



**ОБЪЕДИНЕННЫЙ ИНСТИТУТ ЯДЕРНЫХ ИССЛЕДОВАНИЙ**

10-98-297

На правах рукописи  
УДК 519.246 + 681.3.06

3-681

**ЗЛОКАЗОВ**  
Виктор Борисович

**МАТЕМАТИЧЕСКИЕ МЕТОДЫ  
И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ  
ДЛЯ КОМПЬЮТЕРНОГО АНАЛИЗА  
СПЕКТРОПОДОБНЫХ РАСПРЕДЕЛЕНИЙ**

Специальность: 05.13.16 — применение  
вычислительной техники, математического моделирования  
и математических методов для научных исследований

Автореферат диссертации на соискание ученой степени  
доктора физико-математических наук

Дубна 1998

Работа выполнена в Лаборатории вычислительной техники и автоматизации Объединенного института ядерных исследований, г.Дубна.

Официальные оппоненты:

Доктор физико-математических наук,  
профессор

И.М.ИВАНЧЕНКО

Доктор физико-математических наук,  
профессор

А.И.ЧУЛИЧКОВ

Доктор физико-математических наук

В.И.ЛУЩИКОВ

Ведущая научно-исследовательская организация: Российский Научный Центр "Курчатовский Институт", г.Москва.

Автореферат разослан "\_\_\_\_" \_\_\_\_\_ 1998г.

Защита диссертации состоится "\_\_\_\_" \_\_\_\_\_ 1998г. в "\_\_\_\_" часов на заседании Диссертационного совета Д 047.01.04 при Лаборатории вычислительной техники и автоматизации по адресу: г.Дубна, Московской области, ОИЯИ ЛВТА.

С диссертацией можно ознакомиться в библиотеке ОИЯИ.

Ученый секретарь Диссертационного совета  
кандидат физико-математических наук З.М.Иванченко

Диссертация посвящена как разработке методов и алгоритмов решения достаточно полной совокупности типовых задач, возникающих в процедурах анализа аппаратурных одномерных и многомерных спектров ядерных и молекулярных излучений, и программной реализации этих методов и алгоритмов в автоматическом и интерактивном режимах, так и математическому осмыслению этих задач, их систематизации и изложению адекватным формальным языком, математическому обоснованию методов их решения и обобщению на случай распределений, формально эквивалентных физическим спектрам; адаптации созданных программных комплексов к реальным физическим экспериментам, непрерывному развитию этих комплексов, теоретическому и практическому изучению вопросов тестирования алгоритмов анализа данных, вопросам точности и надежности такого анализа.

#### Актуальность проблемы.

Бурный прогресс техники автоматической регистрации информации дал возможность одновременно продемонстрировать как большую мощь сложившегося математического аппарата анализа данных, так и его упрощенность, недостаточность и неадекватность колоссальному разнообразию регистрируемых данных и задач их анализа.

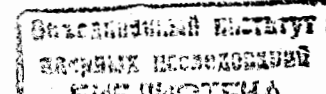
Измерение дает нам функцию  $f(x)$ . Если анализ таких функций осуществляется в задачах, семантически до конца не понимаемых наукой, то сразу же выясняется вся сложность и неоднозначность таких понятий как информативность функции, ее внутренняя структура и т.д.

Разумеется, огромные объемы данных, представляемых функциями  $f(x)$  и то обстоятельство, что  $f(x)$  может описываться огромным количеством чисел, лишь усугубляют упомянутую сложность.

В традиционных подходах анализа данных, строго говоря, не было проблемы определения понятия информационной структуры функции или распределения. Обычно считалось, что если некоторая функция  $f(x)$  задана, то тем самым задана и осознана ее внутренняя структура.

Математическая статистика тоже понимает функцию распределения как целостность, заданную с точностью до некоторых признаков, описываемых параметрами, -  $f(x, p)$ , и свою задачу видит в том, чтобы оптимальным в том или ином смысле способом оценить неизвестные  $p$  или, если  $p$  заданы, проверить гипотезу об адекватности  $f(x, p)$  данным. Вопросы "Как строить модель  $f(x, p)$ , что параметризовать и как параметризовать?" как правило не обсуждаются.

Но и новые разделы математики, такие как автоматическая классификация или автоматическое распознавание образов, формулируют свои задачи



как задачи построения отображений классов заданных (а тем самым и явно описанных) объектов классам явных же дескрипторов.

Однако в современной практике анализа, данная задача часто ставится с максимальной степенью неопределенности и некорректности:

задана  $f(x)$ ; требуется выявить ее информационную структуру и помочь тем самым экспериментатору не просто получить ответ на интересующие его вопросы, но и поставить эти вопросы.

Типичным и важнейшим примером такой ситуации является безусловно спектрометрия, или, говоря более обще, методики измерения выходов продуктов физических реакций. Их можно описать следующим образом.

Ядерные реакции в принципе не могут наблюдаться визуально. Поступают так. Пучок частиц источника (заряженных частиц, нейтронов, фотонов и т.д.) попадает на мишень, вступает с ее веществом в общем случае в многоканальное и каскадное взаимодействие, порождающее множество вторичных, третичных и т.д. частиц (осколков ядер, заряженных частиц, нейтронов, гамма-квантов, продуктов атомных излучений и т.д.). Регистрирующий детектор представляет собой ловушку для этих частиц и способен при достаточном усложнении своего устройства осуществлять ограниченную селекцию этих частиц по типам. "Пойманные" частицы вступают опять-таки в многоканальное и каскадное взаимодействие с веществом детектора и вместе с порожденными ими частицами-продуктами могут сообщить детектору некоторые свои совместные (т.е. в общем случае искаженные) характеристики: энергию (полную или какую-то часть) или долю уменьшения этой энергии, угол или время прилета и более или менее (чаще менее) точно измеренную длину пути пробега в веществе детектора.

Эти характеристики (обозначим их вектором  $X$ ) являются аргументом измерения. Число зарегистрированных частиц  $N$ , отнесенное к определенному значению  $X$ , образует аппаратное спектральное распределение  $N(X)$ . Истинное теоретическое распределение продуктов изучаемого канала реакции пучка и мишени (назовем его  $K(X)$ ) связано с  $N(X)$  очень длинной и запутанной цепочкой не всегда ясных качественно и известных количественно трансформаций. Но задача анализа  $N(X)$  состоит в том, чтобы, отталкиваясь от  $N(X)$ , получить как можно лучшее приближение к  $K(X)$ , ибо информацию для физика содержит только  $K(X)$ .

Таковы задачи анализа одно- и многомерных спектров первичного или вторичного гамма-излучения, оптических спектров и спектров рентгеновского излучения, сечений нейтронных взаимодействий и одно- и многомерных дифрактограмм, одно- и многомерных спектров распада и фрагмен-

тации ядер и многих других. Технический прогресс в области аппаратуры экспериментов создает новые, уникальные возможности для физической науки. Например, появление нейтронной дифрактометрии в реальном времени позволяет регистрировать временную зависимость  $N(X, t)$ , которое через цепь сложных зависимостей связано с динамикой  $K(X, t)$  истинного процесса. Естественно, задача анализа  $N(X, t)$  имеет больший порядок сложности и нетривиальности, чем обычно. Применение методологии ядерных измерений в смежных с физикой науках, а то и в удаленных от нее, часто означает еще большее усложнение задачи анализа спектров. Например, в последнее время распространена точка зрения, что изменения концентрации почвенного радона связаны с колебаниями напряженности земной коры, а тем самым с землетрясениями. Но разобраться в зарегистрированном распределении эманаций радона  $N(t)$  при пока еще недостаточно высоком уровне знаний о механизме землетрясений, а тем более увязать  $N(t)$  с землетрясениями хотя бы связью корреляционного типа - весьма непростая задача. Во всех таких задачах явственно ощущается недостаточность традиционного математического аппарата для извлечения информации из нетрадиционных данных.

Специфический характер математического анализа спектров долгое время не осознавался в литературе. Преобладали частные, эмпирические подходы, в которых задачи анализа спектров трактовались как задачи статистической оценки параметров наугад подобранных, более или менее удачно, функций регрессии. Однако резкий рост объемов получаемых спектральных данных и разнообразия их видов (появление многомерных данных, рентгено- и нейтронограмм поликристаллов, данных космической и аэрофотосъемки в инфракрасном диапазоне и т.д.), а также широкое распространение методологии спектроподобных (включающих спектры как частный случай) измерений (применение ядерных детекторов в геологии, биологии, и вообще в любой области, где измерение производит данные, которые необходимо анализировать с целью различения в них внутренней структуры - информативных частей или компонент) сделали очень актуальными следующие задачи:

1. формализация понятий структуры функции и компонент этой функции и создание методов ее алгоритмического распознавания, распознавания специфических образов - образов функций;
2. создание универсальных, прецизионных и в то же время достаточно быстрых, максимально независимых от условий получения данных методов решения задач выявления внутренней структуры наблюдаемых

функций;

3. "интеллектуализация" методик анализа спектроподобных данных как с помощью развития интерактивных возможностей программ, реализующих методы этого анализа, так и (что особенно актуально) с помощью широкого внедрения алгоритмов в управление методиками анализа - область, наименее поддающаяся формализации: автоматическое формирование гипотез и принятие решений, управление нелинейными процессами, решение задач распознавания и классификации;
4. программная реализация применяемых подходов в виде пакетов программ, максимально независимых от компьютеров и компьютерных систем, на которых они используются и рассчитанных на самый широкий класс решаемых задач и анализируемых данных;
5. изучение всех аспектов применения таких пакетов и в первую очередь вопросов точности и надежности анализа.

