 M_1}^{L_1}$, при этом L_1 совсем не обязательно равно $N_1 + M_1 + 1$. Строки, у которых только первый или последний элемент равен единице, а остальные — нулю, обозначим соответственно $(e, 0)$ и $(0, e)$. Последние строки в числителе и знаменателе (3.1) обозначены $(\tilde{z}, 0)$ $(0, \tilde{z})$. В дальнейшем такие строки всегда будут присоединяться к матрицам $C_{N(k), M(l)}^{L(i)}$, поэтому число нулей в строке $(0, \tilde{z})$ всегда равно числу столбцов без множителя f_n , число нулей в строке $(\tilde{z}, 0)$ всегда равно числу столбцов с этим множителем, а степени z в ненулевых элементах этих строк всегда совпадают со степенями z_n соответствующих столбцов тех матриц, к которым они присоединяются. В этих обозначениях нормировка на $d_M^M = 1$ получится, если числитель и знаменатель выражения (3.1) разделить на $|C_{N, M-1}^L|$.

Детерминантное решение позволяет получить не слишком сложные формулы для $P_N(z)$ и $Q_M(z)$ лишь при малых N и M , поскольку ранг соответствующих детерминантов равен $N + M + 2$. Формулы можно несколько упростить, если использовать разложение детерминантов по минорам (теорема Лапласа) и свойства определителей Вандермонда [27]. Используя принятые выше обозначения, выпишем известное соотношение для определителя Вандермонда W_{N+1} :

$$W_{N+1} = |C_{N,0}^{N+1}| = \prod_{1 \leq i < k \leq N+1} (z_i - z_k) \tag{3.2}$$

и определителя, получающегося из него отбрасыванием одного столбца и последней строки [27]:

$$W_N^{(l)} = |C_{N(l),0}^N| = S_{N(l)} \prod_{1 \leq i < k \leq N} (z_i - z_k). \quad (3.3)$$

Здесь $S_{N(l)} = \sum_{\gamma} \prod_{i=1}^{N-l} z_i^{\gamma_i}$, т.е. $S_{N(l)}$ есть сумма всех возможных различных произведений $N-l$ абсцисс z_i из общего числа N абсцисс z_1, z_2, \dots, z_N . В частности, $S_{N(N)} = 1$; $S_{N(N-1)} = \sum_{i=1}^N z_i$; $S_{N(N-2)} = \sum_{1 \leq i < k \leq N} z_i z_k$; $S_{N(0)} = \prod_{i=1}^N z_i$.

Используя теорему Лапласа, разложим определители, входящие в (3.1), по первым $N+1$ столбцам. Миноры, составленные из элементов этих столбцов, являются определителями Вандермонда, дополнительные к ним миноры отличаются от определителей Вандермонда лишь домножением каждой строки на соответствующее f_i , поэтому, используя (3.2), получаем:

$$\left. \begin{aligned} P_N(z) &= \sum_{\alpha} \prod_{1 \leq i < k \leq N} (z_{\alpha_i} - z) \prod_{l=1}^{M+1} f_{\tilde{\alpha}_l} (z_{\alpha_i} - z_{\alpha_k}) \prod_{1 \leq n < m \leq M+1} \times \\ &\times (z_{\tilde{\alpha}_n} - z_{\tilde{\alpha}_m}) (-1)^{L + \sum_{i=1}^N \alpha_i + B}; \\ Q_M(z) &= \sum_{\alpha} \prod_{1 \leq i < k \leq N+1} (z_{\alpha_i} - z_{\alpha_k}) \prod_{l=1}^M f_{\tilde{\alpha}_l} (z_{\tilde{\alpha}_l} - z) \prod_{1 \leq n < m \leq M} \times \\ &\times (z_{\tilde{\alpha}_n} - z_{\tilde{\alpha}_m}) (-1)^{\sum_{i=1}^{N+1} \alpha_i + B} \end{aligned} \right\} \quad (3.4)$$

Здесь B от α_i не зависит; α_i и $\tilde{\alpha}_n$ — номера строк, которые входят в миноры, составленные из первых $N+1$ столбцов и из остальных $M+1$ столбцов соответственно. Суммирование по α ведется по различным разбиениям индексов $1, 2, \dots$ на две группы α и $\tilde{\alpha}$ из N и $M+1$ индексов в случае $P_N(z)$ и из $N+1$ и M индексов в случае $Q_M(z)$. Для иллюстрации выпишем $P_N(z)$ и $Q_M(z)$ в случае $N = M = 1$:

$$\left. \begin{aligned} P_1(z) &= -(z_1 - z)f_2 f_3(z_2 - z_3) + (z_2 - z)f_1 f_3(z_1 - z_3) - \\ &- (z_3 - z)f_1 f_2(z_1 - z_2); \\ Q_1(z) &= (z_1 - z)f_1(z_2 - z_3) - (z_2 - z)f_2(z_1 + z_3) + \\ &+ (z_3 - z)f_3(z_1 - z_2). \end{aligned} \right\} (3.5)$$

Коэффициенты полиномов числителя и знаменателя (p_N^n и q_M^m) аппроксиманты $f^{[N, M]}(z)$ получаются при разложении соответствующих детерминантов выражения (3.1) по элементам последних строк. Используя при этом теорему Лапласа и формулы (3.2) и (3.3), получаем для p_N^n :

$$p_N^n = \sum_{\alpha, \beta} \prod_{i=1}^{N-n} z_{\beta_i} \prod_{1 \leq l < k \leq N} (z_{\alpha_l} - z_{\alpha_k}) \prod_{m=1}^{M+1} f_{\tilde{\alpha}_m} \prod_{1 \leq t < s \leq M+1} \times \\ \times (z_{\tilde{\alpha}_t} - z_{\tilde{\alpha}_s}) (-1)^{L-n + \sum_{k=1}^N \alpha_k + B}, \quad (3.6)$$

где α — набор N различных индексов из L , соответствующий выбору строк минора из первых N столбцов; β — набор $N - n$ индексов из N индексов набора α ; $\tilde{\alpha}$ — набор $M + 1$ индексов из L , дополнительный к α . Суммирование по α — это разложение детерминанта по теореме Лапласа, суммирование по β появляется из-за использования формулы (3.3).

Аналогично для q_M^m получим:

$$q_M^m = \sum_{\alpha, \beta} \prod_{1 \leq i < k \leq N+1} (z_{\alpha_i} - z_{\alpha_k}) \prod_{l=1}^{M-m} z_{\beta_l} \prod_{n=1}^M f_{\tilde{\alpha}_n} \prod_{1 \leq s < t \leq M} \times \\ \times (z_{\tilde{\alpha}_s} - z_{\tilde{\alpha}_t}) (-1)^{m + \sum_{k=1}^{N+1} \alpha_k + B} \quad (3.7)$$

Здесь α — набор $N + 1$ индексов из L ; $\tilde{\alpha}$ — набор M индексов из L , дополнительный к α ; β — набор $M - m$ индексов из M индексов набора $\tilde{\alpha}$.

Выражения (3.3), (3.6), (3.7) довольно компактны, симметричны и удобнее для пользования, чем хорошо известная формула (3.1). Их можно рассматривать как аналог интерполяционной формулы Лагранжа, используемой в полиномиальной интерполяции. Однако с ростом N и M число слагаемых быстро увеличивается, и для получения приближений высоких порядков они все равно неудобны.

3.2. ПРОИЗВОДНЫЕ ПО ОПОРНЫМ ОРДИНАТАМ

Используя свойства детерминантов, можно получить соотношения для элементов рациональной интерполяции, которые понадобятся нам в дальнейшем. Прежде всего, выразим $Q_M(z_i)$ через детерминант ранга $L - 1$. Для этого в знаменатель формулы (3.1) подставим $z = z_i$, вычтем из i -й строки последнюю, домноженную на f_i , затем из каждого столбца, кроме $N + 1$ -го и $N + 2$ -го, вычтем соседний, домноженный на z_i , и из каждой строки, кроме i -й, вынесем $Z_k - z_i$, где k — номер строки. Разложим полученный детерминант по i -й и последней строкам. В результате получим:

$$Q_M(z_i) = \prod_{k \neq i} (-1)^{L+i} (z_k - z_i) |C_{N-1, M-1}^{L(i)}|. \quad (3.8)$$

Исключительно полезно при работе с детерминантами рациональной интерполяции следующее тождество:

$$\begin{vmatrix} A & | & A \\ a & | & c \\ b & | & d \end{vmatrix} = \begin{vmatrix} A & | & A \\ a & | & c \\ d & | & b \end{vmatrix} + \begin{vmatrix} A & | & A \\ a & | & b \\ c & | & d \end{vmatrix}. \quad (3.9)$$

Здесь A — прямоугольная матрица с $L - 1$ строкой и $L + 1$ столбцом; a , b , c и d обозначают строки $(a_1, a_2, \dots, a_L, a_{L+1})$ и т.д. Для доказательства этого соотношения надо применить тождество Сильвестра [27] к каждому из входящих в него детерминантов и привести подобные члены.

Простейшим примером использования этого соотношения может служить вывод формулы для производной рациональной интерполяции по значению функции $f(z)$ в узле интерполяции z_i :

$$\frac{\partial f^{[N, M]}(z)}{\partial f_i} = \left[\frac{\partial P_N(z)}{\partial f_i} Q_M(z) - \frac{\partial Q_M(z)}{\partial f_i} P_N(z) \right] [Q_M(z)]^{-2}. \quad (3.10)$$

Поскольку от f_i зависят только i -е строки детерминантов $P_N(z)$ и $Q_M(z)$, дифференцируя их по f_i и применяя затем тождество (3.9), которое, естественно, справедливо и в том случае, когда детерминанты различаются любыми двумя строками, а не обязательно последними, получаем:

$$\frac{\partial f^{[N, M]}(z)}{\partial f_i} = \frac{\begin{vmatrix} C_{N, M}^L & | & C_{N, M}^{L(i)} \\ 0, \tilde{z}_i & | & (i) \tilde{z}, 0 \\ & | & 0, \tilde{z} \end{vmatrix}}{Q_M^2(z)}. \quad (3.11)$$

Здесь $\begin{vmatrix} C_{N, M}^{L(i)} \\ (i) z, 0 \end{vmatrix}$ — матрица, которая получается из $C_{N, M}^L$ заменой

i -й строки строкой $(\tilde{z}, 0)$. Первый из входящих в (3.11) детерминантов есть $Q_M(z_i)$, а второй получается из него заменой в i -й и последней строках $z_i \rightarrow z$, поэтому, используя (3.8), имеем

$$\frac{\partial f^{[N, M]}(z)}{\partial f_i} = \frac{Q_M^2(z_i)}{Q_M^2(z)} \frac{\prod_{k \neq i} (z - z_k)}{\prod_{k \neq i} (z_i - z_k)}. \quad (3.12)$$

Последнее выражение можно получить и из общих соображений, вообще не рассматривая детерминанты. Именно, поскольку от опорных ординат f_i зависят лишь коэффициенты полиномов $P_N(z)$ и $Q_M(z)$, в знаменателе производной стоит $Q_M^2(z)$, а в числителе — полином степени $N + M$. Этот полином должен обращаться в нуль во всех остальных узлах интерполяции, поскольку изменение одной опорной ординаты не меняет других и $\partial f(z_k)/\partial f_i = \delta_{ik}$. Следовательно, данный полином имеет вид $\text{const} \prod_{k \neq i} (z - z_k)$. Определяя константу из условия

$\partial f(z_i)/\partial f_i = 1$, получаем окончательно (3.12).

В некоторых случаях, например при использовании для оптимизации аппроксиманты метода Ньютона — Рафсона или других методов второго порядка, необходимо уметь вычислять вторые производные аппроксиманты по параметрам. Дифференцируя (3.12) по f_k , находим:

$$\frac{\partial f^{[N, M]}(z)}{\partial f_k \partial f_i} = 2 \frac{\prod_{n \neq i} (z - z_n)}{\prod_{n \neq i} (z_i - z_n)} \frac{Q_M(z_i)}{Q_M^3(z)} \left[\frac{\partial Q_M(z_i)}{\partial f_k} Q_M(z) - \frac{\partial Q_M(z)}{\partial f_k} Q_M(z_i) \right]. \quad (3.13a)$$

Применив к выражению в квадратных скобках соотношение (3.9), получим произведение $Q_M(z_k)$ и детерминанта, который отличается от $Q_M(z_i)$ тем, что вместо k -й строки $(\tilde{z}_k, f_k \tilde{z}_k)$ в нем стоит $(0, \tilde{z})$. Действуя так же, как при выводе формулы (3.8), получим при $i > k$ следующее выражение для второй производной:

$$\frac{\partial f^{2[N, M]}(z)}{\partial f_k \partial f_i} = (-1)^{i+k+N} \cdot 2 \frac{\prod_{n=1}^L (z_n - z)}{(z_i - z_k)} \times \frac{Q_M(z_i) Q_M(z_k)}{Q^3(z)} \left| \begin{array}{c} C_{N-1, M-1}^{L(i, k)} \\ 0, \tilde{z} \end{array} \right|. \quad (3.13b)$$

Последний детерминант — это знаменатель выражения для рациональной интерполяции по $L - 2$ точкам без узлов z_i и z_k . Таким образом, получено довольно компактное выражение для второй производной, однако оно не всегда пригодно при совместном использовании с рекуррентной процедурой построения рациональной интерполяции, так как при рекуррентном построении воспользоваться уже полученными приближениями ранга $L - 2$ для вычисления (3.13б) можно лишь в том случае, когда z_i и z_k — это два последних узла, т.е. z_L и z_{L-1} . Во всех остальных случаях, а их порядка L^2 , необходимы другие цепочки рекуррентного построения, и объем вычислений при больших L резко возрастает.

Поэтому вместо использования сравнительно простой формулы (3.13б) приходится переходить в (3.13а) к величинам, которые уже были получены при рекуррентном построении. Это сложная задача, требующая многократного применения тождества (3.9), соотношений типа (3.8) и некоторой изобретательности. Приведем здесь лишь окончательные формулы:

$$\left. \begin{aligned} \frac{\partial^2 f^{[N, M]}(z)}{\partial f_i \partial f_k} &= 2 \frac{Q_{(L)}(z_i) Q_{(L)}(z_k)}{Q_{(L)}^3(z)} \frac{\prod_{n \neq i} (z - z_n)}{\prod_{n \neq i} (z_i - z_n)} \times \\ &\times \frac{\prod_{n \neq L} (z_n - z_L)}{\prod_{n \neq k} (z_n - z_k)} \frac{\beta_{ik}}{Q_{(L)}(z_L) [-P_{L-1}(z_L) + f_L Q_{(L-1)}(z_L)]}; \\ \beta_{ik} &= \frac{z - z_L}{z - z_k} Q_{(L)}(z_i) [Q_{(L)}(z) Q_{L-1}(z_k) - Q_{L-1}(z) Q_L(z_k)] - \\ &- \frac{z_i - z_L}{z_i - z_k} Q_{(L)}(z) [Q_{(L)}(z_i) Q_{(L-1)}(z_k) - Q_{(L-1)}(z_i) Q_{(L)}(z_k)], \quad i \neq k; \\ \beta_{kk} &= Q_{(L)}(z_k) \frac{z - z_L}{z - z_k} [Q_{(L)}(z) Q_{(L-1)}(z_k) - Q_{(L-1)}(z) Q_L(z_k)] - \\ &- (z_k - z_L) Q_{(L)}(z) [Q_{(L-1)}(z_k) \frac{\partial}{\partial y} Q_{(L)}(y) \Big|_{y=z_k} - Q_{(L)}(z_k) \frac{\partial}{\partial y} \times \\ &\times Q_{(L-1)}(y) \Big|_{y=z_k}]. \end{aligned} \right\} \quad (3.14)$$

Поскольку коэффициенты полиномов $Q_L(z)$ и $Q_{(L-1)}(z)$ известны, вычисление производных в выражении для β_{kk} не представляет труда. Для проверки этих формул в частном случае $k = L - 1$, $i = L$ можно

использовать (3.136) и аналогичное выражение для $i = k = L$. Отметим, что первая производная по опорным ординатам зависит лишь от знаменателя рационального интерполяционного выражения и положения узлов интерполяции, но не зависит от числителя.

3.3. РЕКУРРЕНТНЫЕ СООТНОШЕНИЯ ДЛЯ РАЦИОНАЛЬНОЙ ИНТЕРПОЛЯЦИИ

Лучше всего исследованы частные случаи рациональной интерполяции и аппроксимации, в которых используются лишь приближения с одинаковыми или различающимися не более чем на единицу степенями числителя и знаменателя. Дело в том, что эти приближения являются так называемыми подходящими дробями непрерывных (цепных) дробей, теория и практика применения которых имеет многовековую историю*. В [2, с. 375] приведена аннотированная библиография алгоритмов построения приближений Паде. В большинстве этих алгоритмов коэффициенты рекуррентных соотношений выражаются через элементы цепных дробей, которые сами по себе для построения рациональных приближений не нужны. Не останавливаясь на них, приведем более простые алгоритмы для непосредственного построения приближений Паде.

Основные тождества, связывающие между собой смежные приближения Паде-1 с произвольными разностями степеней числителя и знаменателя, были получены Фробениусом еще в 1881 г. (см. [1], там же приведены рекуррентные алгоритмы построения приближения Паде-1, в основе которых лежат эти тождества).

В настоящем параграфе для элементов приближения Паде-2 будут получены соотношения, аналогичные тождествам Фробениуса для Паде-1. Это четыре трехчленных соотношения, связывающие смежные интерполяции $[N-1, M-1]$, $[N-1, M]$, $[N, M-1]$ и $[N, M]$, а также соотношения, связывающие между собой приближения внутри каждой из следующих троек: $[N, M]$, $[N-1, M]$, $[N-2, M]$; $[N, M]$, $[N, M-1]$, $[N, M-2]$; $[N, M]$, $[N-1, M-1]$, $[N-2, M-2]$; $[N, M]$, $[N-1, M+1]$, $[N+1, M-1]$. В рекуррентных соотношениях этого параграфа учтены особенности машинного построения приближений высоких рангов.

Рассмотрим прежде всего вывод таких соотношений в детерминантном подходе. Используя тождество (3.9) и конкретную структуру матриц $C_{N,M}^L$, можно получить следующее соотношение между детерминантами:

$$\begin{vmatrix} C_{N,M}^L \\ 0, \tilde{z} \end{vmatrix} \begin{vmatrix} C_{N,M}^{L-1} \\ 0, \tilde{z}_{L-1} \\ e, 0 \end{vmatrix} = \begin{vmatrix} C_{N,M}^L \\ e, 0 \end{vmatrix} \begin{vmatrix} C_{N,M}^{L-1} \\ 0, \tilde{z}_{L-1} \\ 0, \tilde{z} \end{vmatrix} + \begin{vmatrix} C_{N,M}^L \\ 0, \tilde{z}_{L-1} \end{vmatrix} \begin{vmatrix} C_{N,M}^{L-1} \\ 0, \tilde{z} \\ e, 0 \end{vmatrix} \quad (3.15)$$

*Современное состояние теории и приложений непрерывных дробей см. в книге Джоунс У., Трон В. Непрерывные дроби: Пер. с англ. М.: Мир, 1985.

Перейдя от матричных обозначений к элементам рациональной интерполяции, получим:

$$\begin{aligned} Q_{N,M}(z) Q_{N-1,M}(z_{L-1}) &= (z - z_{L-1}) \prod_{i=1}^{L-2} (z_i - z_{L-1}) [f_L Q_{N-1,M-1}(z) - \\ &- P_{N-1,M}(z_L)] Q_{N-1,M-1}(z) - \prod_{i \neq L-1} (z_i - z_{L-1}) [f_L Q_{N-1,M-1}(z_L) - \\ &- P_{N-1,M-1}(z_L)] Q_{N-1,M}(z), \end{aligned}$$

т.е. рекуррентное соотношение для знаменателей приближений Паде-2. Замена строки $(0, \tilde{z})$ строкой $(\tilde{z}, 0)$ в левой части (3.15) приведет к такой же замене лишь в тех детерминантах правой части, которые зависят от z , и для числителей получится рекуррентное соотношение с теми же коэффициентами.

Применяя (3.9) к произведениям детерминантов

$$\left| \begin{array}{c} C_{N,M}^L \\ 0, \tilde{z} \end{array} \right| \left| \begin{array}{c} C_{N,M}^{L-1} \\ 0, \tilde{z}_{L-1} \\ 0, e \end{array} \right|, \left| \begin{array}{c} C_{N,M}^L \\ 0, e \end{array} \right| \left| \begin{array}{c} C_{N,M}^{L-1} \\ 0, \tilde{z} \\ e, 0 \end{array} \right|, \left| \begin{array}{c} C_{N,M}^{L-1} \\ 0, \tilde{z} \\ 0, e \end{array} \right| \left| \begin{array}{c} C_{N,M}^{L-1} \\ 0, \tilde{z}_{L-1} \\ e, 0 \end{array} \right|,$$

аналогичным образом можно получить оставшиеся три соотношения, связывающие смежные интерполяции. Каждое из этих соотношений станет рекуррентным, если все входящие в него коэффициенты выразить через величины, полученные на предыдущих этапах построения.

Хотя формально нормировка числителя и знаменателя интерполяционного выражения роли не играет, так как их всегда можно домножить на одно и то же число, при машинном построении приближений высоких порядков отнюдь не безразлично, каковы значения числителя и знаменателя в отдельности. Исходя из этих соображений, а также для удобства сравнения приближений разных рангов перейдем к полиномам $\tilde{P}_N(z)$ и $\tilde{Q}_M(z)$, нормированным условием $d_M^M = 1$. Основываясь на единственности нормированного рационального интерполяционного выражения и предполагая существование соответствующего рекуррентного соотношения, можно определить его коэффициенты из условий на коэффициенты нормированных числителей и знаменателей и из условия прохождения построенного приближения через узлы интерполяции. Сравнение этого подхода с детерминантным позволяет получать соотношения между детерминантами, доказательство которых иным путем весьма затруднительно.

В наших расчетах основными являются рекуррентные соотношения, которые позволяют при увеличении ранга (числа узлов интерполяции) на единицу увеличивать на единицу степень числителя или знаменателя

приближения. Первое из них запишем в виде

$$\frac{\tilde{P}_{N,M}(z)}{\tilde{Q}_{N,M}(z)} = \frac{\alpha \tilde{P}_{N-1,M}(z) + \beta(z - z_{L-1})\tilde{P}_{N-1,M-1}(z)}{\alpha \tilde{Q}_{N-1,M}(z) + \beta(z - z_{L-1})\tilde{Q}_{N-1,M-1}(z)}. \quad (3.16)$$

Условие равенства единице старших коэффициентов всех $\tilde{Q}(z)$ дает сразу $\alpha = 1 - \beta$. Функция (3.16) при любых α и β проходит через все узлы интерполяции, кроме z_L . Полагая $f^{[N,M]}(z_L) = f(z_L)$, получаем

$$\beta = \tilde{c}_{N-1,M}(z_L) / (\tilde{c}_{N-1,M}(z_L) - (z_L - z_{L-1})\tilde{c}_{N-1,M-1}(z_L)), \quad (3.17)$$

где

$$\tilde{c}_{N,M}(z) = f(z)\tilde{Q}_{N,M}(z) - \tilde{P}_{N,M}(z);$$

$$\tilde{c}_{N-1,M}(z_L) = \left| C_{N-1,M}^L \right| \left/ \left| C_{N-1,M-1}^{L-1} \right| \right|;$$

$$\tilde{c}_{N-1,M-1}(z_L) = \left| \begin{array}{c} C_{N-1,M-1}^{L-1} \\ \tilde{z}_L, f_L \tilde{z}_L \end{array} \right| \left/ \left| C_{N-1,M-2}^{L-2} \right| \right|.$$

Если предыдущее приближение было не $f^{[N-1,M]}(z)$, а $f^{[N,M-1]}(z)$, получим $\beta = 1$ и

$$\tilde{R}_{N,M} = (z_{L-1} - z_L) \frac{\tilde{c}_{N-1,M-1}(z_L)}{\tilde{c}_{N,M-1}(z_L)} \tilde{R}_{N,M-1} + (z - z_{L-1}) \tilde{R}_{N-1,M-1} \quad (3.18)$$

Для того чтобы пользоваться этими рекуррентными соотношениями, необходимо выбрать два начальных приближения [28, с. 267]. Формальные начальные условия $f^{[-1,0]}(z) = 0/1$, $f^{[0,0]}(z) = f_1/1$ позволяют строить диагональные и околодиагональные приближения, у которых $N = M - 1$ чередуется с $N = M$; условия $f^{[0,-1]}(z) = 1/0$ и $f^{[0,0]}(z) = f_1/1$ приводят к чередованию $N = M + 1$ с $N = M$. Приближения такой структуры, но с другими нормировками получаются в теории цепных дробей. Соответствующие разным вариантам определения этих дробей рекуррентные соотношения для подходящих дробей получаются, если вместо $q_M^M = 1$ положить в (3.16) $\alpha = 1$ или $\beta = 1$. В каждом из этих вариантов вместо (3.16), (3.18) получается одно рекуррентное соотношение для наращивания степени и числителя, и знаменателя. Сравнивая эти соотношения с приведенными выше, можно найти связь получающихся полиномов с полиномами, нормированными условием $q_M^M = 1$.

Для построения приближений, у которых модуль разности степеней числителя и знаменателя больше единицы, в качестве начальных прибли-

жений выбирают последовательные полиномиальные приближения $f^{[l,0]}(z)$ и $f^{[l+1,0]}(z)$ в случае, когда $N - M \geq l$, и приближения $f^{[0,l]}(z)$ и $f^{[0,l+1]}(z)$, если $M - N \geq l$. Последние легко получить, строя полиномиальные интерполяции $Q_{0,M}(z)$ функции $1/f(z)$, поскольку вследствие единственности рационального интерполяционного выражения $f^{[0,M]}(z) = 1/Q_{0,M}(z) = 1/[q_M^M \tilde{Q}_{0,M}(z)]$.

Получим теперь два соотношения, совместное использование которых позволяет, не наращивая числа узлов интерполяции, переходить от приближений $f^{[L,0]}(z)$ или $f^{[0,L]}(z)$ к приближениям $f^{[N,L-N]}(z)$ с ненулевыми степенями полиномов и в числителе, и в знаменателе. Использование этих соотношений может оказаться удобнее последовательного наращивания степеней в том случае, когда известен ранг $[N, M]$ конечного приближения и есть простая процедура построения полиномиальной интерполяции, так как оно позволяет почти вдвое уменьшить число обращений к рекуррентным формулам.

Первое из соотношений запишем в виде

$$\tilde{R}_{N-1,M} = \alpha \tilde{R}_{N,M} + \beta \tilde{R}_{N,M-1}.$$

Условие нормировки знаменателей дает $\alpha = 1$, условие равенства нулю коэффициента при z^N в числителе дает $\beta = -\tilde{p}_{N,M}^N / \tilde{p}_{N,M-1}^N$. В результате получим

$$\tilde{R}_{N-1,M} = \tilde{R}_{N,M} - \tilde{p}_{N,M}^N \tilde{R}_{N,M-1} / \tilde{p}_{N,M-1}^N. \quad (3.19)$$

Аналогичные рассуждения приводят к следующему выражению для второго соотношения:

$$\tilde{R}_{N-1,M} = (z - z_{L-1}) \tilde{R}_{N-1,M-1} - (\tilde{p}_{N-1,M-1}^{N-1} / \tilde{p}_{N,M-1}^N) \tilde{R}_{N,M-1}. \quad (3.20)$$

Ясно, что таким методом легко получить коэффициенты соотношений, связывающих те же приближения, но понижающих степень знаменателя и повышающих степень числителя.

Четыре приведенных выше рекуррентных соотношения (3.16)–(3.20) являются исчерпывающими в том соотношении, что с их помощью, взяв в качестве начальных любые два смежных приближения, можно построить любое приближение с тем же или большим числом узлов.

Для рекуррентного построения приближений с постоянной степенью полинома в числителе можно использовать соотношение

$$\tilde{R}_{N,M} = \left[(z - z_L) + (z_L - z_{L-1}) \frac{\tilde{p}_{N,M-1}^N \tilde{c}_{N,M-2}(z_L)}{\tilde{p}_{N,M-2}^N \tilde{c}_{N,M-1}(z_L)} \right] \tilde{R}_{N,M-1} -$$

$$- (z - z_{L-1}) \frac{\tilde{p}_{N,M-1}^N}{\tilde{p}_{N,M-2}^N} \tilde{R}_{N,M-2}. \quad (3.21)$$

В том случае, когда постоянна степень знаменателя, рекуррентное соотношение имеет вид

$$\tilde{R}_{N,M} = \alpha \left\{ \left[(z - z_L) + (z_L - z_{L-1}) \frac{\tilde{c}_{N-2,M}(z_L)}{\tilde{c}_{N-1,M}(z_L)} \right] \tilde{R}_{N-1,M} - \right. \\ \left. - (z - z_{L-1}) \tilde{R}_{N-2,M} \right\}; \\ \alpha = \left\{ \tilde{q}_{N-1,M}^{M-1} - \tilde{q}_{N-2,M}^{M-1} + \left[(z_L - z_{L-1}) \frac{\tilde{c}_{N-2,M}(z_L)}{\tilde{c}_{N-1,M}(z_L)} + z_{L-1} \right] \right\}^{-1}. \quad (3.22)$$

Разные приближения, проходящие через одни и те же узлы интерполяции, связаны соотношением

$$\tilde{R}_{N-2,M+2} = \left(z - \frac{\tilde{p}_{N-1,M+1}^{N-2}}{\tilde{p}_{N-1,M+1}^{N-1}} + \frac{\tilde{p}_{N,M}^{N-1}}{\tilde{p}_{N,M}^N} \right) \tilde{R}_{N-1,M+1} - \\ - \frac{\tilde{p}_{N-1,M+1}^{N-1}}{\tilde{p}_{N,M}^N} \tilde{R}_{N,M}. \quad (3.23)$$

Наконец, приведем рекуррентное соотношение, связывающее интерполяционные выражения с одинаковой четностью числа узлов. Это соотношение может оказаться полезным в том случае, когда известна разность степеней числителя и знаменателя искомого приближения, хотя и не известно оптимальное число параметров:

$$\tilde{R}_{N,M} = [\alpha(z - z_{L-1}) + \beta] \tilde{R}_{N-1,M-1} + \\ + (1 - \alpha)(z - z_{L-2})(z - z_{L-3}) \tilde{R}_{N-2,M-2}. \quad (3.24)$$

Коэффициенты α и β нетрудно выразить через элементы предыдущих приближений, подставляя (3.24) в систему линейных уравнений:

$$f(z_{L-1}) \tilde{Q}_{N,M}(z_{L-1}) - \tilde{P}_{N,M}(z_{L-1}) = 0; \\ f(z_L) \tilde{Q}_{N,M}(z_L) - \tilde{P}_{N,M}(z_L) = 0.$$

3.4. АЛГОРИТМ ПОСТРОЕНИЯ ДВУХ АППРОКСИМАНТ С ОДИНАКОВЫМИ ЗНАМЕНАТЕЛЯМИ

Сечения различных ядерных реакций, проходящих через одно и то же составное ядро, имеют одинаковую резонансную структуру — один и тот же резонанс в различных реакциях может проявляться с разной силой, интерферировать с другими, но его положение и полная ширина неизменны и характеризуют состояние составного ядра при данной энергии. Обычно обработка данных по сечениям различных реакций ведется раздельно, и согласовывать полученные результаты непросто — при попытке использовать результаты статистической обработки одного сечения для описания другого иногда даже появляются физически бессмысленные отрицательные слагаемые. Совместная обработка всей информации с учетом статистического веса каждого измерения была бы наиболее корректной, но она пока технически неосуществима из-за вычислительных трудностей при традиционных методах обработки экспериментальных данных. Нам представляется, что использование метода перебора и рекуррентной процедуры, которая будет изложена в этом параграфе, может существенно облегчить выполнение такой программы.

Простейшая задача этого класса — совместная обработка данных по сечениям двух реакций — применением метода перебора сводится к следующей интерполяционной задаче: функции $f(x)$ и $g(y)$ известны в L_1 точках x_i и L_2 точках y_k соответственно; надо построить три полинома — $P_N(z)$, $T_K(z)$ и $Q_M(z)$ таких, что

$$\left. \begin{aligned} f^{[N, M]}(x_i) &= P_N(x_i)/Q_M(x_i) = f(x_i), \quad i = 1, 2, \dots, L_1; \\ g^{[K, M]}(y_k) &= T_K(y_k)/Q_M(y_k) = g(y_k), \quad k = 1, 2, \dots, L_2, \end{aligned} \right\} \quad (3.25)$$

т.е. построить два рациональных интерполяционных выражения с одинаковыми знаменателями, что соответствует одинаковой резонансной структуре двух сечений. Заметим, что в отличие от условия $L = N + M + 1$, связывающего число узлов и степени полиномов обычной рациональной интерполяции, в этом случае $L_1 \leq N + M + 1$, $L_2 \leq K + M + 1$, $L = L_1 + L_2 = K + M + N + 2$. Поскольку рекуррентные соотношения для всех трех полиномов одинаковы, введем для них общее обозначение $R_{N, K, M}^{L_1, L_2}$.

Прежде всего, схематически изобразим один из множества возможных вариантов наращивания ранга интерполяции:

$$\begin{aligned} R_{N, K, M}^{L_1, L_2} &\rightarrow R_{N, K+1, M}^{L_1, L_2+1} \rightarrow R_{N+1, K+1, M}^{L_1+1, L_2+1} \rightarrow R_{N+1, K+1, M+1}^{L_1+1, L_2+2} \rightarrow \\ &\rightarrow R_{N+1, K+2, M+1}^{L_1+2, L_2+2} \rightarrow R_{N+2, K+2, M+1}^{L_1+2, L_2+3} \end{aligned} \quad (3.26)$$

В качестве нормировки можно использовать условие $q_M^M = 1$.

Все рекуррентные соотношения в этой задаче можно записать в виде

$$R_{(L+1)} = \alpha R_{(L)} + \beta R_{(L-1)} + \gamma(z - z_{L-1})R_{(L-2)}. \quad (3.27)$$

Коэффициенты α , β , γ определяются из условия нормировки знаменателя и двух линейных уравнений, которые следуют из условий $\varphi^{L+1}(z_L) = \varphi(z_L)$, $\varphi^{L+1}(z_{L+1}) = \varphi(z_{L+1})$, где $\varphi(z)$ обозначает $f(x)$ или $g(y)$ в зависимости от того, к какой функции относятся узлы z_L , z_{L+1} .

Рассмотрим, например, наращивание степени знаменателя в цепочке приближений (3.26):

$$\begin{aligned} R_{N+1, K+1, M+1}^{L_1+1, L_2+2} &= \alpha R_{N+1, K+1, M}^{L_1+1, L_2+1} + \beta R_{N, K+1, M}^{L_1, L_2+1} + \\ &+ (z - y_{L_2+1}) R_{N, K, M}^{L_1, L_2}. \end{aligned} \quad (3.28)$$

Система линейных уравнений для α и β в этом случае имеет вид:

$$\left. \begin{aligned} P_{N+1, K+1, M+1}^{L_1+1, L_2+2}(x_{L_1+1}) - f(x_{L_1+1}) Q_{N+1, K+1, M+1}^{L_1+1, L_2+2} &= 0; \\ T_{N+1, K+1, M+1}^{L_1+1, L_2+2}(y_{L_2+2}) - g(y_{L_2+2}) Q_{N+1, K+1, M+1}^{L_1+1, L_2+2}(y_{L_2+2}) &= 0. \end{aligned} \right\} (3.29)$$

Ее решение дает

$$\begin{aligned} \alpha &= \delta^{-1} [(y_{L_2+2} - y_{L_2-1}) a^{(L+1)}(x_{L_1+1}) b^{(L)}(y_{L_2+2}) - \\ &- (x_{L_1+1} - y_{L_2+2}) a^{(L)}(x_{L_1+1}) b^{(L+2)}(y_{L_2+2})]; \\ \beta &= \delta^{-1} [(x_{L_1+1} - y_{L_2+1}) a^{(L)}(x_{L_1+1}) b^{(L+2)}(y_{L_2+2}) - \\ &- (y_{L_2+2} - y_{L_2+1}) a^{(L+2)}(x_{L_1+1}) b^{(L)}(y_{L_2+2})]. \end{aligned} \quad (3.30)$$

Здесь

$$\left. \begin{aligned} a^{(L)}(x) &= f(x) Q_{(L)}(x) - P_{(L)}(x); \\ b^{(L)}(y) &= g(y) Q_{(L)}(y) - T_{(L)}(y); \\ \delta &= a^{(L+2)}(x_{L_1+1}) b^{(L+1)}(y_{L_2+2}) - \\ &- a^{(L+1)}(x_{L_1+1}) b^{(L+2)}(y_{L_2+1}). \end{aligned} \right\} (3.31)$$

Аналогично определяются коэффициенты рекуррентных соотношений для наращивания степени одного из числителей в цепочке (3.25), а также для других цепочек изменения верхних индексов, т.е. для других последовательностей наращивания числа узлов интерполяции.

