

к-616

ОБЪЕДИНЕННЫЙ ИНСТИТУТ ЯДЕРНЫХ ИССЛЕДОВАНИЙ

11-88-704

УДК 519.254 + 681.3.06

КОЛТУН

Ирина Аркадьевна

**ДЕКОМПОЗИЦИОННЫЕ МЕТОДЫ АНАЛИЗА
ЭКСПОНЕНЦИАЛЬНЫХ ЗАВИСИМОСТЕЙ
И ИХ ПРИМЕНЕНИЕ**

**Специальность 05.13.16 - применение вычислительной
техники, математического моделирования
и математических методов в научных исследованиях**

**Автореферат диссертации на соискание ученой степени
кандидата физико-математических наук**

Дубна 1988

Работа выполнена на кафедре математической физики факультета вычислительной математики и кибернетики Московского ордена Ленина, ордена Октябрьской революции и ордена Трудового Красного Знамени государственного университета им. М.В.Ломоносова.

Научный руководитель
доктор физико-математических наук
профессор

Зайкин П.Н.

Официальные оппоненты:
доктор технических наук

Цупко-Ситников В.М.

кандидат физико-математических наук

Пергамент А.Х.

Ведущая организация: Ленинградский институт ядерной физики им. Б.П.Константинова АН СССР.

Защита состоится " " 198 г.
в часов на заседании специализированного совета
Д 047.01.04 Лаборатории вычислительной техники и автоматизации ОИЯИ, г. Дубна (Московской области).

С диссертацией можно ознакомиться в библиотеке ОИЯИ.

Автореферат разослан " " 198 г.

Ученый секретарь
специализированного совета
кандидат
физико-математических наук

Иванченко З.М.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы. При исследовании таких физических процессов, как анализ дискретных спектров излучения в спектрометрии, в активационном анализе и др. с учетом широкого освоения персональных компьютеров профессионального типа наряду с задачами построения программного обеспечения вплоть до физической интерпретации результатов большое внимание уделяется проблеме построения гибких вычислительных систем, обрабатывающих данные как в диалоговом, так и в автоматизированном режимах. Эти задачи входят в проблему компьютеризации рабочих мест экспериментаторов. В этом аспекте при создании систем обработки экспериментальных данных предъявляются новые требования к качеству алгоритмов и устойчивости используемых методов, надежности создаваемой программной продукции, способам организации систем математического обеспечения, простоте их использования, визуализации данных.

Например, при декомпозиции экспоненциальных зависимостей, часто встречающейся в задачах автоматизации научных исследований, важное место занимают проблемы устойчивого оценивания количества компонент и разделения близких, тем более, что исходные данные обычно известны с погрешностями. Эта задача была решена для данных, измеренных на равномерных сетках. Однако для многих экспериментов, в частности, ряда медико-биологических исследований, требование неравномерности сетки измерений является существенным.

Учетом далее, что типовая задача устойчивой декомпозиции спектротрических данных сводится к анализу экспоненциальных зависимостей. В связи с этим, а также о появлении в последнее время инструментальных средств для конструирования пакетов прикладных программ (ППП) на персональных электронно-вычислительных машинах (ПЭВМ), реальной стала возможность создания базового ППП для анализа спектротрических данных, который стал бы удобным и эффективным средством для построения индивидуальных систем обработки в конкретных экспериментах.

Цель работы заключалась в разработке математического, программного и информационно-экспертного обеспечения пакетов прикладных программ для анализа экспоненциальных зависимостей и задач декомпозиции спектротрических данных.

Научная новизна работы

Для данных, измеренных на неравномерных сетках, разработан устойчивый итерационный алгоритм автоматизированной декомпозиции экспоненциальных зависимостей в рамках приближения к главному квазирешению. Для решения задачи разделения близких компонент предложен устойчивый одношаговый метод.

Создана библиотека прикладных программ для анализа экспоненциальных зависимостей на ПЭВМ.

В рамках современного понятия предметной области создано математическое, программное и информационно-экспертное обеспечение базового пакета прикладных программ для решения задачи декомпозиции спектрометрических данных на ПЭВМ.

Разработаны алгоритмы и программное обеспечение для анализа данных в некоторых классах экспериментов - ИПИ "ОНКО-1" во Всесоюзном онкологическом научном центре, ИПИ "СПЕКТР-2" в ЛНФ ОИЯИ, ориентированные на эксплуатацию на персональных компьютерах профессионального типа класса ЕС-1840, ЕС-1841, ПРАВЕЦ-16, IBM PC AT/XT.