Методы ядерной спектроскопии и аналогичные им подходы продолжают в нарастающих масштабах охватывать все большие области научной и прикладной деятельности и это делает перечисленные задачи все более актуальными. Решение этих задач потребовало глубокого проникновения в математическую суть перечисленных проблем. Выяснилось, что задачи спектроскопии имеют множество особенностей, не укладывающихся в рамки классического математического анализа и статистики. Создание адекватного математического формализма, расширение области применения создаваемых в его рамках методов, создание математического аппарата "интеллектуализации" процессов решения задач анализа структур функций означали возникновение в современной отрасли науки - анализе данных - новой, самостоятельной и обширной области математической спектроскопии, которая по насыщенности математикой - от вычисления интегралов до кластер-анализа, от решения дифференциальных уравнений до методов топологии - относится к лидирующим областям прикладной математики. Современная программа обработки спектров так же отличается от первоначальных программ для этих же целей, как современный ускоритель от ускорителя 40-х годов. Создаваемые подчас в условиях жесткой конкуренции, эти программы, для того, чтобы отстоять свое право на жизнь, а тем более получить широкое признание, должны не только доказать свою эффективность в обработке данных реальных экспериментов, но и продемонстрировать высокий уровень и мощь на анализе международных тестовых материалов.

## Цели и задачи исследования.

Формализация задачи автоматического распознавания структуры произвольной функции (или распределения) в простейшем подходе может быть осуществлена следующим образом.

Пусть  $K_i, i = 1, 2, \dots, k$  - классы функций, связанных с выходами наблюдаемого процесса  $s(x)$  взаимно-независимыми ПРИЧИННЫМИ связями; пусть далее  $E$  - класс функций помехи измерения, возмущающих  $s(x)$ ; тогда  $Q$  - оператор измерения - есть отображение прямого произведения  $K_i$  и  $E$  на  $S$  - класс наблюдаемых функций:

$$\hat{Q}K_1 * K_2 * \dots * K_k * E \rightarrow S \quad (0.1)$$

и  $s(x) = \hat{Q}g(x) * e(x)$ , где  $g(x) \in K_1 * K_2 * \dots * K_k$ , а  $e(x) \in E$ .

Решение задачи анализа функции  $s(x)$  есть в общем случае не столько нахождение оператора, обратного к  $Q$ , сколько решение задачи выявления структуры  $s(x)$ , т.е., пусть  $M_i, i = 1, 2, \dots, n$  - классы функций, могущих быть связанными с выходами  $s(x)$  связями КОРРЕЛЯЦИОННОГО характера; следует определить эти классы и найти элемент  $m : m \in M_1 * M_2 * \dots * M_n$ , близкий к элементу  $g(x)$  в смысле какого-либо функционала (критерия близости)  $C$ .

Определение функционала  $C$  является частью и важным аспектом исходной задачи. Другая ее часть - это постулирование определений фундаментальных информационно-емких с содержательной точки зрения компонент, из которых складывается структура функции или распределения  $s(x)$ .

Эти определения могут быть сделаны только на основе обобщения большого опыта содержательного анализа данных в конкретных физических экспериментах и с учетом основных идей и концепций аналитической математики. И, наконец, самую обширную часть задачи составляют методы, алгоритмы и программы ее решения и анализ качества результатов их работы.

Сказанное определяет цели, стоявшие перед данной работой:

1. теоретическое исследование комплекса задач анализа спектров и спектроподобных распределений, создание адекватного математического формализма для их описания;
2. разработка и математическое обоснование универсальных методов и алгоритмов для решения этих задач и "интеллектуализация" этих методов и алгоритмов;

3. создание набора программных универсальных комплексов, реализующих эти методы и алгоритмы для решения самого широкого класса задач анализа спектроподобных распределений и максимально независимых от вида данных и используемых ЭВМ;
4. непрерывное поддержание и развитие этих комплексов, теоретическое и практическое изучение вопросов точности и надежности результатов их работы.

Для достижения этих целей необходимо было решить следующие конкретные задачи:

- исследование проблемы декомпозиции функций и гистограмм - основной операции по определению их информационной структуры - и создание универсальных и практически реализуемых методов решения этой проблемы;
- систематизация, изучение особенностей и оптимизация формализма операций с дискретными функциями - гистограммами, основной формой данных в спектрометрии;
- разработка концепций информации в спектрометрии и исследование математических вопросов, связанных с особенностями этих концепций;
- создание универсальных и практически реализуемых методов нелинейной параметризации и нелинейной аппроксимации функций;
- создание универсальных и практически реализуемых методов нелинейной фильтрации (пико- и периодосохраняющей) данных с резонансными трендами;
- создание методов обобщенной нелинейной подгонки функций, в частности, с помощью фильтрующей и/или робастной минимизации нелинейных функционалов;
- разработка широкого набора алгоритмов и приемов для "интеллектуализации" наиболее неформальных этапов анализа данных: построения гипотез, принятия решений и управления алгоритмами.

Научная новизна и значимость работы.

Хотя простые методы анализа спектров использовались давно, по крайней мере, с момента зарождения спектрометрии, точный анализ с применением сложных математических методов оказался возможным только лишь

после появления и широкого распространения ЭВМ и средств интерактивного взаимодействия с ними, и понадобился достаточный опыт компьютерного анализа спектров, чтобы перейти, не теряя в качестве, к полному автоматическому и "интеллектуализированному" этапам. Математическая спектрометрия, возникшая как результат автоматизации и компьютеризации ядерно-физических экспериментов, долгое время была и в значительной мере остается и ныне вне поля зрения академической математики.

В истории науки часто специфический характер задач способствовал зарождению новых математических подходов и нового формализма. Это в полной мере относится и к спектрометрии. Основная задача спектрометрии, которую на самый общий взгляд можно определить как задачу качественного и количественного определения структуры функций, имеет множество особенностей, не рассматривавшихся в классической математике.

Создание адекватного формализма потребовало строгой и корректной постановки задач, широкого осмысливания и переосмысливания существующих методов их решения и создания существенно новых методов.

Задача анализа спектроподобных распределений была сформулирована (повидимому, впервые в математической литературе) как задача декомпозиции сложных функций на формально недоопределенные компоненты.

Были разработаны формализм и универсальные методы решения задач декомпозиции, дающие принципиальные возможности решать все известные автору задачи анализа спектроподобных распределений.

Новыми явились методы, имеющие общематематическое значение: методы нелинейной параметризации функций, методы нелинейной аппроксимации функций на основе обобщения понятия аппроксиманта, методы обобщенной, в том числе робастной, подгонки функций, методы фильтрации нестационарных рядов с резонансными трендами, методы распознавания и кластеризации образов функций и множеств.

Значимость работ определяется прежде всего ролью и масштабами применения методологии спектроподобных измерений. Созданные автором математический формализм и аппарат представляет собой универсальное средство корректного решения самого широкого класса задач, связанных с этой методологией.

Общематематическая часть диссертации может быть с успехом применена и для решения иных математических задач, таких, которые используют для своего решения процедуры нелинейной минимизации, фильтрации и кластеризации.

Созданное программное обеспечение, адаптированное к реальным экспериментам и непрерывно развиваемое, является программным средством очень

широкого практического назначения.

### Практическая полезность работы.

Спектрометрическая и эквивалентные ей методологии имеют как научное, так и прикладное, в том числе и народнохозяйственное значение (активационный и элементный анализ, нейтронографические методы исследования твердого тела, методы ИК-спектрометрии и т.д.).

Программное обеспечение, в котором были реализованы созданные автором математические методы, входит в стандартные библиотеки программ всех основных ЭВМ ОИЯИ, использовалось и используется в очень большом числе научных и научно-прикладных работ, проводимых в рамках темплана ОИЯИ.

Основные программные комплексы входят в международную библиотеку физических программ при журнале Computer Physics Communications, а также в библиотеку программ обработки кристаллографических данных и в республиканский фонд алгоритмов. Они используются в большом числе научных центров как СССР, так и стран-участниц ОИЯИ.

Следующие основные результаты диссертации выносятся на защиту.