Приведем один из вариантов построения интерполяции низших рангов:

$$\left. \begin{aligned}
 L = 1, \quad \frac{P_{-1}}{Q_0} &= \frac{0}{1}, \quad \frac{T_0}{Q_0} = \frac{g_1}{1}; \\
 L = 2, \quad \frac{P_{0,0,0}^{1,1}}{Q_{0,0,0}^{1,1}} &= \frac{f_1}{1}, \quad \frac{T_{0,0,0}^{1,1}}{Q_{0,0,0}^{1,1}} = \frac{g_1}{1}; \\
 L = 3, \quad \frac{P_{0,0,1}^{1,2}}{Q_{0,0,1}^{1,2}} &= \frac{f_1}{1 + \alpha + \beta(x - y_1)}, \quad \frac{T_{0,0,1}^{1,2}}{Q_{0,0,1}^{1,2}} = \frac{(1 + \alpha)g_1}{1 + \alpha + \beta(y - y_1)}; \\
 \alpha &= (y_1 - x_1)\beta; \quad \beta = \left(\frac{g_2}{g_1} - 1 \right) [y_1 - x_1 - \frac{g_2}{g_1}(y_2 - x_1)]^{-1}.
 \end{aligned} \right\} \quad (3.32)$$

При обобщении на большое число обрабатываемых по такой методике (с общим знаменателем) кривых глубина рекурсии по L на единицу больше числа обрабатываемых функций, и, возможно, для вычисления коэффициентов удобнее численное решение соответствующей линейной системы. Все слагаемые, кроме последнего, в правой части такого рекуррентного соотношения соответствуют приближениям, отличающимся от искомого (левая часть) тем, что каждое из них не проходит через один из узлов интерполяции какой-либо кривой. Они входят в соотношение с постоянными коэффициентами. Последнее слагаемое соответствует приближению, которое описывает на две точки одной из кривых меньше, чем искомое. Это слагаемое умножается на $z - z_n$, где z_n — та из двух точек, которая всеми остальными слагаемыми описывается.

Использование разных цепочек построения делает возможным перебор не только узлов интерполяции каждой из функций, но и узлов, соответствующих разным функциям. Если у какой-нибудь функции исчерпаны все узлы интерполяции, то через несколько шагов в рекуррентных соотношениях не будет слагаемых, которые не обеспечивали бы прохождение через все эти узлы, и глубина рекурсии уменьшится на единицу. Можно, наоборот, увеличивать число обрабатываемых функций. Все эти алгоритмы идейно просты, но слишком громоздки в записи, чтобы их здесь приводить.

3.5. ТРИГОНОМЕТРИЧЕСКАЯ МЕРОМОРФНАЯ ИНТЕРПОЛЯЦИЯ

Тригонометрическим полиномом степени N называется выражение

$$A_0 + \sum_{n=1}^N (A_n \cos n\varphi + B_n \sin n\varphi),$$

где хотя бы один из коэффициентов A_N или B_N отличен от нуля. Тригонометрическая интерполяция применяется для периодических функций. Если функция к тому же четна, то переходят к переменной $z = \cos \varphi$ и далее все проводят, как для обычных полиномов или рациональных функций, если же она нечетна — можно положить $z = \sin \varphi$. В общем случае для функции без определенной симметрии задача тригонометрической интерполяции полиномом минимальной степени имеет единственное решение лишь тогда, когда число узлов нечетно. В случае же четного числа узлов один из коэффициентов тригонометрического полинома можно выбрать произвольно (например, положить $A_N = 0$ или $B_N = 0$). Точно такая же ситуация будет, если в качестве интерполирующей функции взять отношение тригонометрических полиномов.

Если перейти к переменным $z_i = e^{i\varphi_k}$, где φ_k — узлы тригонометрической интерполяции с нечетным числом узлов $L = N_1 + N_2 + M_1 + M_2 + 1$ вида

$$f^L(\varphi) \equiv f^{[N_1, N_2, M_1, M_2]}(\varphi) = \frac{P_{(L)}(\varphi)}{Q_{(L)}(\varphi)} =$$

$$= \frac{A_0 + \sum_{n=1}^{N_1} A_n \cos n\varphi + \sum_{n=1}^{N_2} B_n \sin n\varphi}{1 + \sum_{m=1}^{M_1} C_m \cos m\varphi + \sum_{m=1}^{M_2} D_m \sin m\varphi} \quad (3.33)$$

при $N_1 = N_2, M_1 = M_2$, можно дать детерминантное решение, аналогичное решению задачи о полиномиальной интерполяции (см. § 3.1). Несколько сложнее случай четных L , на котором останавливаться не будем, а дадим рекуррентную процедуру интерполяции отношением тригонометрических полиномов.

Наметим сначала построение такой цепочки, в которой нечетному числу узлов соответствует отношение полных тригонометрических полиномов, т.е. $N_1 = N_2 = N, M_1 = M_2 = M$. Для четного числа узлов строятся две интерполяции: $f^{L+1}(\varphi) = f^{[N+1, N, M, M]}(\varphi)$ и $f''^{L+1}(z) = f''^{[N, N+1, M, M]}(\varphi)$, затем строится $f^{L+2}(\varphi) = f^{[N+1, N+1, M, M]}(\varphi)$ и опять две интерполяции: $f^{(L+3)} = f^{[N+1, N+1, M+1, M]}(\varphi)$ и $f''^{L+3}(\varphi) = f''^{[N+1, N+1, M, M+1]}(\varphi)$ и т.д.

Интерполяции с нечетным числом узлов строят так:

$$R_{(L+2)} = aR'_{(L+1)} + (1-a)R''_{(L+1)}. \quad (3.34)$$

Условие прохождения аппроксиманты через $f(\varphi_{L+2})$ дает для a :

$$a^{-1} = 1 - \frac{P''_{(L+1)}(\varphi_{L+2}) - f_{L+2} Q''_{(L+1)}(\varphi_{L+2})}{P'_{(L+1)}(\varphi_{L+2}) - f_{L+2} Q'_{(L+2)}(\varphi_{L+2})}. \quad (3.35)$$

Для построения интерполяции с четным числом узлов используем соотношения типа

$$\begin{aligned} R_{(L+1)} = & (1 + a \cos \varphi_L + a_2 \sin \varphi_L + \beta_1 \cos \varphi_L + \beta_2 \sin \varphi_L) R_{(L)} + \\ & + [a_1 (\cos \varphi - \cos \varphi_L) + a_2 (\sin \varphi - \sin \varphi_L)] R'_{(L-1)} + \\ & + [\beta_1 (\cos \varphi - \cos \varphi_L) + \beta_2 (\sin \varphi - \sin \varphi_L)] R''_{(L-1)}. \end{aligned} \quad (3.36)$$

Рассмотрим подробнее случай $f^{L+1}(\varphi) = f^{[N+1, N, M, M]}(\varphi)$, $f^{L-1}(\varphi) = f^{[N, N, M, M-1]}(\varphi)$, $f^{L-1}(\varphi) = f^{[N, N, M-1, M]}(\varphi)$. Коэффициенты a_1 , a_2 , β_1 , β_2 определяются из системы четырех линейных уравнений, два из которых получаются, если приравнять нулю коэффициенты при $\cos(M+1)\varphi$ и при $\sin(M+1)\varphi$ в соотношении (3.36) для знаменателя $Q_{(L+1)}(\varphi)$, третье уравнение следует из условия $f_{(L+1)}(\varphi_{L+1}) = f(\varphi_{L+1})$, а четвертое — из условия равенства нулю коэффициента при $\sin(N+1)\varphi$ в числителе аппроксиманты. Аналогично получаются системы уравнений для определения коэффициентов рекуррентных соотношений при построении остальных приближений с четным числом узлов. Во всех этих случаях можно найти явный вид решения и выразить коэффициенты a_1 , a_2 , β_1 , β_2 через коэффициенты тригонометрических полиномов предыдущих приближений, однако эти формулы слишком громоздки, чтобы их здесь приводить, хотя и вполне пригодны для программирования.

Для построения аппроксиманты с помощью приведенных выше соотношений можно выбрать, например, следующие начальные приближения:

$$f^{[0,0,0,0]}(\varphi) = \frac{f_1}{1}; \quad f^{[0,0,1,0]}(\varphi) = \frac{a}{1 + b \cos \varphi};$$

$$f^{[0,0,0,1]}(\varphi) = \frac{c}{1 + d \sin \varphi};$$

$$a = \frac{f_1 f_2 (\cos \varphi_2 - \cos \varphi_1)}{f_2 \cos \varphi_2 - f_1 \cos \varphi_1}; \quad b = \frac{f_2 - f_1}{f_1 \cos \varphi_1 - f_2 \cos \varphi_2};$$

$$c = \frac{f_1 f_2 (\sin \varphi_2 - \sin \varphi_1)}{f_2 \sin \varphi_2 - f_1 \sin \varphi_1}; \quad d = \frac{f_2 - f_1}{f_1 \sin \varphi_1 - f_2 \sin \varphi_2}.$$

Приведем также алгоритм построения тригонометрической интерполяции, при котором используются лишь приближения с нечетным числом узлов и не соблюдается условие нормировки:

$$\begin{aligned} R_{N,N,M,M} &= R_{(2l+1)} = R_{(2l-1)} + \\ &+ a \sin \frac{\varphi - \varphi_{2l-1}}{2} \sin \frac{\varphi - \varphi_{2l-2}}{2} R_{(2l-3)} + \\ &+ \beta \frac{\sin \varphi - \sin \varphi_{2l-1}}{2} \sin \frac{\varphi - \varphi_{2l-3}}{2} R'_{(2l-3)}. \end{aligned} \quad (3.37)$$

Здесь $l = N + M$. Для того чтобы с помощью этого алгоритма построить приближение ранга $2l + 1$, необходимо на предыдущих этапах получить два приближения ранга $2l - 3$, различающихся одним из узлов интерполяции, а именно приближение $f^{2l-3}(\varphi)$, которое проходит через точки $\varphi_1, \varphi_2, \dots, \varphi_{2l-4}, \varphi_{2l-3}$, и приближение $f'^{2l-3}(\varphi)$, описывающее узлы $\varphi_1, \varphi_2, \dots, \varphi_{2l-4}, \varphi_{2l-2}$, т.е. отличающееся от предыдущего последним узлом. Коэффициенты a и β определяются системой двух линейных уравнений, следующих из условий $f^{2l+1}(\varphi_{2l}) = f(\varphi_{2l})$; $f'^{2l+1}(\varphi_{2l+1}) = f(\varphi_{2l+1})$. Начальные приближения можно выбрать следующим образом:

$$\begin{aligned} f^1(\varphi) &= \frac{f_1}{1}, \quad f'^1(\varphi) = \frac{f_2}{1}; \\ f^3(\varphi) &= \frac{a}{1 + b \cos \varphi + c \sin \varphi}; \quad f'^3(\varphi) = \frac{a'}{1 + b' \cos \varphi + c' \sin \varphi}. \end{aligned}$$

Здесь a, b и c определяются из системы

$$f^3(\varphi_1) = f_1; \quad f^3(\varphi_2) = f_2; \quad f^3(\varphi_3) = f_3,$$

а a', b' и c' — из системы

$$f'^3(\varphi_1) = f_1; \quad f'^3(\varphi_2) = f_2; \quad f'^3(\varphi_4) = f_4.$$

Построенные с помощью этих начальных условий аппроксиманты будут иметь структуру $[N-1, N-1, N, N]$ и $[N, N, N, N]$. Выбирая другие начальные условия, можно получить другие соотношения степеней тригонометрических полиномов в числителе и знаменателе аппроксиманты.

3.6. ОСОБЫЕ СЛУЧАИ РАЦИОНАЛЬНОЙ ИНТЕРПОЛЯЦИИ

При любых значениях z_i и $f(z_i)$ можно получить детерминантное решение задачи интерполяции в виде (3.1), однако это отнюдь не означает, что всегда можно построить приближение Паде-2 $f^{[N,M]}(z)$ с $L = N + M + 1$. Может оказаться, что старшие коэффициенты полиномов $Q_M(z)$ или $P_N(z)$ равны нулю, тогда получим приближение с меньшим числом параметров, которое в дальнейшем будем называть приближением неполного ранга. Кроме того, у числителя и знаменателя могут оказаться совпадающие корни. Особого внимания все эти случаи требуют при рекуррентном построении приближений, поскольку приводят к необходимости модификации алгоритмов.

Из единственности рациональной интерполяции следует, что если L точек кривой $f(z)$ лежат на кривой $f^{[N_0, M_0]}(z)$, где $N_0 + M_0 + 1 = L_0 < L$, то вместо всех рациональных приближений $f^{[N_1, M_1]}(z)$ ранга L_1 , где $L_0 \leq L_1 \leq L$, получим при $N_1 > N_0$ и $M_1 \geq M_0$ приближение $f^{[N_0, M_0]}(z)$. Однако могут быть построены и приближения, у которых или $N_1 < N_0$, или $M_1 < M_0$. Так, если L точек кривой $f(z)$ описываются полиномом $P_{N_0}(z)$ степени N_0 , то все рациональные приближения $P_{N_1 > N_0}(z)/Q_{M_1}(z)$ будут просто совпадать с $P_{N_0}(z)$, но приближения вида $P_{N_0-1}(z)/Q_M(z)$ существуют.

При детерминантном подходе случаи равенства нулю старших коэффициентов полиномов $P_N(z)$ и $Q_M(z)$ ничем не выделены. Гораздо интереснее следующий вопрос: могут ли быть числитель и знаменатель не взаимно простыми полиномами? Предположим, что у них есть совпадающий корень z_α , который не совпадает ни с одним из узлов интерполяции z_i . Сократив числитель и знаменатель на общий множитель $z - z_\alpha$, получим приближение, описывающее все $L = N + M + 1$ узлов, ранг которого на два меньше. Поскольку при этом $f^{[N-1, M]}(z_L) = f^{[N-1, M-1]}(z_L) = f(z_L)$, детерминант $\left| C_{N-1, M}^L \right|$, равный $P_{N-1, M}(z_L) - f_L Q_{N-1, M}(z_L)$, будет равен нулю, а он является старшим коэффициентом полинома в числителе приближения $f^{[N, M]}(z)$ при детерминантном подходе, т.е. степень числителя равна $M - 1$. Таким образом, предложение о том, что у числителя и знаменателя есть общий корень z_α , должно быть отвергнуто, поскольку приводит к противоречию.

Допустим, что такой корень совпадает с одним из узлов интерполяции $z_\alpha = z_i$. Используя (3.8), т.е. $Q_M(z_i) = \prod_{k \neq i} (z_k - z_i) \left| C_{N-1, M-1}^{L(i)} \right| \times (-1)^{L+i}$, получим, что в этом случае $P_{N, M-1}^N = q_{N-1, M}^M = \left| C_{N-1, M-1}^{L(i)} \right| = 0$. Следовательно, $f^{[N-1, M]}(z)$ и $f^{[N, M-1]}(z)$, не описывающие $f(z_i)$, были приближениями неполного ранга и равнялись $f^{[N-1, M-1]}(z)$.

И наоборот, если предыдущие приближения были неполного ранга и $f^L(z_i) \neq f^{L-1}(z_i)$, то $f^L(z)$ — приближение полного ранга и у $P_N(z)$ и $Q_M(z)$ есть общий множитель $z - z_i$.

Рассмотрим теперь, как проявляются эти особые случаи при рекуррентном построении

$$R_{(L)} = \alpha R_{(L-1)} + \beta(z - z_{L-1})R_{(L-2)}, \quad (3.38)$$

коэффициенты которого определяются формулами (3.17) или (3.18) в зависимости от того, применяется оно для повышения степени числителя или знаменателя. В обоих случаях, если $f(z_L) = f^{L-2}(z_L)$, т.е. если точка $f(z_L)$ лежит на кривой $f^{L-2}(z)$, получим $\alpha = 0$ и $\beta = 1$, и использование $(z - z_{L-1})P_{(L-2)}(z)/(z - z_{L-1})Q_{(L-2)}(z)$ в качестве $f^L(z)$ (без сокращения общего множителя) позволяет строить дальнейшие приближения без изменения алгоритма. Смысл такого приближения мы рассмотрим ниже, а сейчас отметим, что в любом случае, когда $L - 1$ точка из L точек кривой $f(z)$ лежит на кривой $f^{L-2}(z)$, а одна точка $f(z_i)$ на этой кривой не лежит, приближение ранга L имеет вид $(z - z_i)P_{N-1}(z)/(z - z_i)Q_{M-1}(z)$. Если не искать корни числителя и знаменателя, то такую особенность интерполирующей функции можно и не заметить.

Когда $f(z_L) = f^{L-1}(z_L)$, т.е. добавляемая точка описывается только что построенным приближением, получим $\alpha = 1$ и $\beta = 0$, если надо повысить степень числителя, и $\alpha = \infty$, если надо повысить степень знаменателя. В обоих случаях это означает, что получается приближение неполного ранга $f^L(z) = f^{L-1}(z)$. Дальнейшее построение с помощью (3.38) невозможно. Если следующая точка $f(z_{L+1})$ не описывается этим же приближением, то построение удобно проводить, поменяв ее местами с одной из точек z_i , например рассмотрев последовательность узлов $z_1, z_2, \dots, z_{L-1}, z_{L+1}, z_L$. В результате получим

$$R'_{(L)} = \alpha R_{(L-1)} + \beta(z - z_{L-1})R_{(L-2)}; f'^{L+1}(z) = \frac{(z - z_{L+1})P_{(L-1)}(z)}{(z - z_{L+1})Q_{(L-1)}(z)},$$

и дальнейшее построение идет обычным образом с использованием $R_{(L)}(z)$ и $R'_{(L+1)}(z)$. Если $f^{L-2}(z_L) \neq f(z_L)$, но

$$f^{L-1}(z_L) - f(z_L) = \frac{(z_{L-1} - z_L)Q_{(L-2)}(z_L)}{Q_{(L-1)}(z_L)} [f^{L-2}(z_L) - f(z_L)],$$

то получим $\alpha = \infty$ в случае, когда наращивается степень числителя. Это означает, что следующее приближение невозможно нормировать условием $q_M^M = 1$. Отказавшись от этого условия, получим в качестве сле-

дующего приближения:

$$f^{[N, M]}(z) = f^{[N, M-1]}(z) = \frac{\alpha' [P_{N-1, M}(z) - (z - z_{L-1}) P_{N-1, M-1}(z)]}{\alpha' [Q_{N-1, M}(z) - (z - z_{L-1}) Q_{N-1, M-1}(z)]}.$$

Здесь α' — нормировочный коэффициент. Вместо $f^{[N, M]}(z)$ имеем функцию неполного ранга $f^{[N, M-1]}(z)$ — возросла степень числителя, но уменьшилась степень знаменателя. Аналогично, если $\alpha = -P_{N, M-1}^N / P_{N-1, M-1}^{N-1}$, то при наращивании степени знаменателя можно получить приближение неполного ранга, у которого уменьшилась степень числителя. В обоих этих случаях дальнейшее построение можно вести обычным образом.

Комбинируя рассмотренные особые случаи, можно несколько усложнить картину, но ничего принципиально нового при этом не возникнет. Надо пользоваться симметрией приближения относительно перестановок узлов и при дальнейшем построении не забывать о том, что требование нормировки $q_M^M = 1$ имеет смысл только тогда, когда степень полинома в знаменателе действительно равна M .

При обработке экспериментальных данных, значение которых в точке определяется суммой гладкой функции и случайной погрешности, рассмотренные выше особые случаи маловероятны. Гораздо чаще оказывается, что значение обрабатываемой физической величины в некоторой точке близко к значению в этой точке одного из последних построенных приближений, хотя с ним и не совпадает. Исследуем поведение особенностей следующих приближений в этом случае.

Обозначим $\Delta_L(z) = f(z) - f^L(z)$ погрешность описания в точке z и будем наращивать ранг приближения с помощью рекуррентного соотношения (3.38), положив в нем для простоты $\alpha = 1$, т.е. отказавшись от условия нормировки $q_M^M = 1$. Определив β из условия описания точки $f(z_L)$, получим:

$$R_{(L)} = R_{(L-1)} - \frac{z - z_{L-1}}{z_L - z_{L-1}} \frac{\Delta_{L-1}(z)}{\Delta_{L-2}(z)} \frac{Q_{(L-1)}(z_L)}{Q_{(L-2)}(z_L)} R_{(L-2)}. \quad (3.39)$$

Если $|\Delta_{L-1}(z)| \ll |\Delta_{L-2}(z)|$, т.е. приближение $f^{L-1}(z)$ в отличие от $f^{L-2}(z)$ "почти" описывает значение $f(z)$ в точке z_L , то $\beta \ll 1$ и приближение $f^L(z)$, построенное согласно (3.39), фактически совпадает с $f^{L-1}(z)$ в широкой области $|z| \ll |\beta^{-1}|$. В зависимости от того, наращивается на данном этапе степень числителя или степень знаменателя, приближение $f^L(z)$ в пределе $\beta \rightarrow 0$ отличается от $f^{L-1}(z)$ смещением особенностей и множителем $1 - c_1 \beta z$ в числителе или знаменателе, т.е. имеет нуль или полюс при больших по модулю $z \approx (c_1 \beta)^{-1}$ (здесь c_1 — константа).

С точностью до членов первого порядка малости по β находим

$$R_{(L+1)} = R_{(L-1)} \left[\frac{z_{L+1} - z}{z_{L+1} - z_L} - \frac{z - z_L}{z_{L+1} - z_L} c_2 \frac{\beta}{\Delta_{L-1}(z_{L+1})} \right] - (z - z_{L+1})\beta R_{(L-2)}, \quad (3.40)$$

где c_2 — константа. При $\beta = 0$ получим рассмотренный ранее особый случай $f^{L+1}(z) = (z - z_{L+1})P_{(L-1)}(z)/(z - z_{L+1})Q_{(L-1)}(z)$. Учет последнего слагаемого в (3.40) приводит к малым смещениям особенностей аппроксиманты, различным для числителя и знаменателя. В частности, вместо совпадающих корней числителя и знаменателя в точке $z = z_{L+1}$ получим при $|\Delta_{L-1}(z_L)| \ll |\Delta_{L-1}(z_{L+1})|$ близкие к этой точке корни, расстояние между которыми пропорционально β . Если еще и $\Delta_{L-1}(z_{L+1})$ мало, т.е. погрешности описания функции $f(z)$ аппроксимантой $f^{L-1}(z)$ малы как при $z = z_L$, так и при $z = z_{L+1}$, то эта пара особенностей удаляется от точки z_{L+1} . Если же $|\Delta_{L-1}(z_L)| \gg |\Delta_{L-1}(z_{L+1})|$, то эти особенности будут вблизи точки z_L , в которой погрешность описания больше.

Ясно, что если все узлы интерполяции, кроме z_i , достаточно хорошо описываются приближением неполного ранга, то при любом выборе цепочки рекуррентного построения вблизи этого узла будут сосредоточены корни числителя и знаменателя аппроксиманты. Такие особенности получили название шумовых*, поскольку они позволяют выявить узел интерполяции z_i , в котором значение $f(z_i)$ таково, что не может быть описано с помощью гладкой вблизи этого узла интерполирующей рациональной кривой ранга L (или тем более неполного ранга) с заданной разностью степеней числителя и знаменателя.

Это свойство приближения Паде-2 полезно при анализе экспериментальных данных, так как дает возможность выявлять точки, имеющие экстремально большую погрешность, при некоторых предположениях о числе параметров и соотношении степеней рациональной аппроксиманты (см. § 5.3). Дальнейшее увеличение ранга приближения, естественно, может привести к построению рациональных приближений с большим числом параметров, гладких вблизи узлов интерполяции.

*Эти особенности называют также дефектами, см. книгу Бейкер Дж., Грейвс-Моррис П. Аппроксимации Паде: Пер. с англ. М.: Мир, 1986.

ГЛАВА 4. ПРИБЛИЖЕНИЕ ПАДЕ ПЕРВОГО РОДА И ЕГО ОБОБЩЕНИЯ. ОСНОВНЫЕ АЛГОРИТМЫ

4.1. АЛГОРИТМЫ ПОСТРОЕНИЯ ПРИБЛИЖЕНИЯ ПАДЕ ПЕРВОГО РОДА

Приближение Паде-1, основной областью применения которого является суммирование степенных рядов, оказывается полезным при обработке экспериментальных данных, соответствующих экспоненциальным зависимостям, а также при построении погрупповых констант. В этой главе изложены основные алгебраические свойства приближения Паде-1, дан простой вывод алгоритмов его рекуррентного построения, а также рассмотрены особые случаи и некоторые обобщения.

Присоединяя к системе (2.4) уравнение

$$\sum_{n=0}^M p_N^n z^n - \left(\sum_{m=0}^M q_M^m z^m \right) f^{[N, M]}(z) = 0$$

и приравнивая нулю детерминант получающейся системы $L + 2$ линейных уравнений, после несложных операций с детерминантами можно получить известное решение задачи рациональной аппроксимации

$$f^{[N, M]}(z) = \frac{\begin{vmatrix} a_{N+1} & a_N & \cdots & a_{N+1-M} \\ a_{N+2} & a_{N+1} & \cdots & a_{N+2-M} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \sum_{n=0}^N a_n z^n & \sum_{n=0}^N a_{n-1} z^n & \cdots & \sum_{n=0}^N a_{n-M} z^n \end{vmatrix}}{\begin{vmatrix} a_{N+1} & a_N & \cdots & a_{N+1-M} \\ a_{N+2} & a_{N+1} & \cdots & a_{N+2-M} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ a_{N+M} & a_{N+M-1} & \cdots & a_N \\ 1 & z & \cdots & z^M \end{vmatrix}}. \quad (4.1)$$

Здесь и везде далее в этом параграфе $a_k = 0$, если $k < 0$.

Матрицы

$$\mathbf{H}_{N, M} = \begin{pmatrix} a_{N+1} & a_N & \dots & a_{N+1-M} \\ a_{N+2} & a_{N+1} & \dots & a_{N+2-M} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ a_{N+M+1} & a_{N+M} & \dots & a_{N+1} \end{pmatrix}$$

известны как матрицы Ганкеля, их детерминанты $h_{N, M} = |\mathbf{H}_{N, M}|$ называются циркулянтами, частично их свойства описаны в [27]. В отличие от детерминантов, которые возникали при рациональной интерполяции, ранг этих матриц от N не зависит и равен $M + 1$. Выпишем некоторые довольно очевидные выражения, связывающие эти детерминанты с элементами рациональной аппроксимации:

$$q_{N, M+1}^{M+1} = \sum_m a_{L-m} q_{N, M}^m = h_{N, M};$$

$$p_{N, M}^N = (-1)^{M+1} h_{N-1, M}; \quad Q_{N, M}(0) = (-1)^M h_{N-1, M-1}. \quad (4.2)$$

Исходя из общих соотношений между детерминантами и частного вида матриц, используемых в задаче, можно получить рекуррентные соотношения между приближениями с различными N и M . Примененный ниже метод получения таких соотношений фактически позволяет их доказать методом математической индукции. Аналогично случаю рациональной интерполяции будем нормировать знаменатель условием $\tilde{q}_M^M = 1$, т.е. будем использовать полиномы $\tilde{R}_{N, M} = R_{N, M}/h_{N, M-1}$. Пусть $f^{L-2}(z)$ и $f^{L-1}(z)$ описывают соответственно $L - 2$ и $L - 1$ первых членов ряда $\sum_{n=0}^{\infty} a_n z^n$. Для построения $f^L(z)$ рассмотрим рекуррентное соотношение

$$\tilde{R}_{N, M} = \alpha \tilde{R}_{N, M-1} + \beta z \tilde{R}_{N-1, M}. \quad (4.3)$$

При любом α и β такое построение обеспечивает равенство нулю коэффициентов при z^k с $k < L < 1$ в выражении

$$\left(\sum_{n=0}^{\infty} a_n z^n \right) \tilde{Q}_{N, M}(z) - \tilde{P}_{N, M}(z) = O(z^L),$$

которое почти всегда эквивалентно определению приближения Паде-1

ранга L (исключения будут рассмотрены в § 4.4)*. Приравнявая нулю коэффициент при z^{L-1} в этом выражении и учитывая, что условие нормировки знаменателя дает $\beta = 1$, получаем

$$\alpha = - \frac{\sum_{m=0}^{M-1} a_{L-2-m} \tilde{q}_{N-1, M-1}^m}{\sum_{m=0}^{M-1} a_{L-1-m} \tilde{q}_{N, M-1}^m} = - \frac{\tilde{h}_{N-1, M-1}}{\tilde{h}_{N, M-1}}. \quad (4.4)$$

Для последовательного наращивания ранга приближения необходимо еще соотношение

$$\tilde{R}_{N, M} = \alpha \tilde{R}_{N-1, M} + (1 - \alpha) z \tilde{R}_{N-1, M-1}, \quad (4.5)$$

где $\alpha = \tilde{h}_{N-1, M-1} / (\tilde{h}_{N-1, M-1} - \tilde{h}_{N-1, M})$.

Аналогично для приближений, нормированных условием $\bar{Q}_M(0) = 1$, т.е. $R_{N, M} = (-1)^M h_{N-1, M-1} \bar{R}_{N, M}$, получим рекуррентное соотношение:

$$\bar{R}_{(L)} = \bar{R}_{(L-1)} + \alpha z \bar{R}_{(L-2)}, \quad \alpha = -\bar{h}_{(L-1)} / \bar{h}_{N-1, M-1}, \quad (4.6)$$

известное в теории цепных дробей и единое для приближений рангов $[N, M-1]$, $[N-1, M]$ на предыдущем шаге. Это соотношение обеспечивает нормировку $q_M^0 = q_0^0 = 1$. Рекуррентные соотношения для коэффициентов полиномов p_N^n и q_M^n , которые следуют из (4.3), (4.5) и (4.6), имеют вид

$$r_{(L)}^n = \alpha r_{(L-1)}^n + \beta r_{(L-2)}^{n-1}. \quad (4.7)$$

Приведенные выше рекуррентные соотношения вида (4.3) являются основными в задачах типа суммирования рядов, например рядов теории возмущений, когда последовательное наращивание ранга необходимо для практической оценки сходимости приближений.

Известный из теории цепных дробей выбор начального приближения $f^{[-1,0]} = 0/1$, $f^{[0,0]} = a_0/1$ позволяет строить аппроксиманты, у которых одинаковые степени полиномов в числителе и знаменателе чередуются с $N = M - 1$; начальные условия $f^{[-1,0]} = 1/0$, $f^{[0,0]} = a_0/1$ приводят к чередованию приближений с $N = M$ и $N = M + 1$.

*Заметим, что при построении приближений, в которых требуется совпадение не коэффициентов разложения в ряд Тейлора, а разложений по ортогональным полиномам, такой подход некорректен и приводит к неверному результату [28, с. 230]. Корректное решение этой задачи, которую называют нелинейным приближением Паде, дано в [29].

Приближения, у которых $N - M = N_0$ чередуется с $N - M = N_0 + 1$, где $N_0 \geq 1$, получаются, если в качестве начальных выбрать полиномы

$$f^{[N_0, 0]}(z) = \sum_{n=0}^{N_0} a_n z^n; \quad f^{[N_0+1, 0]}(z) = \sum_{n=0}^{N_0+1} a_n z^n, \quad (4.8)$$

т.е. просто соответствующее число членов исходного ряда. Для построения приближений, у которых $M - N = M_0$ чередуется с $M - N = M_0 + 1$, где $M_0 \geq 1$, начальными служат аппроксиманты $f^{0, M_0}(z) = 1/S_{M_0}(z)$ и $f^{[0, M_0+1]}(z) = 1/S_{M_0+1}(z)$. Их знаменатели должны удовлетворять уравнениям

$$S_K(z) \sum_{n=0}^{\infty} a_n z^n = O(z^K). \quad (4.9)$$

Здесь $K = M_0$ или $K = M_0 + 1$.

Решение этих уравнений, т.е. фактически построение ряда Тейлора функции $1/f(z)$, дает для коэффициентов этих полиномов:

$$s_K^0 = 1/a_0; \quad s_K^m = - \sum_{l=0}^{m-1} a_{m-l} s_K^l / a_0. \quad (4.10)$$

Как и в случае рациональной интерполяции, рекуррентные соотношения для полиномов с различной нормировкой можно получать друг из друга, используя свойства детерминантов. В частности, чтобы перейти от (4.5) к (4.6), достаточно доказать, что

$$h_{N-1, M-1}^2 - h_{N-1, M-2} h_{N-1, M} = h_{N, M-1} h_{N-2, M-1}, \quad (4.11)$$

а это есть просто тождество Сильвестра для $h_{N-1, M-2} h_{N-1, M}$, если $h_{N-1, M}$ представить в виде

$$h_{N-1, M} = \begin{vmatrix} a_N & a_{N-1} & \cdots & a_{N-M} \\ a_{N+1} & \mathbf{H}_{N-1, M-2} & & a_{N-M+1} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ a_{N+M} & \dots\dots\dots & & a_N \end{vmatrix}. \quad (4.12)$$

Остальные рекуррентные соотношения здесь приводить не будем. Отметим лишь, что они совпадают с аналогичными соотношениями для рациональной интерполяции, если в последних произвести замену $z - z_i \rightarrow z, z_i - z_k \rightarrow 1, \tilde{c}_{N, M}(z_i) \rightarrow \tilde{h}_{N, M}$.

Исключением из этого правила служат лишь выражения для коэффициентов соотношения, аналогичного (3.24), в которых кроме $\tilde{h}_{N,M}$ появляются еще величины $\tilde{h}'_{N,M} = \sum_m a_{N+M+2-m} q_{N,M}^m$.

4.2. ИНТЕРПОЛЯЦИЯ С КРАТНЫМИ УЗЛАМИ

В § 3.3 были приведены алгоритмы построения рациональных функций, проходящих через заданные точки кривой $f(z)$, а в § 4.1 — алгоритмы построения рациональных функций, описывающих заданное число производных функции $f(z)$ в точке $z = 0$. Теперь рассмотрим задачу интерполяции с кратными узлами, т.е. построение рациональных приближений, которые проходят через несколько точек кривой $f(z)$ и в этих точках имеют несколько первых производных, совпадающих с производными функции $f(z)$. Такое приближение иногда называют многочотечной паде-аппроксимацией, приближение Паде-1 — одноточечной.

Рациональную функцию, разложение которой в каждой из Λ точек z_i , $i = 1, 2, \dots, \Lambda$ в ряд по степеням $z - z_i$ совпадает с n_i первыми членами ряда Тейлора функции $f(z)$ относительно точки z_i , обозначим

$$f^L(z) = f^{(n_1, n_2, \dots, n_\Lambda)}(z) = P_{(\Lambda)}(z)/Q_{(\Lambda)}(z). \quad (4.13)$$

Здесь $L = N + M + 1 = \sum_{i=1}^{\Lambda} n_i$.