Практическая значимость работы. Работа проводилась в соответствии с заданием ГЗ "Разработка пакетов прикладных программ по планированию эксперимента и обработке" по проблеме 2.2.2 "Создание систем автоматизации исследований и экспериментов (АСНИ)" КИ НТП стран-членов СЭВ до 2000 года.

Предложенные алгоритмы решения задач анализа экспоненциальных зависимостей и декомпозиции спектрометрических данных реализованы в виде программных комплексов на языке ФОРТРАН для персональных компьютеров профессионального типа класса ЕС-1840, ЕС-1841, ПРАВЕЦ-16, IBM PC AT/XT, а также для ЭВМ БЭСМ-6. На их основе разработаны ИПИ "СПЕКТР-1" для анализа спектров нейтронов в ЛЯП ОИЯИ, "СПЕКТР-2" для анализа рентгеновских спектров в задаче определения микроэлементного состава веществ в ЛНФ ОИЯИ, "ОНКО-1" для анализа экспоненциальных зависимостей в задачах Всесоюзного онкологического научного центра, "TUTOR" для практикума на биологическом факультете МГУ.

Апробация работы. Основные результаты работы докладывались автором на заседании Рабочей группы по координации работ в области обработки спектрометрической информации и использования банка оцененных ядерных данных при Секции математического обеспечения Совета по автоматизации научных исследований при Президиуме АН СССР (г. Дубна, 1985 г.), конференции молодых ученых факультета ВМиК МГУ им. М.В. Ломоносова (1985 г.), обсуждались на научных семинарах биологического факультета МГУ (1987 г.), ЛНФ ОИЯИ (1987 г.), ЛЯП ОИЯИ (1987 г.), ЛВТА ОИЯИ (1988 г.).

Публикации. Основные результаты диссертации опубликованы в работах /Г-3/.

Структура работы. Диссертация состоит из введения, четырех глав основного текста, заключения, списка литературы из 140 наименований, пяти приложений.

Во введении сформулированы общие постановки задач анализа экспоненциальных зависимостей и декомпозиции спектрометрических данных, показана их связь, обоснована актуальность работы, дано ее краткое содержание по главам, и изложены основные результаты.

Анализ и интерпретацию экспериментальных спектрометрических данных часто проводят в рамках зависимостей вида:

$$F_g(x) = \sum_{j=1}^M A_j \psi(x, \theta_j) + \xi(x), \quad (1)$$

- где
- M - количество компонент,
 - ψ - заданная форма линий,
 - $\theta_j, j = \overline{1, M}$ - положения линий,
 - $A_j, j = \overline{1, M}$ - интенсивности линий,
 - $\xi(x)$ - шум, $\|\xi(x)\| \ll \|F_g(x)\|$.

Общая задача декомпозиции состоит в восстановлении параметров разложения $\{M; \theta_j, A_j, j = \overline{1, M}\}$ по известным значениям экспериментальных данных $F_g(x)$ и оценкам ошибок их измерения $\Delta F_g(x)$. Она отличается от традиционной наличием среди оцениваемых параметров M - количества компонент в представлении (1).

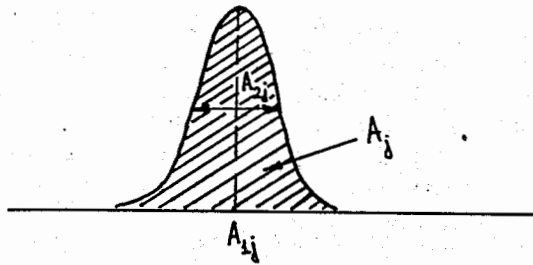
В первых трех главах работы рассмотрены экспоненциальные классы функций ($\theta_j = \lambda_j, j = \overline{1, M}$):

$$\mathcal{F}_1 = \left\{ F: F(x) = \sum_{j=1}^M A_j e^{-\lambda_j x}, a \leq x \leq b, 0 < M < \infty, \lambda_i \neq \lambda_j \text{ при } i \neq j, |\lambda_j| < \infty, |A_j| < \infty; j = \overline{1, M} \right\}. \quad (2)$$

