**Тетуев Руслан Курманбиевич. Алгебра спектральных преобразований в задачах обработки данных : диссертация ... кандидата физико-математических наук : 05.13.17 / Тетуев Руслан Курманбиевич; [Место защиты: Вычисл. центр РАН].- Пущино, 2007.- 111 с.: ил. РГБ ОД, 61 07-1/1590**

РОССИЙСКАЯ АКАДЕМИЯ НАУК ИСТИТУТ МАТЕМАТИЧЕСКИХ ПРОБЛЕМ БИОЛОГИИ

07-1/1590 ***На правах рукописи***

ТЕТУЕВ РУСЛАН КУРМАНБИЕВИЧ

**АЛГЕБРА СПЕКТРАЛЬНЫХ ПРЕОБРАЗОВАНИИ В ЗАДАЧАХ ОБРАБОТКИ ДАННЫХ**

Специальность: 05.13.17 - Теоретические основы информатики

Диссертационная работа на соискание ученой степени кандидата физико-математических наук



Пущи но - 2007

Руководитель: д.т.н. проф. Ф.Ф. Дедус

**ОГЛАВЛЕНИЕ**

[**ВВЕДЕНИЕ 3**](#bookmark3)

[**ГЛАВА I Аналитические преобразования над спектром 10**](#bookmark9)

1. [*Общие свойства классических ортогональных систем 10*](#bookmark10)
2. [*Основы спектрального представления функций 16*](#bookmark12)
3. *Предпосылки к развитию новых методов спектральной обработки... 18*
4. [*Понятие спектрального преобразования функций 21*](#bookmark16)
5. [*Задача поиска методов быстрых спектральных преобразований 22*](#bookmark21)
6. [*Ранние попытки решения задачи. Матричный способ 23*](#bookmark22)

[**ГЛАВА II Быстрые аналитические преобразования над спектром 29**](#bookmark30)

1. [*Достаточное условие существования быстрых алгоритмов спектрального преобразования 29*](#bookmark31)
2. [*Классические ортогональные полиномы 32*](#bookmark35)
3. [*Быстрые преобразования для оператора вида* ***A(f) = xf(x).*** *Лемма о суперпозиции линейных операторов 36*](#bookmark39)
4. *Быстрые преобразования для операторов вида* ***A(f****) = f(x)/x. Теорема обратимости линейных операторов*
5. *Быстрые преобразования для оператора дифференцирования*
6. [*Быстрые преобразования для оператора интегрирования 42*](#bookmark49)
7. [*Нахождение рекуррентных соотношений для различных модификаций ортогональных базисов 43*](#bookmark50)

**ГЛАВА III Общая вычислительная схема для реализации быстрых**

**спектральных преобразований 46**

1. [*Определение спектрального каскада и диффузии 46*](#bookmark58)
2. [*Метод спектрального каскада и диффузии 48*](#bookmark61)
3. [*Пример применения метода каскада-диффузии 49*](#bookmark62)
4. [*Сравнительный анализ результатов расчетов и основных характеристик алгоритмов 51*](#bookmark64)

**ГЛАВА IV Обобщения алгебры спектральных преобразований 56**

* 1. [*Быстрые нелинейные спектральные преобразования 57*](#bookmark68)
	2. [*Быстрые спектральные преобразования для ортогональных систем дискретного аргумента 63*](#bookmark72)
	3. [*Сверхбыстрые спектральные преобразования на системах с параллельной архитектурой вычислений 68*](#bookmark74)

**ГЛАВА V Применение быстрых спектральных преобразований 75**

1. [*Применение в задачах распознавания образов 75*](#bookmark80)
2. [*Применение в задачах биоинформатики 81*](#bookmark82)

[**ЗАКЛЮЧЕНИЕ 88**](#bookmark87)

[**ВЫВОДЫ 89**](#bookmark88)

[**ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ 92**](#bookmark89)

**ЛИТЕРАТУРА 94**

**ПРИЛОЖЕНИЯ 108**

**ВВЕДЕНИЕ**

Во введении обосновывается актуальность проблемы, сформулированы цели и задачи, дан краткий обзор содержания диссертации, перечислены полученные новые результаты исследования. На примере рассматриваемых систем ортогональных полиномов будут показаны основные приемы обработки и аналитического преобразования функций посредством изменения их спектральных представлений и рассмотрены перспективные подходы практической реализации подобных преобразований.

**Актуальность темы**

Во второй половине двадцатого столетия стремительное развитие вычислительной техники ознаменовало начало новой эры для всей математической теории и различных ее приложений. Возможность проведения множества элементарных операций на скоростных вычислителях требовала развития и адаптации математических методов решения задач к новым создавшимся условиям. Весьма скоро среди вычислительных машин наметился явный лидер - электронные вычислительные машины (ЭВМ) с целочисленным бинарным представлением величин и последовательным ходом выполнения операций. Это впоследствии всецело определило развитие соответствующей математической базы.