Поскольку $n_i = 0$ означает, что приближение через точку вообще не проходит, для упрощения будем нумеровать узлы в порядке их первого появления при рекуррентном построении, при этом Λ — число использованных узлов и все индексы $n_1, n_2, \dots, n_\Lambda$ не меньше единицы. Тогда следующее приближение может быть либо $f^{(n_1, \dots, n_\Lambda, 1)}$, либо $f^{(n_1, \dots, n_{k+1}, \dots, n_\Lambda)}$. Если предыдущее приближение было $f^{(\dots, n_{k-1}, \dots)}$, то при $k < \Lambda$ должно быть $n_k > 1$ и лишь для последнего узла, т.е. при $k = \Lambda$, может быть $n_k = 1$.

Прежде всего, научимся прибавлять еще один узел интерполяции. Для этого используем обычное рекуррентное соотношение рациональной интерполяции

$$R_{(L+1)}^{(n_1, \dots, n_\Lambda, 1)} = \alpha R_{(L)}^{(n_1, \dots, n_\Lambda)} + \beta(z - z_k) R_{(L-1)}^{(n_1, \dots, n_{k-1}, \dots, n_\Lambda)} \quad (4.14)$$

Учитывая условие нормировки и подставляя в $f^{L+1}(z)$ значение функции в добавляемом узле $z_{\Lambda+1}$, легко получить коэффициенты этого рекуррентного соотношения. Очевидно, что такое построение не нарушает равенства

$$P_{(L+1)}(z) - Q_{(L+1)}(z) \sum_{n=0}^{\infty} a_n^{(i)} (z - z_i)^n = O[(z - z_i)^{n_i}]$$

во всех остальных точках, в том числе и в точке z_k , т.е. сохраняет описание производных во всех остальных узлах интерполяции.

Если же следующее приближение соответствует добавлению не еще одного узла, а производной в одном из уже учтенных узлов z_i , то перейдем к переменным $y = z - z_i$. Тогда

$$\left. \begin{aligned} Q_{N,M}(z) = S_{N,M}(y); \quad s_{N,M}^m &= \sum_{l=0}^{M-m} q_{N,M}^{l+m} \frac{(l+m)!}{l! m!} z_i^l; \\ P_{N,M}(z) = T_{N,M}(y); \quad t_{N,M}^n &= \sum_{l=0}^{N-n} p_{N,M}^{l+n} \frac{(l+n)!}{l! n!} z_i^l. \end{aligned} \right\} \quad (4.15)$$

Рассмотрим рекуррентное соотношение

$$\begin{aligned} R_{N,M}^{(\dots, n_i^{+1}, \dots)}(y) &= \alpha R_{N,M-1}^{(\dots, n_i, \dots, n_k, \dots)}(y) + \\ &+ (y + z_i - z_k) R_{N-1,M-1}^{(\dots, n_i, \dots, n_k-1, \dots)}(y). \end{aligned} \quad (4.16)$$

Построенное с его помощью приближение $f^{[N,M]}$ при любом α описывает столько же производных в каждом узле, сколько и предыдущее приближение ранга $[N, M-1]$. Потребовав выполнения соотношения

$$T_{N,M}(y) - S_{N,M}(y) \sum_{n=0}^{\infty} a_n^{(i)} y^n = O(y^{n_i+1}),$$

найдем при $i \neq k$

$$\begin{aligned} \alpha &= (z_i - z_k) \sum_{l=0}^{\min(n_i, M-1)} (s_{N,M-1}^l a_{n_i-l}^{(i)} - \\ &- t_{N,M-1}^{n_i}) / \sum_{l=0}^{\min(n_i, M)} (s_{N,M}^l a_{n_i-l}^{(i)} - t_{N,M}^{n_i}). \end{aligned} \quad (4.17)$$

Если $k = i$, т.е. если надо учесть еще один член ряда Тейлора в той же точке, что и на предыдущем этапе, получим

$$\alpha = - \sum_{l=0}^{n_i-1} (s_{N,M-1}^l a_{n_i-1-l}^{(i)} - t_{N,M-1}^{n_i-1}) / \sum_{l=0}^{n_i} (s_{N,M}^l a_{n_i-l}^{(i)} - t_{N,M}^{n_i}). \quad (4.18)$$

Последнее выражение отличается от (4.4), полученного при построении приближения Паде-1, тем, что в него входят коэффициенты числителя. Дело в том, что степень полинома в числителе многоточечной паде-аппроксиманты может быть даже больше, чем число членов ряда Тейлора, которые мы в этом узле используем. Такова же причина появления $t_{N,M}^{n_l}$ в (4.17). Заметим еще, что в этих формулах $s_{N,M}^l = 0$ при $l > M$ и $t_{N,M}^n = 0$ при $n > N$.

Аналогично можно получить коэффициенты рекуррентных соотношений в том случае, когда предыдущим этапом было приближение ранга $[N-1, M]$, т.е. для повышения степени полинома в числителе. Построение нормированных приближений не меняет структуры рекуррентных соотношений и полностью совпадает с построением при рациональной интерполяции.

Полученные рекуррентные формулы позволяют, используя лишь два предыдущих приближения, добавлять новые узлы интерполяции, а также увеличивать на единицу число учтенных производных в любом из уже описанных узлов, поэтому они могут быть использованы в методе перебора.

4.3. ПОСТРОЕНИЕ МНОГОТОЧЕЧНОЙ АППРОКСИМАНТЫ С УЧЕТОМ РАЗЛОЖЕНИЯ ПО ОТРИЦАТЕЛЬНЫМ СТЕПЕНЯМ

В задаче определения подгрупповых констант теории переноса нейтронов, которая рассматривается в гл. 7, возникает необходимость построения рационального приближения, которое не только принимает определенные значения и имеет заданные производные по z в некоторых точках z_i , но и описывает в точке z_1 первые n^- членов разложения в ряд по степеням $t = (z - z_1)^{-1}$. Это разложение имеет вид $\sum_{n=1}^{\infty} c_{-n} t^n$,

т.е. в отличие от разложений по положительным степеням z начинается не с нулевой степени z , а с t^1 . Если же $n^- = 0$, то получается рассмотренная в предыдущем параграфе задача интерполяции с кратными узлами.

Обозначим искомое приближение

$$f_{N,M}^{(n^-; n_1, \dots, n_\Lambda)}(z) = \frac{P_N^{(n^-; n_1, \dots, n_\Lambda)}(z)}{Q_M^{(n^-; n_1, \dots, n_\Lambda)}(z)} \quad (4.19)$$

Отличие от предыдущего параграфа состоит лишь в том, что появился индекс n^- , описывающий разложение в ряд по отрицательным степеням $z - z_1$. В дальнейшем положим $z_1 = 0$, что упростит изложение, не уменьшая его общности. Сделав в (4.19) замену переменной $t = 1/z$,

получим

$$P_N(z)/Q_M(z) = t^M W_N(t)/t^N V_N(t). \quad (4.20)$$

Поскольку разложение по степеням t начинается с первой степени, только рациональные приближения с $N=M-1$ могут быть решением поставленной задачи.

Рассмотрим теперь разложение по отрицательным степеням вблизи точки $z = z_k$, т.е. разложение по положительным степеням $u = (z - z_k)^{-1}$. Эта переменная связана с t преобразованием Эйлера, поскольку $u = t/(1 - z_k t)$, а диагональное приближение $t W_{M-1}(t)/V_M(t)$, как показано в § 2.4, инвариантно относительно преобразования Эйлера. Следовательно, разложение в ряд по отрицательным степеням в одной точке однозначно определяет такое же разложение в любой другой точке. Учет этого факта позволяет упростить алгоритм по сравнению с изложенным в [30].

Приближения с $N = M - 1$ при построении по обычным правилам можно получить лишь при $L = n^- + \sum_i n_i = 2M$, т.е. при четном числе

описываемых величин. Для того чтобы и при $L = 2M - 1$ обеспечить описание отрицательных степеней, будем для нечетных L строить приближение избыточного ранга, т.е. вместо $f^{L-2} = f^{[M-1, M-1]}(z)$ будем строить $f^{L-2} = f^{[M-1, M]}(z)$. Это можно осуществить не единственным образом, поскольку еще один из коэффициентов полиномов $P'_{M-1}(z)$ и $Q'_M(z)$, кроме нормировочного, можно выбрать произвольно.

Например, приближение минимального ранга $f^1 = f^{[0,0]}(z)$ с $L = 1$, $N = 0$, $M = 0$, описывающее прохождение аппроксиманты через $f(z_i)$, есть $f(z_i)/1$, а приближение избыточного ранга $f'^1 = f^{[0,1]}(z)$ с $L = 1$, $N = 0$, $M = 1$ есть $f(z_i)/(a+bz)$, где $a = 1 - bz_i$. При рекуррентном построении выбор начальных приближений и цепочки построения позволяет на каждом из следующих этапов получать одно приближение, даже если оно избыточного ранга. Выберем в качестве таких начальных условий $f^0 = 0/1$ и $f'^1 = f(0)/(1+z)$ и в дальнейшем получим приближения $f^L(z) = P \left(\frac{L-1}{2} \right) (z)/Q \left(\frac{L+1}{2} \right) (z)$, где $E(A)$ — целая часть числа A .

Несколько сложнее построить алгоритм, который позволял бы в этой задаче использовать метод перебора, т.е. при увеличении L на единицу давал бы возможность, исходя из тех же предыдущих приближений, получать приближения с увеличенным на единицу любым n_i , а также n^- . Поскольку разложение по отрицательным степеням достаточно учесть в одной точке, удобнее использовать более простой алгоритм — строить функции с определенными n^- , а затем осуществлять перебор сравнением вариантов с различными $n_i + 1$.

Для построения приближений, описывающих ряд $\sum_{n=1}^{n^-} c_{-n} t^n$, воспользуемся обычным рекуррентным соотношением приближения Паде-1, применяя его к $f^{n^-}(t) = W_{(n^-)}(t)/V_{(n^-)}(t)$:

$$\frac{W_{(n^-+1)}(t)}{V_{(n^-+1)}(t)} = \frac{W_{n^-}(t) + \alpha t W_{(n^- - 1)}(t)}{V_{(n^-)}(t) + \alpha t V_{(n^- - 1)}(t)}, \quad (4.21)$$

где

$$\alpha = - \frac{\sum_{k=0}^{n^-} c_{k-n^- - 1} w_{n^-}^k}{\sum_{k=0}^{n^-} c_{k-n^-} w_{n^- - 1}^k}. \quad (4.22)$$

В качестве начальных приближений возьмем $0/1$ и $c_{-1}/(1+t)$. Подставляя в (4.21) $t = 1/z$, перейдем к привычным полиномам $P_L(z)$ и $Q_{(L)}(z)$. При таком построении получаются приближения, нормированные условием $q_m^M = 1$. Приближения $n^- \leq 4$, которых, как правило, достаточно в задаче определения подгрупповых констант, при этом имеют вид:

$$\begin{aligned} f^{1^-}(z) &= \frac{c_{-1}}{1+z}; \quad f^{2^-}(z) = \frac{c_{-1}}{z - c_{-2}/c_{-1}}; \\ f^{3^-}(z) &= \frac{c_{-1} [z + (c_{-2}^2 - c_{-3}c_{-1})/c_{-1}(c_{-2} + c_{-1})]}{z^2 - z(c_{-3} + c_{-2})/(c_{-2} + c_{-1}) + (c_{-2}^2 - c_{-3}c_{-1})/c_{-1}(c_{-2} + c_{-1})}; \\ f^{4^-}(z) &= \frac{c_{-1}z + c_{-2} - c_{-1}(c_{-4}c_{-1} - c_{-3}c_{-2})/(c_{-3}c_{-1} - c_{-2}^2)}{z^2 + z(c_{-3}c_{-2} - c_{-4}c_{-1})/(c_{-3}c_{-1} - c_{-2}^2) + (c_{-4}c_{-2} - c_{-3}^2)/(c_{-3}c_{-1} - c_{-2}^2)}. \end{aligned}$$

Следующий шаг — это учет значения функции в точке z_k . Если n^- — нечетное, т.е. последнее приближение было избыточного ранга, то построение осуществляется следующим образом:

$$R_{(n^-, n_k=1)} = R'_{(n^-)} + \alpha R_{(n^- - 1)}. \quad (4.23)$$

Переходя к $t = 1/z$, нетрудно убедиться, что при любом α такое приближение описывает n^- членов разложения в ряд по отрицательным степеням. Требование описания $f(z_k)$ определяет α :

$$\alpha = - \frac{P_{(n^-)}(z_k) - f(z_k)Q_{(n^-)}(z_k)}{P_{(n^- - 1)}(z_k) - f(z_k)Q_{(n^- - 1)}(z_k)}. \quad (4.24)$$

Если же n^- — четное и надо построить приближение избыточного ранга, 56

то воспользуемся соотношением

$$R'_{n^-, n_k=1} = (z + a_k)R_{(n^-)} + \beta_k R'_{(n^- - 1)}. \quad (4.25)$$

В этом соотношении два коэффициента, это значит, что приближение избыточного ранга можно выбрать не единственным образом. Нам удобно при $z_k = z_1 = 0$ положить $\beta_1 = 1$, тогда

$$a_1 = - \frac{P_{(n^- - 1)}(0) - f(0)Q_{(n^- - 1)}(0)}{P_{(n^-)}(0) - f(0)Q_{(n^-)}(0)}. \quad (4.26)$$

При $z_k \neq 0$ положим $a = 0$ и получим

$$\beta_k = - \frac{z_k [P_{(n^-)}(z_k) - f(z_k)Q_{(n^-)}(z_k)]}{P_{(n^- - 1)}(z_k) - f(z_k)Q_{(n^- - 1)}(z_k)}. \quad (4.27)$$

Рекуррентные соотношения, связывающие приближения с одним и тем же значением n^- , несмотря на использование приближений избыточного ранга, полностью совпадают с соотношениями § 4.2.

Отметим еще, что в этом случае использование метода перебора может оказаться не столь эффективным, как при рациональной интерполяции, поскольку младшие производные, описанные на предыдущих этапах построения аппроксиманты, варьировать уже нельзя.

4.4. ОСОБЫЕ СЛУЧАИ ПРИБЛИЖЕНИЯ ПАДЕ ПЕРВОГО РОДА

При построении приближения Паде-1 всегда можно записать решение (3.1) в виде отношения детерминантов. Как и при рациональной интерполяции, это решение может оказаться приближением неполного ранга, если старшие коэффициенты числителя или знаменателя равны нулю. Особую опасность представляет обращение в нуль младшего коэффициента знаменателя $q_M^0 = 0$ — формальное решение системы уравнений при этом существует, но радиус сходимости ряда Тейлора аппроксиманты ранга $[N, M]$ равен нулю! Дело в том, что эта система уравнений и ее детерминантное решение получаются в результате перехода от первоначальной постановки задачи о совпадении L членов рядов Тейлора аппроксиманты и функции $f(z)$ к требованию $P_N(z) = Q_M(z) f(z) + O(z^L)$. Согласно теореме Абея об умножении рядов, такой переход допустим лишь в той области, в которой сходится ряд Тейлора функции $1/Q_M(z)$.

В детерминантном подходе, используя тождество Сильвестра, можно показать, что если приближение ранга $[N_0, M_0]$ описывает $L_0 + n$ коэффициентов ряда Тейлора функции $f(z)$, то все приближения рангов $[N_1, M_1]$ с $N_1 \geq N_0$, $M_1 > M_0$, $L_0 \leq L_1 \leq L_0 + n$ будут приближениями

неполного ранга и совпадут с приближениями ранга $[N_0, M_0]$, а $(1/2)n(n-1)$ следующих приближений рангов $[N_2, M_2]$ с $N_0 + n \geq N_2 \geq N_0$, $M_0 + n \geq M_2 \geq M_0$ вообще будут отсутствовать, так как при формальном их построении $q_{M_2}^0 = 0$. Не останавливаясь на этом, рассмотрим рекуррентное построение приближений в особых случаях.

При рекуррентном построении приближения Шаде-1 возможны те же особенности коэффициентов, что и при рациональной интерполяции. Чаще всего используется построение с помощью рекуррентных соотношений (4.3) – (4.5), т.е. с чередующимся увеличением степеней числителя и знаменателя. Эти соотношения можно записать в виде $R_{(L+1)} = \alpha R_{(L)} + \beta z R_{(L-1)}$. Если $\sum_{m=0}^M a_m q_{(L)}^{L-m} = 0$, т.е. приближение ранга $[N, M]$ описывает еще и коэффициент при z^L в разложении функции $f(z)$ в ряд Тейлора, а не только a_i с $i \leq L-1$, то приближение $f^{L+1}(z)$ совпадает с $f^L(z)$. При этом для приближений, нормированных условием $q_M^M = 1$, в случае увеличения степени знаменателя $\alpha = \infty$, а в случае увеличения степени числителя $\alpha = 1, \beta = 0$. Попытка построить с помощью этих же соотношений следующее приближение дает $\alpha = 0$, т.е. формально приближение имеет вид $f^{L+2}(z) = z P_{(L)}(z)/z Q_{(L)}(z)$. Если коэффициент при z^{L+1} действительно описывается приближением $f^L(z)$, то можно сократить общий множитель z , т.е. $f^{L+2}(z) = P_{(L)}(z)/Q_{(L)}(z)$. В противном случае не существует не только приближения $f^{L+2}(z)$, но и эквивалентных ему приближений неполного ранга.

Есть еще возможность (как и при рациональной интерполяции) уменьшения степеней числителя или знаменателя на некотором этапе построения. Например, если строить рациональные аппроксимации полинома $P_N(z)$ степени N , то после обычного построения приближений рангов $[N-1, M-1]$, $[N-1, M]$ получим $f^{[N, M]}(z) = f^{[N, 0]}(z) = P_N(z)$.

Рассмотрим подробнее следующий случай: пусть до ранга $[N-1, M + \lambda]$ построение шло по обычным формулам (4.3) – (4.5). Следующий шаг дал вместо $f^{[N, M+\lambda]}(z)$ приближение ранга $[N, M]$. Пусть n – число коэффициентов ряда Тейлора функции $f(z)$, которые, сверх $L = N + M + 1$, описываются приближением ранга $[N, M]$. Очевидно, что $n \geq \lambda$. Для построения приближения ранга

$$\left[N + E\left(\frac{n + \kappa + \lambda}{2}\right), M + \lambda + E\left(\frac{n + \kappa - \lambda + 1}{2}\right) \right],$$

где $E(A)$ – целая часть числа A , рассмотрим рекуррентное соотношение

$$f_{L+n+\kappa}(z) = \frac{\left(E\left(\frac{n+\kappa-\lambda}{2}\right) \sum_{m=0}^{\infty} \alpha_m z^m \right) P_N(z) + \beta z E\left(\frac{n+\kappa-\lambda+1}{2}\right) P_{N-1}(z)}{\left(E\left(\frac{n+\kappa-\lambda}{2}\right) \sum_{m=0}^{\infty} \alpha_m z^m \right) Q_M(z) + \beta z E\left(\frac{n+\kappa-\lambda+1}{2}\right) Q_{M+\lambda}(z)}. \quad (4.28)$$

Учитывая, что $P_N(z) = f(z)Q_M(z) + O(z^{L+n})$ и $P_{N-1}(z) = f(z)Q_{M+\lambda}(z) + O(z^{L+\lambda-1})$, и рассматривая аналогичное уравнение для

$$P_{L+n+\kappa}(z) = f(z)Q_{L+n+\kappa}(z) + O(z^{L+n+\kappa}),$$

из условия равенства между собой коэффициентов при z^k с $k \leq L + n + \kappa$ в правой и левой его частях получим систему зацепляющихся линейных уравнений для определения коэффициентов рекуррентного соотношения (4.28).

Решив эту систему уравнений, получим, что при $\kappa \leq n - \lambda$ всегда $\beta = 0$ и все α_m , кроме α_m с $m \geq \kappa$, также равны нулю. Следовательно, таким образом можно получить формальное решение вида $z^\kappa P_N(z) / z^\kappa Q_M(z)$, о котором уже говорилось, но построить аппроксиманту, удовлетворяющую определению приближения Паде-1, невозможно. При $\kappa = n - \lambda + 1$ все коэффициенты α_m и β отличны от нуля и получается приближение полного ранга $f^{L+2n+1-\lambda}(z) = P_{N+n-\lambda}(z) / Q_{M+n+1}(z)$. Если оно вдобавок описывает еще k лишних коэффициентов ряда Тейлора, то для дальнейшего построения надо использовать рекуррентное соотношение

$$R_{(L+2n+2k+1-\lambda)} = \left(\sum_{m=0}^k \alpha_m z^m \right) R_{(L+2n+1-\lambda)} + \beta z^{n+k+1} R_{(L+\lambda-1)}. \quad (4.29)$$

После этого при $k = 0$ можно возвращаться к использованию обычных рекуррентных соотношений (4.3) – (4.5). Аналогичным образом можно рассмотреть случай, когда приближение ранга $[N, M]$ получается при добавлении еще одного узла к приближению ранга $[N + \lambda, M - 1]$.

На основании проведенных выше рассуждений можно более подробно рассмотреть вопрос о том, при каких N и M получаются приближения неполного ранга, а при каких построение рационального приближения вообще невозможно. Оказывается, что в таблице, клетки которой занумерованы числами N и M , этим особым случаям соответствуют квадраты со стороны n , где n – число лишних коэффициентов, описываемых приближением ранга $[N, M]$. Если в такой таблице N и M

увеличиваются слева направо и сверху вниз, соответственно то клетка (N, M) находится в левом верхнем углу квадрата. Этот результат был получен еще Паде, а таблица называется таблицей Паде.

Поскольку есть рекуррентные соотношения, которые позволяют менять степень только числителя или только знаменателя, т.е. двигаться по вертикалям или горизонталям таблицы Паде, может оказаться удобнее использовать их для обхода опасных областей таблицы, а не применять сравнительно сложное рекуррентное построение, изложенное выше в этом параграфе.

Рассмотрим теперь более реалистичский случай, когда последнее построенное приближение $f^L(z)$ "почти" описывает коэффициент a_L ряда Тейлора функции $f(z)$, но не описывает коэффициент a_{L+1} . Введем обозначение $\sum_{m=0}^{\infty} a_m q_{(L-1)}^{L-1-m} = c_{L-1}$ и для простоты положим $\alpha = 1$, т.е. откажемся от нормировки $q_M^M = 1$. При этом $\beta = -c_L/c_{L-1}$ и $R_{L+1} = R(L) - z(c_L/c_{L-1})R_{(L-1)}$. Учитывая, что $q_{(L+1)}^m = q_{(L)}^m - (c_L/c_{L-1})q_{(L-1)}^{m-1}$, для приближений следующего ранга получаем

$$R_{(L+2)} = \left(1 + \frac{c_L}{c_{L-1}} z - \frac{c_{L+1}}{c_L} z \right) R_{(L)} - \frac{c_L}{c_{L-1}} z R_{(L-1)}.$$

Таким образом, при $|c_L| \ll |c_{L-1}|$ везде, кроме очень больших z , приближение $f^{L+1}(z)$ фактически совпадает с $f^L(z)$, а при $z \approx c_{L-1}/c_L$ получаем корень числителя, если наращивалась его степень, и корень знаменателя при переходе от ранга $[N, M]$ к рангу $[N+1, M+1]$. Если, кроме того, $|c_L| \ll |c_{L+1}|$, то приближение $f^{L+2}(z)$ при $c_L \rightarrow 0$ переходит в $z P_{(L)}(z)/z Q_{(L)}(z)$, т.е. происходит слияние близких к нулю корней числителя и знаменателя. Если же c_{L+1} тоже мало и одного порядка с c_L , то получим, что приближение $f^{L+2}(z)$ фактически совпадает с $f^L(z)$, но имеет близкие корни числителя и знаменателя при $z = c_L/c_{L+1}$.

Такие предельные случаи можно рассмотреть и тогда, когда приближение f^L "почти" описывает большее число лишних коэффициентов исходного ряда Тейлора. Общий вывод таков: если с помощью $f^L(z)$ "почти" описано n лишних коэффициентов, то при построении следующих приближений будем получать фактически то же приближение во всей плоскости комплексного переменного z , кроме точек $z = \infty$, $z = 0$ и точек, определяемых отношениями отклонений полученного описания от точного описания этих коэффициентов ряда Тейлора. В этих точках будут скапливаться нули числителей и знаменателей таких приближений. Только после того, как число коэффициентов $a_L, a_{L+1}, \dots, a_{L+k+1}$, не описанных приближением $f^{L+k}(z)$, превысит число "почти"

описанных ими коэффициентов, получим существенно новое приближение. Таким образом, получается некая инертность — если по коэффициентам ряда Тейлора удалось предсказать еще n коэффициентов, то отклонения в описании следующих n коэффициентов при движении вдоль диагонали таблицы Паде воспринимаются как шум и приводят к появлению шумовых особенностей. Отметим, что такая же инертность есть и при рациональной интерполяции — если приближение $f^L(z)$ описывает n лишних узлов, то следующими приближениями полного ранга без шумовых полюсов будут $f^{L+2n+1}(z)$ и приближения, соответствующие обходу опасной области таблицы Паде.

ГЛАВА 5. РАЦИОНАЛЬНАЯ АППРОКСИМАЦИЯ ЭКСПЕРИМЕНТАЛЬНЫХ ЗАВИСИМОСТЕЙ

5.1. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

В гл. 3 были описаны способы построения рациональной интерполяции с помощью различных рекуррентных соотношений. Однако непосредственно интерполяция как способ приближенного описания применяется по большей части лишь к аналитическим функциям (достаточно широко распространенным исключением является, пожалуй, лишь метод сплайнов, см. [31]). Если же значения аппроксимируемой функции включают заметную погрешность, то, как правило, используются аппроксимирующие функции с числом параметров, гораздо меньшим числа аппроксимируемых точек, и аппроксимация осуществляет, таким образом, сглаживание поточечной зависимости с учетом информации с статистических свойствах погрешностей исходных значений. Эксперимент, поставленный с целью определения гладкой "истинной" функции, называют регрессионным, функцию, параметры которой надо определить, — регрессионной моделью или функцией регрессии, а задачу обработки экспериментальных данных в такой постановке — регрессионным анализом [32]. Эта задача и рассматривается в настоящей главе.

Пусть числа $F_1, F_2, \dots, F_{N_{\text{ex}}}$ — результаты эксперимента ("исходные значения"), полученные при значениях аргумента $z_1, z_2, \dots, z_{N_{\text{ex}}}$. Задача аналитической аппроксимации экспериментальных данных обычно сводится к нахождению функции регрессии, принадлежащей к определенному L -параметрическому семейству и минимизирующей некоторый функционал S , зависящий от исходных значений F_i , значений аппроксиманты при тех же значениях аргумента $f^L(z, \{p\})$, а также от матрицы (или вектора), определяемой свойствами статистического распределения погрешности исходных значений. На практике в подавляющем большинстве случаев S выбирают в виде квадратичной формы по элементам

вектора отклонений аппроксиманты от исходных значений:

$$S = (1/2) [f^L - F]^T V^{-1} [f^L - F], \quad (5.1)$$

где $f^L - F$ — вектор-столбец с компонентами $f^L(z_i) - F_i$; V — корреляционная матрица исходных значений; T обозначает транспонирование. Функционал S , определенный выражением (5.1), называется *статистической суммой*. Если экспериментальные значения можно считать статистически независимыми и нормально распределенными, то $V_{ik} = \sigma_i^2 \delta_{ik}$, где σ_i^2 — дисперсии соответствующих нормальных распределений, и выражение для S упрощается:

$$S = \frac{1}{2} \sum_{i=1}^{N_{\text{ex}}} [f^L(z_i) - F_i]^2 / \sigma_i^2. \quad (5.2)$$

Аппроксимация на основе минимизации статистической суммы называется *методом наименьших квадратов* (МНК). Исследованию МНК посвящено множество работ, а его описание содержится практически во всех руководствах по статистическим методам обработки данных (см., например, [33]), поэтому здесь мы приводим лишь необходимые краткие сведения.

Согласно сказанному МНК — нахождение минимума S как функции многих переменных, а именно параметров аппроксиманты $\{p\}$. Необходимые условия этого минимума — обращение в нуль производных по параметрам, т.е. выполнение равенств

$$\frac{\partial S}{\partial p_k} = \sum_{i=1}^{N_{\text{ex}}} \frac{1}{\sigma_i^2} [f^L(z_i, \{p\}) - F_i] \frac{\partial f^L(z_i, \{p\})}{\partial p_k} = 0 \quad (5.3)$$

для всех k .

Если f — линейная функция параметров p_l (задача линейной регрессии):

$$f^L(z, \{p\}) = \sum_{k=1}^L p_k \varphi_k(z), \quad (5.4)$$

то (5.3) сводится к системе линейных уравнений относительно p_l :

$$\sum_{l=1}^L p_l \sum_{i=1}^{N_{\text{ex}}} \varphi_l(z_i) \varphi_k(z_i) / \sigma_i^2 = \sum_{i=1}^{N_{\text{ex}}} F_i \varphi_k(z_i) / \sigma_i^2. \quad (5.5)$$

Если функции $\varphi_l(z)$ ортогональны на дискретной совокупности точек z_i с весом $1/\sigma_i^2$, т.е.

$$\sum_{i=1}^{N_{\text{ex}}} \varphi_l(z_i) \varphi_k(z_i) / \sigma_i^2 = \delta_{lk}, \quad (5.6)$$

то система (5.5) сводится к выражениям для

$$p_k = \sum_{i=1}^{N_{\text{ex}}} F_i \varphi_k(z_i) / \sigma_i^2. \quad (5.7)$$

Линейный МНК с разложением по ортогональным функциям (чаще всего полиномам) широко используется, поскольку он чрезвычайно удобен.

Однако в задаче нелинейной регрессии условия минимума S приводят к системам уравнений, нелинейных относительно p_k . Положение резко осложняется, сколько-нибудь универсальных методов решения этой задачи не существует, и даже использование проблемно-ориентированных численных методов сильно затруднено, если число параметров велико.

Метод наименьших квадратов тесно связан с *методом максимального правдоподобия* (ММП) — в нем в качестве оценок истинных значений параметров выбираются те, которые максимизируют функцию правдоподобия $\mathcal{L} = \prod_i \psi[f(z_i), F_i]$. Здесь $\psi[f(z_i), F_i]$ — значение плотности вероятности распределения погрешности эксперимента для выборочного (т.е. измеренного) значения F_i . Если $\psi(f, F)$ — нормальное распределение и если дисперсия его в каждой точке от оцениваемых параметров не зависит, то ММП сводится к МНК. В общем случае оценки МНК в асимптотике (при $N_{\text{ex}} \rightarrow \infty$) совпадают с оценками ММП, и математические теоремы, полученные для одного из этих методов, можно использовать для другого, чем мы будем неоднократно пользоваться в дальнейшем.

Итак, при заданных числе параметров и соотношении степеней числителя и знаменателя рациональной аппроксиманты для выбора оптимального приближения мы используем МНК. Простейший метод определения оптимального числа параметров основан на свойствах остаточной суммы квадратов, т.е. значения статистической суммы, которое получается при подстановке в нее значений аппроксиманты. Если погрешности эксперимента в разных точках независимы и подчиняются нормальному распределению, то сумма их квадратов, т.е. квадратов отклонений экс-

периментальных данных от истинных значений $S_{\text{ост}} = \sum_{i=1}^{N_{\text{ex}}} [f(z_i) - F_i]^2$, подчиняется χ^2 -распределению с N_{ex} степенями свободы. В случае линейной регрессии сумма квадратов отклонений эксперименталь-

ных данных от L -параметрической оценки по МНК $S_{\text{ост}}^L = \sum_{i=1}^{N_{\text{ex}}} [f^L(z_i) - F_i]^2$ должна подчиняться χ^2 -распределению с числом степеней свободы $N_{\text{ex}} - L$. При нелинейной регрессии это утверждение справедливо для $N_{\text{ex}} \rightarrow \infty$. Вычисляя остаточную сумму квадратов, можно, задавшись некоторым доверительным интервалом, определить, согласуется ли описание при данном числе параметров с предположением о независимом нормальном распределении погрешностей, т.е. значимо ли отличие остаточной суммы квадратов от значения, предсказываемого χ^2 -распределением с соответствующим числом степеней свободы. Оптимальное число параметров L_0 можно найти, сравнивая с критерием χ^2 и между собой значения $S_{\text{ост}}^i / (N_{\text{ex}} - L_i)$, т.е. вычисленные при разном числе параметров значения остаточной суммы квадратов, деленной на число степеней свободы, и используя статистические критерии значимости. Можно привлекать более сложные критерии, например t -статистику Стьюдента или F -статистику Фишера [32–33].

Все эти статистические методы, реализованные в виде стандартных программ во многих прикладных пакетах, несложно включить в программу обработки экспериментальных данных для ЭВМ. Практически число резонансов при анализе сечений почти всегда известно; если его превысить, появятся шумовые полюсы, и дальнейшее наращивание числа параметров смысла не имеет. Если же экспериментальные данные столь неточны, что возможно их описание аппроксимантами с различным числом параметров, и используемый нами критерий $S_{\text{ост}}^i / (N_{\text{ex}} - L_i)$ принимает близкие значения для нескольких L_i , то предпочтение отдается описанию с меньшим числом параметров. Описание с большим числом параметров рассматривается как допустимое. Такая ситуация указывает на необходимость дальнейшего уточнения экспериментальных данных в рассматриваемой области. Можно даже, следуя теории планирования эксперимента, определить точку или область, в которой максимально различие описаний и уточнение экспериментальных данных наиболее информативно для выбора оптимального числа параметров.

5.2. ДИСКРЕТНАЯ ОПТИМИЗАЦИЯ ПАДЕ-АППРОКСИМАНТЫ (МЕТОД ПЕРЕБОРА)

В любом варианте выбора параметров рациональной аппроксиманты (2.1) попытка найти их обычным методом наименьших квадратов, т.е. использованием условий (5.3), приводит к весьма сложным системам нелинейных уравнений.

В то же время, как детально показано в гл. 3, построение рациональной аппроксиманты при $L = N_{\text{ex}}$ осуществляется с помощью простых рекуррентных соотношений без численного обращения матриц высокого порядка. С учетом этого обстоятельства в [8–10, 34–35] был предложен и реализован в виде программ для ЭВМ способ рациональной ап-

проксимации одномерных зависимостей, основанный на дискретной оптимизации методом перебора, который и будет описан в этом параграфе.

Суть метода сводится к следующему: из экспериментальных точек некоторым образом выбирают *начальный набор* из L опорных точек $z_k, F_k, 1 \leq k \leq N_{\text{ex}}$, где все k различны, и для них с помощью рекуррентных алгоритмов, описанных в § 3.3 решают задачу рациональной интерполяции, т.е. строят рациональную аппроксиманту, проходящую через все L опорных точек. После этого вычисляются значения $f^L(z_i)$ для всех $i = 1, 2, \dots, N_{\text{ex}}$ и значение минимизируемого функционала, которыми определяется начальное приближение. Затем начинают итерационный процесс, на i -м шаге которого одна из опорных точек заменяется другой из экспериментальных, строят новую аппроксиманту, для нее вычисляют значение функционала S^i , и если $S^i < S_{\min}^{i-1} = \min \{S^1, S^2, \dots, S^{i-1}\}$, то i -й набор опорных точек запоминается в качестве *текущего оптимального*.

Способы получения начального отбора опорных точек, выбора заменяемой и заменяющей опорных точек и определения момента окончания итерационного процесса можно варьировать в широких пределах. Описанные ниже варианты выбраны на основе многолетнего опыта эксплуатации различных алгоритмов паде-аппроксимации.