Последняя глава посвящена анализу более общего класса данных спектрометрического эксперимента:

$$\mathcal{F}_2 = \left\{ f: f(x) = \sum_{j=1}^M A_j \psi(x - A_{1j}, A_{2j}), a \leq x \leq b, 0 < M < \infty, A_{1i} \neq A_{1j} \text{ при } i \neq j, |A_{1j}| < \infty, |A_{2j}| < \infty, |A_j| < \infty; j = \overline{1, M} \right\}, \quad (3)$$

где ψ - локализованные, одномодальные, симметричные, непрерывные функции, параметры которых носят следующий смысл (A_{1j} - положения, A_{2j} - полуширины, A_j - интенсивности, $j = \overline{1, M}$):



В первой главе проводится сравнение известных методов решения задачи (I), (2), в частности:

- 1) графического анализа экспоненциальной кривой и ее производных,
- 2) обработки данных с фиксированным набором показателей экспонент,
- 3) класса методов типа наименьших квадратов, в том числе с привлечением регуляризирующих подходов,
- 4) интегро-дифференциальных методов и др.

Показано, что наиболее перспективной для дальнейшего анализа является группа интегро-дифференциальных методов. Преимущество их заключается в достаточно простой формализации способов получения оценок параметров $\{\lambda_j, A_j, j=1, \overline{M}\}$.

Характерная схема решения задачи этими методами имеет вид:

1) решение некоторой системы линейных алгебраических уравнений (СЛАУ) для получения оценок вспомогательного вектора $C = (c_1, \dots, c_e)$ $C = P(M)$:

$$BC = b, \quad B = B(F(x), M), \quad b = b(F(x), M) \quad (4)$$

2) вычисление корней полинома:

$$P_M(C, y) = y^M + \sum_{j=1}^M (-1)^j c_j y^{M-j} \quad (5)$$

с помощью которых определяются параметры $\lambda_j, j=1, \overline{M}$:
 - для данных, измеренных на равномерных сетках с шагом H , в методе Прони

$$\lambda_j = \frac{1}{H} \ln y_j, \quad j=1, \overline{M}, \quad (6)$$

- для остальных методов

$$\lambda_j = y_j, \quad j=1, \overline{M}, \quad (7)$$

3) оценивание $A_{j, j=1, \overline{M}}$ из СЛАУ

$$RA = z, \quad (8)$$

$$R_{ij} = e^{-\lambda_j x_i}, \quad z_i = F(x_i); \quad i=1, \overline{N}, \quad j=1, \overline{M}.$$

Как известно, интегро-дифференциальные методы были построены для анализа точных значений. Наличие погрешностей и конечная разрядность ЭВМ не позволяют непосредственно применить эти алгоритмы. В самом деле, в этом случае, например, СЛАУ из первого (4) и третьего (8) этапов становятся плохообусловленными и даже вырожденными, среди корней полинома (5) могут появляться комплексно-сопряженные, что приводит к необходимости разработки новых подходов. Поэтому в последние годы был создан ряд модификаций интегро-дифференциальных методов.

Вторая глава посвящена построению и обоснованию устойчивого алгоритма декомпозиции экспоненциальных зависимостей для данных, измеренных на неравномерных сетках.

Уточним постановку задачи. Принято называть G - множеством формальных решений задачи анализа экспоненциальных зависимостей

$$G = \{ \langle M_\epsilon, \vec{\lambda}_\epsilon, \vec{A}_\epsilon \rangle : S(F(x), F_\epsilon(x)) \leq \epsilon; F(x), F_\epsilon(x) \in \mathcal{F}_1 \}, \quad (9)$$

$\langle M'_\epsilon, \vec{\lambda}'_\epsilon, \vec{A}'_\epsilon \rangle$ - ее главным квазирешением:

$$\langle M'_\epsilon, \vec{\lambda}'_\epsilon, \vec{A}'_\epsilon \rangle = \operatorname{arg\,min} \left\{ \min_{M_\epsilon} \left\{ \min_{\vec{\lambda}_\epsilon, \vec{A}_\epsilon} S(F(x), F_\epsilon(x)) \right\} \right\} \quad (10)$$

Меру близости главного квазирешения к точному решению определим так:

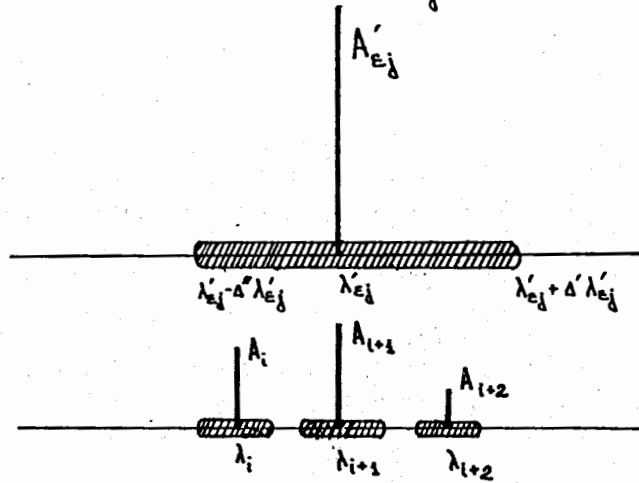
$$S_{\lambda, A}^2 = S_\lambda^2 + S_A^2, \quad \text{где} \quad (II)$$

$$S_\lambda^2(\vec{\lambda}_\epsilon, \vec{\lambda}) = \sum_{j=1}^{M_\epsilon} \sum_{i \in I'_j} |\lambda'_{\epsilon j} - \lambda_i|$$

$$I'_j = \{ i : |\lambda'_{\epsilon j} - \lambda_i| \leq \Delta \lambda'_{\epsilon j} \}$$

$$S_A^2(\vec{A}_\epsilon, \vec{A}) = \sum_{j=1}^{M_\epsilon} |A_{\epsilon j} - \sum_{i \in I'_j} A_i|,$$

$\Delta \lambda'_{\epsilon j}$ - заданные диапазоны изменения $\lambda'_{\epsilon j}$:



Для данных, измеренных на равномерных сетках, известна устойчивая к возмущениям модификация алгоритма типа Прони решения задачи (I), (2), (9) - (II).

В главе нами предложен алгоритм автоматизированной декомпозиции экспоненциальных зависимостей для данных, определенных на неравномерных сетках, позволяющий находить решение задачи в виде устойчивого приближения к главному квазирешению.

Основные шаги алгоритма заключаются в следующем:

1. Определяется \bar{M} - нижняя оценка количества компонент в экспоненциальной модели из вычисления ранга упорядоченной системы функций

$$S = \{1, x, x^2, \dots, x^{M'-1}, I_1, I_2, \dots, I_{M'}\} \quad (I2)$$

$$I_j(x) = \int_0^x dx_{[1]} \int_0^{x_{[1]}} dx_{[2]} \dots \int_0^{x_{[j-1]}} F(x_{[j]}) dx_{[j]}, \quad j = \overline{1, M'}; \quad M' \geq M.$$

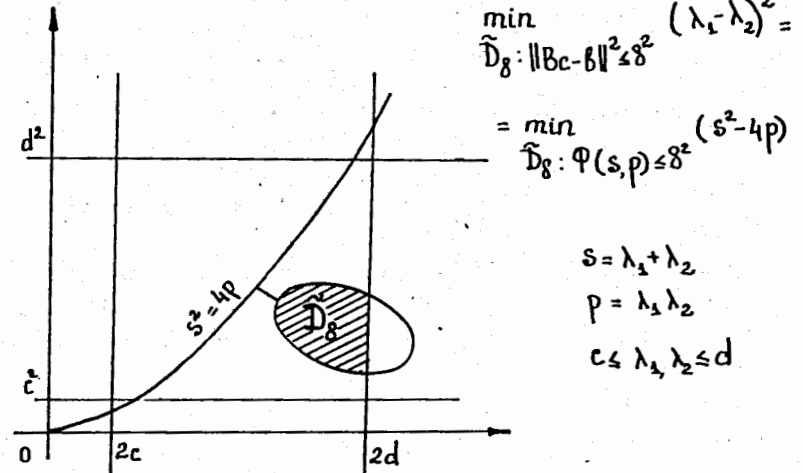
2. Из (4), (5), (7), (8) уже для хорошо обусловленных СЛАУ определяются параметры разложения $\{\lambda_j, A_j, j = \overline{1, \bar{M}}\}$ и погрешности их оценивания $\{\Delta \lambda_j, \Delta A_j, j = \overline{1, \bar{M}}\}$.

Построенная модель подвергается проверке на соответствие экспериментальным данным. Если согласие с экспериментом достигнуто, то решение задачи считаем завершенным, иначе переходим к следующему шагу.