1. Изложение адекватным математическим языком задач анализа широкого класса данных - одномерных и многомерных экспериментальных спектров; задача анализа таких распределений была сформулирована как задача декомпозиции сложных функций на в общем случае формально недоопределенные компоненты; декомпозиция является весьма универсальной операцией решения широкого класса задач, в том числе и многих общематематических. Создание спектроориентированного математического формализма - аппарата декомпозиции функций и гистограмм и изучение его различных аспектов.
2. Математическое обоснование используемой в спектрометрии помимо статистической детерминистской интерпретации данных и методов их анализа, а также обоснование характерных для спектроподобных распределений трактовки понятий информации.
3. Создание универсальных методов для решения задач анализа спектроподобных распределений. Эти методы, там где возможно, были обобщены до уровня общематематических, поэтому самостоятельными результатами, выносимыми на защиту, являются -

- методы малопараметрической нелинейной аппроксимации функций;
- методы универсальной нелинейной параметризации функций и гистограмм
- методы обобщенной подгонки функций;
- методы робастной минимизации нелинейных функционалов;
- методы фильтрации нестационарных данных с резонансными трендами;
- методы кластеризации множеств и множеств функций, основанной на преобразованиях переменных;
- методы распознавания специфических образов - образов функций.
- методы декомпозиции дискретных функций и гистограмм на локализованные и периодические компоненты; методы выявления скрытых ангармонических периодичностей.

4. Создание программных комплексов UPEAK, DOMUS, ACTIV, DECAN+NTIME, SPEVA, ERTQU, FIMDA, MRJA, AUTOX и их модификаций для анализа спектров и спектроподобных распределений как в интерактивном, так и в полностью автоматическом режимах, адаптация их к реальным физическим экспериментам, непрерывное развитие этих комплексов.
5. Теоретическое и практическое изучение вопросов тестирования алгоритмов анализа спектроподобных распределений, вопросов надежности и точности такого анализа.

### Аппробация работы.

Материалы, вошедшие в диссертацию, докладывались на научных семинарах в ЛВТА, ЛНФ, ЛЯР ОИЯИ, ИЯФ МГУ, ИПМ АН СССР, ЦИРИ (Лейпциг, ГДР), НМИ (Берлин, ФРГ), на заседаниях рабочей группы по спектрометрии при САНИ АН СССР, на международных и всесоюзных конференциях (Ташкент 1974, Киев 1976, Дубна 1977, Дубна 1978, Алмата 1978, Краков 1978, Дубна 1980, Дубна 1982, Дубна 1984, Харьков 1985, Ташкент 1987, Лейпциг 1988, Мюнхен 1990, Берлин 1990, Лейпциг 1992, Вена 1993, Атланта (США) 1994, Дрезден 1994, Честер (Англия) 1995, Парма (Италия) 1997, Будапешт 1998.

### Публикации.

### Краткое содержание диссертации.

Во введении дана общая характеристика проблематики работы и кратко перечислены основные результаты.

Второй раздел посвящен описанию физического содержания задач анализа спектроподобных распределений. Перечислены основные черты и особенности таких задач, применяемые методы и алгоритмы их решения, указаны достоинства и недостатки таких методов и намечены цели и пути как построения адекватного задачам формализма, так и универсальной методики их решения.

В третьем разделе "Алгебра спектральных гистограмм" систематизирован, расширен и обобщен формализм дискретных функций: одно- и многомерных гистограмм, в форме которых анализируются спектроподобные распределения. Важнейшие задачи, возникающие здесь, следующие:

1. влияние дискретизации на точность представления функции гистограммами (задачи биннинга);
2. применение дискретного преобразования Фурье (ДПФ) к спектрограммам;
3. оптимизация дискретизации;
4. изучение свойств различных статистик от спектрограмм, подчиненных пуассоноподобным распределениям.

Измерение осуществляет гистограммирование - отображение теоретической непрерывной функции  $s(x)$  на пару множеств  $[X_k, x_k]$ ,  $k = 1, 2, \dots, n$ , где  $n$  число ячеек гистогрaммы ( $X_k$ ), а  $(x_k)$  - узлы этих ячеек.

Показано, что оптимальное гистограммирование будет даваться ячейками равного веса (минимум погрешности 1-го рода), а в качестве узлов следует брать точки, в которых функция  $s(x)$  равна своему интегральному (метрика  $L_2$ ) или амплитудному (метрика  $C$ ) среднему (минимум погрешности 2-го рода).

Рассмотрены вопросы о погрешности биннинга в ряде важнейших случаев конкретных функций  $s(x)$ : экспоненты, гауссиана, лоренциана. Для ангармонических периодических гистограмм  $g(i)$ ,  $i = 0, 1, \dots, m$ , доказана важная теорема: если  $T$  - период и  $m + 1 = TL$ , то ДПФ от  $g(i)$  будет периодической ангармонической гистограммой с периодом  $L$ . Эта теорема

может быть использована при поиске скрытых ангармонических периодичностей.

Важнейшую роль в описании спектрограмм играют пуассоноподобные распределения: распределения неотрицательных сумм независимых событий, когда средние этих распределений равны (точно или приближенно) их дисперсиям (например, биномиальное, пуассоновское и т.д.). Фундаментальную роль при анализе спектрограмм, подчиненных таким распределениям, играет следующая статистика

$$r = \frac{s_1 - s_2}{s_1 + s_2},$$

определенная для двух произвольных пуассоноподобных случайных величин  $s_1$  и  $s_2$ . Доказана следующая теорема, дающая возможность обоснованно использовать эту статистику в процедурах статистической проверки гипотез: если  $s_1$  и  $s_2$  суть суммы  $n$  событий, подчиненные одному и тому же пуассоноподобному распределению, то предельным распределением статистики  $r$  при  $n \rightarrow \infty$  будет нормальное распределение  $N(0, 1)$ . Эта статистика играет большую роль в различных процедурах идентификации информационных частей произвольного распределения.

Рассмотрен также ряд других аспектов операций с гистограммами.

Четвертый раздел посвящен формализации и анализу операции декомпозиции - основной операции при изучении экспериментальных распределений аддитивной структуры.

Задачи декомпозиции функций могут быть систематизированы в рамках следующего формализма. Пусть

$$f(x) = \sum_{i=1}^k f_i(x) + e(x), \quad (0.2)$$

где  $f_i$  и  $e$  - функции, взаимно линейно-независимые. Функция  $e$  выделяется особо и называется погрешностью функции  $f$ . О  $e$  предполагается известной только  $b(x)$  - верхняя грань значений ее модуля в каждой точке  $x$  (далее называемая нормой  $e(x)$  в каждой точке  $x$ ).

Пусть  $\rho(f_1, f_2)$  - метрика на  $A$ -пространстве рассматриваемых функций - и рассмотрим неотрицательный выпуклый функционал  $c$

$$c = c(\rho(f_i, g_j)) \quad (0.3)$$

где  $f_i \in A$ ,  $g_j \in A$ ,  $i, j = 1, 2, \dots, \max(k, n)$ , обращающийся в минимум только при  $k = n$ ,  $\rho(f_i, g_j) = 0$ .

Выражение  $c$  есть критерий групповой близости наборов элементов  $f_i$  и  $g_j$ .



**ОПРЕДЕЛЕНИЕ 1.** Качественным этапом декомпозиции называется отображение

$$f \rightarrow (n, [M_i]), \quad i = 1, 2, \dots, n$$

где  $M_i \in A$  - непересекающиеся классы функциональных элементов, такие, что  $f_i \in M_i$  для некоторых индексов  $i$ .

**ОПРЕДЕЛЕНИЕ 2.** Количественным этапом декомпозиции называется отображение

$$f \rightarrow (n, [g_i(x)]), \quad i = 1, 2, \dots, n,$$

где  $g_i \in M_i$ .

**ОПРЕДЕЛЕНИЕ 3.** Оптимальной декомпозицией элемента  $f$  на компоненты называется нахождение набора  $(g_j)$  такого, что

$$(g_i) = ARG \ MIN \ c(\rho(f_i, g_j))$$

Можно указать следующие модификации, исчерпывающие понятие декомпозиции: декомпозиция в сочетании с гистограммированием, факторная декомпозиция, классификационная (дискриминантная и кластеризационная), со связями и в мультикритериальной формулировке.

Наиболее важная и нетривиальная область действия процедуры декомпозиции:

распознавание аддитивной структуры функций, выделение в них компонент, важных для нас в информационном или классификационном смысле. Если обобщить понятие функциональной компоненты, выделяемой при визуальном анализе графиков функций, то можно формализовать следующие типы аддитивных компонент:

1. почти-сосредоточенные, т.е. функции, интегрируемые в заданной степени на всей прямой, т.е. такие, что для любого  $\epsilon$  и заданного положительного  $n$  существует конечная область  $A$ ,  $A \in R$ , такая, что

$$\left| \int_A |f(x)|^n dx - \int_R |f(x)|^n dx \right| < \epsilon;$$

2. почти периодические функции, т.е., такие, что для любого  $\epsilon \gg 0$  существует положительное число  $L$  такое, что в любом интервале длины  $L$  найдется хотя бы одно  $T$  такое, что при любом  $x$  имеет место

$$|f(x+T) - f(x)| < \epsilon;$$

3. распределенные компоненты, т.е. не интегрируемые на всей оси функции.