Характеристики итерационного процесса, как оказалось, слабо зависят от начального выбора опорных точек, и в качестве таковых обычно берут эквидистантную последовательность: $k_i = (i - 1)E(N_{\text{ex}}/L) + 1, i = 1, 2, \dots, L$.

Перебор опорных точек организуется следующим образом. Точки с номерами k_1, k_2, \dots, k_{L-1} фиксируются, а k_L поочередно принимает все значения $N_{\text{ex}}, N_{\text{ex}} - 1, \dots, 1$, пропуская только уже "занятые" опорные точки, т.е. $k_L \neq k_1, k_2, \dots, k_{L-1}$. Для каждого набора k_i строят паде-аппроксиманту, вычисляют и сравнивают с минимальным из предыдущих минимизируемый функционал. После выполнения этих вычислений $N_{\text{ex}} - L$ раз заканчивают *малую итерацию*. Затем последней опорной точке присваивают номер, соответствующий минимальному из полученных значений функционала, и эта точка делается первой из опорных, а все остальные опорные сдвигаются на одно место вправо, малая итерация повторяется с очередной крайней правой точкой. Цикл из L малых итераций называют *большой итерацией*. Набор, полученный в результате очередной большой итерации, становится исходным для следующей большой итерации в том же смысле, в каком исходным для первой итерации был начальный набор.

Описанная процедура существенным образом основана на том благоприятном обстоятельстве, что для "перестройки" паде-аппроксиманты при изменении только одной последней опорной точки требуется сравнительно небольшое число операций: получение коэффициента рекуррентного соотношения и вычисление значений аппроксиманты по форму-

лам (3.16), (3.18) с использованием оставшихся неизменными значений полиномов $L-2$ -го и $L-1$ -го приближений.

Безусловным признаком окончания итерационного процесса служит получение в результате очередной большой итерации в точности того же набора $\{k_i\}$, что и после предыдущей большой итерации. Принципиальная сходимость описанной итерационной процедуры очевидна, поскольку полное число доступных для перебора вариантов $N_{ex}! / [L! (N_{ex} - L)!]$ конечно, хотя и очень велико. Практическая сходимость оказалась вполне приемлемой, число больших итераций в редких случаях превышает 10. Поэтому наряду с указанным безусловным признаком используются два условных: итерации заканчиваются, если на очередном "большом" шаге относительное изменение $S_{ост}$ не превышает некоторого малого значения (обычно 10^{-3}) или если число больших итераций достигло заданного максимума (обычно 10).

В необходимых случаях можно добавлять *дополнительные точки* к совокупности, из которой выбираются опорные, за счет, например, равномерной линейной интерполяции между исходными точками (на всем рассматриваемом интервале или его части). Следует подчеркнуть, что такое добавление не означает расширения совокупности аппроксимируемых точек. Последняя всегда совпадает с исходной, и минимизируемый функционал вычисляется только по отклонению аппроксиманты от исходных точек.

К достоинствам метода дискретной оптимизации по сравнению с непрерывным МНК можно отнести возможность использования без каких-либо затруднений сравнительно *разнообразных минимизирующих функционалов*. Например, вес σ_i^2 в статистической сумме (5.2) можно сделать функцией значения аппроксиманты, а не только номера i или исходных значений F_i . Эта возможность введения "гладкого веса" существенна в случае больших экспериментальных погрешностей аппроксимируемых значений. Принципиально ничего не меняется в итерационном процессе и при вычислении статистической суммы с учетом корреляций экспериментальных значений (5.1), хотя объем вычислений при этом возрастет примерно пропорционально числу "ненулевых" диагоналей корреляционной матрицы экспериментальных значений. Совсем прост и не вызывает увеличения объема вычислений переход от минимизации статистической суммы к чебышевскому критерию — минимизации максимума взвешенного отклонения.

Если в итерационном процессе сохранять не только статистически оптимальный, но еще и несколько близких к нему по качеству минимизации S наборов опорных точек $\{k_i\}$, то в результате можно получать не единственную аппроксиманту, а "пучок" *статистически эквивалентных кривых*, для которых значения S лежат в пределах подающей оценке погрешности этой величины. Из набора этих равноценных результатов можно выбрать один оптимальный уже не по статистическим, а физическим или аналитическим критериям. Эта возможность может

оказаться особенно полезной, если абсолютно лучшее в смысле минимизации S приближение получилось "раскачанным", т.е. принимающим бессмысленно большие по модулю значения между аппроксимируемыми точками. В случае рациональной аппроксимации раскачка обычно имеет особенно резкую форму — появляются так называемые шумовые дублеты (см. § 5.3), а такая неединственность оптимизации методом перебора позволяет найти статистически удовлетворительную гладкую аппроксиманту.

Недостатком метода является отсутствие его полного теоретического обоснования. Однако теоретические оценки, учитывающие свойства рациональных аппроксимант, свойства порядковых статистик, т.е. упорядоченных выборок случайно распределенных величин, и алгоритм перебора, показывают, что среднее квадратическое отклонение аппроксиманты, найденной методом перебора, от точного решения задачи МНК при $N_{ex} \gg L$ в $a\sqrt{N_{ex}/L}$ раз меньше, чем отклонение последнего от истинной кривой. Здесь a порядка единицы. Если отношение N_{ex}/L достаточно велико, то аппроксиманта, полученная дискретной оптимизацией, статистически эквивалентна решению задачи МНК, поэтому мы везде используем для нее результаты статистической теории, полученные для оценок МНК.

Если же N_{ex}/L не превышает нескольких единиц, то решение задачи можно уточнить, минимизируя статистическую сумму нелинейной задачи известными численными методами*. Поскольку для производных по опорным ординатам известен явный вид (3.12), (3.14), удобно использовать градиентные методы. Однако в практике авторов этой книги многочисленные попытки такого улучшения аппроксиманты после дискретной оптимизации никогда не приводили к статистически значимому улучшению статистической суммы. Подробнее особенности метода перебора для минимизации статистической суммы при небольших N_{ex}/L рассмотрены в [35] для простейшего случая — одномерного нормального распределения.

Еще одним недостатком описанного выше алгоритма перебора, как и других алгоритмов случайного поиска, в которых не используется полученная на предыдущих этапах информация о предпочтительном направлении поиска, является быстрый рост объема вычислений с увеличением числа параметров L и экспериментальных точек N_{ex} . Общее число операций на одну большую итерацию растет примерно пропорционально $L^2(N_{ex} - L)^2$, поэтому практически при обработке резонансных кривых максимальные значения числа параметров и числа экспериментальных точек были $L \approx 40$ и $N_{ex} \approx 500$. Примеры использования этого алго-

*Заметим, что аппроксимацию экспериментальной зависимости с помощью рациональной функции, при которой на первом этапе строится приближение Паде-2, а затем опорные ординаты или коэффициенты полиномов рассматриваются в качестве свободных параметров и численными методами оптимизируется статистический критерий, в [4,36] называют приближением Паде третьего рода (Паде-3).

ритма в модельных и практических задачах аппроксимации экспериментальных данных приведены и проанализированы в § 5.4.

5.3. ПОСТРОЕНИЕ И АНАЛИЗ РАЦИОНАЛЬНОЙ АППРОКСИМАНТЫ. ШУМОВЫЕ ПОЛЮСЫ

В том случае, когда оптимальное число параметров аппроксиманты заранее не известно, удобно пользоваться рекуррентным построением с последовательным наращиванием ранга приближения. Рекуррентные соотношения (3.16) – (3.18) дают такую возможность. Для каждого из вариантов интерполяции, сравниваемых в методе перебора, при вычислении статистической суммы необходимо иметь значения аппроксиманты не только в опорных точках, но и при всех z_i , для которых известны экспериментальные данные.

Коэффициенты рекуррентных соотношений, с помощью которых наращивается ранг приближения, выражаются через уже вычисленные на предыдущих этапах значения элементов аппроксимант рангов $L - 1$ и $L - 2$ и через экспериментальное значение в точке z_L . Поэтому для вычисления значения аппроксиманты $f^L(z)$ в каждой из N_{ex} точек z_i достаточно один раз (для всех z) найти коэффициенты рекуррентного соотношения и подставить в него полученные на предыдущих этапах значения $P_{(L-1)}(z_i)$, $Q_{(L-1)}(z_i)$, $P_{(L-2)}(z_i)$, $Q_{(L-2)}(z_i)$. Таким образом можно, используя критерий минимизации статистической суммы и значения аппроксиманты в точках z_i , определить число параметров и выбрать опорные точки оптимального приближения, не зная явного вида $P_N(z)$ и $Q_M(z)$.

Для анализа полученного приближения необходимо знать коэффициенты полиномов p_N^n и q_M^m , а не только их значения в отдельных точках. Если построение ведется с помощью рекуррентного соотношения типа $R_{(L)} = \alpha R_{(L-1)} + \beta(z - z_{L-1})R_{L-2}$, то коэффициенты полиномов p_N^n , q_M^m связаны с коэффициентами полиномов предыдущих приближений рекуррентным соотношением $r_{(L)}^n = \alpha r_{(L-1)}^n + \beta r_{(L-2)}^{n-1} - \beta z_{L-1} r_{(L-2)}^n$. Таким образом, определив методом перебора оптимальный набор узлов и использовав уже полученные для него коэффициенты рекуррентных соотношений, можно найти явный вид аппроксиманты именно для этого случая, не затрачивая времени на нахождение коэффициентов полиномов при других наборах узлов интерполяции. Можно поступить еще проще: поскольку значения полиномов оптимального приближения $P_N(z)$ и $Q_M(z)$ известны в N_{ex} точках, где $N_{\text{ex}} \gg \gg N + M + 1$, взяв любые N из них для $P_N(z)$ и любые M для $Q_M(z)$, воспользоваться простыми алгоритмами полиномиальной интерполяции, например интерполяционной формулой Лагранжа (1.9), и построить эти полиномы.

Следующий этап — поиск корней числителя и знаменателя аппроксиманты — осуществляется с помощью стандартных численных процедур. Наиболее эффективным оказалось использование в качестве начального алгоритма Мюллера с последующим уточнением методом Ньютона. При больших степенях полиномов эта задача может оказаться неустойчивой, что накладывает ограничения на максимальный ранг аппроксиманты, примерно совпадающие с условием $L \approx 40$, которое возникает из-за ограниченного быстродействия алгоритма перебора.

Зная корни числителя и знаменателя, можно перейти к эквивалентному представлению аппроксиманты

$$f^L(z) = p_{(L)}^N \prod_{\alpha=1}^N (z - z_{\alpha}) / \prod_{\beta=1}^M (z - z_{\beta}),$$

которое позволяет анализировать ее поведение не только в точках z_i , но и между ними. Еще удобнее для этого полюсное разложение аппроксиманты

$$f^L(z) = A_{N-M}(z) + \sum_{\beta} a_{\beta} / (z - z_{\beta}) \equiv A_{N-M}(z) + \sum_{i=1}^{l_1} \frac{a_i}{z - p_i} + \sum_{k=1}^{l_2} \frac{a_k(z - \epsilon_k) + \beta_k}{(z - \epsilon_k)^2 + \gamma_k^2}, \quad (5.8)$$

которое получается, если из $f^L(z)$ выделить целую часть (она имеется при $N \geq M$), при этом коэффициенты определяются соотношением

$$a_{\beta} = [P_N(z) - A Q_M(z)] / \partial [Q_M(z)] / \partial z |_{z_{\beta}}.$$

При кратных корнях знаменателя полюсное разложение сложнее.

Даже в случае использования линейного МНК или других методов аналитической аппроксимации, для которых теоретически гарантирована единственность решения, при численной реализации этих методов вследствие ограниченной точности вычислений или по иным причинам полученное решение иногда "раскачивается". При рациональной аппроксимации эта раскачка носит совершенно специфический характер, обусловленный наличием у аппроксиманты полюсных особенностей, которых нет у аппроксимируемых функций других классов. Рассмотрим это явление подробнее.

Пусть, постепенно наращивая ранг L , мы строим паде-аппроксиманту для поточечно заданной функции, которая сама является рациональной, т.е. отношением полиномов $P_{N_0}(z)/Q_{M_0}(z)$. Дойдя до $L_0 = N_0 + M_0 + 1$, мы восстановим исходную функцию по любым опорным точкам. При дальнейшем увеличении ранга приближения L мы все равно должны получить $f^L(z) = P_{N_0}(z)/Q_{M_0}(z)$, т.е. приближение неполного ранга. Имен-

но к такому результату приводили бы описанные выше алгоритмы паде-аппроксимации при неограниченной точности вычислений. Если же вычисления ведутся с ограниченной, например машинной, точностью, то приближение $f^{L_0}(z)$ будет с этой же точностью проходить только через точки, по которым оно построено, в остальных будут малые отклонения восстановленных значений от истинных.

Как показано в § 3.6, в этом случае увеличение ранга приближения на единицу вызывает появление шумовых особенностей аппроксиманты на бесконечности, а увеличение ранга на две единицы приводит к построению аппроксиманты $(Az + B)P_{N_0}(z)/(A'z + B')Q_{M_0}(z)$, где A и A' , B и B' мало отличаются друг от друга. Но эта небольшая количественная разница обуславливает качественно иной результат — биномы не сокращаются и у аппроксиманты появляется "шумовой дублет", близкие нуль и полюс, — она обращается в нуль при $z = -B/A$ и в бесконечность при $z = -B'/A'$.

Найдя корни знаменателя аппроксиманты, ее можно представить в виде полюсного разложения (1.1). В этом разложении шумовой полюс будет присутствовать в виде слагаемого $a/(z - b)$, где a — малая величина, а $b = -B'/A'$. Сокращение почти одинаковых биномов $Az + B$ и $A'z + B'$ в числителе и знаменателе соответствует отбрасыванию этого слагаемого в полюсном разложении. Таким образом, несмотря на бесконечную раскачку, ее локальный характер [слагаемое $a/(z - b)$ при малых a заметно отлично от нуля только при z , близких к b] позволяет весьма просто с ней бороться.

Это рассуждение в общем справедливо и тогда, когда шумовой полюс появляется вследствие наличия случайных погрешностей у исходных данных. Если при этом погрешность в одной точке заметно превышает погрешности в других точках, то, как показано в § 3.6, шумовой дублет возникает вблизи точки, выпадающей из общей закономерности. Шумовые полюсы в большинстве случаев появляются, как это и показывают рассуждения, приведенные в начале параграфа, когда ранг аппроксиманты уже превысил оптимальное значение. Наряду со статистическими критериями это обстоятельство служит независимым указанием на то, что аналитическая информация, содержащаяся в аппроксимируемой совокупности точек, исчерпана. Довольно часто отсутствует надежная информация о статистической погрешности аппроксимируемых значений и выполнение критериев типа χ^2 непосредственно проконтролировать нельзя, тогда возникновение шумовых полюсов может быть единственным объективным указанием на то, что оптимальное значение L уже достигнуто и превзойдено. Это, например, относится к случаю, когда аппроксимируются не экспериментальные, а расчетные данные, для которых априорная оценка погрешности затруднительна или невозможна.

5.4. ПРИМЕРЫ ИСПОЛЬЗОВАНИЯ ДИСКРЕТНОЙ ОПТИМИЗАЦИИ ДЛЯ АППРОКСИМАЦИИ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Алгоритмы рациональной аппроксимации, описанные выше, применялись и совершенствовались в течение примерно десяти лет. Ниже для иллюстрации их возможностей и ограничений приведены примеры получаемых результатов, относящихся к тем случаям, когда не требовалось оценивать погрешности параметров аппроксиманты. Аппарат для вычисления таких погрешностей изложен в гл. 6, там же даны и соответствующие примеры.

А. Начнем с простого модельного примера, на котором наглядно иллюстрируются особенности работы алгоритма и получаемых результатов. Величину $\sigma(E)$ брали в виде суммы двух резонансных слагаемых брейт-вигнеровского типа:

$$\sigma(E) = \frac{\beta_1}{(E - \epsilon_1)^2 + \Gamma_1^2/4} + \frac{\beta_2}{(E - \epsilon_2)^2 + \Gamma_2^2/4} \quad (5.9)$$

со значениями параметров $\epsilon_1 = 5$, $\Gamma_1 = \beta_1 = 1$, $\epsilon_2 = 10$, $\beta_2 = \Gamma_2 = 4,542$. Последнее число было задано датчиком случайных чисел. Интервал изменения аргумента $0 \leq E \leq 15$ с шагом $\Delta E = 0,1$ (всего 151 точка). "Экспериментальный" разброс моделировался с помощью датчика случайных чисел из нормального распределения с равным нулю средним. Связь относительной погрешности с сечением в точке была выбрана в виде

$$\Delta(E) = \Delta_0 \sqrt{\sigma_{\max}/\sigma(E)}.$$

Это выражение описывает погрешность, промежуточную между пуассоновской (тогда в этом выражении стоял бы квадратный корень) и постоянной относительной погрешностью. Амплитуда $\Delta_0 = 2\%$, что соответствует $2\% \leq \Delta(E) \leq 5\%$ и средней относительной погрешности $\bar{\Delta} = 3,44\%$ (рис. 5.1). Критерием оптимизации аппроксиманты была выбрана минимизация среднего квадратического относительного отклонения, т.е. в программе использовалась искаженная информация относительно статистических свойств экспериментальной погрешности. Аппроксиманта представлялась в виде

$$\sigma(E) = C + \sum_i \frac{a_i}{E - p_i} + \sum_k \frac{a_k (E - \epsilon_k) + \beta_k}{(E - \epsilon_k)^2 + \gamma_k^2},$$

где все постоянные действительные результаты обработки при $L = 8, 9$, 10 приведены в табл. 5.1.

Идея метода дискретной оптимизации подсказана свойствами порядковых статистик. Если $L \ll N_{\text{ex}}$, то среди N_{ex} точек, случайно разбросанных относительно истинной кривой, есть "наилучшие", наименее ук-

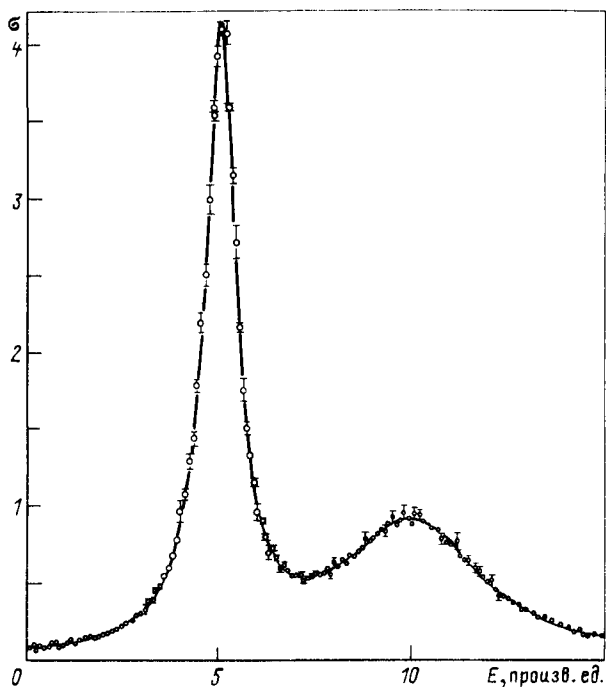


Рис. 5.1. Результат обработки модельной кривой [функция (5.7)]:

точки разбросаны датчиком случайных чисел с $\bar{\Delta} = 3,44\%$; сплошная линия — аппроксиманта, соответствующая оптимальному числу параметров $L = 9$. Значения ее параметров приведены в табл. 5.2

лонившиеся, распределение отклонений которых гораздо уже полного. Естественно ожидать, что опорные точки, полученные в результате описанного выше итерационного процесса, при прочих равных условиях должны отбираться именно из них. На примере модельной кривой справедливость такого предположения может быть проверена. Кроме того, восстанавливаемую кривую можно сравнить не только с исходными точками, но и с истинной функцией. Функционалы, характеризующие перечисленные средние квадратические относительные отклонения, также приведены в таблице наряду с полученными параметрами аппроксиманты: $\bar{\Delta}_{\text{эксп}}$ и $\bar{\Delta}_{\text{ист}}$ — отклонения аппроксиманты от "экспериментальных" и истинных точек, $\Delta_{\text{опор}}$ — отклонение опорных точек от истинных.

Данные, приведенные в табл. 5.1, показывают следующее:

1) $\bar{\Delta}_{\text{ист}}$ заметно (в оптимальном варианте примерно в 6 раз) меньше $\bar{\Delta}_{\text{эксп}}$. Это уменьшение погрешностей качественно согласуется с простейшей оценкой, полученной в предположении, что дисперсию параметров (опорных ординат) результирующей кривой можно вычислять как дисперсию среднего по N_{ex}/L значениям;

Таблица 5.1. Результаты обработки модельной кривой

Характеристика аппроксиманты	Точные значения параметров	$L = 8$	$L = 9$	$L = 10$
C	0	0	0,0019	0
a_1	0	-0,0081	-0,0009	0,0022
β_1	1	0,9828	0,9858	0,9733
γ_1	0,5	0,4957	0,4975	0,4939
ϵ_1	5	5,001	4,998	4,991
a_2	0	0,0060	0,0075	-0,015
β_2	4,542	4,591	4,549	4,630
γ_2	2,271	2,278	2,276	2,292
ϵ_2	10	9,988	9,977	10,008
a_1	-	-	-	$0,263 \cdot 10^{-4}$
p_1	-	-	-	1,598
$\bar{\Delta}_{\text{экср, \%}}$	-	3,328	3,352	3,344
$\bar{\Delta}_{\text{ист, \%}}$	-	0,589	0,552	0,775
$\bar{\Delta}_{\text{опор, \%}}$	-	0,466	0,580	3,560

2) $\bar{\Delta}_{\text{опор}}$ близко к оптимальному $\bar{\Delta}_{\text{ист}}$, т.е. вся восстановленная кривая отклоняется от истинной в среднем так же, как выбранные опорные точки, а эти последние действительно лежат гораздо ближе к истинной кривой, чем каждая из N_{ex} точек в среднем;

3) аппроксиманта при $L = 10$ после отбрасывания шумовых полюсов сильнее отклоняется от истинной кривой, чем оптимальная аппроксиманта при $L = 8$, и полученные параметры сильнее отличаются от истинных;

4) точность восстановления ϵ_i и $\gamma_i = \Gamma_i/2$ координат полюсов в комплексной плоскости заметно выше, чем для a_i и β_i [в модели (5.9) $a_i = 0$];

5) шумовой полюс появляется, когда L на две единицы превышает "истинное" значение $L = 8$, при этом в качестве одной из опорных была выбрана точка при $E = 1,6$, для которой в "экспериментальное" значение датчиком случайных чисел была внесена наибольшая относительная погрешность - 9,6%.

Результаты обработки многих других модельных задач, в том числе и более сложных, чем приведенная, неизменно качественно подтверждали выводы 1 - 5.

Б. Сечения реакций ${}^6\text{Li}(n, \alpha)$, ${}^{10}\text{B}(n, \alpha_0)$, ${}^{10}\text{B}(n, \alpha_1)$ и ${}^{10}\text{B}(n, \alpha)$ широко используются в качестве стандартов при экспериментальных исследованиях взаимодействия нейтронов с ядрами. Наиболее полной и теоретически обоснованной современной оценкой этих сечений является библиотека ENDF/B-V [12]. Результаты этой оценки приводятся в табличном виде, что не всегда удобно для использования. Поэтому в [37] зависимости $\sqrt{E} \sigma(E)$ для этих реакций были аппроксимированы ра-

Таблица 5.2. Значения параметров аппроксимации величин $\sqrt{E} \sigma_{n\alpha}(E)$, 10^{-28} м²

Характеристика аппроксиманты	⁶ Li(<i>n</i> , α)	¹⁰ B(<i>n</i> , α ₀)	¹⁰ B(<i>n</i> , α ₁)	¹⁰ B(<i>n</i> , α)
<i>C</i>	8,4361	1,1931	2,3696	5,8468
<i>a</i> ₁	-2,4420 · 10 ⁴	-286,25	40,314	26,020
<i>p</i> ₁	-5,0045 · 10 ³	1,3381 · 10 ³	-18,296	-14,492
<i>a</i> ₁	-1,3978 · 10 ³	183,62	-2,5382 · 10 ³	3,5630 · 10 ³
<i>β</i> ₁	5,4164 · 10 ⁶	2,3854 · 10 ⁴	1,8669 · 10 ⁶	1,3885 · 10 ⁶
<i>γ</i> ₁	1,1053 · 10 ³	130,11	345,76	319,58
<i>ε</i> ₁	2,4509 · 10 ³	467,90	230,49	287,93
<i>a</i> ₂	414,54	49,034	465,45	893,84
<i>β</i> ₂	9,0840 · 10 ⁴	1,7806 · 10 ⁴	4,7728 · 10 ⁴	7,2963 · 10 ⁴
<i>γ</i> ₂	44,533	152,92	124,88	138,18
<i>ε</i> ₂	238,01	245,03	449,79	447,16
<i>Δ</i> _{отн} , %	0,69	0,42	0,24	0,21
<i>N</i> _{ex}	170	69	72	68
<i>E</i> _{min} , эВ	10 ⁻⁷	10 ⁻⁴	10 ⁻⁷	10 ⁻⁴

циональными функциями с помощью описанных выше алгоритмов; результаты приведены в табл. 5.2. Поскольку случайный разброс точек отсутствует и критерий χ^2 непосредственно использован быть не может, число параметров аппроксимации *L* выбиралось из условия достижения заранее намеченной относительной точности описания в интервале до *E* = 900 кэВ, а именно погрешности 0,5 %, что заведомо меньше средней погрешности оцененного стандартного сечения, но по порядку близко к ней в рассмотренных случаях. Оптимальное число параметров *L*_{опт} оказалось равным 11 во всех рассмотренных случаях при *N* = *M* = 5. Поскольку при поточечном задании оцененного сечения необходимо задавать абсциссы и ординаты, полное число подлежащих хранению значений в этих случаях в 10 – 20 раз превышает число параметров аналитической аппроксимации.

Следует отметить, что при обработке результатов по сечению реакции ⁶Li(*n*, α) в табличных данных с помощью программы была обнаружена одна выпадающая точка – значение, соответствующее *E* = 750 кэВ и равное 0,2462 · 10⁻²⁸ м². Она смещена больше чем на 5 % вниз от любой гладкой интерполяции по соседним точкам, и программа паде-аппроксимации неизменно использовала для ее описания шумовой полюс. Предположительно, в библиотеке ENDF/B-V [12] допущена опечатка, и правильное значение равно 0,2642.

В. Аналогичные расчеты были выполнены при переводе в аналитическую форму данных по оцененным сечениям пороговых реакций под действием нейтронов (библиотека БОСПОР [38]). Были обработаны

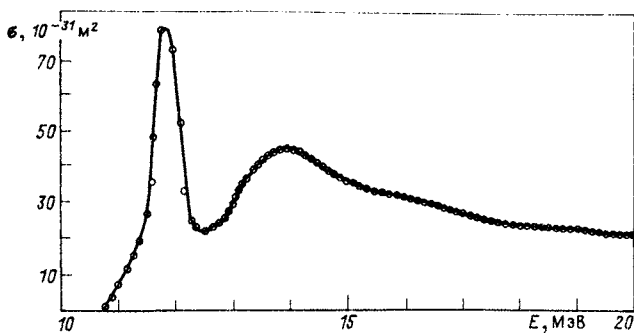


Рис. 5.2. Пример аппроксимации оцененных данных из библиотеки БОСПОР – сечение реакции $^{16}\text{O}(n, p)^{16}\text{N}$ (см. [38])

в общей сложности 144 кривые с результирующим примерно 20-кратным сокращением подлежащей хранению числовой информации. На рис. 5.2. приведен один из типичных случаев такой аппроксимации.

Г. На рис. 5.3 представлен сравнительно сложный по числу параметров случай ($L = 28$).

Д. На рис. 5.4 приведены результаты обработки полных нейтронных сечений изотопов железа ^{54}Fe и ^{56}Fe , измеренных методом времени пролета в интервале энергий $E = 1 \div 70$ кэВ [40]. Результаты анализа в виде значений резонансных параметров даны в табл. 5.3.

Константа C и слагаемое $a/(E - p)$, соответствующее действительному полюсу, описывают суммарный вклад потенциального рассеяния и "хвосты" резонансов, лежащих за пределами рассматриваемого интервала, информация о которых недостаточна для определения полного набора параметров $\epsilon_k, \gamma_k, \alpha_k, \beta_k$. Наибольший интерес представляют слагаемые такого типа, соответствующие резонансам при отрицательных энергиях нейтронов, близких к нулю. Обычно их описывают так же, как и резонансы при положительных энергиях. Следует, однако, отметить, что и без того некорректная задача аналитического продолжения функции осложняется наличием в этом случае у аппроксимируемой функции

Таблица 5.3. Параметры аппроксимации полных нейтронных сечений ^{54}Fe и ^{56}Fe

Нуклид	$a, 10^{-28} \text{ м}^2 \cdot \text{кэВ}$	$\beta, 10^{-28} \text{ м}^2 \cdot \text{кэВ}^2$	$\gamma, \text{ кэВ}$	$\epsilon, \text{ кэВ}$	Другие параметры
^{54}Fe	47,28	105,33	0,582	7,71	$C = 17,18$
	29,7	73,141	1,43	52,53	
	23,1	$5,796 \cdot 10^4$	60,65	45,5	
^{56}Fe	29,1	51,46	0,791	27,6	$C = 0$ $a = 65,7$ $p = -6,40$
	-15,15	1689	24,7	42,5	

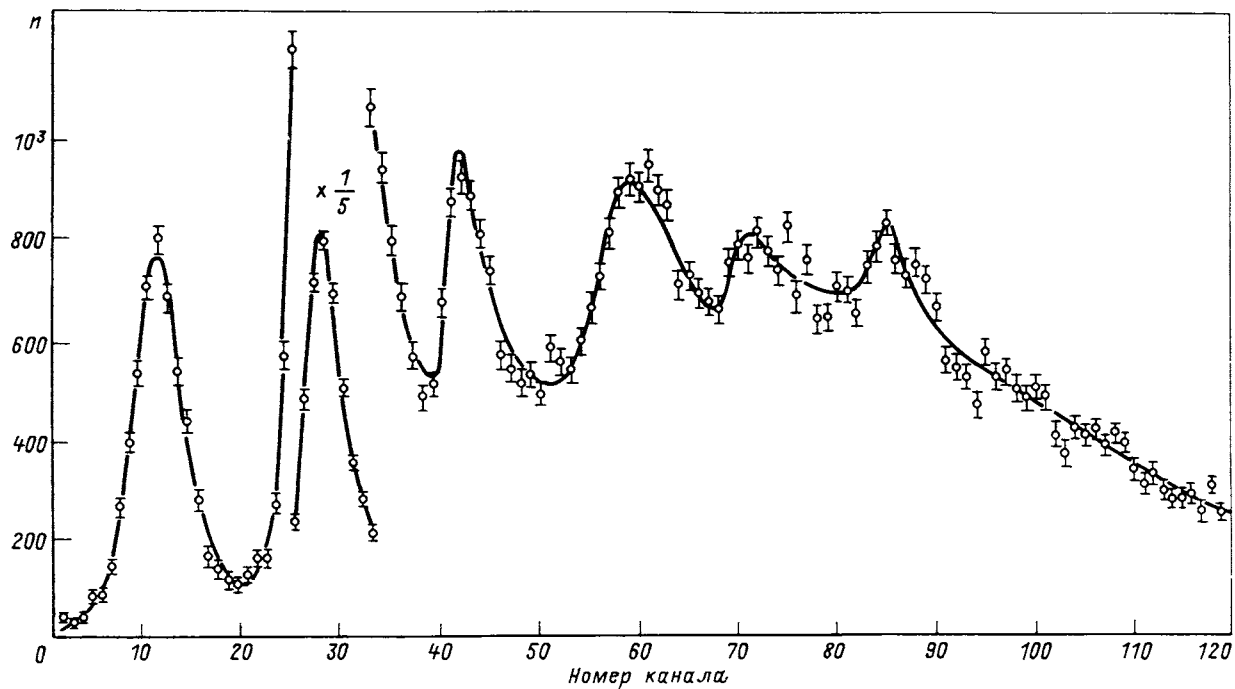


Рис. 5.3. Результат аппроксимации временного спектра при измерении сечения реакции $^{52}\text{Cr}(p, n)$ [39] для числа параметров $L = 28$

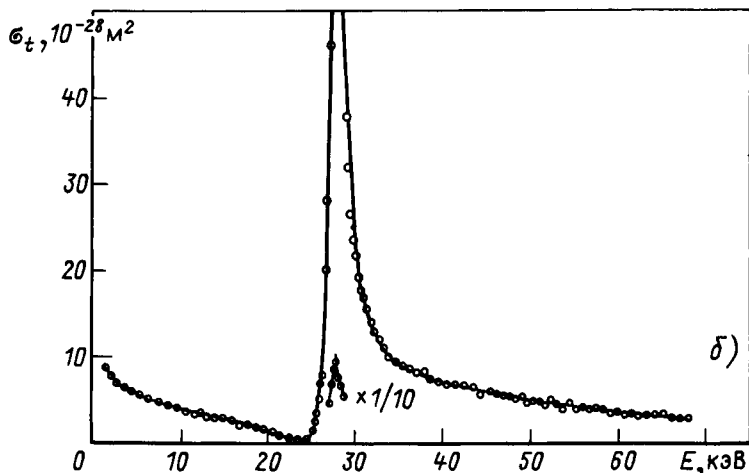
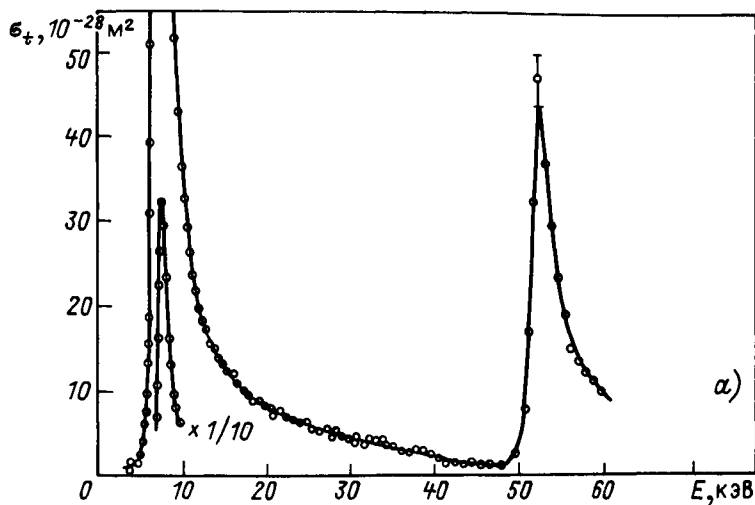


Рис. 5.4. Результаты измерения полного нейтронного сечения для ^{54}Fe (а) и ^{56}Fe (б) (точки) и аппроксимации полученных данных рациональной функцией (сплошная кривая)

точки ветвления (вклад $1/\sqrt{E}$), что чаще всего лишает полученные параметры физического смысла. Отсюда — известные различия при определении их разными авторами. Несомненно, наличие отрицательного уровня энергии у ^{56}Fe и его вклад в виде слагаемого $a/(E - p)$ определяются в нашем анализе, однако интерпретировать p как энергию резонанса при этом нужно уже с осторожностью. Особый вопрос — описание вклада потенциального рассеяния. Для обоих изотопов железа анализ дает в выражении для полного сечения слагаемое, имеющее формально

резонансный характер, однако ширина этого резонанса примерно равна или больше всего рассматриваемого интервала, и его вклад слабо зависит от энергии, поэтому мы относим его к потенциальному рассеянию.

5.5. РЕЗОНАНСНЫЙ АНАЛИЗ НА ОСНОВЕ РАЦИОНАЛЬНОЙ АППРОКСИМАЦИИ

Как уже упоминалось в гл. 1, рациональная аппроксимация как метод аналитического продолжения позволяет при обработке энергетической зависимости сечений ядерных реакций определять параметры полюсов S -матрицы, имеющие физический смысл резонансных параметров. В настоящем параграфе математические аспекты этого вопроса рассматриваются несколько подробнее.