Специфика ЭВМ диктовала в то время свои условия, но целочисленные представления лишь отчасти, приближенно отвечали развитым к тому времени математическим, аналитическим понятиям и представлениям. Классическая техника математических вычислений привыкла оперировать понятиями вещественных чисел, аналитических функций, операций дифференцирования, интегрирования и т.д., однако попытки реализовать эти представления на вычислительных машинах приводили к необходимости решения довольно сложных, нетривиальных задач.

Представления функций в «электронном» виде и методы осуществления на ЭВМ элементарных аналитических операций над функциями (интегрирования, дифференцирования и т.д.) начали развиваться приблизительно в одно время, однако развитие этих задач протекало во многом независимо друг от друга, несмотря на их очевидное математическое «родство». С другой стороны можно предположить ряд факторов, видимо, послуживших этому причиной. Приведем ниже предпосылки, по нашему мнению, приведшие к такому разделению подходов и методов:

* матричное представление функций в виде массива дискретных значений, изначально оказалось наиболее естественным для ЭВМ представлением и наиболее удобным для реализации большинства арифметических операций над функциями;
* было предложено немало численных методов для осуществления более сложных, аналитических преобразований. С возникновением ЭВМ эти методы получили развитие и распространение. Явным преимуществом таких подходов явилась простота вычислений и скорость - обычно затраченное вычислительное время и объемы требуемой памяти были пропорциональны величине массива данных;
* спектральное представление произвольных сигналов как некоторых аналитических функций стало практически использоваться на ЭВМ намного позже, лишь после того, как в зарубежной и отечественной науке была разрешена основополагающая задача о быстром преобразовании Фурье (БПФ);
* до сих пор не было предложено универсальных и достаточно быстрых схем вычислений для осуществления аналитических преобразований над спектральными представлениями функций на практике. Предложенные ранее подходы сложны и сильно уступают по времени выполнения аналогичным преобразованиям функций, например, на основе численных методов.

Спектральное представление сегодня широко применяется, но большей частью для передачи, хранения и анализа данных, изначально полученных в виде массивов значений функций. Трудно переоценить влияние спектрального представления функций на развитие современной прикладной математики, однако на практике спектральное представление довольно редко используется для проведения определенных аналитических преобразований функций. В случае необходимости приходилось восстанавливать функцию в виде массива значений, а затем, проведя требуемые преобразованиями численно, снова «вернуть» функцию в спектральное представление. Однако, такая ситуация на наш взгляд не оправдана и как показано ранее, в ряде работ [111, 106], тем не менее, существует принципиальная возможность осуществления аналитических преобразований, при использовании лишь самих спектральных представлений функций или, иначе, их *спектральных преобразований.*

В перечисленных работах показано, что такой подход к тому же приводит к более приемлемым с практической точки зрения результатам в сравнении с теми, что были получены при осуществлении аналитических преобразований численными методами. Этот факт объяснен авторами свойствами высокой адаптивности метода при выборе базиса из семейства классических ортогональных полиномов, гладкостью аппроксимативного представления, заданного ограниченным ортогональным рядом и т.д [117].

Действительно, существенным недостатком численных методов является то, что практическая ценность результатов сильно зависит от вида преобразования и характера функции. К примеру, оказывается весьма сложным даже приближенное нахождение второй производной от функции, сильно искаженной помехами. В свою очередь, основным недостатком спектральных преобразований функций является то, что вычислительный процесс оказывается намного более сложным и громоздким. Алгоритмы оказались хуже на порядок по временной сложности, а также требующими стадии предварительных вычислений и резервирования больших объемов памяти под вспомогательные коэффициенты.

Указанные недостатки являются существенным ограничением для внедрения спектральных преобразований функций на практике в задачах, требующих проведения быстрых, устойчивых и точных аналитических преобразований функций, сигналов и т.д. Несмотря на положительные результаты, при решении практических задач лишь отставание во времени способно вынудить разработчиков программ отдать предпочтение в пользу более грубых, но быстрых алгоритмов [95, 101].

Вышеописанное положение демонстрирует потребность в создании альтернативного простого и точного прикладного математического аппарата для осуществления быстрых спектральных преобразований функций, соответствующих ряду основных аналитических преобразований функций и групп их суперпозиций. В качестве ортогональных систем в работе рассмотрен каждый базис из семейства классических ортогональных полиномов (Лагерра, Якоби, Эрмита, Лежандра, Сонина-Лагерра, Гегенбауэра, Чебышева первого и второго рода) и некоторые основные модификации этих базисов. В главе IV показано, что без нарушения общности основные вычислительные схемы легко переносятся на случай применения классических ортогональных полиномов дискретного аргумента: Хана, Мейкснера, Шарлье, Кравчука и Чебышева.

**Цель работы**

Решение проблемы создания математического аппарата, для осуществления на вычислительных машинах быстрых аналитических преобразований функций, заданных на практике конечными ортогональными рядами. В качестве ортогональных базисов рассматриваются системы функций, построенные на классических ортогональных полиномах. В качестве операторов преобразования - основные, часто используемые преобразования: умножение/деление на аргумент, дифференцирования/интегрирования и т.д.