В произвольной функции могут присутствовать любые комбинации компонент перечисленного типа. Эти фундаментальные компоненты представляют собой наиболее универсальную типизацию элементов функций аддитивной структуры, которыми обычно описываются распределения, информационные с точки зрения экспериментальных наук, в частности, спектроподобные распределения.

Функции мультипликативной структуры приводятся к аддитивным с помощью логарифмического преобразования.

Процедура решения задачи расшифровки внутренней структуры произвольной функции сводится к декомпозиции этой функции на фундаментальные компоненты. При этом естественным образом выделяются следующие этапы.

1. Анализ модуля преобразования Фурье распределения, нахождение периодичностей гармонических или ангармонических.
2. Поиск и обнаружение почти-сосредоточенных компонент; при этом будут различаться случаи, когда носители этих компонент будут содержать неперекрывающиеся части (например, анализ пиков), и случаи, когда таких частей не будет (например, анализ экспонент);

Эти операции могут быть проделаны и над производными распределений (если информация заключена в динамике изменения их функций) или над другими их трансформациями, усиливающими особенности информационной структуры распределения.

В работе рассмотрены всевозможные теоретические вопросы и аспекты этой процедуры и описаны алгоритмы ее реализации.

В работе подробно изучены используемые методы решения задач декомпозиции функций - аппроксимационный и фильтрационный методы, их обе разновидности - параметрическая и непараметрическая.

Установлен ряд теорем (теоремы 1 - 8), выясняющих, какая параметризация и какое гистограммирование являются достаточным условием того, что параметрическая аппроксимация в квадратичной метрике даст минимальную погрешность декомпозиции.

Пятый раздел посвящен методам нелинейной селективной аппроксимации и параметризации функций и гистограмм. Важность этих вопросов определяется следующими соображениями:

1. параметрическая селективная аппроксимация - самый естественный и наиболее доступный для реализации аппарат решения задач декомпозиции и, в частности, декомпозиционной фильтрации;



2. для того, чтобы аппроксимант не только приближал функцию как целое, но и обеспечивал адекватность структур, своей и приближаемого объекта, надо, чтобы части аппроксимирующего выражения - субаппроксиманты - обладали важным свойством селективной аппроксимации - способностью аппроксимировать только "свои" части приближаемой функции и при этом не аппроксимировать (дискриминировать) другие части функции; это диктует необходимость новых подходов, не рассматриваемых классической теорией аппроксимации;

3. проблема общей параметризации функции является одной из важнейших, (хотя, похоже, самых необсуждаемых в литературе проблем прикладной математики) ввиду огромной роли параметрических методов в решении множества задач не только декомпозиции функций, но и из других областей.

Аксиоматически созданный метод селективной аппроксимации и параметризации можно изложить следующим образом. Вводится понятие обобщенного аппроксиманта: это тройка функций  $Q_k(x)$ ,  $m(x)$ ,  $P_n(x)$ , образующих выражение

$$Q_k(x) \cdot m(P_n(x)) \quad (0.4)$$

где  $Q_k$  и  $P_n$  классические аппроксиманты (полиномы алгебраические, дробно-рациональные, тригонометрические, сплайны и т.д.), зависящие от соответственно  $k$  и  $n$  параметров,  $m(x)$  - функтор селективной аппроксимации. При  $m(x) \equiv x$  и  $Q_k(x) \equiv 1$  мы получаем, как частный случай, классический аппроксимант  $P_n(x)$ . Выражение (0.4) естественным образом обобщается на случай многомерного аргумента  $x$ ; при этом под  $P_n(x)$  будет пониматься многомерный вектор, где каждая компонента представляет собой полиномиальное (например) преобразование пространства значений аргумента  $x$ ;  $n$ - степень - тоже будет величиной векторной.

Функтор  $m(x)$  (иначе, грубая модель) для приближаемой произвольной непрерывной (в общем случае многомерной) функции  $f(x)$  вводится так. Пусть задана непрерывная функция  $f(x)$ ,  $x \in X$ , и пусть  $P_n(x)$  и  $Q_k(x)$  являются функциями, способными аппроксимировать при  $n, k \rightarrow \infty$  произвольную непрерывную функцию.

**ОПРЕДЕЛЕНИЕ.** Непрерывную, отличную от константы функцию  $m(x)$  мы назовем грубой моделью функции  $f(x)$ , если существуют 2 такие непрерывные функции  $r(x)$  и  $h(x)$  (наборы  $[r(x), h(x)]$  в многомерном случае), что справедливо представление

$$f(x) = r(x) \cdot m(h(x)) \quad (0.5)$$

Имея грубую модель  $m(x)$  на основе (0.5) и набор аппроксимантов  $Q_k(x)$  и  $P_n(x)$ , мы можем искать приближение для функции  $f(x)$  с помощью (0.4):

$$f(x) \simeq Q_k(x) \cdot m(P_n(x))$$

Обоснованием такого подхода служат две следующие теоремы.

**ТЕОРЕМА 1.** Любая, не равная тождественно константе, аналитическая функция  $m(x)$ ,  $x \in X$ , является грубой моделью любой непрерывной функции, определенной на  $X$ .

**ТЕОРЕМА 2.** Если  $m(x)$  - грубая модель непрерывной функции  $f(x)$ , то выражение  $Q_k(x) \cdot m(P_n(x))$ , где  $Q_k(x)$  и  $P_n(x)$  - полиномы степеней  $k$  и  $n$  со свободными коэффициентами, будет при  $k \rightarrow \infty$  и  $n \rightarrow \infty$  аппроксимировать  $f(x)$  в метриках  $C$ ,  $L_1$  и  $L_2$ .

Теорема 2 естественным образом распространяется на случай произвольных аппроксимантов (помимо полиномов)  $Q_k$  и  $P_n$ .

Описанный формализм дает следующие преимущества по сравнению с классическим подходом:

1. легкость построения аппроксиманта (0.5) по сравнению с эмпирическим подбором приближающих функций произвольного класса;
2. небольшое число параметров (при правильном выборе модели  $m(x)$ ) и в то же время высокая точность аппроксимации;
3. избирательность аппроксимации при декомпозиции функций, т.е. способность добиться того, что при приближении суммы компонент функции суммой субаппроксимантов каждый субаппроксимант приближает лишь свою компоненту и дискриминирует остальные компоненты.

Выражение (0.4) вместе с тем задает универсальный способ нелинейной параметризации произвольных функций  $m(x)$ , заданных произвольным образом, в том числе и табличным, и графическим.

При этом вводимые параметры помимо упомянутых аппроксимирующих свойств, могут также быть объектами аналитических операций: минимизации, дифференцирования, интегрирования и т.д.

В шестом разделе описаны непараметрические методы нелинейной фильтрации данных с трендами резко-нестационарного типа (например, резонансного). Фильтры, которые строятся на основе этих методов, могут кратко быть охарактеризованы как пико- и периодосохраняющие.

Задачи, на которые ориентированы эти методы, могут быть типизированы следующим образом.

1. Основная задача - задача пико- и периодосохраняющего сглаживания (квадратичного или робастного) регрессии, функция математического ожидания которой по своему частотному составу слишком сильно перекрывается с помехой.
2. Задача декомпозиционной фильтрации, часто возникающая в процедурах опознавания образов функций, - задача выделения в сумме неотрицательных компонент одной из этих компонент, например, распределенной по всей области определения, или задача построения функции, огибающей регрессию снизу.

Классический аппарат фильтрации, например, линейные оптимальные фильтры, здесь неэффективен. Гораздо более мощными оказываются нелинейные фильтры вариационного типа, которые строятся следующим образом.

Для произвольной функции  $f(x)$  вводится понятие ее меры осцилляций. Поточечной (в точке  $x$ ):

$$\mu(f, x) = \frac{f''(x)}{\sqrt{1 + f'(x)^2}},$$

и интегральной ( на отрезке  $[a, b]$  ):

$$\mu(f, a, b) = \int_a^b \mu(f, x)^2 dx$$

Эти меры осцилляций могут быть обобщены на случай комплекснозначных функций. Кроме того в знаменателе поточечной меры можно вместо корня квадратного использовать и другие степени  $\beta$ .

Для построения огибающей снизу вместо меры осцилляций вводится понятие меры изменчивости функции (тоже поточечной и интегральной):

$$\mu(f, x) = f'(x), \quad \mu(f, a, b) = \int_a^b \mu(f, x)^2 dx$$

В многомерном случае поточечные варианты этих мер строятся как:

1. отношение квадрата суммы 2-х производных к элементу поверхности, задаваемой функцией  $f(x)$  (для 1-й задачи);
2. сумма квадратов 1-х производных (для 2-й задачи).