Одна из принципиальных трудностей описания экспериментальных данных аналитическими функциями состоит в том, что физически наблюдаемыми являются не сами аналитические функции $f(z)$, которые могут принимать комплексные значения, а действительные величины $|f(z)|$, $\text{Re} f(z)$, $\text{Im} f(z)$. В том случае, когда эти величины известны при всех комплексных значениях аргумента, можно однозначно определить аналитическую функцию во всей комплексной плоскости [18], но экспериментальные данные известны лишь при некоторых действительных значениях аргумента.

Так, сечение упругого рассеяния $\sigma_e^{(l)}$, сечение реакции $\sigma_r^{(l)}$ и полное сечение $\sigma_t^{(l)}$ для частиц с заданным орбитальным моментом l следующим образом выражаются через S_l аналитическую функцию энергии или волнового вектора k [41]:

$$\left. \begin{aligned} \sigma_e^{(l)} &= (\pi/k^2) (2l + 1) |1 - S_l|^2; \\ \sigma_r^{(l)} &= (\pi/k^2) (2l + 1) (1 - |S_l|^2); \\ \sigma_t^{(l)} &= (\pi/k^2) (2l + 1) (1 - \text{Re} S_l). \end{aligned} \right\} \quad (5.10)$$

Если известны два из трех сечений, то известно $\text{Re} S_l$ и, с точностью до знака $\text{Im} S_l$ на действительной оси, т.е. при действительных $z = k$. Знак $\text{Im} S_l$ определяется из физических соображений (см., например, [4]). Используя условие Коши–Римана, нетрудно убедиться, что если $f(z)$ — аналитическая функция, то $f^*(z)$ — неаналитическая в любой точке комплексной плоскости, кроме тех (если они существуют), где $du/dx = \partial v/\partial y = \partial u/\partial y = \partial v/\partial x = 0$. Поэтому функции $|f(z)|^2 = f(z)f^*(z)$, $\text{Re} f(z) = (1/2)[f(z) + f^*(z)]$ и $\text{Im} f(z) = (1/2i)[f(z) - f^*(z)]$ также неаналитические. Аналитической функцией будет $f^*(z^*)$, и можно построить аналитические функции $f^+(z) = (1/2)[f(z) + f^*(z^*)]$ и $f^-(z) = (1/2i)[f(z) - f^*(z^*)]$, которые на действительной оси совпадают с неаналитическими $\text{Re} f(z)$ и $\text{Im} f(z)$.

Как уже отмечалось, задание аналитической функции на некотором отрезке полностью определяет ее во всей комплексной плоскости, поэтому, используя $f^+(z) + i f^-(z)$, получим аналитическую функцию, совпадающую с $f(z)$ на действительной оси и, следовательно, во всей комплексной плоскости. При этом аналитическая функция $f^+(z) - i f^-(z)$ совпадает на действительной оси с неаналитической функцией $f^*(z)$. Если потребовать, чтобы аналитической функцией была $f^*(z)$, то $f(z)$ будет неаналитической — выбор одной из этих функций в качестве S_I определяется теми же соображениями, что и знак $\text{Im } S_I$.

Гораздо чаще при обработке экспериментальных данных известен только модуль или только действительная или мнимая часть аналитической функции на действительной оси. В любом из этих случаев можно построить единственную аналитическую функцию $\varphi(z)$, принимающую заданные действительные значения на действительной оси. Как следует из теории аналитических функций, $\varphi(z)$ будет принимать комплексно-сопряженные значения в комплексно-сопряженных точках, поэтому ее можно представить в любом из следующих видов: $\varphi(z) = f(z)f^*(z^*)$, $\varphi(z) = (1/2)[f(z) + f^*(z^*)]$, $\varphi(z) = (1/2i)[f(z) - f^*(z^*)]$. Однако в каждом случае такое разбиение можно сделать не единственным образом — функция $\varphi(z)$ определена с точностью до множителя $c_1(z) = 1/c_1^*(z^*)$ в первом случае, с точностью до слагаемого $c_2(z) = -c_2^*(z^*)$ во втором и с точностью до слагаемого $c_3(z) = c_3^*(z^*)$ в третьем. Рассматривая значения этих слагаемых и множителя на действительной оси и требуя их аналитичности во всей плоскости, нетрудно убедиться, что в первом случае $c_1 = e^{i\delta}$, во втором $c_2 = ia$, в третьем $c_3 = b$, где δ, a, b — действительные постоянные. Однако основное затруднение при переходе от $\varphi(z)$ к $f(z)$ заключается не в этих несущественных постоянных, а в том, что любую из пары комплексно-сопряженных особых точек, которые имеются у $\varphi(z)$, можно отнести к $f(z)$, а сопряженную ей — к $f^*(z^*)$. Для проведения такого разбиения необходимо пользоваться дополнительной информацией, например привлекать известные из теории аналитические свойства S_I . При анализе резонансных сечений, таким образом, можно однозначно определить положения и ширины резонансов, которым соответствуют полюсы аналитической функции S_I .

В теории ядерных реакций общее выражение для энергетической зависимости сечения в резонансной области имеет вид

$$\sigma(E) = \sum_{k=1}^m \left| c_k + \sum_{j=1}^{n_k} \frac{\rho_{kj} e^{i\varphi_{kj}}}{E - E_{kj} + i\Gamma_{kj}/2} \right|^2. \quad (5.11)$$

В нем учитывается как интерференция резонансного рассеяния с потенциальным, так и межрезонансная интерференция. При паде-аппроксими-

матрицы сечение восстанавливается в виде

$$\sigma(E) = \frac{P_{2n}(E)}{Q_{2n}(E)} = c + \sum_{k=1}^m \sum_{j=1}^{n_k} \left(\frac{\alpha_{kj}}{E - E_{kj} + i\Gamma_{kj}/2} + \frac{\alpha_{kj}^*}{E - E_{kj} - i\Gamma_{kj}/2} \right), \quad (5.12)$$

где $n = \sum_{k=1}^m n_k$; α_k — комплексные постоянные.

Сравнивая (5.11) и (5.12), получаем

$$\left. \begin{aligned} C &= \sum_{k=1}^m c_k^2; \\ \alpha_{kj} &= \rho_{kj} e^{i\varphi_{kj}} \left(c_k + \sum_l \frac{\rho_{kl} e^{-i\varphi_{kl}}}{E_{kj} - E_{kl} + i(\Gamma_{kj} + \Gamma_{kl})/2} \right). \end{aligned} \right\} \quad (5.13)$$

Если в описании участвует N резонансов, то для определения $2N + m$ величин (ρ_{kj} и φ_{kj} для каждого уровня и m постоянных c_k) есть $2N + 1$ уравнение (5.13), считая сопряженные, и задача не имеет однозначного решения. Поэтому ограничимся качественными выводами и рассмотрим частные случаи. Простейший случай — резонансное рассеяние на одном уровне, интерферирующее с потенциальным рассеянием, когда

$$\sigma(E) = \left| c + \frac{\rho e^{i\varphi}}{E - E^0 + i\Gamma/2} \right|^2 = c^2 + \frac{\gamma + i\delta}{E - E^0 + i\Gamma/2} + \frac{\gamma - i\delta}{E - E^0 - i\Gamma/2}, \quad (5.14)$$

где $\gamma + i\delta = a$. Используя (5.13), находим:

$$\left. \begin{aligned} \rho^4 - \rho^2 \Gamma^2 (c^2 + 2\delta/\Gamma) + \Gamma^2 (\delta^2 + \gamma^2) &= 0; \\ \cos \varphi &= \gamma/c\rho. \end{aligned} \right\} \quad (5.15)$$

Однако на практике не всегда известно, какая доля потенциального рассеяния интерферирует с рассеянием на данном уровне, поэтому представляет интерес следующее ограничение снизу на соответствующую величину, которое получается из очевидного условия $\rho^2 \geq 0$:

$$c^2 \geq (2\delta/\Gamma) (\sqrt{1 + \gamma^2/\delta^2} - 1). \quad (5.16)$$

Когда $C = c_k = 0$, т.е. нет интерференции резонансного рассеяния с по-

тенциальным, уравнения (5.13) в принципе разрешимы относительно ρ_{kj} и $\varphi_{kj} - \varphi_{kl}$. Так, для двух интерферирующих резонансов получим

$$\left. \begin{aligned} \rho_1^4 + \rho_1^2 \left[-2\delta_1\Gamma_1 - \frac{\Gamma_1^2\Gamma_2(\delta_1 + \delta_2)}{\epsilon^2 - \Gamma_1\Gamma_2} \right] + \frac{(\gamma_1^2 + \delta_1^2)\Gamma_1^2\epsilon^2}{\epsilon^2 - \Gamma_1\Gamma_2} &= 0; \\ \rho_2 &= -\Gamma_2(\delta_1 - \delta_2 - \rho_1^2/\Gamma_1); \operatorname{tg}(\varphi + \beta) = \delta_1/\gamma_1 + \rho_1^2/\gamma_1\Gamma_1; \\ \epsilon^2 &= (E_1 - E_2)^2 + (\Gamma_1 + \Gamma_2)^2/4; \\ \operatorname{tg}\beta &= -(\Gamma_1 + \Gamma_2)/2(E_1 - E_2); \varphi = \varphi_2 - \varphi_1. \end{aligned} \right\} (5.17)$$

Решение этой системы дает два набора резонансных параметров. Следует отметить, что результаты анализа с помощью выражений (5.17) чувствительны к сравнительно малым изменениям наблюдаемых величин.

Если $C = c_k = 0$, то, суммируя по j в правой части (5.13), можно найти

$$\sum_j (a_{kj} + a_{kj}^*) = \sum_j \gamma_{kj} = 0. \quad (5.18)$$

Выражение (5.18) дает способ разбиения резонансов при отсутствии интерференции резонансного рассеяния с потенциальным на группы с межрезонансной интерференцией, внутри каждой из которых выполняется равенство (5.18), а γ_{kj} определяются с помощью приближения Паде.

Ситуация упрощается, если потенциальное рассеяние отсутствует, т.е. речь идет о сечении реакции, идущей через составное ядро, например о сечении деления, и построена паде-аппроксиманта, у которой $N = M - 2 = 2(m - 1)$. После нахождения корней числителя и знаменателя ее можно разложить на два комплексно-сопряженных множителя:

$$\sigma(E) = \frac{P_{2(m-1)}(E)}{Q_{2m}(E)} = A \frac{\prod_{k=1}^{m-1} (E - \epsilon'_k + i\gamma'_k)(E - \epsilon'_k - i\gamma'_k)}{\prod_{j=1}^m (E - \epsilon_j + i\gamma_j)(E - \epsilon_j - i\gamma_j)}. \quad (5.19)$$

Один из множителей, например тот, полюсы и нули которого лежат в верхней полуплоскости, и приравнивается к полюсному разложению амплитуды, не содержащему постоянных слагаемых c_k .

Заметим, что при практическом использовании данных по нейтронным сечениям и оценке этих данных все чаще под резонансным анализом понимают представление зависимости $\sigma(E)$ в формализме Адлер-Адлера [42]. Это представление есть не что иное, как рациональная аппроксимация функций $\sqrt{E}\sigma(E)$ с параметризацией (5.8) полюсного разложения полученной аппроксиманты.

Методы анализа экспериментальных данных по рассеянию, основанные на использовании теоретических и феноменологических результатов физики ядерных взаимодействий, подробно изложены в [4].

ГЛАВА 6. ОЦЕНКА ПОГРЕШНОСТЕЙ ПАДЕ-АППРОКСИМАНТЫ

6.1. ИНФОРМАЦИОННАЯ И КОВАРИАЦИОННАЯ МАТРИЦЫ ПРИ ПАРАМЕТРИЗАЦИИ РЕЗОНАНСНЫМИ ПАРАМЕТРАМИ И ОПОРНЫМИ ОРДИНАТАМИ

В предыдущей главе был рассмотрен этап обработки экспериментальных данных, целью которого было получение рациональной функции, аппроксимирующей истинную зависимость. Любой критерий выбора такой аппроксиманты, в частности используемый здесь МНК, и любой метод ее получения, будь то решение системы уравнений, численные методы поиска экстремумов или метод перебора, приводят к построению функции, параметры которой определяются случайными величинами — экспериментальными данными и в этом смысле являются случайными. Например, если распределение погрешностей эксперимента в точке считать нормальным, то значение найденной с помощью МНК аппроксимирующей функции в этой точке является оценкой среднего этого нормального распределения, т.е. оценкой значения истинной функции. Эта оценка сама подчиняется некоторому статистическому распределению, и ее дисперсия определяет погрешность описания в данной точке. В этой главе рассмотрено определение погрешностей параметров аппроксиманты и их корреляций.

Для задачи линейной регрессии, т.е. при линейной зависимости истинной функции от оцениваемых параметров, имеется хорошо развитая теория, которая позволяет по распределениям погрешностей эксперимента получать распределения оценок параметров и их характеристики. В этом случае оценки подчиняются нормальному распределению, а ковариационная матрица погрешностей параметров аппроксиманты V получается обращением информационной матрицы A — матрицы Фишера (см., например, [32 — 33]). Матрица Фишера при многомерном нормальном распределении погрешностей эксперимента следующим образом выражается через математическое ожидание производных статистической суммы (5.1) по параметрам аппроксиманты

$$A_{\mu\nu} = (V^{-1})_{\mu\nu} = \frac{\partial S}{\partial p_\mu} \frac{\partial S}{\partial p_\nu} . \quad (6.1)$$

В случае нескоррелированных погрешностей исходных данных, когда выражение для статистической суммы упрощается и сводится к (5.2),
82

подстановка (5.2) в (6.1) дает

$$A_{\mu\nu} = \sum_{i=1}^{N_{\text{ex}}} \frac{1}{\sigma_i^2} \frac{\partial f(z_i)}{\partial p_\mu} \frac{\partial f(z_i)}{\partial p_\nu}. \quad (6.2)$$

В дальнейшем будем рассматривать именно этот случай.

Наряду с ковариационной используется корреляционная матрица с элементами

$$\rho_{\mu\nu} = V_{\mu\nu} / \sqrt{V_{\mu\mu} V_{\nu\nu}}. \quad (6.3)$$

Ее диагональные элементы равны единице, а недиагональные — коэффициентам корреляции соответствующих параметров.

В задачах нелинейной регрессии, к числу которых относится аппроксимация экспериментальных зависимостей с помощью рациональных функций, гораздо сложнее не только сам процесс получения оценки, но и определение погрешностей аппроксиманты при небольшом числе экспериментальных точек N_{ex} . Оценка МНК в этой задаче лишь асимптотически нормальна, т.е. распределение параметров аппроксиманты стремится к нормальному при $N_{\text{ex}} \rightarrow \infty$. Асимптотическое разложение ковариационной матрицы параметров имеет вид

$$A^{-1} + O(N_{\text{ex}}^{-1}),$$

где A определено формулой (6.2). В дальнейшем будем предполагать, что N_{ex} достаточно велико, для того чтобы в качестве ковариационной матрицы и в этой задаче использовать просто A^{-1} .

Как упоминалось выше, параметры паде-аппроксиманты можно выбирать различными способами. Удобными для резонансного анализа, например, представляются параметры полюсного разложения (5.8). Производные аппроксиманты по этим параметрам, входящие в (6.1), (6.2), элементарно вычисляются прямым дифференцированием:

$$\left. \begin{aligned} \frac{\partial f}{\partial \alpha_k} &= \frac{z - \epsilon_k}{Q_k}; & \frac{\partial f}{\partial \beta_k} &= \frac{1}{Q_k}; \\ \frac{\partial f}{\partial \gamma_k} &= -2 \frac{P_k \gamma_k}{Q_k^2}; \\ \frac{\partial f}{\partial \epsilon_k} &= -\frac{\alpha_k}{Q_k} + 2 \frac{P_k(z - \epsilon_k)}{Q_k^2}, \end{aligned} \right\} \quad (6.4)$$

где $Q_k = \gamma_k^2 + (z - \epsilon_k)^2$; $P_k = \alpha_k(z - \epsilon_k) + \beta_k$.

Таблица 6.1. Оцененные значения резонансных параметров, их погрешностей и коэффициентов корреляции

Параметр	c	α_1	β_1	γ_1	ϵ_1	α_2	β_2	γ_2	ϵ_2
Точное значение	1	0,2	1	0,3	0,5	0	1	0,5	-0,5
Оценка	1,03	0,193	1,01	0,302	0,501	-0,03	0,974	0,497	-0,494
Абсолютная погрешность	0,14	0,051	0,08	0,008	0,004	0,22	0,062	0,011	0,033
ρ_{ik}	1	-0,54	-0,88	-0,86	0,35	0,74	-0,58	-0,17	-0,69
		1	0,46	0,44	-0,92	-0,54	0,57	0,26	0,57
			1	0,99	-0,33	-0,93	0,18	-0,23	0,88
				1	-0,32	-0,90	0,16	-0,24	0,85
					1	0,44	-0,37	-0,11	-0,47
						1	-0,87	0,30	-0,99
							1	0,87	0,84
								1	-0,28
									1

Рассмотрим пример, иллюстрирующий свойства ковариационной матрицы этих параметров. В качестве модельной была выбрана функция

$$F(z) = 1 + \frac{1}{(z + 0,5)^2 + 0,5^2} + \frac{1 + 0,2z}{(z - 0,5)^2 + 0,3^2}, \quad (6.5)$$

заданная на отрезке $[-1,1]$ с шагом 0,05 (всего 41 точка). С помощью датчика случайных чисел, выбранных из нормального распределения, к значениям функции (6.5) добавлялись случайные отклонения с диспер-

Таблица 6.2. Оцененные значения опорных ординат, их погрешностей и коэф-

$\bar{\Delta} = 1\%$	z_i	-1	-0,6	-0,45	-0,25
	f_i	3,35	5,506	5,871	5,656
	$\Delta f_i / f_{i_0} \%$	0,80	0,45	0,47	0,48
$\bar{\Delta} = 5\%$	z_i	-1	-0,9	-0,25	0,1
	f_i	3,05	3,766	5,687	6,638
	$\Delta f_i / f_i$	0,12	0,027	0,021	0,011
	$\rho_{ik}, \%$ кроме диагональной	1	-0,08	0,5	-0,4
		10,2	1	1,5	-0,7
		13,7	0,4	1	0,6
		-23,0	0,35	-3,9	1
		17	0,00	1,5	-0,6
		-23,5	0,00	-3,5	3,7
		-9,3	-1,1	-3,7	3,0
		-4,4	1,3	2,0	-1,0
		-9,4	-1,0	-3,1	2,7

Примечание. Справа от диагонали приведены ρ_{ik} для $\bar{\Delta} = 1\%$, слева -

сией, соответствующей постоянной относительной погрешности. Обработывались два набора: с $\bar{\Delta} = 1\%$ и $\bar{\Delta} = 5\%$. Исходная кривая является рациональным выражением с $L = 9$. Оптимальные аппроксиманты для описания значений функции с учетом этих случайных отклонений также получились при $L = 9$ в обоих случаях со средними квадратическими относительными отклонениями 1,03 и 4,11%.

Результаты аппроксимации и вычисления ковариационных матриц приведены в табл. 6.1 для $\bar{\Delta} = 1\%$. Для $\bar{\Delta} = 5\%$ матрица Фишера оказалась плохо обусловленной, что привело к бессмысленным значениям элементов корреляционной матрицы. Из результатов для $\bar{\Delta} = 1\%$ обращают на себя внимание близкие по модулю к единице значения коэффициентов корреляции многих пар параметров, в первую очередь β и γ для одного резонанса, а также большие погрешности в определении значений α_i .

Выше было получено общее выражение для производных аппроксиманты по опорным ординатам (3.12), позволяющее при подстановке его в (6.2) вычислять для этих параметров матрицу Фишера и корреляционную матрицу. В табл. 6.2 приведены результаты таких вычислений для той же модельной кривой (6.5), что и в табл. 6.1, но для параметризации опорными ординатами. При этом в качестве опорных точек взяты те, которые выбрала в оптимальном варианте программа паде-аппроксимации методом перебора, описанным в § 5.2. Видно, во-первых, что информационная матрица хорошо обусловлена и при $\bar{\Delta} = 5\%$ в отличие от случая параметризации резонансными параметрами. Во-вторых, результирующая корреляционная матрица близка к диагональной — наибольший по модулю коэффициент корреляции параметров равен 0,033 для $\bar{\Delta} = 1\%$ и 0,235 для $\bar{\Delta} = 5\%$. Если учесть, что для уже построенной

коэффициентов корреляции

0,05	0,15	0,35	0,5	0,8
6,261	7,327	11,48	13,93	7,98
0,61	0,59	0,51	0,52	0,57
0,15	0,5	0,65	0,85	1,0
7,292	14,15	11,73	7,086	4,868
0,008	0,019	0,01	0,01	0,023
0,5	0,4	-0,3	0,3	-0,7
0,6	0,5	-0,4	0,3	-0,6
0,06	0,15	-0,05	-0,06	-0,3
-0,6	-0,8	0,4	-0,2	0,7
1	3,3	-0,4	-0,03	-0,9
-2,9	1	1,0	-0,7	-0,09
-2,2	14,4	1	1,5	0,9
0,8	4,3	-2,4	1	-2,6
-2,1	11,9	2,9	-0,4	1

ρ_{ik} для $\bar{\Delta} = 5\%$.

аппроксиманты положение задающих ее L опорных точек можно выбирать в принципе произвольно, то близость полученной ковариационной матрицы к диагональной приводит к естественному вопросу: а нельзя ли, варьируя абсциссы опорных точек, получить диагональную информационную (и, следовательно, корреляционную) матрицу для опорных ординат? Оказывается, если погрешности аппроксимирующих значений статистически независимы, то такая возможность имеется. Она обсуждается в следующем параграфе.

6.2. ДИАГОНАЛИЗАЦИЯ ИНФОРМАЦИОННОЙ МАТРИЦЫ ОПОРНЫХ ОРДИНАТ

Согласно (6.2) для диагональности ковариационной матрицы опорных ординат как параметров аппроксиманты необходимо и достаточно, чтобы выполнялись равенства

$$A_{\mu\nu} = \sum_{i=1}^{N_{\text{ex}}} \frac{\partial f(z_i)}{\partial f_{\mu}} \frac{\partial f(z_i)}{\partial f_{\nu}} = \lambda_{\mu} \delta_{\mu\nu}, \quad (6.6)$$

т.е. чтобы производные аппроксиманты по опорным ординатам как функции z были диагональны на дискретном множестве значений аргумента z_i с весом σ_i^{-2} . Тогда дисперсии опорных ординат аппроксиманты статистически независимы и равны

$$\overline{(\Delta f_{\nu})^2} = 1/\lambda_{\nu}. \quad (6.7)$$

Согласно (3.12)

$$\frac{\partial f}{\partial f_{\mu}} = \frac{Q_M^2(z_{\mu})}{Q_M^2(z)} \prod_{\nu \neq \mu} \frac{(z - z_{\nu})}{(z_{\mu} - z_{\nu})}, \quad (6.8)$$

где $Q_M(z)$ — знаменатель паде-аппроксиманты, поэтому (6.6) эквивалентно требованию ортогональности полиномов, стоящих в числителе (6.8) с весом, который легко определяется из (6.6) — (6.8) и равен $1/\sigma_i^2 Q_M^4(z_i)$, на той же совокупности значений аргумента $z_1, z_2, \dots, z_{N_{\text{ex}}}$.

Из теории ортогональных полиномов (см., например, [21]) известно следующее их свойство. Пусть $P_0(z), P_1(z), \dots, P_L(z)$ — система полиномов степеней до L , ортогональных на множестве точек $\{z_i\}$ с весом $w(z)$, а z_1, z_2, \dots, z_L — корни полинома $P_L(z)$. Тогда L различных полиномов одной и той же степени $L - 1$, определяемых выражением

$$l_{\nu}(z) = \prod_{\mu \neq \nu} (z - z_{\mu}) / \prod_{\mu \neq \nu} (z_{\nu} - z_{\mu}), \nu = 1, 2, \dots, L, \quad (6.9)$$

попарно ортогональны на том же множестве с тем же весом. Таким образом, если выбрать в качестве веса $w(z) = 1/Q_M^4(z) \sigma^2(z)$, построить систему полиномов степеней до L , ортогональных с этим весом на множестве N_{ex} абсцисс исходных точек, и взять в (6.8) в качестве опорных абсцисс корни полинома $P_L(z)$, то соотношения (6.6) будут выполняться.

Приведем формулы, позволяющие построить нужную систему ортогональных полиномов [21]. Введем по определению для произвольной функции $\varphi(z)$ взвешенную сумму значений на множестве $\{z\}_{N_{\text{ex}}}$:

$$\langle \varphi \rangle = \sum_{i=1}^{N_{\text{ex}}} \varphi(z_i) / Q_M^4(z_i) \sigma_i^2. \quad (6.10)$$

Полином будем нормировать на единичный коэффициент при старшей степени. Тогда, опуская для краткости аргумент полиномов, можно записать следующее рекуррентное соотношение:

$$P_n = \left(z - \frac{\langle z P_{n-1}^2 \rangle}{\langle P_{n-1}^2 \rangle} \right) P_{n-1} - \frac{\langle z P_{n-1} P_{n-2} \rangle}{\langle P_{n-2}^2 \rangle} P_{n-2}. \quad (6.11)$$

С учетом начальных условий $P_{-1} = 0$, $P_0 = 1$ рекуррентное соотношение (6.11) позволяет решить поставленную задачу.

6.3. ПОГРЕШНОСТИ АППРОКСИМАНТЫ

Если ковариационная матрица опорных ординат диагональна, т.е. $\overline{\Delta f_\mu \Delta f_\nu} = (\Delta f_\nu)^2 \delta_{\mu\nu}$, то в предположении малости дисперсий $\overline{\Delta^2(z)}$ средний квадрат погрешности аппроксиманты в произвольной точке z будет равен

$$\overline{\Delta^2(z)} = \sum_{\nu} \left[\frac{\partial f^L(z)}{\partial f_\nu} \right]^2 \overline{(\Delta f_\nu)^2}, \quad (6.12)$$

а коэффициент корреляции значений $f^L(z_1)$ и $f^L(z_2)$ в двух произвольных точках z_1, z_2

$$\rho(z_1, z_2) = \frac{\sum_{\mu=1}^L \frac{\partial f^L(z_1)}{\partial f_\mu} \frac{\partial f^L(z_2)}{\partial f_\mu} \overline{(\Delta f_\mu)^2}}{\sqrt{\overline{\Delta^2(z_1)} \overline{\Delta^2(z_2)}}}, \quad (6.13)$$

где $\overline{\Delta^2(z)}$ определяется выражением (6.12).

Итак, вычисление погрешностей построенной аппроксиманты как функции z (построение "коридора ошибок") производится следующим образом:

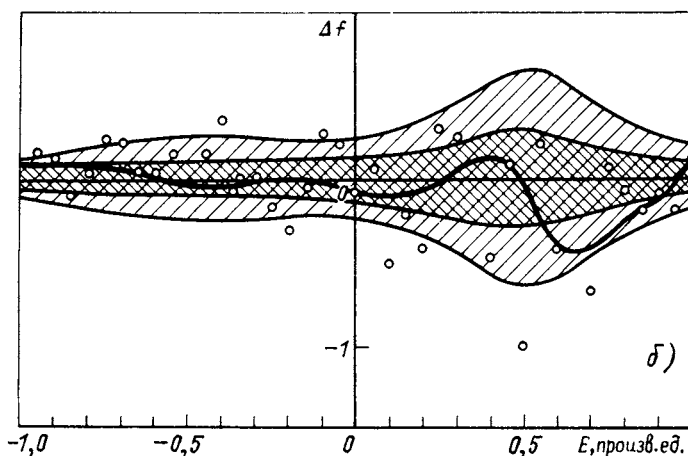
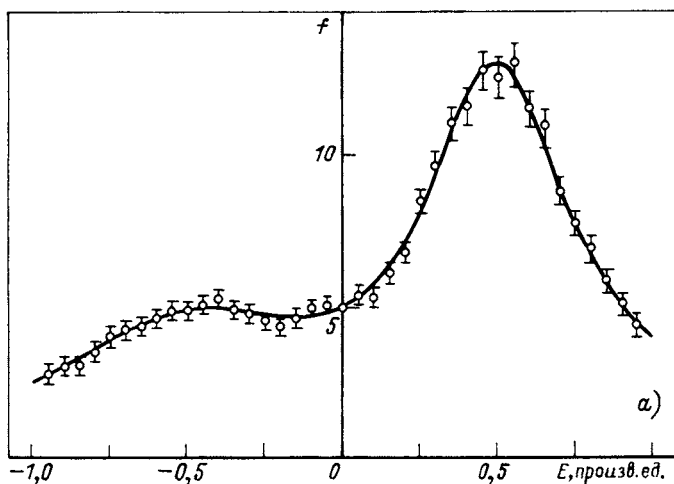


Рис. 6.1. Результаты обработки модельной задачи с 5%-ным разбросом точек относительно кривой (6.5):

a – паде-аппроксиманта; *б* – оценка погрешности аппроксиманты по (6.13) (двойная штриховка); одинарная штриховка – экспериментальная погрешность; жирная кривая – отклонение аппроксиманты от истинной кривой; точки – “эксперимент”

1. По формуле (6.11) строят систему полиномов, ортогональных на множестве $\{z\}_{N_{\text{ex}}}$ с весом $1/\sigma^2(z) Q_M^4(z)$.

2. Численными методами находят корни полинома $P_L(z)$. В силу ортогональности полинома все корни действительны и лежат внутри интер-

вада аппроксимации. Значения этих корней z_{ν} выбирают в качестве опорных абсцисс.

3. По формуле (6.7) с использованием (6.6) вычисляют $\overline{\Delta f_{\nu}}$ – дисперсию погрешностей опорных ординат. Подчеркнем, что $\overline{\Delta f_{\nu}}$ – это не дисперсия экспериментальной погрешности в точке z_{ν} , а дисперсия параметра аппроксиманты f_{ν} .

4. В выражение (6.12) подставляют полученные значения $\overline{\Delta f_{\nu}}$ и $\partial f/\partial f_{\nu}$.

Приведем примеры использования этого метода.

В качестве модельной задачи рассмотрим кривую (6.5) без постоянного слагаемого, т.е. сумму двух интерферирующих резонансов при $\overline{\Delta} = 5\%$ (рис. 6.1, а, б). Как и следовало ожидать, точки аппроксиманты укладываются во второй, примерно вдвое более узкий, коридор ошибок с вероятностью, близкой к 70%.

На рис. 6.2,а представлены результаты подобной же обработки для оценки сечения реакции $^{238}\text{U}(n, 2n)$ в интервале энергий от пороговой до 19 МэВ (пример взят из [43], где указаны источники экспериментальной информации). Отсчет на рис. 6.2,б ведется от значений аппроксиманты. Хотя число параметров в этом случае такое же, как в модельном примере (восемь), резкие скачки в зависимости экспериментальной погрешности от энергии (обрабатывались совместно данные многих работ с разной точностью) ухудшают условия численного построения нужной системы ортогональных полиномов (однако это осложнение не вызывает в данном случае затруднений). Числовые данные, как и для модельной задачи, приведены в табл. 6.3. Как видно, погрешность оценки везде заметно меньше погрешностей экспериментальных значений.