3. Для последующего уточнения модели (2) анализируются корни полинома (5) - показатели экспоненциальной функции с учетом погрешностей их определения:

- показатели, не различимые в пределах полученных диапазонов, объединяются;
- в рамках специально разработанного устойчивого одношагового метода "изолированные" корни, чьи диапазоны не пересекаются с соседними, разделяются на два-три близких.

При таком анализе поиск компонент, максимально близких друг к другу в пределах погрешностей измерения, сводится к решению задачи на условный экстремум:



В работе решение этой задачи получено в явном виде.

4. Размерность экспоненциальной модели увеличивается на 1. При этом разделению на две составляющих подлечит та компонента, разбиение которой вызывает максимальное уменьшение невязки СЛАУ (8). Далее повторяется пункт 2.

В рамках обоснования предложенного алгоритма в работе доказаны:

1. Теорема. Пусть в каждой точке измерения данных известны значения экспоненциальной функции $F(x)$ и M' - оценка сверху количества ее компонент. Тогда $\bar{M} = \text{rang } S - M'$, где $\text{rang } S$ - ранг упорядоченной системы функций (I2).

2. Теорема. Функционал невязки СЛАУ (4) $\Delta = \Phi(s, p)$ имеет вид эллиптического параболоида в пространстве (s, p, Δ) .

Третья глава посвящена описанию общего пакета прикладных программ для анализа экспоненциальных зависимостей для персональных компьютеров профессионального типа класса ЕС-1840, ЕС-1841, ПРАВЕЦ-16, IBM PC AT/XT. Соответствующая библиотека реализована также на ЭВМ БЭСМ-6.

В общем ППП на ПЭВМ реализованы два основных способа решения задачи: анализ данных в рамках интегро-дифференциальных подходов с привлечением устойчивых алгоритмов типа Прони, описанных в предыдущей главе модификаций типа Фосса, Мура, и в рамках *Grid*-метода.

Приведем основные шаги анализа экспоненциальных зависимостей и соответствующие им вычислительные модули в ППП.

Первый подход включает следующие этапы:

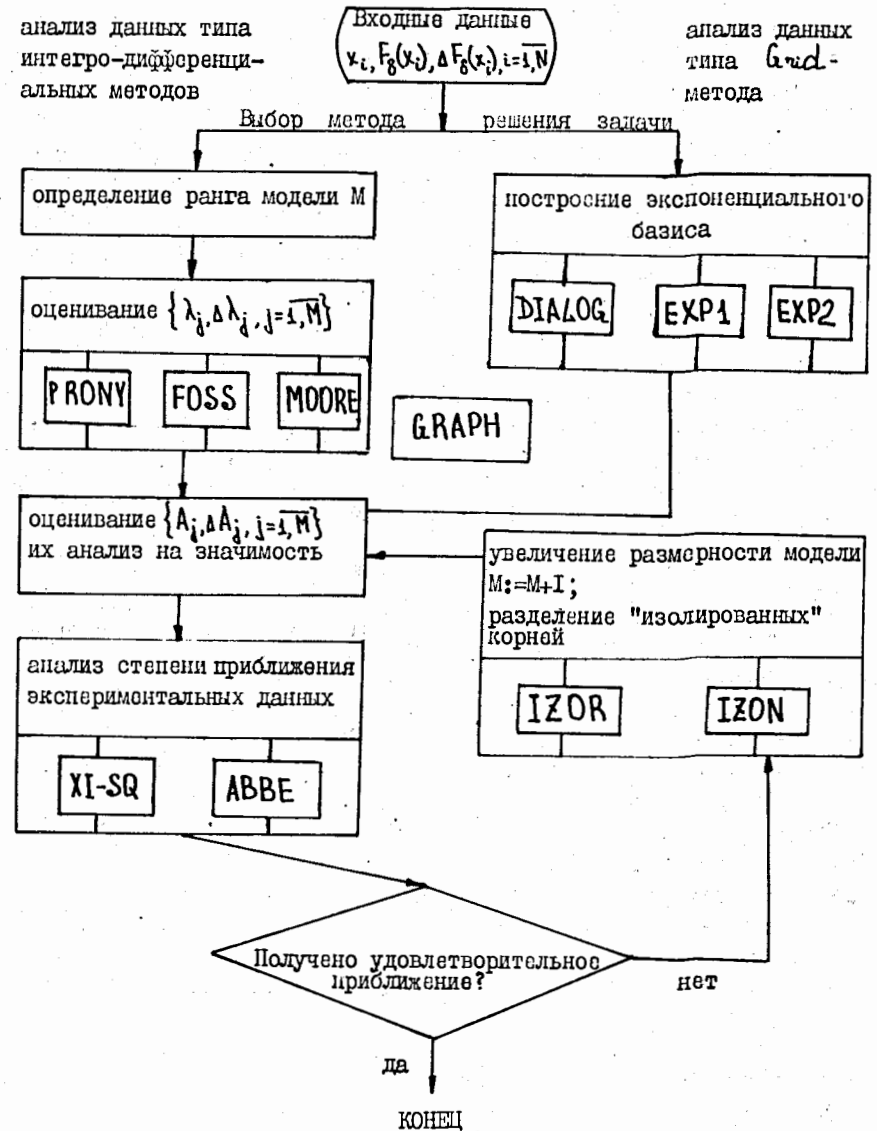
- 1) определение ранга модели M на основе анализа сингулярного разложения матриц данных,
- 2) оценивание $\{\lambda_j, j=1, \overline{M}\}$ с помощью модулей
 - PRONY** - анализ данных, измеренных на равномерных сетках, с помощью метода типа Прони,
 - FOSS** - анализ данных, определенных на неравномерных сетках, на основе описанного в главе 2 модифицированного алгоритма Фосса,
 - MOORE** - анализ небольшого количества компонент с помощью метода Мура,
- 3) оценивание $\{A_j, j=1, \overline{M}\}$ и их анализ на значимость,
- 4) оценивание меры соответствия экспоненциальной кривой (2) с найденными значениями показателей экспериментальным данным на основе модулей
 - XI-SQ** - анализ кривой взвешенных остатков с помощью критерия χ^2 ,
 - ABBE** - анализ кривой взвешенных остатков с помощью критерия Аббе,
- 5) переход к модели большей размерности с использованием программ
 - IZOR** - решение задачи разделения близких компонент для данных, измеренных на равномерных сетках,
 - IZON** - анализ данных, определенных на неравномерных сетках.