При различных  $\beta$  мера осцилляций  $f(x)$  будет приближенно-инвариантна относительно амплитуды  $f(x)$  ( $\beta = 1$ ) или ее ширины ( $\beta = 2$ ).

Далее выбирается метрика, в которой будет оцениваться близость отфильтрованной компоненты  $g$  к исходным данным  $f$ ; обозначим ее  $\rho(f, g)$ .

Для решения первой задачи, т.е. построения сглаживающего пикосохраняющего фильтра, это будет обычная квадратичная или робастная метрика. Для решения второй задачи, т.е. построения огибающей функции снизу, это будет взвешенная квадратичная метрика, где веса обратно пропорциональны амплитуде функции (или линейным комбинациям амплитуды и 2-й производной и т.д.)

После этого искомое решение задачи - отфильтровать из данных  $f(x)$  компоненту  $g(x)$  - будет решением следующей вариационной задачи: найти

$$\text{MIN} \mu(f, a, b) + \lambda \cdot \rho(f, g)$$

при граничных условиях

$$f(a) = g(a) \quad f(b) = g(b) \quad f''(a) = 0 \quad f''(b) = 0$$

или

$$f(a) = g(a) \quad f(b) = g(b) \quad f'(a) = 0 \quad f'(b) = 0$$

Периодосохраняющий фильтр строится точно так же, только вместо  $f(x)$  берется ее преобразование Фурье, а мера  $\mu$  берется комплекснозначная.

К результату комплексной фильтрации применяется обратное преобразование Фурье, дающее сглаженную  $f(x)$ , у которой будут сохранены периоды. Численное решение этих задач осуществляется следующим образом.

1-ая производная  $f'$  в знаменателе меры осцилляций заменяется на  $f_0'$ , где  $f_0'$  - априорная оценка сглаженной 1-й производной, в качестве которой можно взять, например, 1-ю производную от  $g$ , сглаженную с помощью некоторого простого фильтра. После этого обе задачи становятся линейными и могут быть решены с помощью метода прогонки.

Коэффициент  $\lambda$  является заданным. Он регулирует степень сглаживания и определяется с помощью тестов.

Описанный глобальный фильтр может быть упрощен; в частности, на основе его идей можно построить локальные (для 5 или 7 точек) фильтры, которые ценой некоторой потери в качестве фильтрации работают быстрее. На многомерный случай данные фильтры могут быть распространены следующим образом:

1. сглаживание по каждой координате одномерных сечений каждой функции;
2. построение локальных многомерных фильтров.

Фундаментальным преимуществом описанных фильтров перед обычными фильтрами является то, что они, эффективно подавляя шумовую компоненту данных, в максимальной степени сохраняют пикообразные компоненты

(по амплитуде или ширине).

В седьмом разделе рассмотрены методы обобщенной аппроксимации функций, а также робастной подгонки, являющейся частным случаем обобщенной.

Метод обобщенной подгонки наиболее целесообразно применять для решения задач декомпозиции или декомпозиционной фильтрации распределений. Пусть решается задача параметрической аппроксимации со связями: найти

$$\text{MIN} \sum_{j=1}^m w(x_j)(y(x_j) - f(x_j, P))^2 \quad (0.6)$$

при связях

$$r(f(x, P)) \geq 0 \quad (0.7)$$

Эти связи определяют аддитивную компоненту функции  $y(x)$ , подлежащую выделению и могут задаваться весьма неточным, качественным образом: например, пусть известна априорная оценка решения  $f(x, P_0)$  и предполагается, что  $r$  в (0.7) есть оператор либо малых в метрическом смысле деформаций  $f(x, P_0)$ , либо простых, легко описываемых синтаксически, трансформаций  $f(x, P_0)$  (сдвигов, уширений и т.д.)

Мы можем записать такое представление

$$f(x, P) = f_1(x, P) + f_2(x, P),$$

где функция  $f_1(x, P)$  удовлетворяет связям (0.7), а  $f_2(x, P)$  есть разность функций  $f$  и  $f_1$ .

При таком подходе исходная задача оказывается задачей декомпозиционной фильтрации: выделить эффективным способом из  $y(x)$  компоненту  $f_1$ .

Итак, рассматривается задача: дана

$$y(x) = f_1(x) + f_2(x) + e(x),$$

$y(x) \in K_1$ ,  $f_1(x) \in K_2$ ,  $e(x)$  - погрешность,  $K_1, K_2$  - классы функций; требуется: построить по  $y(x)$  на классе функций  $K_3$   $g(x)$  - оценку функции  $f_1(x)$  - так, чтобы

$$g(x) = \text{ARG MIN} \|g(x) - f_1(x)\| \quad (0.8)$$

Задача допускает эффективное решение, в частности, при выполнении следующих условий:

1. существует параметрическое семейство  $K_3 = [g(x, P), P \in R^n]$ ,  $P$  - параметр, такое, что  $K_2$  принадлежит  $K_3$ , но  $y$  не принадлежит  $K_3$ ;
2. можно построить метрику  $\rho$  на  $K_3$  и  $K_1$  так, что имеет место

$$\text{MIN} \rho(y, g) \rightarrow \text{MIN} \|f_1 - g\|.$$

Рассмотрим задачу: для дискретного множества  $X$  найти

$$\text{MIN} \sum_{x \in X}^w (x)[y(x) - f(x)]^2 \quad (0.9)$$

при связи

$$g(f(x)) = 0. \quad (0.10)$$

Здесь  $w, y, f, g$  - функции из некоторых классов. Пусть  $K_1$  - класс функций  $[f_1]$ , для которых имеет место (0.10), и пусть  $K_2$  - класс функций  $[f_2]$  таких, что  $y = f_1 + f_2$ .

Определим скалярное произведение функций  $u$  и  $g$ :

$$(g, u) = \sum_{x \in X}^g (x)u(x).$$

Имеет место следующая

**ТЕОРЕМА 3.** Если  $w(x)$  такова, что

1.  $w$  ортогональна любой  $f_2 \in K_2$ ;
2. из  $(w, f_1 - f) = 0$  при любой  $f_1 \in K_1$  следует  $f = f_1$ , то глобальный минимум (0.9) по  $f \in (K_1 \cup K_2)$  будет давать элемент  $f = f_1$ .

Если  $y(x)$  задана с некоторой погрешностью  $e(x)$ , то легко видеть, что теорема распространяется на этот случай со следующей оговоркой: глобальный минимум (0.9) будет достигаться на элементе  $t(x)$ , который будет мнк-оценкой  $f_1$ .

Условия теоремы будут выполняться для таких важных случаев фильтрации:

1. распределенной компоненты из ее суммы с локализованными, если объединение интегральных носителей последних не покрывает всю область определения первой;
2. локализованной из ее суммы с другими локализованными, если объединение интегральных носителей последних не покрывает носитель первой.

Итак, будем считать, что интегральным носителем компоненты  $f_1(x)$  в (0.8) является множество  $A$ ,  $x \in A$ , а компоненты  $f_2(x)$  - множество  $B$ ,  $x \in B$ , причем  $A$  содержит  $B$ , а  $A - B = 0$ .

Параметрическое семейство  $K_1$  - семейство оценок компоненты  $f_1$  - мы можем построить на основе метода аппроксимации, изложенном выше:

$$K_1 = [Q_k(x)m(P_n(x))],$$

где  $m$  - грубая модель функции  $f_1$ , а  $Q_k$  и  $P_n$  - полиномы степеней  $k$  и  $n$ . Эту модель и степени полиномов мы можем выбрать так, чтобы  $K_1$  содержало  $f_1$ , но не содержало  $y(x)$ .

Обозначим произвольный элемент из  $K_1$  как  $f(x, P)$ , где  $P$  - вектор параметров размерности  $k+n$ . Искомая метрика строится так: пусть задан вектор  $P_0$  - вектор априорных оценок параметров  $P$  и дискретное множество  $X$ ,  $X \in A$ . Построим квадратичное выражение

$$\sum_{x \in X} w(x, P_0) [y(x) - f(x, P)]^2 \quad (0.11)$$

где

$$w(x, P_0) = \begin{cases} 1/\|e(x)\|^2 & \text{если } |h(x)| < c; \\ (1 + \beta)/(\|e(x)\|^2((h(x)/c)^2 + \beta)) & \text{иначе;} \end{cases} \quad (0.12)$$

$h(x) = y(x) - f(x, P_0)$ ,  $c$  и  $\beta$  - заданные константы. Минимизация (0.11) дает  $f(x, P^*)$  - оценку  $f_1(x)$ .