Еще один практический пример. В [44 – 46] методом активации были измерены сечения радиационного захвата нейтронов ядром ^{236}U для $E = 0,15 \div 1,15$ МэВ. Полная погрешность, полученная в экспериментах с использованием сечения захвата ядром ^{197}Au быстрых нейтронов в качестве стандарта, находится в пределах 10,7 – 5%, а погреш-

Таблица 6.3. Опорные абсциссы E_{μ} , ординаты f_{μ} , обеспечивающие диагональность ковариационной матрицы, и погрешности опорных ординат $\overline{\Delta f_{\mu}}$

μ	Модельная задача (рис. 6.1)			Сечение реакции $^{238}\text{U}(U, 2n)$ (рис. 6.2)		
	E_{μ}	f_{μ}	$\overline{\Delta f_{\mu}}$	E_{μ} , МэВ	$f_{\mu} \cdot 10^{-31} \text{ м}^2$	$\overline{\Delta f_{\mu}} \cdot 10^{-31} \text{ м}^2$
1	-0,936	2,650	0,0665	5,55	86,5	4,2
2	-0,669	4,184	0,0870	7,09	508	13,2
3	-0,368	4,838	0,0995	8,24	1209	22,5
4	-0,080	4,652	0,0965	9,75	1405	34,5
5	0,193	7,006	0,1503	11,62	1447	54,1
6	0,432	12,375	0,2935	13,61	1052	19,4
7	0,656	10,254	0,2268	14,60	720	11,8
8	0,927	4,790	0,1101	17,29	335	16,7

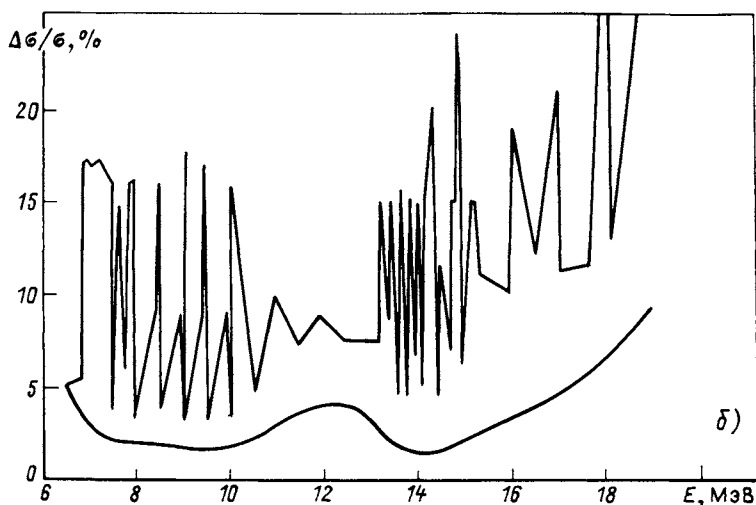
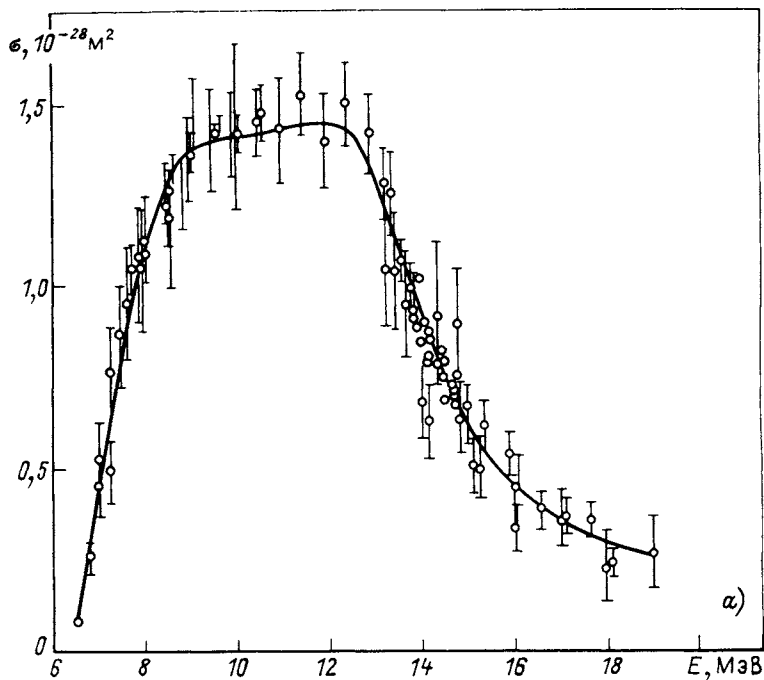


Рис. 6.2. Результаты построения оцененной кривой $\sigma(E)$ для реакции $^{238}\text{U}(n, 2n)$ по совокупности экспериментальных данных, приведенных в [43]:

a – паде-аппроксиманта; *б* – оценка погрешности аппроксимации (плавная кривая); значения экспериментальных погрешностей показаны ломаной линией

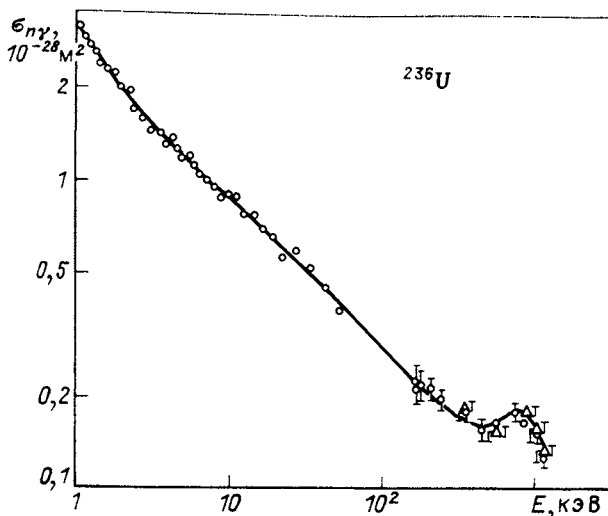


Рис. 6.3. Сравнение экспериментальных данных [44 – 46] с оцененной кривой: данные [45] обозначены треугольниками; экспериментальные погрешности в левой части рисунка не указаны, так как они там примерно соответствуют размеру точек графика

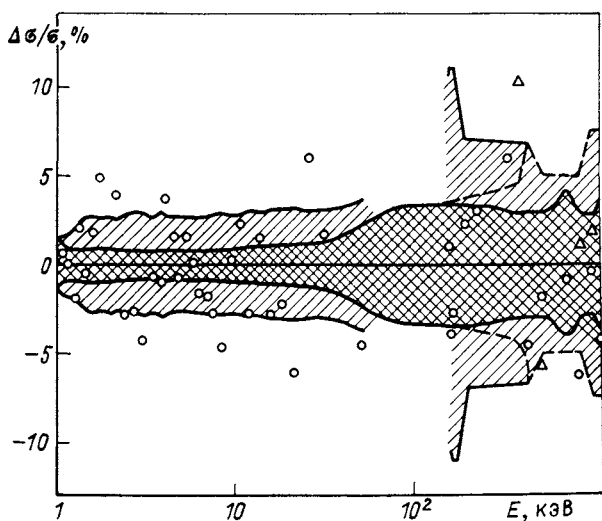


Рис. 6.4. Сравнение погрешностей оцененной кривой (коридор с двойной штриховкой) с экспериментальными погрешностями (коридор с одинарной штриховкой): пунктирная граница в правой части рисунка соответствует оценке погрешности без учета данных [45], полученных позже других результатов. Ось абсцисс, от которой отсчитано значение погрешности, соответствует оцененной кривой на рис. 6.3

ность, найденная с использованием сечения (n, p) -рассеяния, составляет 4,1 – 3,2%. По сравнению с этими результатами оценка ENDF/B-V представляется заметно завышенной, поэтому полученные данные оценены методом паде-аппроксимации, и результаты приведены на рис. 6.3, 6.4. Оптимальная кривая вида (5.12) соответствовала $L = 10$ при $C = 0$. Подстановка в (5.12) числовых значений полученных параметров дает следующую формулу для энергетической зависимости $\sigma_{n\gamma}$ (сечение – в 10^{-28} м², энергия – в кэВ):

$$\sigma_{n\gamma}(E) = \frac{62,01}{E+316,2} + \frac{12,45}{E+14,87} + \frac{1,749}{E-0,25} + \frac{15,99(E-767,1) + 18854}{(420,5)^2 + (E-767,1)^2}. \quad (6.14)$$

Из рис. 6.4 видно, что оцененная погрешность на большей части интервала измерений примерно вдвое ниже экспериментальной. Рисунок иллюстрирует также влияние добавления информации на оцененную погрешность, добавление данных уменьшает погрешность оценки. Ниже приведены значения E_ν и $\Delta\sigma_\nu$, позволяющие по формулам (6.8), (6.12) вычислять значение оцененной погрешности сечения реакции $^{235}\text{U}(n, \gamma)$:

E_ν , кэВ	1,139	2,270	6,837	23,48	73,72
$\Delta\sigma_\nu$, 10^{-31} м ²	26,3	16,1	8,25	6,87	10,0
E_ν , кэВ	280,3	515,3	717,8	924,5	1125,0
$\Delta\sigma_\nu$, 10^{-31} м ²	5,92	4,65	7,38	5,04	4,17

6.4. СТАТИСТИЧЕСКИ ОПТИМАЛЬНАЯ РАЦИОНАЛЬНАЯ ИНТЕРПОЛЯЦИЯ

Рассмотрим следующую задачу. Пусть известно, что измеряемая зависимость описывается L -параметрической рациональной функцией, и пусть измерения проводятся при значениях аргумента z_μ , $\mu = 1, 2, \dots, L$, которые можно выбирать произвольно в пределах заданного интервала, а абсолютная точность измерений фиксирована и не зависит от z . Это означает, что измеренные значения

$$F_\mu = f(z_\mu) + \Delta_\mu, \quad \mu = 1, 2, \dots, N_{\text{ex}} = L, \quad (6.15)$$

где $f(z_\mu)$ – истинные значения рациональной функции; Δ_μ – случайные независимые одинаково нормально распределенные величины. Интерполирующая функция $f^L(z)$ строится по L узлам (z_μ, F_μ) с помощью рекуррентных соотношений (3.26), (3.18) (без перебора, $L = N_{\text{ex}}$). Тогда значения $f^L(z)$ в произвольных точках рассматриваемого интервала представляют собой случайные величины, распределение которых зависит от выбора абсцисс узлов интерполяции z_μ . В качестве критерия оптимизации выбора узлов возьмем минимизацию $\max \Delta^2(z)$ –

максимума математического ожидания квадрата отклонений значений проходящей через эти узлы интерполирующей функции от истинной на всем рассматриваемом интервале, не задаваясь целью минимизации статистической суммы. Если измерения проводятся только в узлах, то такая аппроксиманта дает оценку решения задачи МНК, и приведенный выше критерий является критерием G -оптимальности в теории планирования эксперимента [32]. В задаче линейной регрессии с равноточными измерениями этот критерий эквивалентен критерию D -оптимальности, т.е. требованию минимизации объема эллипсоида рассеяния параметров. Спектр D -оптимального плана, т.е. набор оптимальных опорных точек, для полиномиальной аппроксимации известен в явном виде — это корни присоединенного полинома Лежандра. Рассмотрим случай рациональной аппроксимации, т.е. нелинейной регрессии. Для того чтобы проиллюстрировать, к чему на практике приводит такой выбор узлов интерполяции, при интерполяции полиномами проводится сравнение с другими вариантами выбора опорных точек.

Итак, используя предложения о независимости экспериментальных погрешностей в разных точках и о малости этих погрешностей, для математического ожидания среднего квадратического отклонения интерполирующей функции в точке z от истинного значения получаем

$$\overline{\Delta^2(z)} = \Delta^2 \sum_{\mu=1}^L [\partial f^L(z)/\partial f_{\mu}]^2, \quad (6.16)$$

где $\Delta^2 \equiv \overline{\Delta_{\mu}^2}$.

Потребуем, чтобы $\overline{\Delta^2(z)} \leq \Delta^2$ при $z \neq z_{\mu}$. Необходимым условием оптимальности выбора опорных точек, при котором это требование может быть удовлетворено, будет справедливость системы уравнений

$$\left. \frac{\partial}{\partial z} \frac{\partial f^L(z)}{\partial f_{\mu}} \right|_{z=z_{\mu}} = 0, \quad \mu = 1, 2, \dots, L, \quad (6.17)$$

где $\partial f^L(z)/\partial f_{\mu}$ определяются выражениями (6.8). Из (1.14) и (6.16) следует, что при нарушении этого условия в μ -й опорной точке всегда найдется такая ее окрестность, где $\overline{\Delta^2(z)} > \Delta^2$.

Введем обозначение для "фундаментального полинома"

$$\mathcal{P}(z, z_1, z_2, \dots, z_L) = \prod_{\mu=1}^L (z - z_{\mu}). \quad (6.18)$$

В дальнейшем все аргументы, кроме z , будем для краткости опускать. Полином $\mathcal{P}(z)$ обладает следующими свойствами (см. [21]):

$$\left. \frac{\partial \mathcal{P}(z)}{\partial z} \right|_{z=z_{\nu}} = \prod_{\mu \neq \nu} (z_{\nu} - z_{\mu}); \quad \left. \begin{array}{l} \vdots \\ \vdots \end{array} \right\}$$

$$\left. \frac{\partial^2 \mathcal{P}(z)}{\partial z^2} \right|_{z=z_\nu} = 2 \frac{\partial}{\partial z_\mu} \prod_{\mu \neq \nu} (z - z_\mu) \Big|_{z=z_\nu} \quad (6.19)$$

Подставляя (6.8) в (6.17) и используя (6.19), получаем

$$\mathcal{P}'''(z_\mu) Q_M(z_\mu) - 4 Q_M'(z_\mu) \mathcal{P}'(z_\mu) = 0, \quad \mu = 1, 2, \dots, L, \quad (6.20)$$

что эквивалентно дифференциальному уравнению

$$\mathcal{P}'''(z) Q_M(z) - 4 Q_M'(z) \mathcal{P}'(z) - R(z) \mathcal{P}(z) = 0, \quad (6.21)$$

где R — полином степени $M - 2$. Очевидно, что если $N > M$, то $|\partial f^L(z)/\partial f_i| \rightarrow \infty$ при $z \rightarrow \infty$ и крайние точки следует выбирать на границах интервала аппроксимации, а система (6.18) или уравнение (6.21) определяет лишь внутренние статистически оптимальные точки. Отметим, что положение этих точек полностью определяется их числом и параметрами полинома $Q_M(z)$, но не зависит от $P_N(z)$. При $Q_M(z) \equiv 1$ получается полиномиальная интерполяция и система имеет решение, определяемое только степенью полинома и не зависящее от значений аппроксимируемой функции (см. также ниже).

Рассмотрим электростатическую модель нашей задачи. Пусть на плоскости xu расположена система зарядов, взаимодействующих по логарифмическому закону: каждому корню z_α полинома $Q_M(z)$ соответствует заряд $+2$, неподвижно закрепленный в точке (x_α, y_α) , где $z_\alpha = x_\alpha + iy_\alpha$, а каждому корню $\mathcal{P}(x)$, т.е. каждой опорной точке x_μ , — подвижный заряд -1 , расположенный в этой точке оси x . Потенциал взаимодействия имеет вид

$$U_{ik} = e_i e_k \ln |z_i - z_k|, \quad (6.22)$$

где e_i — заряд i -й точки; z_i — ее радиус-вектор. Следует отметить, что взаимодействие, описываемое этим потенциалом, не соответствует в точности двумерному закону Кулона: такие заряды создают двумерный кулоновский потенциал, но взаимодействуют с ним, как точечные заряды трехмерного пространства. Однако сохраним за этой моделью традиционное название "электростатическая". Потенциал, действующий на заряд -1 в точке x_i со стороны всех остальных зарядов системы, равен

$$U_i(x_i) = -\ln \left| \prod_{k \neq i} (x_i - x_k) / Q^2(x_i) \right|, \quad (6.23)$$

а потенциальная энергия всей системы (без учета взаимного отталкивания неподвижных положительных зарядов) есть

$$U(x) = -\frac{1}{2} \ln \prod_{i < k} \frac{(x_i - x_k)^2}{Q^4(x_i)}. \quad (6.24)$$

Таблица 6.4. Статистически оптимальные узлы для простейших случаев рациональной интерполяции

$f^L(z)$	Число параметров ("зарядов")	Интерпретация	$(z_i - \epsilon)/\gamma$
$1/[\gamma^2 + (z - \epsilon)^2]$	2	Симметричный резонанс с фиксированной амплитудой	$\pm \gamma^{-1/2}$
$\beta/[\gamma^2 + (z - \epsilon)^2]$	3	Симметричный резонанс с произвольной амплитудой	0; $\pm (3/5)^{1/2}$
$\frac{\alpha(z - \epsilon) + \beta}{\gamma^2 + (z - \epsilon)^2}$	4	Асимметричны резонанс	$\pm (1 \pm 2/\sqrt{5})^{1/2}$
$C + \frac{\alpha(z - \epsilon) + \beta}{\gamma^2 + (z - \epsilon)^2}$	5	Асимметричный резонанс с фоном	0; $\pm (5 \pm 2\sqrt{5})^{1/2}$

Равновесное положение такой системы зарядов, соответствующее минимуму потенциальной энергии, должно удовлетворять системе уравнений $\partial U/\partial x_i = 0$, которая, как легко убедиться, совпадает с (6.20).

Корни знаменателя рациональной аппроксиманты обычно соответствуют резонансам аппроксимируемой кривой, поэтому в соответствии с описанной электростатической аналогией оптимальные опорные точки "притягиваются" к резонансам в отличие от случая полиномиальной аппроксимации [$Q(x) = 1$], когда расположение этих точек соответствует равновесной конфигурации свободной системы одноименных зарядов на оси x (с закрепленными крайними зарядами).

В общем случае систему (6.20) необходимо решать численно. Результаты для простейших случаев, когда ее решение находится в явной форме, приведены в табл. 6.4. Статистически оптимальная интерполяция полиномами подробно рассмотрена в [35, 47]; для статистически оптимальных узлов интерполяции полиномом степени n

$$\overline{\Delta^2(z)} = \Delta^2 \left\{ 1 - (1 - z^2) [\partial P_{n-1}(z)/\partial z] / [n(n+1)] \right\}, \quad (6.25)$$

где $P_n(z)$ — полином Лежандра. Сравнение с другими вариантами выбора узлов интерполяции при $n = 6$ приведено на рис. 6.5.

Вернемся к рациональной интерполяции. Выражение (6.17) представляет собой необходимое условие для статистической оптимальности ее узлов. Нетрудно убедиться, что при $N \geq M - 1$ оно является и достаточным. Прежде всего, используя (6.16), получаем

$$\overline{\Delta^2(z)} = R(z)/Q_M^4(z); \quad (6.26)$$

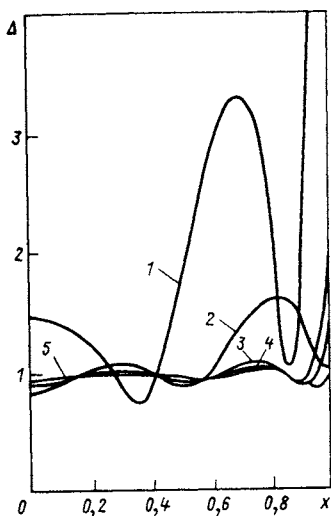


Рис. 6.5. Статистические погрешности $\Delta(x) = \sqrt{\overline{\Delta^2(x)}}$ полиномиальной интерполяции для различных вариантов выбора узлов:

1 – узлы чебышевской квадратурной формулы, $\overline{\Delta} = 3,11$; 2 – эквидистантные узлы, $\overline{\Delta} = 1,12$; 3 – узлы гауссовой квадратурной формулы, $\overline{\Delta} = 1,03$; 4 – узлы чебышевской интерполяции, $\overline{\Delta} = 0,97$; 5 – оптимальный выбор узлов $[\overline{\Delta^2(x)}$ определяется выражением (6.25)], $\overline{\Delta} = 0,95$ ($\overline{\Delta}$ – средние значения по интервалу $[0,1]$)

где $R(z)$ – полином степени $2(N+M)$. В то же время $\overline{\Delta^2(z_\mu)} = \Delta^2$ и $\partial \overline{\Delta^2(z)}/\partial z|_{z=z_\nu} = 0$ (последнее при $N > M$ верно лишь для внутренних узлов). Следовательно, при $N > M$

$$\overline{\Delta^2(z)} = \Delta^2 [1 - \alpha(z - z_1)(z - z_L) \prod_{\nu \neq 1, L} (z - z_\nu)^2 / Q_M^4(z)], \quad (6.27)$$

так как степень полинома в числителе правой части (после приведения к общему знаменателю) равна степени $R(z)$, т.е. $2(N+M)$.

Вычисляя $\partial \overline{\Delta^2(z)}/\partial z|_{z=z_1}$ из (6.27), находим

$$\alpha = 2 \frac{\partial}{\partial z} \frac{\partial f}{\partial f_1} \Big|_{z=z_1} Q_M^4(z_1) / (z_L - z_1) \prod_{\nu \neq 1, L} (z_1 - z_\nu). \quad (6.28)$$

При $N > M$ имеем

$$\frac{\partial}{\partial z} \frac{\partial f}{\partial f_1} \Big|_{z=z_1} < 0,$$

поэтому $\alpha < 0$ и при $z_1 < z < z_L$ будет $\overline{\Delta^2(z)} \leq \Delta^2$. Аналогичные рассуждения при $N \leq M$ приводят к выражению

$$\overline{\Delta^2(z)} = \Delta^2 [1 - \psi(z) \prod_{\nu=1} (z - z_\nu) / Q_M^4(z)], \quad (6.29)$$

где $\psi(z)$ не имеет полюсов в точках z_ν . При $N = M$ выражение $Q^4(z) - \psi(z) \prod_{\nu} (z - z_\nu)^2$ будет полиномом степени $4M$ лишь для $\psi(z) \equiv 0$, поскольку степень $\sum_{\nu=1}^L (z - z_\nu)^2$ равна в этом случае $4M + 2$. Поэтому в диагональном приближении при статистически оптимальном выборе узлов $\Delta^2(z) \equiv \Delta^2$, т.е. погрешность интерполяции постоянна и равна погрешности опорных значений.

При $N = M - 1$ функция $\psi(z)$ в (6.29) должна быть такой, чтобы выражение $Q_M^4 - \psi(z) \prod_{\nu=1}^L (z - z_\nu)^2$ было полиномом степени $4M - 2$. Это возможно лишь при $\psi(z) \equiv 1$ (вспомним, что $q_M^M = 1$) и тогда

$$\Delta^2(z)/\Delta^2 = 1 - \prod_{i=1}^L (z - z_i)^2 / Q_M^4(z) \leq 1.$$

В заключение обратим внимание на то, что упоминавшиеся в гл. 1 "функции отсчета" (1.5) удовлетворяют равенствам

$$\left. \frac{\partial}{\partial z} \frac{\partial f(z)}{\partial f_\nu} \right|_{z=z_\nu} = 0, \quad \left. \frac{\partial f(z)}{\partial f_\mu} \right|_{z=z_\nu} = \delta_{\mu\nu} \quad (6.30)$$

и, кроме того,

$$\sum_{i=-\infty}^{\infty} \frac{\sin^2[\pi(z-i)]}{[\pi(z-i)]^2} \equiv 1. \quad (6.31)$$

Следовательно, бесконечная последовательность эквидистантных опорных точек представляет собой статистически оптимальную для интерполяции функций с ограниченной шириной фурье-спектра с помощью разложения (1.5), при этом статистическая погрешность интерполяции постоянна, если постоянна погрешность в опорных точках, и равна последней.

6.5. ОЦЕНКА КОЛИЧЕСТВА ИНФОРМАЦИИ, ПОЛУЧЕННОЙ ПРИ ОБРАБОТКЕ И АНАЛИЗЕ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

В основе любой количественной оценки передаваемой информации лежит сравнение с ранее накопленными знаниями и требование ее аддитивности. Классическим является определение количества информации, полученной при приеме сообщения о событии

$$I_i = \ln[P_{\text{out}}(a_i)/P_{\text{in}}(a_i)], \quad (6.32)$$

где $P_{\text{out}}(a_i)$ и $P_{\text{in}}(a_i)$ — вероятности события a_i после и до приема

сообщения соответственно. В теории передачи сообщений, основанной на этом определении Шеннона, известно событие a_i и вероятности P_{out} и P_{in} . Основной величиной служит энтропия $H = \sum_{i=1}^N I_i/N$ (среднее количество информации в одном сообщении), зависящая от набора возможных событий и характеристик передающей системы. Непосредственно использовать эту схему в задаче обработки данных невозможно, поскольку если в качестве "события" рассматривать F_i — результат измерения величины $f(z, \{p\})$ в точке z_i , то $P_{in}(F_i)$ и $P_{out}(F_i)$ неизвестны и можно лишь делать некоторые предположения о них. Кроме того, конечная цель обработки эксперимента заключается, как правило, в определении вектора параметров физических величин $\{p\}$, а не характеристик передающего (измерительного) устройства.

Поэтому рассмотрим следующую постановку задачи: известны результаты измерений на выходе измерительного устройства и сравниваются две гипотезы (α и β) о вероятностях этих результатов. В определении Кульбака [48, с. 15] количество "различающей" информации определяется как логарифм отношения соответствующих плотностей вероятности $\psi_\alpha(F_i)$ и $\psi_\beta(F_i)$:

$$I(\alpha : \beta) = \sum_i I_i(\alpha : \beta) = \sum_i \ln \frac{\psi_\alpha(F_i)}{\psi_\beta(F_i)}. \quad (6.33)$$

Одним из достоинств такого определения "различающей" информации является возможность сравнения гипотез α и β , которые могут различаться не только значениями параметров функции $f(z, \{p\})$, но и их числом, а также структурой функций. Это дает возможность не только выбрать оптимальный набор параметров для $f(z, \{p_\alpha\})$ и $f(z, \{p_\beta\})$, но и определить оптимальную из двух функций $f^\alpha(z, \{p_\alpha\})$ и $f^\beta(z, \{p_\beta\})$. Очевидно, для тех z_i , при которых $\psi_\alpha(F_i) > \psi_\beta(F_i)$, количество информации I_i будет положительным, и это определение действительно позволяет выбрать из двух гипотез наиболее правдоподобную по всей совокупности экспериментальных результатов. В такой постановке задачи погрешности эксперимента пока никак не использовались.

В качестве гипотезы α естественно принять обычное предположение о том, что погрешности эксперимента независимы и в каждой точке подчиняются нормальному распределению с известной дисперсией:

$$\psi_\alpha(F_i) = \frac{1}{\sqrt{2\pi}\sigma_i^\alpha} \exp \left\{ - \frac{|f^\alpha(z_i, \{p_\alpha\}) - F_i|^2}{2(\sigma_i^\alpha)^2} \right\}. \quad (6.34)$$

Неизвестны лишь параметры $\{p_\alpha\}$ функции $f^\alpha(z_i, \{p_\alpha\})$. Очевидно, что при любом выборе структуры и числа параметров этой функции ко-

личество "различающей" информации будет максимально, если эти параметры положить равными параметрам $\{\hat{p}_\alpha\}$ оценки максимального правдоподобия:

$$\frac{\partial}{\partial p} \prod_i \psi_\alpha(F_i) \Big|_{\hat{p}_\alpha} = 0.$$

Для распределения (6.34) эта оценка совпадает с оценкой МНК.

Если в качестве гипотезы β принять априорные знания, например определить ψ_β как распределение с параметрами, оцененными по результатам предыдущих экспериментов, то значение $I(\alpha : \beta)$ естественно рассматривать как прирост информации в результате последнего эксперимента, которому соответствует гипотеза α . Отметим, что если нет основания подвергать сомнению результаты предыдущих экспериментов, т.е. гипотезу β , то имеет смысл вместо гипотезы α рассматривать гипотезу $\alpha' = \alpha + \beta$, получающуюся при совместной обработке накопленных ранее и полученных в результате последнего эксперимента (гипотеза α) сведений. Ясно, что чем точнее эксперимент и чем больше число измерений, тем больше прирост информации; это согласуется с обычным представлением об информации, извлекаемой из эксперимента.

Кроме количества информации, полученной в результате конкретного эксперимента, обычно, как и в дискретном случае, когда набор возможных событий образует счетное множество, рассматривают еще энтропию

$$H = \int \prod_i \psi_\alpha(F_i) \ln \prod_i \frac{\psi_\alpha(F_i)}{\psi_\beta(F_i)} dF_i. \quad (6.35)$$

Эта величина связана с I следующим образом: $H = \lim_{n \rightarrow \infty} (I/n)$, где

$n = n_i$ — число измерений, проводимых в каждой точке z_i . Энтропия представляет собой среднее количество "различающей" информации, которое получается при проведении измерений в точках z_i . В качестве параметров $\{p_\alpha\}$, $\{p_\beta\}$ при этом обычно берут их оценки, полученные в конкретном эксперименте, т.е. неусредненные.

Обычно число параметров, которые извлекаются из эксперимента, гораздо меньше числа измерений. Так, если в точке z_i проводится n_i измерений и оцениваемым параметром, как обычно, служит среднее нормального распределения $\psi(F_{i,k})$, где $F_{i,k}$ — результат k -го измерения, то оценкой величины $F_{i,k}^0 = \mu$ является значение $\frac{1}{n_i} \sum_{k=1}^{n_i} F_{i,k}$.

Из этого примера видно, что для оценки параметров совсем не обязательно знать результаты всех измерений. Необходимая для получения оценки информация содержится в некоторых их комбинациях, которые

называются достаточными статистиками. В рассмотренном примере это

$\hat{\mu} = \frac{1}{n} \sum_{k=1}^{n_i} F_{i,k}$. Для функции правдоподобия получим:

$$\begin{aligned} & \left(\frac{1}{\sqrt{2\alpha} \sigma_i} \right)^{n_i} \exp \left[- \frac{\sum_{k=1}^{n_i} (\mu - F_{i,k})^2}{2\sigma_i^2} \right] = \\ & = \left(\frac{1}{\sqrt{2\pi} \sigma_i} \right)^{n_i} \exp \left[- \frac{n_i (\mu - \hat{\mu})^2}{2\sigma_i^2} \right] \exp \left[- \frac{\sum_{k=1}^{n_i} (\hat{\mu} - F_{i,k})^2}{2\sigma_i^2} \right], \end{aligned}$$

т.е. функция правдоподобия факторизуется и для оценки значения μ необходимо знать лишь достаточную статистику $\hat{\mu}$. Такая факторизация лежит в основе определения достаточных статистик.

В общем случае будем обозначать достаточную статистику, необходимую для оценки векторного параметра $\{p\}$ функции $f(z, \{p\})$, как $\{\hat{p}\}$. Дисперсии и корреляции этих компонент вектора $\{p\}$, а также прочие статистики (комбинации результатов измерений), которые информации о значениях параметров $\{p\}$ не несут, обозначим $\{s\}$. Параметры $\{s\}$ и их достаточные статистики $\{\hat{s}\}$ характеризуют надежность полученных оценок параметров $\{p\}$, свойства измерительного устройства и статистику эксперимента. Общее число введенных таким образом параметров p_i и s_k равняется числу экспериментальных измерений N_{ex} .

Перейдем теперь от распределения $\Pi \psi [f(z_i, \{p\}), F_i] dF_i$ к распределению $\psi' (\{p\}, \{s^i\}, \{\hat{p}\}, \{\hat{s}\})$. В случае линейной регрессии последнее распределение факторизуется, $\psi' = \varphi(\{p\}, \{\hat{p}\}) \rho(\{s\}, \{\hat{s}\})$; в случае нелинейной регрессии это утверждение носит асимптотический характер, т.е. верно при $N_{\text{ex}} \rightarrow \infty$. Используя свойство факторизации, энтропию (6.35) можно представить в виде суммы двух слагаемых:

$$\begin{aligned} H &= \int \varphi_\alpha \ln \frac{\varphi_\alpha}{\varphi_\beta} d\{\hat{p}\} + \int \rho_\alpha \ln \frac{\rho_\alpha}{\rho_\beta} d\{\hat{s}\} = \\ &= H_p(\alpha : \beta) + H_s(\alpha : \beta), \end{aligned} \tag{6.36}$$

первое из которых будем рассматривать как определение полезной (аналитической) информации, извлекаемой из эксперимента. Второе слагаемое дает информацию о "шумах" в широком смысле слова. Если эти шумы одинаковы в обеих гипотезах, т.е. $f^{(\alpha)}(z, \{p_\alpha\})$ и $f^{(\beta)}(z, \{p_\beta\})$

различаются только значениями параметров $\{p_\alpha\}$ и $\{p_\beta\}$, то второе слагаемое равно нулю. Если шумы разные, то может получиться, что $H_p > H$, а это означает, что отброшенное слагаемое было отрицательным, т.е. уровень шумов выше в эксперименте, которому соответствует гипотеза α , и необходимо дополнительное исследование.

Как уже указывалось, при нелинейной регрессии распределение оценок параметров аппроксиманты $\{p\}$ асимптотически нормально с ковариационной матрицей A^{-1} , где A – матрица Фишера (6.2).

Предположим теперь, что эксперимент проводится впервые. В этом случае удобно выбрать гипотезу β так, чтобы выражение для энтропии принимало наиболее простой вид. Возьмем в качестве такой гипотезы предположение о равновероятных значениях параметров $\{p\}$ на интервале $[-T, T]$, достаточно широком для того, чтобы вне его можно было считать $\psi_\beta(\{p\})$ равным нулю. Если $\psi_\alpha(\{p\}, \{\hat{p}\})$ – многомерное нормальное распределение с ковариационной матрицей A , т.е.

$$\psi_\alpha(\{p\}, \{\hat{p}\}) = \frac{(\det A)^{1/2}}{\sqrt{2\pi}} e^{-\frac{1}{2}\{p - \hat{p}\} A \{p - \hat{p}\}^T},$$

то для количества полезной информации получим

$$H_p(\alpha : \beta) = \frac{1}{2} \ln(\det A) 2\pi e + \nu \ln 2T = H_{p,\alpha} - H_{p,\beta}, \quad (6.37)$$

где ν – размерность вектора $\{p\}$.

При переходе от параметров $\{p\}$ к параметрам $\{t\}$ плотность вероятностного распределения преобразуется очевидным образом:

$$\psi_\alpha(\{p\}) d\{p\} = \psi_\alpha(\{p(t)\}) \left| \frac{dp}{dt} \right| d\{t\} = \kappa_\alpha(\{t\}) d\{t\}.$$

Здесь $|dp/dt| = J$ – якобиан преобразования параметров. Энтропия при этом не меняется: $H_p(\alpha : \beta) = H_t(\alpha : \beta)$, но меняются слагаемые $H_{p,\alpha}$ и $H_{p,\beta}$:

$$\begin{aligned} H_{t,\alpha} &= \int \kappa_\alpha(\{t\}) \ln \kappa_\alpha(\{t\}) d\{t\} = \\ &= H_{p,\alpha} + \ln \left| \frac{dp}{dt} \right|. \end{aligned} \quad (6.38)$$

Если в качестве параметров выбраны опорные ординаты, то в случае линейной регрессии якобиан J будет равен

$$\left| \prod_{i < k} (z_{p_i} - z_{p_k}) / \prod_{i < n} (z_{t_i} - z_{t_n}) \right|.$$

При любом выборе опорных точек $z \{t\}$ этот якобиан принимает максимальное значение, если $z \{p\}$ — статистически оптимальные точки. И наоборот, при любом выборе $z \{p\}$ этот якобиан минимален, если $z \{t\}$ — статистически оптимальные точки. Можно показать, что такая связь экстремумов якобиана со статистически оптимальным выбором опорных точек сохраняется и для рациональной функции регрессии $f^L(z) = P_N(z)/Q_M(z)$.

Предположим теперь, что априорная информация отсутствует и в качестве гипотезы β о распределении параметров аппроксиманты — опорных ординат при любом выборе опорных абсцисс принимается гипотеза об их равномерном распределении на интервале $[-T, T]$. Тогда, как следует из (6.37) и (6.38), количество "различающей" информации для двух различных наборов опорных абсцисс подчиняется следующему соотношению: $H_t(\alpha : \beta) = H_p(\alpha : \beta) + \ln J$. Если измерения проводятся в точках z_{p_i} и распределение погрешностей эксперимента от z_p не зависит, то количество различающей информации $H_p(\alpha : \beta)$ будет одним и тем же при любом выборе абсцисс $z \{p\}$. В этом случае количество "различающей" информации $H_t(\alpha : \beta)$ для параметров $\{t\}$ будет максимальным при таком выборе $z \{p\}$, которому соответствует максимальное значение якобиана, т.е. когда z_{p_i} — статистически оптимальные точки. В то же время при одном и том же наборе $z \{p\}$ количество "различающей" информации для параметров $\{t\}$ будет минимальным для тех $z \{t\}$, при которых якобиан минимален, т.е. когда z_{t_i} — статистически оптимальные точки.

Суммируя сказанное, можно сделать следующие выводы: максимум информации о значениях опорных ординат получается тогда, когда измерения проводятся в статистически оптимальных точках. При проведении измерений с произвольным набором точек распределение опорных ординат, соответствующих статистически оптимальным точкам, труднее всего отличать от равномерного, и их можно задавать менее точно. В этом смысле статистически оптимальные точки наиболее информативны.

ГЛАВА 7. НЕКОТОРЫЕ ПРИЛОЖЕНИЯ ПАДЕ-АППРОКСИМАЦИИ

7.1. РАСЧЕТ ПОДГРУППОВЫХ КОНСТАНТ ТЕОРИИ ПЕРЕНОСА НЕЙТРОНОВ

Для физического расчета ядерных реакторов желательно в принципе иметь как можно более детальную информацию о нейтронных сечениях. Однако существующие методы расчета с учетом производительности современных ЭВМ не позволяют использовать всю эту информацию

непосредственно, т.е. в виде функций $\sigma(E)$ во всем энергетическом диапазоне. Поэтому разработаны и с успехом применяются расчетные методы, основанные на преобразовании исходных "микроскопических" ядерных данных в интегральную форму. Общей основой этих методов служит использование многогруппового приближения в теории переноса нейтронов. Его современное описание можно найти в монографии [15], существенную часть содержания которой составляет изложение именно методов расчета групповых констант из микроскопических данных. Там же содержится необходимая библиография.

Одна из ветвей этого направления — расчет так называемых погрупповых констант (см. [15]), при котором широко используется аппроксимация рациональными функциями. Настоящий параграф посвящен рассмотрению возможностей описанных выше алгоритмов паде-аппроксимации применительно к этой задаче.

Основной формой интегрального представления информации о резонансной структуре сечений в многогрупповом приближении являются функционалы вида

$$F(\sigma_0, N) = \int_{\Delta E} \frac{\varphi(E) dE}{[\sigma(E) + \sigma_0]^N} \bigg/ \int_{\Delta E} \varphi(E) dE \equiv \left\langle \frac{1}{(\sigma + \sigma_0)^N} \right\rangle; \quad (7.1)$$

$$F_x(\sigma_0, N) = \int_{\Delta E} \frac{\sigma_x(E) \varphi(E) dE}{[\sigma(E) + \sigma_0]^N} \bigg/ \int_{\Delta E} \varphi(E) dE \equiv \left\langle \frac{\sigma_x}{(\sigma + \sigma_0)^N} \right\rangle. \quad (7.2)$$

Здесь σ_0 — параметр, называемый сечением разбавления; $\varphi(E)$ — стандартный спектр нейтронов; $\sigma_x(E)$, $\sigma(E)$ — сечение реакции типа x и полное сечение в зависимости от энергии E ; ΔE — групповой энергетический интервал.