Во втором подходе линейные параметры $\{A_j, j=1, \overline{M}\}$ определяются из СЛАУ (8) по заданному набору показателей экспонент. Имеются три основных модуля для выбора или построения экспоненциального базиса, включенные в общий ППП:

- DIALOG** - значения показателей запрашиваются у пользователя в режиме диалога,
- EXP1, EXP2** - программы для вычисления экспоненциального базиса с равномерным или неравномерным шагом по λ .

Для удобства работы пользователей создана и включена в общий ППП специальная графическая программа **"GRAPH"** для визуализированного

Схема общего ППП для решения задачи декомпозиции экспоненциальных зависимостей



представления данных как на этапе определения ранга модели, так и для проверки соответствия экспоненциальной функции с найденными значениями параметров экспериментальным данным.

На приведенной схеме изображены основные возможности общего ППП для решения задачи анализа экспоненциальных зависимостей.

На основе этого ППП создана его версия "ОНКО-1" для анализа экспоненциальных зависимостей в задачах изучения выведения боросодержащих веществ из опухолей и непораженных тканей организма и др., эксплуатирующаяся во Всесоюзном онкологическом научном центре. Работа "ОНКО-1" обсуждается в заключительном параграфе третьей главы.

В четвертой главе рассмотрена типовая задача декомпозиции спектротрических данных. Уточним постановку задачи. Включим в представление (1) компоненту $b(x)$ - гладкую базовую кривую, характеризующую фоновую подложку спектра. Полагаем также, что полуширины всех линий одинаковы:

$$A_{21} = A_{22} = \dots = A_{2n}.$$

Как известно, характерные задачи анализа и интерпретации спектров включают:

- 1) выделение фоновой подложки,
- 2) разделение области задания спектра на участки, содержащие одиночные линии (синглеты) и мультиплеты,
- 3) оценку параметров синглетов (полуширин, положений и интенсивностей),
- 4) анализ мультиплетности пиков,
- 5) анализ качества приближения экспериментальных данных.

В работе описывается созданный нами базовый ППП "SPECTR" для декомпозиции спектротрических данных, рассчитанный на эксплуатацию на ПЭВМ. В ППП программы хранятся в виде вычислительных процессов (оттранслированных процедур). Связь между модулями в ППП осуществляется через файлы данных. Использование специальных инструментальных средств - Персональной Инструментальной Компьютерной Системы (ПИКС) - позволило нам реализовать в пакете многовариантные стратегии математической обработки данных.