Если фильтруемая компонента должна удовлетворять связи

$$z(f_1) = 0$$

причем на  $A - B$  связь имеет место, то условие  $|h(x)| < c$  в (0.12) при формировании весов  $w(x, P_0)$  заменяется на такие условия:

$$\begin{aligned} |h(x)| < c \\ z(f(x, P)) = 0 \end{aligned}$$

Построение робастного решения задачи подгонки данных является частным случаем описанного формализма, так как полную погрешность можно записать как сумму  $f_2(x) + e(x)$ , где под  $e$  понимается нормальная погрешность, а под  $f_2$  - "выброс". Тогда построение робастного решения сводится к построению оценки компоненты  $f_1$  с помощью вышеописанных методов. Если "выбросы" имеют место лишь в отдельных группах точек, то условия вышеприведенной теоремы выполнены и минимум (0.11) действительно даст робастную оценку компоненты  $f_1$ .

Пусть  $M$  - матрица Грамма, соответствующая истинным весам, а  $D$  - матрица добавок за счет пересчета весов до матрицы Грамма задачи (0.11) и пусть  $P_0$  - вектор априорных оценок параметров  $P$  и выполняются условия:

1. априорная оценка  $P_0$  и константа  $c$  выбраны настолько хорошо, что "выбросы" и нормальные данные идентифицированы правильно;
2. матрица  $M$  достаточно хорошо обусловлена, норма матрицы поправок к весам (0.12) достаточно мала и константа  $\beta$  выбрана удовлетворительно, так что минимальное собственное значение матрицы  $M$  больше максимального собственного значения матрицы добавок.

Тогда ответ на вопрос о преимуществах робастной мнк-оценки перед неробастной дает нижеследующая

**ТЕОРЕМА.** При выполнении условий 1,2 норма погрешности робастной мнк-оценки строго меньше нормы погрешности неробастной мнк-оценки.

В общем случае качество как робастной, так и обобщенной подгонки можно проанализировать с помощью тестов.

Тесты показывают, что при выполнении ряда условий, процедура будет давать требуемое решение с приемлемой точностью. Эти условия такие:

1. достаточно высокое качество априорного решения; в частности, оно само должно удовлетворять связи (0.10);
2. достаточная надежность идентификации точек  $x$ , где весовая функция должна пересчитываться; это зависит от правильности выбора константы  $c$ ;
3. правильный выбор константы  $\beta$ ; уже при  $\beta = 1(c = 1)$  решение может не удовлетворять связи (0.10), но  $\beta = 0$  может увести итерационный процесс к ложным решениям, могущим появиться как побочное следствие пересчета весовой функции на слишком большом множестве точек  $x$ ;
4. правильность задания весовой функции  $w(x)$ ; в частности, правильный масштаб представления данных и согласованность  $w(x)$  с данными.

Установлен ряд теорем и рассмотрены различные вычислительные аспекты практического применения методики обобщенной подгонки - демпфирования матрицы итерационного процесса и построения некоторых теоретических проекторов при решении задачи минимизации с ограничениями, когда не может быть применен обобщенный подход.

8-й раздел посвящен изучению методов параметрической декомпозиции - важнейших при практическом анализе экспериментальных распределений аддитивной структуры.

Рассмотрен вопрос о системах Чебышева для функций, нелинейно зависящих от параметров, и изучены условия, при которых системы функций или гистограмм, непрерывно зависящих от параметров, являются либо системами Чебышева в непрерывной области, либо линейно-независимыми на дискретных множествах.

Подробно проанализирован важный практический, просто реализуемый, прием исследования декомпозиционных свойств используемой параметризации - с помощью межпараметрического коэффициента корреляции  $K(p_i, p_j)$ . Он определяется следующим образом: пусть дана функция, зависящая от

параметров,  $f(x, p_i)$ ,  $x \in X$ , и пусть  $g_j(x, p_i)$  - частная производная от  $f(x, p_i)$  по параметру  $p_j$ ; тогда

$$K(p_i, p_j) = \int_X g_i(x, p_i) \dot{g}_j(x, p_j) / (G_1 G_2) dx,$$

где  $G_i = \|g_i(x, P)\|$ .

Если  $X$  - дискретное множество, то интеграл в  $K$  заменяется суммой. Чем ближе  $K_m = \max |K(p_i, p_j)|$  к нулю, тем лучше параметризация.

С помощью коэффициента  $K$  были проанализированы наиболее широкоупотребляемые параметризации для описания различных компонент в спектроподобных распределениях и для важнейших из них (гауссовой, лоренцевой и экспоненциальной) найдены оптимальные параметрические формулы, при которых  $K_m = 0$ .

Рассмотрено влияние на качество декомпозиции таких факторов как:

- изменение параметризации;
- наложение связей на параметры;
- оптимизация использования пространства независимых переменных;
- переход к другим метрикам при аппроксимации.

Рассмотрен ряд подходов линейной декомпозиции и, в частности, показана интерпретация компонент распределений в терминах компонентного и факторного анализа.

В 9-м разделе рассмотрены методы качественной декомпозиции спектроподобных распределений, т.е., формальные методы построения функциональных классов ( $M_i$ ) для описания компонент и определения их числа  $n$ :

1. синтаксические, с помощью формальных грамматик, описывающих суммы пикообразных функций;
2. статистические;
3. кластеризационные.

Для одного важного частного случая доказана теорема, позволяющая простым алгоритмом построить точное решение задачи кластер-анализа.

В 10-м разделе рассмотрен дополнительный к статистическому детерминистский подход к понятию погрешности данных. Этот подход важен для спектроскопии, поскольку неопределенность в спектрограммах очень часто не исчерпывается случайной погрешностью и, более того, эта неопределенность осталась бы, даже если детектор давал бы абсолютно точные данные.

Метод наименьших квадратов детерминистски интерпретируется следующим образом: погрешность  $e(x_j)$  наблюдаемой кривой  $y(x_j)$  мы можем считать неизвестной дискретной функцией с известной нормой в каждом  $x_j$ , равной максимально наблюдаемому по модулю значению, и все рассуждения вести не в терминах функций распределений, ковариаций и дисперсий, а в терминах функциональных пространств и определенных в них норм.

В работе было показано, что существует внутреннее единство между статистическим и детерминистским МНК, что важнейшие теоремы статистики: об обратной мнк-матрице как матрице ошибок оценок, Гаусса-Маркова и другие, как в линейном, так и в нелинейном вариантах, могут быть перенесены на случай детерминистских погрешностей и таким образом результаты мнк-анализа не зависят от характера используемой в нем терминологии. При этом детерминистский мнк часто использует более адекватный практике язык и включает в себе меньше субъективизма, поскольку информацию о нормах погрешностей получить значительно легче и проще, чем о полной функции их распределения.

Но разумеется, часть процедур анализа спектров использует и язык статистики.

Дополнительной целью работы было исследование различных способов и форм накопления данными информации. В частности, был рассмотрен вопрос о состоятельности детерминистских мнк-оценок, если под информационной емкостью данных понимать отношение нормы  $y(x)$  к норме  $e(x)$ .

В 11-разделе перечислены базисные программы: GFIT, SMOOS, ROBUS, FILTBK, DSPLIN, SFFT, лежащие в основе всех конкретных программных проблемно-ориентированных пакетов.

В 12-разделе перечислены эти пакеты:

1. Программный комплекс универсального назначения UPEAK для разложения одномерных произвольных распределений аддитивной структуры на компоненты;
2. программный комплекс универсального назначения DOMUS для разложения двумерных распределений аддитивной структуры на двумерные компоненты;
3. специализированный программный комплекс ACTIV для анализа одномерных гамма- и рентгеновских спектров в широком диапазоне режимов, включая и полностью автоматический; ACTIV содержит аппарат, позволяющий использовать его и при обработке спектров в задачах активационного анализа;

4. специализированный программный комплекс DECAN для анализа одномерных распределений на компоненты экспоненциального типа;
5. специализированный программный комплекс SPEVA для анализа временной последовательности нейтрограмм, снимаемых в дифракционных измерениях;
6. специализированный программный комплекс ERTQU для анализа одномерных распределений радонных излучений и формально им эквивалентных распределений в задачах сейсмического анализа;
7. специализированный программный комплекс MRJA для анализа многоспектровых многофазных времяпролетных нейтронно-дифракционных распределений поликристаллов.

Перечислены пакеты, поддерживаемые автором лично. Кроме того на их базе созданы более узко специализированные пакеты, поддержанием которых занимаются сами пользователи.

В процессе развития данных комплексов от ранних версий к более поздним окончательно сложилась их следующая структура и организация. Каждый пакет представляет собой объединение программных модулей, по возможности типизированных, ориентирован на оверлейный способ загрузки в память, и настраивается на размеры памяти, отводимой под данные.

Все программы написаны на Фортране-90 с использованием только стандартных средств языка.

Все это обеспечивает очень большую портативность перечисленных пакетов и дает возможность использовать их на ЭВМ самых разных классов (от суперкомпьютеров до персональных ЭВМ с сохранением всех алгоритмических возможностей этих пакетов).

Краткое описание алгоритмов пакета и инструкция оформлены в специальный модуль HELP, входящий в пакет в качестве составной части.

Пакеты функционируют следующим образом.