Зависимость этих функционалов от σ_0 можно задавать в поточечном представлении в виде таблиц факторов резонансной самоэкранировки [16]. При использовании подгрупповых констант эти зависимости представляются аналитически с помощью аппроксимации рациональными функциями:

$$F_n(\sigma_0, N) = \sum_{i=1}^n a_i / (\sigma_i + \sigma_0)^N; \quad (7.3)$$

$$F_{x,n}(\sigma_0, N) = \sum_{i=1}^n a_i \sigma_{x_i} / (\sigma_i + \sigma_0)^N. \quad (7.4)$$

Здесь n — число подгрупп; a_i , σ_i и σ_{x_i} — подгрупповые параметры, на значения которых накладываются следующие нормировочные ограничения:

$$\sum_i a_i = 1; \quad \sum_i \sigma_{x_i} = \sigma_i; \quad a_i, \sigma_i, \sigma_{x_i} > 0. \quad (7.5)$$

Обычно эти параметры получаются так: располагая информацией о детальной энергетической зависимости нейтронных сечений, можно вычислять значения функций (7.1), (7.2) при различных комбинациях σ_0 и N и затем подставлять полученные значения в соотношения (7.3), (7.4), рассматривая последние как уравнения для определения подгрупповых констант. Однако этот путь сложен. Во-первых, получение моментов сечений высоких порядков связано с заметными вычислительными трудностями. Достаточно заметить, что порядок моментов N в практических расчетах меняется от -5 до 4 , в то же время диапазон изменения сечений — от 10^{-29} до 10^{-23} м². Во-вторых, решение возникающей системы нелинейных уравнений со многими неизвестными при соблюдении условий (7.5) также представляет собой непростую задачу. Эти трудности и методы их преодоления обсуждаются в [15, гл. 2]. Там, в частности, отмечается, что при достаточно точном аналитическом описании зависимости функционала $\langle 1/(\sigma + \sigma_0) \rangle$ от сечения разбавления можно избежать численного расчета моментов высоких порядков, получая их дифференцированием соответствующих аналитических зависимостей

$$\left\langle \frac{1}{(\sigma + \sigma_0)^{N+1}} \right\rangle = \frac{(-1)^N}{N!} \frac{d^N}{d\sigma_0^N} \left\langle \frac{1}{\sigma + \sigma_0} \right\rangle. \quad (7.6)$$

Понятно, что для получения надежных значений дифференцируемую функцию необходимо знать с весьма высокой точностью.

Первым этапом в расчете подгрупповых констант и является задача рациональной аппроксимации зависимости функционала F , определяемого лишь полным сечением, от σ_0 : $F(\sigma_0) \equiv F(\sigma_0, 1) = \langle 1/(\sigma + \sigma_0) \rangle$.

Учитывая, что сечение $\sigma(E)$ положительно, а $\int_0^E \varphi(E') dE'$ — функция неубывающая, и разбивая ΔE на интервалы, в которых $\sigma(E)$ меняется монотонно, можно перейти к интегрированию по σ (см. [15, с. 62]).

В результате получим интеграл Стильтьеса $\int_{\sigma_{\min}}^{\sigma_{\max}} p(\sigma) d\sigma/(\sigma + \sigma_0)$, в

котором $p(\sigma)$ — кусочно-гладкая, положительная на интервале $[\sigma_{\min}, \sigma_{\max}]$ функция, поэтому $P(\sigma) = \int_0^{\sigma} p(\sigma) d\sigma$ — кусочно-гладкая неубывающая функция.

Введение подгрупповых констант основано на аппроксимации этой функции кусочно-постоянной "функцией скачков" $P(\sigma) \approx \sum_{i=1}^n a_i$, $i = 1, 2, \dots, n$, где a_i положительны. При этом $p(\sigma)$ ап-

проксимируется суммой δ -функций $p_n(\sigma) = \sum_{i=1}^n a_i \delta(\sigma - \sigma_i)$, а $F(\sigma_0)$ — рациональной функцией $F_n(\sigma_0)$. Отсюда ясен смысл подгрупповых

констант: функция распределения значений полного сечения в энергетической группе аппроксимируется гистограммой из n ступенек высотой σ_i , причем доля нейтронов в групповом интервале, которым приписывается постоянное значение сечения σ_i , равна a_i .

В предыдущих главах были описаны два основных способа построения рациональной аппроксиманты — приближения Паде-1 и Паде-2, которые основаны на использовании разложения аппроксимируемой функции в степенной ряд и ее значений в дискретном наборе точек соответственно. Ниже оба эти способа последовательно применены для аппроксимации функции $F(\sigma_0)$ и проведено сравнительное обсуждение результатов.

Заменяя в подынтегральном выражении $F(\sigma_0) = \int p(\sigma) d\sigma / (\sigma + \sigma_0)$ дробь $1/(\sigma + \sigma_0)$ сходящимися в соответствующих областях рядами по степеням σ_0/σ и σ/σ_0 , можно сформулировать задачу рациональной аппроксимации как усеченную проблему моментов [49, с. 42] и показать, что получаемые при этом a_i и σ_i положительны [1]. Разложения в ряды и будут использованы для построения приближений Паде-1 с помощью рекуррентных соотношений, описанных в гл. 4. При этом переменные удобно выбирать в несколько более общем виде, а именно $\sigma - \sigma'$ и $(\sigma_0 - \sigma')^{-1}$, где σ' — постоянная. Получим, таким образом, два разложения:

$$F(\sigma_0) = \sum_{l=0}^{\infty} (-1)^l (\sigma_0 - \sigma')^l F(\sigma', l + 1); \quad (7.7)$$

$$F(\sigma_0) = \sum_{k=1}^{\infty} (-1)^{k-1} F(\sigma', -k + 1) / (\sigma_0 - \sigma')^k. \quad (7.8)$$

Полагая $z = \sigma_0 - \sigma'$ и $a_l = (-1)^l F(\sigma', l + 1)$, можно, наращивая число используемых a_l , построить приближение $F_n(\sigma_0) = P_{n-1}(z)/Q_n(z)$. Перейдя затем к полюсному представлению этой аппроксиманты, получим искомые параметры a_i и σ_i . Аналогичным образом, рассматривая (7.8) как ряд по степеням $t = z^{-1}$, можно, используя тот же алгоритм, построить приближение Паде $P_{n-1}(t)/Q_{n-1}(t)$ для функции $t^{-1} F(t^{-1})$. Возвратясь затем к переменной $\sigma_0 - \sigma'$ и перейдя к полюсному представлению, получим параметры a'_i и σ'_i . Они будут совпадать с a_i и σ_i лишь в том случае, если $F(\sigma_0)$ действительно есть рациональная функция. Заметим, что построение паде-аппроксиманты по ряду (7.7) автоматически обеспечивает точное выполнение важного нормировочного условия $\sum_{i=1}^n a_i = 1$.

Итак, вычислив в некоторой точке σ' значения функционалов (7.1) с положительными и отрицательными N , т.е. фактически коэффициенты разложений (7.7), (7.8), можно построить рациональную функцию,

аппроксимирующую $F(\sigma_0)$ во всей области изменения σ_0 . Поскольку при этом $p_n(\sigma) = \sum_{i=1}^n a_i \delta(\sigma - \sigma_i)$, получаются аппроксиманты типа (7.3) и для функционалов с другими N . Однако, как уже упоминалось, сложность достаточно точного вычисления $F(\sigma', N)$ резко увеличивается с ростом $|N|$, поэтому мы рассмотрим еще один способ аппроксимации, позволяющий обойтись сравнительно малым $|N|$ за счет совместного использования первых членов обоих рядов (7.7), (7.8).

Рассмотрим рациональное приближение $F_n(\sigma_0, 1)$, первых n_1 коэффициентов разложения которого по положительным степеням $\sigma_0 - \sigma'$ суть $a_l = (-1)^l F(\sigma', l+1)$, а первых $n_2 = 2n - n_1$ коэффициентов разложения по отрицательным степеням суть $\beta_k = (-1)^{k-1} F(\sigma', -k+1)$, $k \geq 1$, т.е. совпадают с соответствующими коэффициентами разложений (7.7) и (7.8). Связь этих коэффициентов с параметрами полюсного представления аппроксиманты (7.3), (7.4) очевидна:

$$\begin{aligned} a_l &= (-1)^l \sum_{i=1}^n a_i / (\sigma' + \sigma_i)^{l+1}; \quad \beta_k = \\ &= (-1)^{k-1} \sum_{i=1}^n a_i (\sigma' + \sigma_i)^{k-1}. \end{aligned} \quad (7.9)$$

Рассмотрим теперь выражение

$$\begin{aligned} \Phi(x) &= a_{n_1-1} + a_{n_1-2}x + \dots + a_0 x^{n_1-1} - \\ &- \beta_1 x^{n_1} - \beta_2 x^{n_1+1} - \dots - \beta_{n_2} x^{n_1+n_2-1} \end{aligned} \quad (7.10)$$

и построим с помощью алгоритма приближения Паде-1 его аппроксиманту $T_{n-1}(x)/S_n(x) = \sum_{i=1}^n c_i / (x + x_i)$. Сравнивая (7.9) с выражением для параметров этой аппроксиманты

$$a_l = (-1)^{l-n_1-1} \sum_{i=1}^n c_i / x_i^{n_1-l}; \quad \beta_k = (-1)^{n_1+k} \sum_{i=1}^n c_i / x_i^{n_1+k} \quad (7.11)$$

и используя единственность приближения Паде, получаем следующие выражения, которые связывают параметры полюсного представления аппроксиманты $F_n(\sigma_0)$, построенной с учетом положительных и отрицательных моментов a_l , β_k , с параметрами паде-аппроксиманты ряда (7.10)

$$\sigma' + \sigma = 1/x_i; \quad a_l = (-1)^{n_1+1} c_i / x_i^{n_1+1}. \quad (7.12)$$

Таблица 7.1. Значения* параметров аппроксимации для ^{197}Au при различном числе групп

Характеристика аппроксиманты	Паде-2			Паде-1	
	3	4	5	3	4
a_1	$1,66 \cdot 10^{-1}$	$1,09 \cdot 10^{-1}$	$7,86 \cdot 10^{-2}$	$2,26 \cdot 10^{-2}$	$1,17 \cdot 10^{-2}$
σ_1	5903,22	13579	18910	3271,6	4444,5
a_2	$3,20 \cdot 10^{-1}$	$1,60 \cdot 10^{-1}$	$1,05 \cdot 10^{-1}$	$3,98 \cdot 10^{-1}$	$6,52 \cdot 10^{-2}$
σ_2	199,96	935,56	2707,6	38,186	407,75
a_3	$5,05 \cdot 10^{-1}$	$3,40 \cdot 10^{-1}$	$1,78 \cdot 10^{-1}$	$6,79 \cdot 10^{-1}$	$6,73 \cdot 10^{-2}$
σ_3	29,178	91,836	310,413	17,322	34,968
a_4	—	$3,87 \cdot 10^{-1}$	$3,17 \cdot 10^{-1}$	—	$2,50 \cdot 10^{-1}$
σ_4	—	25,768	61,698	—	14,910
a_5	—	—	$3,20 \cdot 10^{-1}$	—	—
σ_5	—	—	24,304	—	—
$S, \%$	0,235	$3,82 \times 10^{-1}$	$4,16 \times 10^{-3}$	2,55	1,26

*Значения σ_i даны в 10^{-28} м^2 , a_i — безразмерные.

Описанным способом (с использованием разложения как по положительным, так и по отрицательным степеням) были рассчитаны параметры a_i и σ_i для полного нейтронного сечения ^{197}Au в одной из энергетических групп $4,65 \leq E \leq 10$ эВ в трех- и четырехподгрупповом представлении. Результаты приведены в левой части табл. 7.1. Кроме значений подгрупповых констант, даны значения S — средней квадратической относительной погрешности описания полученными параметрами кривой $F(\sigma_0)$, рассчитанной при $N_{\text{ex}} = 20$ значениях сечения разбавления. Значения S вычислялись по формуле

$$S = \sqrt{\frac{N_{\text{ex}}}{\sum_{i=1} [(f_{pi} - f_{ai})^2 / f_{pi}]} / (N_{\text{ex}} - 1)},$$

где f_{pi} и f_{ai} — исходные расчетные значения и значения аппроксиманты в i -й точке. При построении паде-аппроксиманты за исходные данные принимались величины

$$\int_{E_{\text{min}}}^{E_{\text{max}}} \varphi(E) dE / [\sigma(E) + \sigma_0]^N, N = -4, -3, \dots, 2, 3,$$

рассчитанные при $\sigma_0 = 10$. При четырех подгруппах была достигнута удовлетворительная точность описания (погрешность порядка 1%). Однако для этого потребовалось вычисление $2n = 8$ моментов, и на том же примере был испытан альтернативный метод построения рациональной аппроксиманты — приближения Паде-2 с помощью алгоритма, описанно-

го в гл. 3. Аппроксимировались при этом те же 20 точек, которые использовались для контроля точности расчетов в первом случае. Полученная точность описания при одинаковом числе подгрупп оказалась гораздо выше, устойчивые результаты были получены и при $n = 5$, в этом случае $S = 4,17 \cdot 10^{-5}$. Нормировочное условие $\sum_i a_i = 1$, хотя теоретически оно и не должно точно выполняться при использовании этого алгоритма, также выполнялось вполне удовлетворительно.

Алгоритм приближения Паде-2 был проверен еще на одном примере с более сложной резонансной структурой, а именно для случая полного нейтронного сечения ^{235}U в двадцатой группе разбиения БНАБ [16] $10,0 \leq E \leq 21,5$ эВ. Значения сечения восстанавливались по резонансным параметрам машинной библиотеки ENDF/B-V [12] при $T = 300$ К. Объектом сравнения служил групповой функционал

$$\bar{\sigma}(\sigma_0) = \int \frac{\sigma(E) \varphi(E) dE}{\Delta E [\sigma(E) + \sigma_0]^2} / \int \frac{\varphi(E) dE}{\Delta E [\sigma(E) + \sigma_0]^2}. \quad (7.13)$$

В качестве спектра выбирался стандартный спектр Ферми $\varphi(E) = 1/E$. Сначала были получены "точные" значения функционала $F(\sigma_0)$ непосредственным вычислением интеграла (7.1) при $N = 1$. Затем построением аппроксиманты Паде-2 определены подгрупповые параметры a_i, σ_i . Достаточная низкая погрешность (около 1%) была получена при $2n = 8$. Далее с помощью подгрупповых параметров восстановлена зависимость от σ_0 по формуле

$$\tilde{\sigma}(\sigma_0) = [\sum_i a_i \sigma_i / (\sigma_i + \sigma_0)^2] / [\sum_i a_i / (\sigma_i + \sigma_0)^2]. \quad (7.14)$$

Таблица 7.2. Зависимость момента $F(\sigma_0) = \langle 1/(\sigma + \sigma_0) \rangle$ от сечения разбавления σ_0 (исходная для определения подгрупповых параметров) и сравнение $\tilde{\sigma}$ - значения, полученного подгрупповой аппроксимацией, с точным блокированным сечением $\bar{\sigma}(\sigma_0) = \int \frac{\sigma(E) \varphi(E) dE}{\Delta E (\sigma(E) + \sigma_0)^2} / \int \frac{\varphi(E) dE}{\Delta E (\sigma(E) + \sigma_0)^2}$ для ^{235}U по данным библиотеки ENDF/B-V [12]; $E = 10,0 \div 21,5$ эВ; $T = 300$ К

σ_0	$F(\sigma_0)$	$\bar{\sigma}(\sigma_0)$	$\tilde{\sigma}$
0	0,02211	33,04	33,12
10	0,01713	35,86	35,91
30	0,01207	39,83	39,92
10^2	$0,6177 \cdot 10^{-2}$	47,90	47,95
$3 \cdot 10^2$	$0,2674 \cdot 10^{-2}$	59,30	59,42
10^3	$0,9184 \cdot 10^{-3}$	76,45	76,52
$3 \cdot 10^3$	$0,3227 \cdot 10^{-3}$	91,86	92,06
10^4	$0,9896 \cdot 10^{-4}$	102,03	102,18

Примечание. Здесь σ_0 и $\tilde{\sigma}$ - в 10^{-28} м².

Результаты приведены в табл. 7.2 и ниже:

a_i	0,0548	0,1719	0,4840	0,2893
$\sigma_i, 10^{-28} \text{ м}^2$	823,2	176,3	52,91	24,35

Устойчивость работы программы, реализующей алгоритм, качество аппроксимации и восстановления на всех этапах оказались удовлетворительными. В настоящее время эта программа включается в существующий программный комплекс по расчету групповых констант [50]. В сложных случаях нормировочные условия выполняются лишь приближенно, но для их уточнения достаточно линейных корректирующих процедур.

Существует более общий по сравнению с изложенным способ построения рациональной аппроксиманты, в котором используются разложения по положительным и отрицательным моментам в различных точках. Он описан в § 4.3, однако в отличие от обсуждавшихся в настоящем параграфе численно реализован не был.

7.2. ОБРАТНАЯ ЗАДАЧА ТЕОРЕТИЧЕСКОЙ СПЕКТРОСКОПИИ

Выше везде предполагалось, что экспериментальные данные отличаются от истинной гладкой кривой лишь вследствие погрешностей эксперимента, которые являются величинами случайными и могут быть учтены только статистической обработкой. Однако на практике чаще всего наблюдаемые зависимости еще и систематически отличаются от тех, которые соответствуют лежащим в их основе физическим закономерностям. Именно такая ситуация имеет место при изучении ядерных реакций. Есть две причины, по которым наблюдаемые сечения отличаются от истинных, и обе связаны не с погрешностью измерения сечения, а с определением значения энергии E .

Энергия E , которая входит в теоретические и аппроксимационные выражения для $\sigma(E)$, — это энергия относительного движения участвующих в реакции частиц. Переход от системы центра масс к лабораторной системе не искажает кривую $\sigma(E)$, а лишь сдвигает начало координат. Однако даже в предположении абсолютной моноэнергетичности пучка налетающих частиц наблюдаемое сечение есть результат физического усреднения по относительной энергии, обусловленного тепловым движением ядер. При отличной от нуля (в абсолютной шкале) температуре такое усреднение не может быть устранено усовершенствованием техники эксперимента.

Другая причина искажения гладкой истинной зависимости связана в основном с немонаэнергетичностью пучка налетающих частиц и обуславливает усреднение по некоторому спектральному распределению. Отсюда при обработке данных эксперимента возникает необходимость учета функции разрешения.

Обе эти причины приводят к тому, что наблюдаемое сечение связано с истинным интегральным преобразованием, и для восстановления истинного сечения надо решать так называемую обратную задачу.

Согласно принятой в математической физике терминологии термин "обратная задача" в широком смысле слова применим ко всем случаям, включающим обращение причинно-следственной связи. Самым распространенным типом таких задач является восстановление сигналов, принятых измерительным прибором, по "отклику" этого прибора, т.е. по сигналу, искаженному регистрирующей, усилительной и анализирующей аппаратурой. Эти задачи в большинстве своем относятся к классу "некорректных" задач ("некорректно поставленных"), в которых малость изменения исходных данных не гарантирует малости изменения решения. В наиболее распространенном случае такие задачи сводятся к решению уравнения

$$\hat{A}\varphi = f, \quad (7.15)$$

где \hat{A} — вполне непрерывный линейный оператор; φ и f — функции, принадлежащие к некоторому линейному пространству, причем f — известная функция, а φ — искомое решение. Сжатое изложение современных методов решения таких задач можно найти в монографиях [51, 52], содержащих подробную библиографию.

В настоящем параграфе речь пойдет о частном, но важном случае обратной задачи, когда f и φ в (7.15) зависят от одной переменной, \hat{A} — оператор свертки и (7.15) сводится к интегральному уравнению Фредгольма первого рода с разностным ядром

$$\int_{-\infty}^{\infty} k(x-y)\varphi(y)dy = f(x). \quad (7.16)$$

Если энергетическая зависимость сечений ядерных реакций, идущих через определенные уровни составного ядра, измерена с конечным разрешением, решение уравнения (7.16) является составной частью резонансного анализа. К нему же относится учет эффекта Доплера — уширения резонансов в результате теплового движения ядер-мишеней.

Рассматриваемый метод относится к аппроксимативным способам решения обратной задачи [52, с. 32], когда $f(x)$ и $k(x)$ аппроксимируются функциями, позволяющими аналитически решить (7.16) с помощью преобразования Фурье. Как известно, уравнению (7.16) соответствует следующая связь между фурье-образами входящих в него функций:

$$K(\omega)\Phi(\omega) = F(\omega). \quad (7.17)$$

Фурье-образы обозначены прописными буквами и определяются как

$$K(\omega) = \int_{-\infty}^{\infty} e^{-i\omega x} k(x) dx \quad (7.18)$$

и т.п. Тогда формальное решение (7.16) записывается в виде

$$\varphi(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{F(\omega)}{K(\omega)} e^{i\omega x} d\omega. \quad (7.19)$$

При использовании аппроксимативного метода функции $K(\omega)$ и $F(\omega)$ должны быть такими, чтобы интеграл (7.19) сходиллся и брался аналитически. Будем использовать паде-аппроксимацию функции $f(x)$ в виде последней суммы полюсного разложения (5.8). Фурье-образ этого выражения есть

$$F(\omega) = \sum_{k=1}^l e^{-i\epsilon_k \omega - \gamma_k |\omega|} \left[\frac{\beta_k}{\gamma_k} - i\alpha_k \operatorname{sign}(\omega) \right]. \quad (7.20)$$

Давно известен пример ядра уравнения (7.16) (см., например, [53]), для которого решение (7.19) имеет очень простой явный вид. Это лоренцево ядро, которое само есть рациональная функция и содержит единственное резонансное слагаемое, т.е.

$$k(x) = \frac{\Gamma}{\pi} \frac{\alpha x + 1}{x^2 + \Gamma^2}; \quad K(\omega) = e^{-\Gamma|\omega|} [1 - i\alpha \operatorname{sign}(\omega)]. \quad (7.21)$$

Тогда после подстановки (7.20), (7.21) в (7.19) получим

$$\varphi(x) = \sum_{k=1}^l \frac{(\alpha_k \gamma_k - \alpha \beta_k \Gamma) (x - \epsilon_k) + (\gamma_k - \Gamma) (\beta_k + \alpha_k \alpha \gamma_k \Gamma)}{\gamma_k (1 + \alpha^2 \Gamma^2) [(x - \epsilon_k)^2 + (\gamma_k - \Gamma)^2]}. \quad (7.22)$$

Видно, что сохраняется число слагаемых в разложении $f(x)$ и $\varphi(x)$, т.е. при решении обратной задачи с лоренцевым ядром происходит переход "резонанса в резонанс" с уменьшением его ширины при некотором искажении формы.

Однако лоренцево ядро (в статистике оно называется распределением Коши) имеет медленно убывающую асимптотику, которая дает, в частности, бесконечную дисперсию, что ограничивает его использование. Приведем примеры ядер уравнения (7.16), допускающих аналитическое решение (7.19) и обладающих более реалистической экспоненциальной или близкой к экспоненциальной асимптотикой:

$$k(x) = \frac{1}{\sigma\sqrt{2}} e^{-\sqrt{2}|x|/\sigma}; \quad (7.23)$$

$$k(x) = 1/2 \sigma \operatorname{ch}(\pi x/2\sigma); \quad (7.24)$$

$$k(x) = \frac{4^{n-1} 2nx}{(2n-1)! \sigma^2 \operatorname{sh}\left(\frac{\pi x}{2} \sqrt{\frac{n}{2}}\right)} \prod_{r=1}^n \left(\frac{x^2 n}{2\sigma^2} + r^2 \right); \quad (7.25)$$

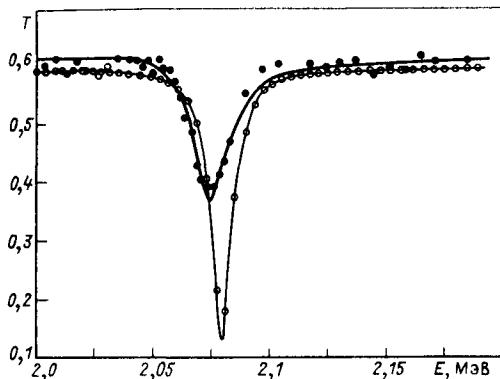


Рис. 7.1. Экспериментальные (темные кружки) и расчетные (светлые кружки) эталонные значения пропускания T на углеороде, использованные для получения функции разрешения: жирная кривая — паде-аппроксиманта

$$k(x) = \frac{2^{2n} \sqrt{2n+1}}{2n! \operatorname{ch}\left(\frac{\pi x}{2} \sqrt{2n+1}\right)} \prod_{r=1}^n \left(\frac{x^2 (2n+1)}{4\sigma^2} + \left(\frac{2r-1}{2}\right)^2 \right); \quad (7.26)$$

$$k(x) = \pi (1 + \cos\beta) \operatorname{ch}(\beta x/a) / a \sin\beta \operatorname{sh}(\pi x/a), \quad (7.27)$$

$$a^2 = \sigma^2 (1 + \cos\beta).$$

Все функции $k(x)$, приведенные выше, нормированы условием $\int_{-\infty}^{\infty} k(x) dx = 1$, а параметр σ выражается через второй момент:

$$\sigma^2 \equiv \int_{-\infty}^{\infty} x^2 k(x) dx. \quad (7.28)$$

Во всех этих примерах, в отличие от лоренцева ядра, не совпадает число членов разложения $f(x)$ и $\varphi(x)$, т.е. каждый "резонанс" в итоге превращается в сумму резонансных слагаемых. Для представления результата в виде стандартного резонансного разложения на последнем этапе решения задачи функцию $\varphi(x)$ можно снова аппроксимировать рациональной функцией с нужным числом полюсов. Ниже будет приведен пример такого рода.

В [54] описанный аппроксимативный метод с лоренцевым ядром использовался для корректировки энергетической зависимости полного нейтронного сечения с целью выяснить влияние учета разрешения на расчетные значения пропускания при прохождении нейтронов через образцы вещества. В качестве примера для обработки было взято полное нейтронное сечение хрома в интервале энергий нейтронов 0,76 — 1,07 МэВ, измеренное в [55] на электростатическом генераторе. Обработка включала следующие этапы:

1. Экспериментально измеренная энергетическая зависимость полного сечения $\sigma_r(E)$ аппроксимировалась рациональной функцией.

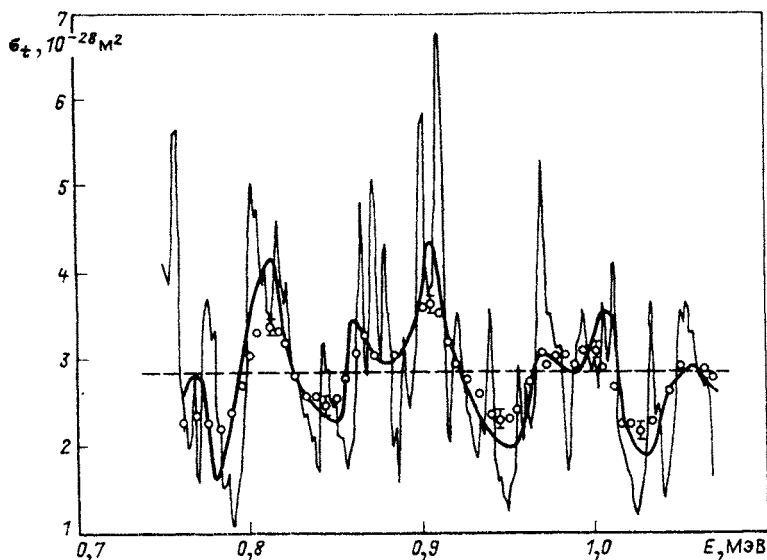


Рис. 7.2. Энергетическая зависимость полного нейтронного сечения хрома в интервале $0,76 < E < 1,07$ МэВ с разной степенью детализации резонансной структуры:

пунктир – постоянное среднее сечение; кружки – экспериментальные данные [56]; жирная кривая – скорректированное сечение; тонкая кривая – детальная структура из файла оцененных данных

2. Для установления вида функции разрешения в [55] были построены приближения Паде для двух эталонных кривых пропускания на резонансе углерода при $E = 2,08$ МэВ: по результатам измерений в условиях [55] и для кривой, рассчитанной по рекомендованным значениям резонансных параметров [56] (рис. 7.1). После этого решалось уравнение (7.16), где в качестве f бралась первая из указанных кривых, а в качестве φ – вторая. Решением была функция разрешения в виде (7.21). Ее полуширина оказалась равной примерно 3,5 кэВ.

3. Полученная функция разрешения использовалась в качестве ядра уравнения (7.16) для получения скорректированной энергетической зависимости $\sigma_t(E)$. В качестве правой части уравнения использовалась аппроксиманта, найденная на первом этапе. На рис. 7.2 изображены исходные экспериментальные точки и скорректированная кривая. Приведены также данные по полному нейтронному сечению хрома, взятые из библиотeki оцененных данных [57]. При их получении использовалось на порядок более высокое разрешение, и они дают более детальную картину зависимости $\sigma(E)$.

4. Все четыре зависимости $\sigma_t(E)$, представленные на рис. 7.2 и в разной степени учитывающие резонансную структуру сечения в указанном интервале, были использованы для расчета среднего по интервалу энергий пропускания нейтронного пучка в зависимости от толщины образца

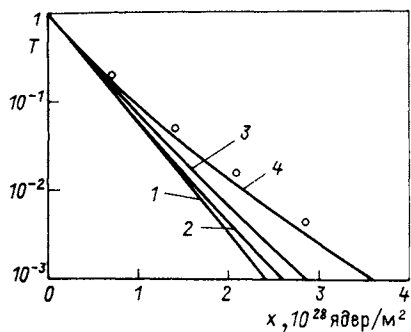


Рис. 7.3. Расчетные кривые и экспериментальные значения (точки) среднего пропускания T в зависимости от толщины образца x :

кривые соответствуют разной степени учета резонансной структуры, т.е. зависимостям $\sigma_f(E)$, приведенным на рис. 7.2; 1 — постоянное среднее сечение; 2 — экспериментальные значения [56]; 3 — скорректированные значения сечения; 4 — файл оцененных данных

хрома. Результаты расчетов представлены на рис. 7.3, где они сравниваются с результатами прямых измерений пропускания на различных толщинах, полученными в [58].

Как и следовало ожидать, каждый шаг в детализации резонансной структуры сечения приводит к увеличению расчетных значений пропускания, которые последовательно приближаются к результатам прямых измерений. В частности, заметное повышение пропускания дает и описанная корректировка, учитывающая разрешение, хотя она, разумеется, не приводит к восстановлению деталей структуры, характерный период которых гораздо меньше ширины линии разрешения.

Эффект Доплера — зависимость наблюдаемого сечения взаимодействия нейтронов с ядрами вещества от температуры последнего — играет важную роль в нейтронной физике и физике реакторов по нескольким причинам. Остановимся на одном аспекте этого вопроса — приведении к нулевой температуре энергетической зависимости сечения в резонансной области, измеренного при не равной нулю температуре T . Интегральное уравнение, связывающее функции энергии $\sqrt{E} \sigma(E, 0)$ и $\sqrt{E} \sigma(E, T)$, в точной форме имеет довольно сложный вид: оно обладает неразностным ядром, однако при не слишком близких к нулю энергиях нейтронов удовлетворительную точность обеспечивает общепринятая газовая модель в приближении "доплеровской ширины", в которой

$$\int_{-\infty}^{\infty} k_G(E - E', T) \sqrt{E'} \sigma(E', 0) dE' = \sqrt{E} \sigma(E, T), \quad (7.29)$$

где гауссово ядро

$$k_G(E, T) = [1/\sqrt{\pi} \Delta(T)] e^{-E^2/\Delta^2(T)}, \quad (7.30)$$

а $\Delta T = \sqrt{2k_B TE/(A+1)}$ — доплеровская ширина; k_B — постоянная Больцмана; A — атомная масса ядра-мишени. Для ядра (7.30) даже формально неприменимо точное решение уравнения (7.29) в форме (7.19), поскольку фурье-образ ядра, имеющий также гауссову зависи-

мость от аргумента, слишком быстро убывает и интеграл (7.19) расходится. Однако аппроксимация правой части (7.29) рациональной функцией и учет того обстоятельства, что искомое решение $\sqrt{E} \sigma(E, 0)$ в нейтронном резонансном анализе по физическим соображениям также хорошо аппроксимируется полюсным разложением, позволяют применять проблемно-ориентированный аппроксимативный метод приближенного решения уравнения (7.29). Он основан на замене ядра (7.30) модельной функцией вида (7.24). Решение (7.19) в этом случае имеет вид

$$\varphi(x) = \frac{1}{2} \left[\frac{\alpha(x - \epsilon) + \beta(1 - \sigma/\gamma)}{(x - \epsilon)^2 + (\gamma - \sigma)^2} + \frac{\alpha(x - \epsilon) + \beta(1 + \sigma/\gamma)}{(x - \epsilon)^2 + (\gamma + \sigma)^2} \right] \quad (7.31)$$

при $f(x) = [\alpha(x - \epsilon) + \beta]/[(x - \epsilon)^2 + \gamma^2]$.

Возникает следующая расчетная схема:

1. Для сечения, измеренного при температуре $T \neq 0$, строится паде-аппроксиманта типа (5.8), число резонансных слагаемых которой может превышать число резонансов истинного решения.

2. По формуле (7.31) каждому из резонансов аппроксиманты сечения при $T \neq 0$ ставится в соответствие два резонанса аппроксиманты при $T = 0$. В результате каждому резонансному слагаемому истинного сече-

Таблица 7.3. Результаты обработки модельной задачи по учету эффекта Доплера для различных температур образца

Характеристика аппроксимант	OK	100 К	273 К	773 К	1473 К
α_1	0,5000	0,4673	0,5121	0,4079	0,4619
β_1	1,3500	1,3864	1,3420	1,3596	1,384
ϵ_1	41,500	41,502	41,493	41,520	41,494
γ_1	0,2500	0,2566	0,2477	0,2550	0,2524
α_2	0,8000	0,8410	0,8232	0,8718	0,8757
β_2	1,5000	1,5136	1,5525	1,4949	1,5615
ϵ_2	40,0000	39,993	39,997	39,972	39,993
γ_2	0,5000	0,5092	0,5160	0,5120	0,5244
L	—	8	8	12	12
$\Delta_L, \%$	—	3,38	3,27	3,26	4,17
$\Delta_f, \%$	—	3,12	2,71	3,31	3,53
$\Delta_\varphi, \%$	—	0,97	1,68	2,80	1,95

Примечание. Здесь L — число параметров аппроксиманты — "экспериментальных" точек; в боковике перечислены восстановленные параметры резонансов (первый столбец, при $T = 0$ — истинные значения). Последние три строки — функционалы, характеризующие точность обработки — средние квадратические относительные отклонения: Δ_L — аппроксиманты от "экспериментальных" точек; Δ_f — результирующей кривой, свернутой с гауссовым ядром, от "экспериментальных" точек; Δ_φ — результирующей кривой от истинной.

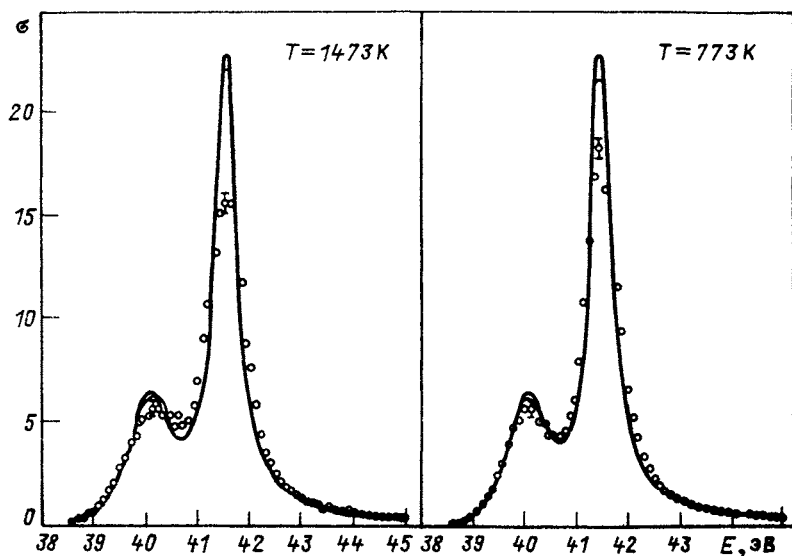


Рис. 7.4. Результаты обработки модельной задачи по учету эффекта Доплера: точки – “экспериментальные” значения доплеровски уширенного сечения; жирная линия “истинное” сечение; тонкая линия – результат восстановления

ния соответствует несколько (обычно два–четыре) резонансных слагаемых приближенного решения обратной задачи.