Структура пакета и его гибкие альтернативные возможности математической обработки данных вместе с рекомендациями эксперта по их реализации в формализованном виде хранятся в ПИКС. Эту схему пользователь может рассматривать лишь как рекомендации и в зависимости от характера конкретных данных менять не только параметры, но и, не прибегая к перепрограммированию, составить свои индивидуальные схемы обработки.

В работе описана предложенная нами многовариантная вычислительная схема предварительной декомпозиции, соответствующая этапам 2, 3 решения задачи анализа спектров.

Автором разработано информационно-экспертное сопровождение ППП для предварительной декомпозиции спектров, включающее в себя:

1) описания вычислительных модулей в рамках правил, покрывающих ЕСИД,

2) описания данных с учетом императивных (связей входа-выхода относительно вычислительных модулей) и семантических (эквивалентности в установленном смысле) отношений,

3) описание пакета, включающее руководство пользователю по обработке данных на каждом этапе и советы по эксплуатации отдельных модулей.

ППП для анализа экспоненциальных зависимостей, описанный в главе 3, также включен в базовый ППП для обработки спектротрических данных на четвертом этапе.

В составе "SPECTR" имеются известные ППП:

- 1) "SMOOTH" - для построения гладкого непараметрического приближения одномерных функций (для выделения фона),
- 2) "RESOLV" - пакет линейной регуляризованной фильтрации для анализа мультиплетности пиков.

Для удобства пользователей нами разработаны специальные графические средства для визуализованного представления и редактирования спектров.

Для машинного моделирования, тестирования программ и установок, в учебных целях в базовый ППП включен пакет "CURSIM" для моделирования квазиреальных данных.

Для анализа рентгеновских спектров и спектров нейтронов в экспериментах ОИЯИ созданы и эксплуатируются специальные версии "SPECTR-1" и "SPECTR-2", работа которых обсуждается в заключительных параграфах этой главы.

Выводы

В результате проведенных исследований и разработок на персональных профессиональных компьютерах класса ЕС-1840, ЕС-1841, ПРАВЕЦ-16, IBM PC AT/XT созданы эффективные программно-алгоритмические средства для широкого класса экспериментов в спектротрии и анализа экспоненциальных зависимостей.

Основные результаты работы состоят в следующем:

1. Для данных, измеренных на неравномерных сетках, построен и обоснован устойчивый итерационный алгоритм восстановления параметров экспоненциального представления функций в рамках поиска приближения к главному квазирешению задачи на основе интегро-дифференциальных методов анализа экспоненциальных зависимостей. Для решения задачи разделения близких компонент предложены устойчивый одношаговый метод.

2. Разработана многовариантная вычислительная схема для решения характерных задач предварительной и полной декомпозиции спектрометрических данных.

3. Разработанные методы, программное и информационное обеспечение реализованы в пакетах прикладных программ для анализа экспоненциальных зависимостей и решения декомпозиционных спектрометрических задач на персональных компьютерах профессионального типа класса ЕС-1840, ЕС-1841, ПРАВЕЦ-16, IBM PC AT/XT. Особенностью пакетов являются гибкая структура и наличие информационно-экспертного сопровождения, разработанного на основе понятия предметной области.

4. Созданы и эксплуатируются версии ШП1 для анализа рентгеновских спектров в задаче определения микроэлементного состава веществ в ЛНФ ОИЯИ, для анализа спектров нейтронов в ЛЯП ОИЯИ и др.

Основные результаты диссертации опубликованы в работах:

1. Колтун И.А., Хорошилова Е.В. Некоторые алгоритмы полной математической обработки спектрометрических данных // Современные проблемы математического моделирования. М.: изд-во МГУ, 1984, с. 13-19.

2. Колтун И.А. О разделении двух компонент в экспоненциальном представлении функций для данных на неравномерных сетках. М.: деп. в ВНИИТИ, 1987, № 9001-В87, 19 с.

3. Ардышев С.Н., Колтун И.А. Информационное и методическое обеспечение процедур предварительной декомпозиции спектрометрических данных в диалоговой системе ШЛКС. М.: деп. в ВНИИТИ, 1987, № 9088-В87, 42 с.

Рукопись поступила в издательский отдел
26 сентября 1988 года.