Входной поток представляет собой последовательность команд, сопровождаемых числовой или/и текстовой информацией. Каждая команда есть указание супервизору пакета совершить какое-либо элементарное действие, на которые разбит весь суммарный алгоритм пакета. Последовательность команд не фиксирована и определяется лишь логикой решения основной задачи. Аппарат команд как правило развит и в таком комплексе как ACTIV достигает нескольких десятков, что позволяет пользователю очень широко вмешиваться в управление алгоритмом. Одновременно очень последовательно претворена концепция умолчания, в силу чего все управляющие

данные имеют начальные значения, определяющие некоторый типовой характер работы алгоритма в автоматическом режиме. Поэтому объем входного потока может варьироваться в очень широких пределах: от минимального до очень большого.

Выходной поток представляет собой последовательность промежуточных или окончательных результатов работы алгоритма. Объем его тоже регулируется пользователем.

Каждый пакет ориентирован на решение соответствующих задач в максимально широком диапазоне условий и целей, связанными с этими задачами. Программный пакет UPEAK. Описание круга задач, решаемых этим комплексом, может быть построено следующим образом. Произвольное одномерное спектроподобное распределение  $s(x)$  может быть представлено следующим образом:

$$s(x) = \sum_{i=1}^n f_i(x) + b(x) + e(x)$$

где  $f_i$  - полезные компоненты,  $b(x)$  - фоновая компонента,  $e(x)$  - помеха.

Для каждого типа компонент  $f_i$  строится своя модель  $m_i(x)$ . Также своя модель строится и для фоновой компоненты  $m_b(x)$ .

Модели параметризуются следующим образом:

$$f_i(x, A, P, W, C) = A \cdot m_i((x - P)/(C \cdot x + W))$$

$$b(x, A, P, W, C) = A \cdot m_b((x - P)/(C \cdot x + W))$$

Решение задачи декомпозиции  $s(x)$  на компоненты состоит в определении оценок параметров  $A, P, W, C$  для каждой компоненты и точности этих оценок.

Эти задачи могут быть разбиты на 3 группы по критерию характеристики, рассматриваемой как дискриминатор отдельной компоненты:

1. дискриминатор - положение  $P$  (например, разложение спектров на пики - альфа-спектры, гамма-спектры, спектры конверсионных электронов, нейтронных резонансов и т.д.);
2. дискриминатор - ширина  $W$  (или комбинация  $C \cdot x + W$ ) (например, анализ спектров радиоактивного распада нескольких источников, бета-спектров и т.д.);
3. дискриминатор - и положение, и ширина (например, нейтронно-дифракционные спектры, где пик упругого рассеяния сидит на пику неупругого, а оба на фоновой подложке).

Возможны и дальнейшие разбиения задач на подгруппы по -

1. числу параметров для описания полезной компоненты:  $A, P, W, C$ ; или  $A, P, W$ ; или  $A, W$ ;
2. знаку амплитуды (+ для спектров излучения или рассеяния, - для спектров поглощения).

Спектр может быть разбит на участки и анализироваться по участкам. На одном участке допускается использование не более 4 моделей полезных компонент и одной модели (неполиномиальной) фона.

Возможно наложение линейных связей (жестких и мягких) на параметры. Широко используются описанные там же алгоритмы пикосохраняющей фильтрации. Программа может использовать метод регуляризации Тихонова при построении оценок параметров в тех случаях, когда мнк-матрица оказывается плохо обусловленной.

Программный пакет для автоматического анализа дву- и трехмерных амплитудных распределений **DOMUS**. Двумерный спектр  $s(x)$  формально записывается следующим образом:

$$s(x, y) = \sum_{i=1}^n f_i(x, y) + b(x, y) + e(x, y)$$

где  $f_i$  -  $i$ -ая полезная компонента,  $b$  - двумерный фон,  $e$  - двумерная погрешность.

Каждому двумерному пику  $f_i$  сопоставляется двумерная модель  $m(x, y)$ , и параметризация осуществляется следующим образом:

$$f_i(x, y, A_i, P_{xi}, P_{yi}, W_x, W_y, S_1, S_2, C_x, C_y) = A_i m(Z_i + S_1 V_i, S_2 Z_i + V_i),$$

где

$$Z_i = \frac{x - P_{xi}}{C_x z + W_x},$$

$$V_i = \frac{y - P_{yi}}{C_y v + W_y},$$

$$z = \frac{2x - x_1 - x_2}{x_2 - x_1},$$

$$v = \frac{2y - y_1 - y_2}{y_2 - y_1},$$

$x_1, y_1$  и  $x_2, y_2$  - левые и правые границы анализируемого участка спектра. Фон описывается двумерным полиномом 2-й степени.

Физический смысл параметров следующий:

$A$  - амплитуда пика,  $P_x$  и  $P_y$  -  $x$ - и  $y$ -координаты положения,  $W_x$  и  $W_y$  -  $x$ - и  $y$ -компоненты полуширины,  $S_1$  и  $S_2$  - коэффициенты, характеризующие поворот пика в плоскости  $(X, Y)$ ,  $C_x$  и  $C_y$  - коэффициенты линейной

зависимости полуширины от  $X$  и  $Y$ , соответственно.

К пакету **DOMUS** относится также набор программ для обработки двумерных спектров, представленных не амплитудными распределениями, а набором зарегистрированных событий.

Программный пакет для обработки гамма- и рентгеновских спектров в широком диапазоне режимов и условий **ACTIV**. По сравнению с другими подобными программами **ACTIV** имеет следующие отличительные черты:

- большое разнообразие режимов работы, простирающееся от полностью автоматического (с минимумом входных данных) до полуручного (для особо нестандартных данных); пакет может взаимодействовать с программами, работающими в интерактивном режиме;
- автоматический поиск пиков и мнк-подгонка используют реальную форму пиков, что повышает надежность и качество декомпозиции и понижает систематическую ошибку операций; алгоритмы всех стадий анализа используют большое множество приемов и методов, описанных ранее, которые значительно улучшают качество работы программы;
- в спектре могут обрабатываться только те части, которые содержат информацию, представляющую интерес и определяемую списком энергий линий или списком изотопов;
- блок подпрограмм для активационного анализа является частью пакета **ACTIV** и позволяет осуществить полный цикл спектрального анализа, включая определение качественной и количественной структуры облученного образца по методу эталонов или на основе формул активации;
- к структуре и функции пакета **ACTIV** относится все то, что было сказано по поводу пакета **UREAK**: модульность, портативность, командный язык взаимодействия и т.д.

Программный пакет для анализа кривых экспоненциального типа. Пакет включает в себя два программных комплекса **DECAN** и **HTIME**. Анализируемые данные могут быть представлены либо в виде распределений, описываемых суммами функций вида

$$y(t) = A \cdot \exp(-t/T)$$

или

$$y(t) = \int R(x - t) \cdot \exp(-x/T) dx$$



в простейшем случае, или функциями вида

$$\frac{AT_2}{T_1 - T_2} (\exp(-t/T_1) - \exp(-t/T_2))$$

при изучении, например, сложных распадов.

Обстоятельствами, сильно осложняющими анализ таких распределений, являются:

- малая статистика измерения;
- отсутствие данных на наиболее информационных участках интервала измерения, например, в начале;
- большая и неравномерная ширина ячейки гистограммы, и ряд других. Программный комплекс NTIME имеет дело с аналогичными задачами, но данные представляются в виде отдельных событий - случайных величин, подчиненных функциям распределения вышеперечисленного типа.

Программный пакет для анализа временной последовательности нейтрограмм, снимаемых в экспериментах по изучению кинетики гетерогенных реакций методами нейтронной дифракции, SPEVA. Каждое измерение представляет собой последовательность из 16 нейтроннодифракционных спектров, характеризующую временную эволюцию изучаемой кристаллической структуры, и задача состоит в извлечении информации как об отдельных спектрах, так и об их эволюции. В отличие от стандартных задач декомпозиции распределений здесь требуется не только выделение компонент из очень большого их множества (все измерение имеет объем 512 кбайт), но еще и автоматическое опознавание их временной эволюции. Программный пакет для анализа данных радоновых измерений в геологических и сейсмографических задачах ERTQU. Из измерения (и/или некоторых его преобразований) извлекаются как периодические, так и пикообразные компоненты, которые далее соотносятся с данными, представляющими интерес.

Программный комплекс FIMDA. Для анализа данных в исследованиях твердого тела с использованием атомного зонда при ионно-полевом микроскопе (Atomic Probe Field Ion Microscope).

Программный комплекс MRJA для анализа многоспектровых многофазных времяпролетных нейтронно-дифракционных распределений поликристаллов, регистрируемых на Фурье-дифрактометре.

Программный комплекс AUTOX для решения задачи автоиндексации многофазных дифракционных спектров от поликристаллов.