3. Для этой сложной функции строится более простая паде-аппроксиманта, параметры которой уже имеют смысл обычных резонансных параметров в представлении Адлер-Адлера [42].

4. Сверткой полученного решения с гауссовым ядром и сравнением результата с исходной правой частью контролируется приемлемость полученных результатов.

Чтобы интеграл (7.19) сходил, необходимо выполнение неравенств $\gamma_k > \sigma$ для всех резонансных слагаемых в аппроксиманте, построенной для правой части.

В табл. 7.3 и на рис. 7.4 приведены результаты использования описанного выше метода на модельном примере. В качестве истинной энергетической зависимости сечения была выбрана функция

$$\sigma(E) = \frac{0,5(E - 41,5) + 1,35}{(E - 41,5)^2 + 0,25^2} + \frac{0,8(E - 40,0) + 1,5}{(E - 400)^2 + 0,5^2} \quad (7.32)$$

(энергия – в эВ), соответствующая двум интерферирующим резонансам, которые отстоят друг от друга на расстояние, равное сумме их ширин. Функция (7.35) свертывалась с гауссовым ядром (7.34) при значениях температуры 1473, 773, 273 и 100 К для $A = 239$. К результату свертки,

вычисленному в 65 эквидистантных точках на интервале $38,6 \leq E \leq 45$ эВ, добавлялась случайная погрешность, выбранная из нормального распределения и соответствующая постоянной относительной погрешности 3%. Полученная совокупность экспериментальных точек обрабатывалась описанным алгоритмом. Данные табл. 7.3, проиллюстрированные рис. 7.4, показывают, что найденное решение вполне удовлетворительно согласуется с истинным. На рис. 7.4 результаты приведены лишь для $T = 1473$ К и $T = 773$ К, поскольку при меньших температурах приближение на графике не отличимо от $\sigma(E)$.

7.3. ЭКСПОНЕНЦИАЛЬНЫЙ АНАЛИЗ

Необходимость представить измеренную на опыте функциональную зависимость (чаще всего временную) в виде суммы экспонент и гармоник часто возникает в самых различных областях науки и техники. При решении задачи в наиболее общей постановке определению подлежат и периоды каждой из экспонент, и коэффициенты разложения по ним.

Подход к решению и сложность задачи существенно зависят от характеристик анализируемой кривой — числа экспериментальных точек, их распределения по оси абсцисс, перепада значений (для экспоненциальных кривых), разницы в периодах экспоненциальных слагаемых, уровня шума и т.д. Практически удобных универсальных методов пока не существует, да и в частных случаях решение затруднено. Часто в той или иной форме используется преобразование Лапласа или его дискретный аналог, так называемое Z -преобразование (см. [59, с. 200]), последнее особенно удобно при эквидистантных абсциссах опорных точек, в том и в другом случае сумма экспонент и гармоник является рациональной функцией. Естественно поэтому попытаться использовать при таком анализе приближение Паде. Z -преобразование по определению применяется к функциям, значения которых f_n известны для бесконечной дискретной последовательности значений аргумента с постоянным шагом, занумерованных индексом n . Z -изображение такой последовательности определяется образом Лапласа $F(s) = L\{f(t)\}$ вспомогательной ступенчатой (кусочно-постоянной) функции $f(t)$, определенной ниже:

$$f(t) \equiv f_n (n \leq t \leq n + 1, n = 0, 1, 2 \dots). \quad (7.33)$$

Тогда

$$L\{f(t)\} \equiv \int_0^{\infty} f(t) e^{-st} dt = \frac{1 - e^{-s}}{s} \sum_{n=0}^{\infty} f_n e^{-ns}. \quad (7.34)$$

В определении Z -преобразования несущественный общий множитель

отбрасывается, и окончательное выражение приобретает вид:

$$D\{f_n\} = \sum_{n=0}^{\infty} e^{-ns} f_n. \quad (7.35)$$

Удобно сделать подстановку $e^{-s} = z$. Тогда этот ряд превращается в степенной:

$$\sum_n f_n z^n \equiv F(z). \quad (7.36)$$

Заметим, что обычно используют замену $e^s = z$ и получают ряд по отрицательным степеням z , однако выражение (7.36) здесь удобнее.

Пусть $f_n = e^{\alpha n}$. Тогда (7.36) представляет собой сумму геометрической прогрессии:

$$F(z) = \sum_{n=0}^{\infty} (e^{\alpha} z)^n = 1/(1 - z e^{\alpha}). \quad (7.37)$$

Таким образом, Z -изображение экспоненты является простейшей рациональной функцией, а суммы экспонент — суммой элементарных выражений типа (7.37), т.е. рациональной функцией, у которой $M = N + 1$. Так как формула суммирования справедлива и в комплексной плоскости, не возникает принципиальных осложнений и в случае применения Z -преобразования к суммам (и произведениям) гармоник и экспонент, когда α может принимать комплексные значения. Тогда

$$f_n = \sum_{j=1}^N a_j e^{-r_j t_n} \sin(\tau_j t_n + \delta_j), \quad (7.38)$$

где $t_n = n\Delta + t_0$ (Δ и t_0 — интервал между измерениями и начало отсчета соответственно), и

$$F(z) = \sum_n f_n z^n = \sum_{j=1}^N \frac{a_j e^{r_j(\Delta - t_0)}}{2} \times \left\{ \frac{e^{i[\tau_j(\Delta - t_0) - \delta_j - \pi/2]}}{z - e^{\Delta(r_j - i\tau_j)}} + \frac{e^{-i[\tau_j(\Delta - t_0) - \delta_j - \pi/2]}}{z - e^{\Delta(r_j - i\tau_j)}} \right\}. \quad (7.39)$$

Итак, если Z -изображение анализируемой функции есть отношение полиномов, то, разлагая известным образом это отношение на элементарные дробно-линейные слагаемые, можно сопоставить каждому слагаемому с действительным корнем знаменателя экспоненциальный член функции-оригинала, а каждой паре комплексно-сопряженных корней — экспоненциальный член, умноженный на гармонику с определенными

частотой и фазой (в частных случаях показатели экспоненты или частоты гармоник могут оказаться равными нулю, что соответствует чисто гармоническим или постоянным слагаемым оригинала).

Несмотря на то, что Z -преобразование применимо к бесконечным последовательностям, рациональная функция, как упоминалось выше, однозначно восстанавливается по первым $N + M + 1$ членам своего разложения в ряд Тейлора (этот факт является основой излагаемого метода), причем приведенные в § 4.1 рекуррентные формулы приближения Паде-1 позволяют сделать это достаточно просто. Схема вычислений тогда сводится к следующему.

1. Значения f_n подлежащей анализу функции, соответствующие эквидистантным значениям аргумента, рассматриваются как коэффициенты степенного ряда (7.36), и по ним с помощью формул (4.3), (4.5) строится приближение Паде-1.

2. Из полюсного разложения этого приближения определяются параметры выражения (7.39).

3. Отбрасываются те слагаемые разложения, у которых значения коэффициентов $|a_j|$ меньше некоторого критического уровня, выбранного в качестве границы уровня шумов.

Прежде чем перейти к примеру анализа экспериментальных данных, рассмотрим близкую ему по характеру модельную задачу.

Пусть

$$f(t) = 50e^{-0,075t} + 5e^{-0,5t} + 0,5 \sin(0,4t + 4)$$

— функция, заданная в 20 точках на действительной оси со случайной погрешностью, распределенной по нормальному закону, причем амплитуда распределения $\Delta(t) = \Delta[f_{\max}(t)/f(t)]^{1/4}$, что соответствует случаю, промежуточному между чисто статистической погрешностью (тогда в этом выражении стоял бы квадратный корень) и постоянной амплитудой

Таблица 7.4. Результаты обработки модельной задачи*

$\Delta_{\text{лучш}}, \%$	$\bar{\Delta}, \%$	$\bar{\Delta}_{\text{эксп}}, \%$	$\Delta_{\text{ист}}, \%$	a_1	r_1
0	0	0	0	50	-0,075
0,5	0,756	0,69	0,66	49,90	-0,07496
1,0	1,532	1,93	2,07	49,79	-0,07489
a_2	r_2	a_3	r_3	r_3	δ_3
5	-0,5	0,5	0	0,4	4
5,42	-0,523	0,5765	-0,003	0,3955	4,055
7,93	-0,687	0,687	-0,007	0,3727	4,71
		1,196	-0,102	-0,551	4,42

*В последнем варианте синусоидальный член расщеплен на два.

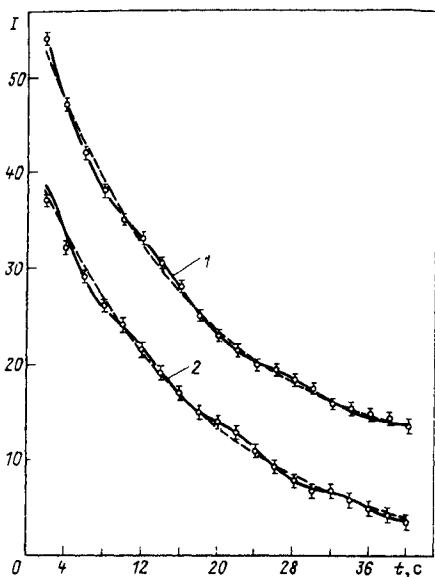


Рис. 7.5. Результаты обработки двух экспериментальных кривых по 20 опорным точкам:

1 — лучшая из всех исследованных в смысле среднего квадратического отклонения кривая (0,192, или 2,46%); 2 — худшая кривая (0,511, или 3,47%). Для наглядности кривая 1 сдвинута вверх на 10 единиц; пунктирные кривые — вклад главного экспоненциального члена

дой разброса. Погрешности в точках задаются датчиком случайных чисел. Результаты анализа таких последовательностей для нескольких значений Δ приведены в табл. 7.4. Качество восстановления оценивается двумя способами — по абсолютному $\bar{\Delta}_{абс}$ и относительному $\bar{\Delta}_{отн}$ сред-

ним квадратическим отклонениям. При рассмотрении задачи имеется возможность сравнить результат восстановления с истинной функцией. С ростом погрешности, как видно из табл. 7.4, надежность выделения компонент, кроме главной, т.е. медленно затухающей экспоненты с большим коэффициентом, быстро падает и синусоидальный член начинает расщепляться на шумовые гармоники.

Практический анализ экспериментальных данных производился на примере набора кривых, изображающих временную зависимость тока электронного охлаждения катода термозмиссионного реактора после переключения нагрузки, по материалам [60]. Всего было обработано 34 кривых, каждая из которых имела от 8 до 20 временных точек, снятых с постоянной абсолютной погрешностью 0,5 единицы. Результаты, характеризующие точность описания экспериментальных данных, приведены ниже:

Число опорных точек	20	18	16	14	12	10	8
Число кривых	11	6	6	7	2	1	1
$\bar{\Delta}_{экс.абс}$	0,43	0,658	0,687	0,537	0,773	0,230	0,577
$\bar{\Delta}_{экс.отн, \%}$	2,62	7,9	6,6	3,1	10,2	2,9	6,1

Для наиболее подробных данных наряду с основным вариантом просчитывались два вспомогательных, по 10 точкам каждый — по четным и по нечетным.

Во всех 34 случаях выделен один главный экспоненциальный член с константой затухания 0,06, а также один или два гармонических члена с максимальной амплитудой на уровне 0,5 – 1,5. В нескольких случаях (примерно третья часть) появлялось второе, чисто экспоненциальное слагаемое с константой затухания, примерно на порядок большей. Вклад всех слагаемых, кроме главного, можно оценить по кривым, изображенным на рис. 7.5. Проведенный анализ позволяет с уверенностью говорить о наличии синусоидальных слагаемых, но данные недостаточно детальны для того, чтобы заключить, соответствует ли этот вклад автоколебаниям с примерно постоянной частотой или шумовым гармоникам.

Образ Лапласа суммы экспонент является рациональной функцией своего аргумента. Если

$$f(t) = \sum_{k=1}^M a_k e^{-p_k t},$$

то

$$F(s) = \int_0^{\infty} e^{-st} f(t) dt = \sum_{k=1}^M \frac{a_k}{s + p_k} = \frac{P_{M-1}(s)}{Q_M(s)}. \quad (7.40)$$

Отсюда вытекает следующий способ разложения функции в сумму экспонент, более общий, чем только что изложенный.

1. Численным интегрированием образ Лапласа $F(s)$ разлагаемой функции $f(t)$ вычисляется в выбранных точках $0 \leq s_n \leq s_{\max}$ действительной оси ($n = 1, 2, \dots, N_s$), полное число их должно быть $N_s \gg 2M + 1$, где M – число искомых экспонент в разложении.

2. Методом дискретной оптимизации этот образ аппроксимируется рациональной функцией, которая разлагается в сумму элементарных слагаемых (7.40). Каждому такому слагаемому сопоставляется экспоненциальная компонента функции-оригинала и по совокупности полученных параметров строится оригинал.

Реально функция $f(t)$ задана в конечном интервале, поэтому перепад экспериментальных значений должен быть достаточным, чтобы можно было пренебречь вкладом неучтенного "хвоста" функции $f(t)$. Пусть $f(t) = e^{-pt}$. Тогда при $s = 0$

$$F(0) = \int_0^{\infty} e^{-pt} dt = 1/p;$$

$$F^*(0) = \int_0^{t_0} e^{-pt} dt = \frac{1}{p} [1 - e^{-pt_0}] = F(0)[1 - e^{-pt_0}].$$

Очевидно, разница между $F(s)$ и $F^*(s^*)$ тем меньше, чем больше произведение pt_0 и чем больше s . Проблема "хвоста" $f(t)$, связанная

с возможностью выделения долгоживущих (с малыми p_k) компонент функции, присутствует в любом методе экспоненциального анализа. На практике эта добавка воспринимается как составляющая шума. Например, при $s = 0$ и $pt_0 \approx 16$ добавка e^{-pt_0} имеет значение порядка 10^{-6} .

Значения параметров, так же как и в других методах анализа, влияют на эффективность их восстановления. Короткоживущие экспоненты с малыми вкладами всегда трудно выделять на фоне долгоживущих с большими коэффициентами. Обычно это свойство задачи учитывается на этапе получения (экспериментального или расчетного) значений исходной функции $f(t)$.

При преобразовании Лапласа поведение функции-оригинала $f(t)$ при малых t определяется поведением $F(s)$ при больших s и наоборот. Отсюда вытекает следующее. Во-первых, величина s_{\max} является параметром метода, и ее оптимальное значение связано с перепадом значений F от $F(0)$ до $F(s_{\max})$ и с эффективностью выделения короткоживущих экспонент. Во-вторых, важную роль играет распределение точек s_n на интервале $[0, s_{\max}]$. Практика показала, что значение s_{\max} должно быть не меньше ожидаемого p_{\max} и его следует выбирать таким образом, чтобы при числе экспонент $M = 2$ перепад от $F(0)$ до $F(s_{\max})$ был не менее десяти, при $M = 5 \div 7$ — не менее $10^3 - 10^4$. Что же касается второго обстоятельства, то точки должны сгущаться к $s = 0$, чтобы точнее восстанавливались долгоживущие экспоненты. На практике желательно варьировать и s_{\max} , и распределение точек на оси s , пользуясь результатом пробного разложения.

Уровень погрешностей исходной функции отражается, конечно, на эффективности восстановления параметров разложения. Он влияет и на результат вычисления образа Лапласа, но $F(s)$, как величина интегральная, является более гладкой функцией своего аргумента, чем исходная функция $f(t)$. В результате точность аппроксимации $F(s)$ выше, чем точность аппроксимации $f(t)$. Если $\bar{\Delta}_f$ и $\bar{\Delta}_F$ — характеристики качества описания $f(t)$ и $F(s)$, то $\bar{\Delta}_F < \bar{\Delta}_f$ и минимальные значения $\bar{\Delta}_f$ не обязательно соответствуют минимальным $\bar{\Delta}_F$. Однако на практике критерием выбора опорных точек при аппроксимации может служить как $\bar{\Delta}_f$, так и $\bar{\Delta}_F$, только для расчета $\bar{\Delta}_f$ требуется существенно большее машинное время.

Проиллюстрируем устойчивость и качество описания на примере анализа экспериментальных кривых пропускания для серы [58] при двух значениях энергии. В обоих случаях оптимальное разложение было получено для двух экспонент. На рис. 7.6 представлены отдельно вклады каждого экспоненциального члена. Для каждого случая были учтены результаты около десятка вариантов обработки, соответствующих различным s_{\max} и разным распределениям точек по оси s . Это дало возможность приблизительно оценить неопределенность вклада каждой экспоненты (заштрихованные области на рисунках). Исходные значения функций

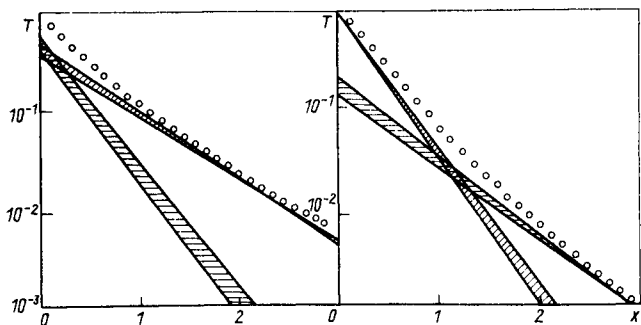


Рис. 7.6. Примеры обработки двух экспериментальных кривых пропускания для серы:

по оси абсцисс – толщина образца в условных единицах; заштрихованные области – вклады каждого экспоненциального члена с учетом его неопределенности

пропускания описывались в пределах экспериментальной погрешности. Видно, что выделение в каждом случае двух главных компонент разложения происходит с достаточной степенью устойчивости. Попытки выделения трех экспонент привели в каждом случае к резкому увеличению неопределенности каждой компоненты, и это означает, что оптимальны разложения на суммы двух экспонент.

Одно из возможных применений описываемого метода – анализ интегральных кривых пропускания, которые используются в физике реакторов для расчета различных интегральных функционалов в рамках подгруппового подхода, так что последующее изложение тесно связано с материалом начала данной главы. Кривой пропускания называют зависимость

$$T_i(x) = \int_{E_1}^{E_2} \varphi(E) \sigma_i(E) e^{-\sigma_t(E)x} dE,$$

где интеграл берется по энергетической группе. Здесь $\varphi(E)$ – функция, описывающая нейтронный спектр; x – толщина образца; σ_i и σ_t – полное и парциальное сечения.

Если приближенно заменить $\sigma_t(E)$ ступенчатой функцией и перейти в интеграле к переменной σ_t , то, поскольку производная от ступенчатой функции есть сумма δ -функций, получим:

$$T_i(x) = \int_{\sigma_{\min}}^{\sigma_{\max}} \varphi \sigma_i e^{-\sigma_t x} \frac{dE}{d\sigma_t} d\sigma_t = \sum_{k=1}^M a_k^i e^{-\sigma_{ik} x}. \quad (7.41)$$

При этом сумму

$$p(\sigma_i) = \sum_{k=1}^M a_k^i \delta(\sigma_i - \sigma_{ik}) \quad (7.42)$$

Таблица 7.5. Результаты анализа кривых пропускания T_f , T_c , T_s для Mn

M	$\bar{\Delta}_F$	a_1	a_2	a_3	a_4	a_5	a_6	a_7
	$\bar{\Delta}_f, \%$	$-r_1$	$-r_2$	$-r_3$	$-r_4$	$-r_5$	$-r_6$	$-r_7$
6	$2,6 \cdot 10^{-7}$	0,00317	0,0105	0,0255	0,06321	0,2369	0,6587	—
	0,67	9,302	14,45	30,04	77,46	244,07	567,0	—
7	$6,2 \cdot 10^{-8}$	0,00406	0,0128	0,0302	0,0698	0,02623	0,6195	0
	1,28	9,577	16,06	35,53	91,94	271,5	581,1	2,266
8	$1,2 \cdot 10^{-8}$	0,00098	0,0068	0,0176	0,0379	0,0816	0,2963	0,5591
	0,96	8,412	11,391	21,03	47,97	121,1	313,6	601,9
9	$2,9 \cdot 10^{-8}$	0,00117	0,0074	0,0165	0,0359	0,07721	0,287	0,5733
	0,66	9,043	12,71	22,58	48,50	117,4	304,8	597,1
7	$7,8 \cdot 10^{-8}$	0,00561	0,0189	0,0416	0,0851	0,2572	0,5904	0
	0,94	9,532	15,65	33,86	87,42	261,45	590,46	—
8	$1,2 \cdot 10^{-8}$	0,04582	0,01523	0,0336	0,0728	0,2148	0,6552	—
	0,66	9,329	14,28	28,12	68,83	213,99	550,68	—
9	$1,1 \cdot 10^{-8}$	0,04252	0,0135	0,0276	0,0526	0,0990	0,2994	0,5034
	0,59	9,265	13,77	25,26	54,35	131,12	325,54	610,7
6	$1,9 \cdot 10^{-5}$	0,08852	0,2011	0,2295	0,1986	0,2771	0	—
	0,74	9,394	15,056	33,88	107,78	446,78	0	—
7	$3,8 \cdot 10^{-6}$	0,1118	0,2121	0,2227	0,1911	0,2593	$2,19 \cdot 10^{-4}$	0
	0,37	9,7508	16,64	39,018	126,73	477,9	5,5929	—
8	$4,0 \cdot 10^{-7}$	0,0611	0,1346	0,1653	0,0302	0,0698	0,2623	0,1753
	0,31	9,0706	12,841	20,628	40,399	92,985	257,99	576,06

Примечание. Здесь M — число экспоненциальных членов в разложении. Три части таблицы по вертикали соответствуют T , T_c и T_s .

можно трактовать как приближенное распределение вероятностей для различных значений сечения в данной энергетической группе [15]. Тогда, разложив $T_i(x)$ на сумму экспонент (7.41), можно с помощью (7.42) просто вычислять различные интегральные величины, зависящие от поведения сечения в пределах группы, например средние сечения

$$\langle \sigma_i \rangle = \sum_{k=1}^M a_k^i \sigma_{ik}$$

и другие функционалы (см. § 7.1). Здесь подгрупповые константы a_k^i и σ_{ik} — параметры разложения (7.41) для полного сечения $T_i(x)$.

В табл. 7.5 приведены результаты анализа кривых пропускания для марганца в пределах энергетической группы $2 \leq E \leq 4$ кэВ, полученных в [61]. Три кривые соответствуют сечениям захвата, рассеяния и полному сечению. Анализируемые кривые пропускания не являются точными суммами конечного числа экспоненциальных слагаемых и в принципе могут быть разложены в бесконечную их сумму с последовательно убывающими вкладками, но практически их число ограничивается точностью расчетов.

Данные табл. 7.5 показывают, что восстановление образа Лапласа с точностью до шестого-седьмого знака обеспечивает восстановление кривых пропускания на интервале восемь-девять декад с относительной погрешностью 0,3 – 0,5 % при выделении шести-семи экспоненциальных слагаемых. Поскольку значение полного сечения на энергетическом интервале группы заключено в пределах от 10^{-27} до $6 \cdot 10^{-26}$ м², то примерно в тех же пределах должны располагаться полученные значения, что, как видно, выполняется. Проверить "качество" полученного набора подгрупповых констант можно, рассчитав с их помощью средние сечения и коэффициенты самоэкранировки и сравнив их со значениями, полученными численным интегрированием с использованием фактической энергетической зависимости $\sigma(E)$. Результаты такого сравнения приведены в табл. 7.6.

Таблица 7.6. Средние сечения $\langle \sigma_f \rangle$ и факторы самоэкранировки F_c при разных значениях сечения разбавления σ_0

Метод расчета	$\langle \sigma_f \rangle, 10^{-28} \text{ м}^2$	F_c		
		$\sigma_0 = 0$	$\sigma_0 = 10^{-27} \text{ м}^2$	$\sigma_0 = 10^{-26} \text{ м}^2$
Численный	167,7	0,1514	0,1995	0,4087
По параметрам экспонент	166,8	0,1504	0,1981	0,4075

Величины обоих типов хорошо согласуются между собой, что и оправдывает введение констант, полученных из экспоненциального разложения.

В силу ограниченности объема данной книги представляется целесообразным в заключение сослаться дополнительно на некоторые работы авторов, в которых либо можно найти более детальное освещение вопросов, затронутых выше, либо рассматриваются вопросы, смежные с затронутыми: [62–64] – построение паде-аппроксиманты, оценка ее погрешностей и практические примеры; [65] – многоуровневый резонансный анализ; [66] – суммирование рядов квантовомеханической теории возмущений в рекурсивной формулировке; [67] – информационная емкость экспериментальных кривых с погрешностями; [68] – получение подгрупповых констант; [69 – 71] – обратная задача; [72, 73] – экспоненциальный анализ.

СПИСОК ЛИТЕРАТУРЫ

1. Baker G. Essentials of Padé approximants. N.Y.: Academic Press, 1975.
2. Lecture Notes on Mathematics 765. Padé approximation and its applications/ Ed. L. Wuytack. Berlin – Heidelberg – N.Y.: Springer-Verlag, 1979.
3. Lecture Notes on Mathematics 888. Padé approximation and its applications/ Ed. M.E. de Bruin, H. van Kossum. Berlin – Heidelberg – N.Y.: Springer-Verlag, 1981.
4. Никитину Ф. Фазовый анализ в физике ядерных взаимодействий: Пер. с рум.: М.: Мир, 1983.
5. Zinn-Justin J. // Phys. Repts. 1971. Vol. 1С. N 3. P. 55 – 102.
6. Basdevant J. // Fortschr. Phys. 1972. Bd 20. S. 283 – 331.
7. Апресян А.А. // Изв. вузов. Сер. Радиофизика. 1979. Т. XXII. № 6. С. 653–674.
8. Виноградов В.Н., Гай Е.В., Работнов Н.С. Препринт ФЭИ-484. Обнинск, 1974.
9. Виноградов В.Н., Гай Е.В., Работнов Н.С. // Вопросы атомной науки и техники. Сер. Ядерные константы. М.: ЦНИИатоминформ, 1975. Вып. 20 (1). С. 13–28.
10. Виноградов В.Н., Гай Е.В., Работнов Н.С. Обзорная информация ОБ-125 ФЭИ. Обнинск, 1981.
11. ENDF/B, Summary documentation/Compiled by D.Garber. BNL-17541 (TNDF-201). Brookhaven, 1975.
12. Nuclear Standard Reference Data: Proc. of advisory meeting. Gell, Belgium, 12–16 Nov., 1984. Vienna: IAEA, 1985.
13. UKNDL-80, Contents and documentation/Compiled by O.Schverer. IAEA-NDC-30 Vienna: IAEA, 1980.
14. JENDL-1, Japanese Evaluated Nuclear Data Library, Version-1. IAEA-NDS-18. Vienna: IAEA, 1979.
15. Многогрупповое приближение в теории переноса нейтронов/М.Н. Николаев, Б.Г. Рязанов, М.М. Савоськин, А.М. Цибуля. М.: Энергоатомиздат, 1984.
16. Групповые константы для расчета реакторов и защиты/Л.П. Абагян, Базаянц Н.О., М.Н. Николаев, А.М. Цибуля. М.: Энергоиздат, 1981.
17. Уиттекер Э.Т., Ватсон Дж.Н. Курс современного анализа: Пер. с англ. М.: Физматгиз, 1963. Т. 1.
18. Лаврентьев М.А., Шабат Б.В. Методы теории функций комплексного переменного. М.: Физматгиз, 1958.
19. Чью Дж. Аналитическая теория S-матрицы: Пер. с англ. М.: Мир, 1968.
20. Голдман С. Теория информации: Пер. с англ. М.: Изд-во иностр. лит., 1957.
21. Сеге Г. Ортогональные многочлены: Пер. с англ. М.: Физматгиз, 1962.
22. Курош А.Г. Курс высшей алгебры. М.: Физматгиз, 1963.
23. Уолш Дж.Л. Интерполяция и аппроксимация рациональными функциями в комплексной области: Пер. с англ. М.: Изд-во иностр. лит., 1961.
24. Маркушевич А.И. Теория аналитических функций. М.: Наука, 1967. Т. 1.
25. Люк Ю. Специальные математические функции и их аппроксимации: Пер. с англ. М.: Мир, 1980.
26. Compton A.K. // J. Phys. 1982. Vol. 15. P. 3665–3667.
27. Мишина А.П., Проскуряков И.В. СМБ. Высшая алгебра. М.: Физматгиз, 1962.
28. Данилов В.Л., Иванова А.Н., Исакова Е.К. и др. СМБ. Математический анализ. М.: Физматгиз, 1962.

29. Fleischer J. // J. Math. Phys. 1973. Vol. 14. N 2. P. 246–248.
30. Гай Е.В. Препринт ФЭИ-1582. Обнинск, 1984.
31. Алберг Дж., Нильсон Э., Уолш Дж.Л. Теория сплайнов и ее приложения. Пер. англ. М.: Мир, 1972.
32. СМБ. Математическая теория планирования эксперимента/С.М. Ермаков, В.З. Бродский, А.А. Жиглявский и др. М.: Наука, 1983.
33. Худсон Д. Статистика для физиков: Пер. с англ. М.: Мир, 1970.
34. Виноградов В.Н., Гай Е.В., Работнов Н.С. Препринт ФЭИ-1328. Обнинск, 1982.
35. Бадиков С.А., Гай Е.В., Работнов Н.С. Препринт ФЭИ-1565. Обнинск, 1984.
36. Краснопольский В.М., Каганов Б.Г., Хлебников С.Ю. // Тезисы докладов XXXIV совещания по ядерной спектроскопии и структуре атомного ядра. Л.: Наука, 1984. С. 399.
37. Виноградов В.Н., Гай Е.В., Кононов В.Н., Работнов Н.С. // Нейтронная физика (материалы 5-й Всесоюз. конф. по нейтронной физике. Киев, 15–19 сент. 1980). М.: ЦНИИАтоминформ, 1980. Ч. 4. С. 58 – 61.
38. Бадиков С.А., Виноградов В.Н., Гай Е.В. и др. // Вопросы атомной науки и техники. Сер. Ядерные константы. М.: ЦНИИАтоминформ, 1982. Вып. 3 (47). С. 66 – 79.
39. Бирюков Н.С., Журавлев Б.В., Рученко А.П. и др. // Ядерная физика. 1982. Т. 3., Вып. 4. С. 814 – 819.
40. Виноградов В.Н., Гай Е.В., Глуховец А.Н. и др. // Там же. 1977. Т. 26, вып. 5. С. 936 – 941.
41. Ландау Л.Д., Лифшиц Е.М. Квантовая механика. М.: Наука, 1974.
42. Adler D.B., Adler F.T. // Nuclear Data for Reactors. Vienna: IAEA, 1970. Vol. 2. P. 77 – 792.
43. Kornilov N.V., Vinogradov V.N., Gay E.V. e.a. // Nuclear Data Science and Technology. Brussels: Inst. Exper. Phys. 1983. P. 679 – 680.
44. Груздевич О.Т., Давлетшин А.Н., Типунков А.О. и др. // Вопросы атомной науки и техники. Сер. Ядерные константы. М.: ЦНИИАтоминформ, 1983. Вып. 2 (51). С. 3 – 15.
45. Давлетшин А.Н., Типунков А.О., Тихонов С.В. и др. // Нейтронная физика (материалы 6-й Всесоюз. конф. по нейтронной физике. Киев, 2–6. окт. 1983). М.: ЦНИИАтоминформ, 1983. Ч. 2. С. 164 – 167.
46. Груздевич О.Т., Давлетшин А.Н., Типунков А.О. и др. // Там же. С. 181 – 184.
47. Виноградов В.Н., Гай Е.В., Работнов Н.С. // Журн. вычисл. матем. и матем. физ. 1981. Т. 21, № 8. С. 1557 – 1559.
48. Кульбак С. Теория информации и статистика: Пер. с англ. М.: Наука, 1967.
49. Ахмезер Н.И. Классическая проблема моментов. М.: Физматгиз, 1961.
50. Синица В.В., Абагян Л.П., Базаянц Н.О. и др. // Ядерно-физические исследования в СССР. М.: ЦНИИАтоминформ, 1979. Вып. 27. С. 31 – 38.
51. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. М.: Наука, 1979.
52. Василенко Г.И. Теория восстановления сигналов. О редукции к идеальному прибору в физике и технике. М.: Сов. радио, 1979.
53. Лукьянов А.А. Структура нейтронных сечений. М.: Атомиздат, 1978.
54. Виноградов В.Н., Гай Е.В., Работнов Н.С., Филиппов В.В. // Вопросы атомной науки и техники. Сер. Ядерные константы. М.: ЦНИИАтоминформ, 1979. Вып. 3 (34). С. 70 – 72.
55. Филиппов В.В., Николаев М.Н. Бюллетень информационного центра по ядерным данным. М.: Атомиздат, 1966 Вып. 111. С. 102 – 107.
56. Mughabhab S., Garber D. Neutron Cross-Sections. BNL-325. – Third Ed. Brookhaven, 1973, Vol. 1.
57. Бычков В.М., Возяков В.В., Манохин В.Н. и др. // Вопросы атомной науки и техники. Сер. Ядерные константы. М.: ЦНИИАтоминформ, 1976. Вып. 23. С. 28 – 39.

58. Филиппов В.В., Николаев М.Н. Доклад АСС-68-23 на англо-советском семинаре "Ядерные константы для расчета реакторов" (Дубна, 1968). INDC-CCP 16/1. Vienna: IAEA, 1971. P. 68 – 73.
59. Деч Г. Руководство к практическому использованию преобразования Лапласа и Z-преобразования. М.: Наука, 1971.
60. Кузнецов В.А., Грязнов Г.М., Артюхов Г.Я. и др. // Атомная энергия. 1974. Т. 36, № 6. С. 450 – 457.
61. Абагян Л.П., Николаев М.Н., Сеница В.В. // Ядерные константы. М.: Атомиздат, 1972. Вып. 9. С. 146 – 175.
62. Бадиков С.А., Виноградов В.Н., Гай Е.В., Работнов Н.С. // Атомная энергия. 1984. Т. 56, вып. 1. С. 20 – 25.
63. Виноградов В.Н., Гай Е.В., Работнов Н.С. // Вопросы атомной науки и техники. Сер. Ядерные константы. М.: ЦНИИатоминформ, 1977. Вып. 25. С. 81 – 84.
64. Бадиков С.А., Блохин А.И., Гай Е.В. и др. // Там же. 1984. Вып. 3 (57). С. 28 – 41.
65. Виноградов В.Н., Гай Е.В., Работнов Н.С. // Там же. 1976. Вып. 21. С. 21 – 31.
66. Будник А.П., Гай Е.В., Работнов Н.С. Препринт ФЭИ-583. Обнинск, 1975.
67. Виноградов В.Н., Гай Е.В., Работнов Н.С. // Прикладная ядерная спектроскопия. М.: Энергоатомиздат. 1982. Вып. 11. С. 128 – 135.
68. Бадиков С.А., Гай Е.В., Работнов Н.С., Сеница В.В. Препринт ФЭИ-1580. Обнинск, 1984.
69. Бадиков С.А., Виноградов В.Н., Гай Е.В., Работнов Н.С. Препринт ФЭИ-1279. Обнинск, 1982.
70. Виноградов В.Н., Гай Е.В., Работнов Н.С. // Вопросы атомной науки и техники. Сер. Ядерные константы. М.: ЦНИИатоминформ, 1978. Вып. 29. С. 37 – 40.
71. Виноградов В.Н., Гай Е.В., Работнов Н.С. // Там же. 1981. Вып. 42. С. 9 – 12.
72. Виноградов В.Н., Гай Е.В., Работнов Н.С. Препринт ФЭИ-513. Обнинск, 1974.
73. Виноградов В.Н., Гай Е.В., Работнов Н.С. Препринт ФЭИ-554. Обнинск, 1975.