#### Список печатных работ.

1. Zloказов V.B. Nuclear Instruments and Methods, 1977, v.143, No 1, p.151-156.
2. Выропаев В.Я., Злоказов В.Б., Кулькина Л.П., Маслов О.Д., Фефилов Б.В. Атомная энергия, 1977, т.43, вып.3, с.187-190.
3. Злоказов В.Б. Computer Physics Communications, 1978, v.13, No 5/6, p.389-398.
4. Artukh A.G. et al. ЯФ, 1978, т.27, вып.1, с.29-36.
5. Зодан Х., Калпакчиева Р., Пенционжкевич Ю.Э., Букланов Г.В., Вакатов В.И., Злоказов В.Б., Саламатина Т.С., Челноков Л.П. ОИЯИ, ЛЯР, Дубна, 1977, P7-10671.
6. Злоказов В.Б. Nuclear Instruments and Methods, 1978, v.151, No 1/2, p.303-305.
7. Балагуров А.М., Длоуга М., Злоказов В.Б., Миронова Г.М. ОИЯИ, ЛНФ, Дубна, 1977, P10-11106.
8. Балагуров А.М., Длоуга М., Злоказов В.Б., Миронова Г.М. Доклады... Дубна, 1977. Дубна, 1978, с.283-287. ОИЯИ, ЛНФ.
9. Балагуров А.М., Длоуга М., Злоказов В.Б., Миронова Г.М. ОИЯИ, ЛНФ, Дубна, 1977, P10-11107.
10. Злоказов В.Б. В кн. "Совещание по программированию и математическим методам решения физических задач". Доклады... Дубна, 1977. Дубна, 1978, с.279-282. ОИЯИ, ЛНФ, Дубна, 1977, D10, 11-11264.
11. Balagurov A.M., Dlouha M., Zloказов V.B., Mironova G.M. Conference on diffraction profile analysis and open meeting of the commission on neutron diffraction. Cracow, Poland, August 14-15, 1978, p.19.
12. Zloказов V.B. Computer Physics Communications, 1979, v.18, No 2, p.281-286.
13. Balagurov A.M., Borca E., Dlouha M., Gheorghiu Z., Mironova G.M., Zloказов V.B. Acta Crystallographica. A35 (1979) 131-136.
14. Демин А.Г., Друин В.А., Злоказов В.Б., Лобанов Ю.В., Оганесян Ю.Ц., Сагайдак Р.Н., Утенков В.К. Тезисы Международного симпозиума по синтезу и свойствам новых элементов. Дубна, 23-29 сентября 1980, стр.25.

15. Злоказов В.Б. Computer Physics Communications,1981,v.21,p.373-383.
16. Злоказов В.Б. Nuclear Instruments and Methods,1982,v.199,No 3,p.509-519.
17. Александрова И.В., Гундорина С.Ф., Злоказов В.Б., Фронтасьева М.В. В кн. "Совещание по использованию новых ядерно-физических методов для решения научно-технических и народно-хозяйственных задач",4-е,Дубна,1981. Доклады.Дубна,1982,с.289-290. ОИЯИ,Дубна,18-82-117.
18. Злоказов В.Б. Computer Physics Communications,1982,v.28,p.27-37.
19. Злоказов В.Б. Nuclear Instruments and Methods,1984,v.227,p.135.
20. Брухертзайфер Х.,Злоказов В.Б. ОИЯИ,ЛВТА,Дубна,1985,Р10-85-333.
21. Бутцева Г.Л.,Воробьева Н.Н.,Говорун Н.Н.,Завьялова А.С., Злоказов В.Б.,Нефедьева Л.С.,Расторгуев А.А.,Рерих Т.С., Салтыков А.И.,Тарасова В.Н.,Ягафарова В.М. ОИЯИ,ЛВТА,Дубна,1985,Р10-85-171.
22. Злоказов В.Б.,Кононенко Г.А., Кузнецов В.В, Фоминых М.И., Цупко-Ситников В.М. ОИЯИ, ЛЯП, Дубна, 1985, Р6-85-606.
23. Брухертзайфер Х.,Айхлер Б.,Эстевес Х.,Круз Т., Вильо Э.,Злоказов В.Б.,Р.дель Портальо,Рюдигер Ю. Радиохимия,1987,т.29,вып.1,с.62-66.
24. Кононенко Г.А.,Гуяш Я.,Злоказов В.Б., Кузнецов В.В, Фоминых М.И., Цупко-Ситников В.М. ОИЯИ,ЛЯП,Дубна,1985,Р6-85-717.
25. Злоказов В.Б. ФЭЧАЯ,1985,т.16,вып.5,стр.1126-1163.
26. Злоказов В.Б. ОИЯИ,ЛВТА,Дубна,1986,Р11-86-135.
27. Злоказов В.Б. ОИЯИ,ЛВТА,Дубна,1986,Р10-86-502.
28. Злоказов В.Б. ОИЯИ,ЛВТА,Дубна,1986,Р10-86-764.
29. Гундорина С.Ф.,Злоказов В.Б.,Островная Т.М., Фронтасьева М.В. Тезисы V-го Всесоюзного совещания по активационному анализу и другим радиоаналитическим методам. Ташкент,апрель 1987.
30. Злоказов В.Б. ОИЯИ,ЛВТА,Дубна,1987,Р10-87-894.
31. Zlokazov V.B. Computer Physics Communications,1989,v.54,p.371-379.

32. Злоказов В.Б. Nuclear Instruments and Methods-A,1989,v.275(2),p.392-396.
33. F.Riedel, H.Bruchertseifer, V.B.Zlokazov. 4th Conference on radioisotope application and radiation processing in industry. Leipzig, GDR, September 19-23, 1988.
34. Злоказов В.Б. Computer Physics Communications,1990,v.59,p.217-225.
35. Балагуров А.М.,Злоказов В.Б.,Миронова Г.М., Хван Чан Ген, Кудряшев В.А.,Трунов В.А., Ульянов В.А.,Антсон О.,Пеюрю Х., Тита А.,Харконен К.,Хинсмьяки П.,Ахти М.,Уониуус Л.,Уллакко К. ОИЯИ,ЛНФ,Дубна,1989,Р14-89-602.
36. Злоказов В.Б., Ишанкулиев Д., Третьякова С.П., Аширов Т.А. ОИЯИ,ЛВТА,Дубна,1989,Р11-89-659.
37. Балагуров А.М.,Злоказов В.Б.,Миронова Г.М., Новожилов В.Е., Островной А.И., Симкин В.Г. ОИЯИ,ЛНФ,Дубна,1989,Р3-89-601.
38. Шабалин Е.П., Анцупов Н.П., Злоказов В.Б., Мельников В.Н., Пепельшев Ю.Н., Попов А.К., Рогов А.Д. ОИЯИ,ЛНФ,Дубна,1989,Р3-90-29.
39. Злоказов В.Б., Чернышев В.В. Journal of Applied Crystallography (1992) 25,447-451.
40. V.B.Zlokazov, V.V.Chernyshev. First European Powder Diffraction Conference, Munich, March 14-16th, 1991, Abstracts. Proceedings, V.1, p.283-288.
41. V.B.Zlokazov Journal of Applied Crystallography,(1992).25,p.69-72.
42. V.B.Zlokazov. GAMM-92. Leipzig, March 24-28th, 1992, Abstracts. p.262.
43. V.V.Chernyshev, V.B.Zlokazov. Report on the ISSI - International Seminar for Structure Investigations, 1-4 September 1992, Dubna.
44. V.B.Zlokazov. Report on the ISSI - International Seminar for Structure Investigations, 1-4 September 1992, Dubna.
45. V.B.Zlokazov,S.P.Tretyakova ОИЯИ,ЛВТА,Дубна,1992,Е-18-92-194.
46. V.B.Zlokazov Nuclear Instruments and Methods-B72,(1992),p.139-142.

47. Zlokazov V.B. A invited lecture at the Radichemical Institute (TU Muenchen)  
22 August 1992, Garching.
48. Zlokazov V.B. Third European Powder Diffraction Conference, Vienna, September 25-28th, 1993, Proceedings, Pt.1,p.67-72.
49. Zlokazov V.B. American Crystallographic Association Annual Meeting, Atlanta, USA, June 25 - July 1, 1994, Abstracts. p.155. Series 2, ISSN 0569-4221, Vol.22.
50. Zlokazov V.B. 15th European Crystallographic Meeting, Dresden, 28.8 - 2.9, 1994, Book of abstracts, p.581.
51. Zlokazov V.B. Computer Physics Communications,1995,v.85,p.415-422.
52. Zlokazov V.B. Fourth European Powder Diffraction Conference, Chester England July 10-14 1995, Book of abstracts, p.112.
53. Zlokazov V.B. Fifth European Powder Diffraction Conference, Parma, Italy - May 25-28 1997, Book of abstracts, p.254.
54. Zlokazov V.B. Journal of Applied Crystallography,(1997).30,p.996-1001.

Рукопись поступила в издательский отдел  
19 октября 1998 года.