В соответствии с поставленной задачей определены **задачи диссертационной работы:**

* 1. Поиск общей вычислительной схемы спектральных преобразований рядов ортогональных полиномов.
	2. Вывод требуемых аналитических соотношений для класса классических ортогональных полиномов и их некоторых модификаций.
	3. Реализация программного комплекса для осуществления спектральных преобразований на практике.
	4. Решение задачи контурного распознавания визуальных объектов средствами разработанной системы аналитических преобразований.
	5. Решение задачи поиска тандемных повторов в геномах средствами разработанной системы аналитических преобразований.

**Постановка научной задачи**

Требуется найти группу алгебраических правил спектральных преобразований, соответствующих основным аналитическим преобразованиям функций, представленных ограниченными рядами, построенными на классических ортогональных полиномов. Найти решение при повышенных требованиях на вычислительную сложность реализуемых алгоритмов [95].

**Научная новизна**

В данной диссертационной работе впервые предложен вычислительный метод, позволяющий существенно понизить временную сложность вычислений над спектральным представлением функций, соответствующих ряду основных аналитических преобразований над функциями.

Сформулирована и доказана Теорема о достаточном условии существования алгоритма быстрого спектрального вычисления.

Показано соблюдение достаточного условия Теоремы для ряда линейных операторов и всех систем классических ортогональных полиномов и функций.

В ходе доказательств получены рекуррентные соотношения особого вида для полиномов Якоби, Гегенбауэра и функций Якоби, Гегенбауэра, Сонина- Лагерра, Лагерра, Чебышева первого и второго рода, Эрмита.

Сформулирована и доказана Теорема об обратимости линейных операторов, а также сформулирована и доказана Лемма о суперпозиции линейных операторов.

На основе применения спектральных преобразований предложено и реализовано вычисление инвариантных геометрических признаков, описанных аналитически с помощью операторов дифференцирования высоких порядков.

Найдены и применены аналитически описанные признаки, инвариантные к выбору начальной точки и вычисляемые на основе отрезков ортогонального ряда.

Реализован программный комплекс "8рес1га1Кеу1$ог", организующий локализацию участков тандемных повторов в ДНК последовательностях на основе вычисления и анализа некоторых первых производных от функций- профилей, построенных на данных генетических последовательностях.

**Практическая значимость работы**

* + 1. Предложенный метод сопоставляет некоторым часто используемым на практике аналитическим преобразованиям функций, набор простых и быстрых вычислений для получения соответствующего преобразования спектрального представления. Это позволяет применять метод в различных прикладных областях, где требуются быстрые, точные и устойчивые аналитические преобразования сигналов различной природы.
		2. Показано, что для сложных преобразований функции, являющихся группой более простых, возможно построение единой вычислительной схемы, составленной на основе объединения соответствующих элементарных схем. При этом, многие частные вычисления в узлах выстраиваемой общей схемы, как показано в работе, могут быть произведены одновременно. Это позволит дополнительно ускорить алгоритмы при реализации на ЭВМ с параллельной архитектурой исполнения команд, что в свою очередь позволит применять данный подход в задачах, сопряженных с необходимостью сверхбыстрой обработки данных:

1) обнаружение и распознавание визуальных образов в реальном времени

на основе контурного восприятия; и) поиск тандемных повторов в сверхдлинных генетических последовательностях, таких, как геном человека = 3 млрд. нуклеотидов и т.д.

3) В ходе выполнения диссертационной работы были получены новые аналитические соотношения для ряда базисов, построенных на классических ортогональных полиномах, что может быть полезным для последующих исследований и разработок в смежных областях науки.

С появлением и развитием систем параллельных вычислений увеличивается разрыв по времени исполнения между еще более ускорившимися, но по- прежнему не прибавившими в точности численными методами расчета с одной стороны и проведением аналитических преобразований функций посредством спектра с другой. В тоже время появился целый ряд успешных работ, посвященных проведению сверхбыстрого преобразования Фурье на системах параллельных вычислений. Однако, задача об ускорении самих спектральных вычислений, соответствующих определенным преобразованиям функций, оставалась по-прежнему актуальной как при реализации на вычислительных системах с последовательной, так и с параллельной архитектурой.

Диссертационная работа посвящена изложению и обсуждению нового вычислительного метода, предложенного на основании проведенных нами исследований и позволяющего существенно понизить временную сложность спектральных вычислений для ряда основных аналитических преобразований функций и групп их суперпозиций. Вычислительный метод основан на знании полученных рекуррентных соотношений определенного вида и приводит к алгоритмам той же временной сложности, что имеют алгоритмы, основанные на использовании численных методов, то есть линейной и логарифмической для систем с последовательными и параллельными вычислениями соответственно. Очевидно, что при этом сохраняются иные преимущества спектрального подхода, такие как широкие адаптивные возможности при аппроксимации функций и более приемлемые с практической точки зрения результаты их преобразований. Заключительная часть работы посвящена изложению результатов применения предложенного метода в некоторых научных областях и сравнительному анализу с альтернативными подходами решения рассматриваемых задач.

Под полиномами Сонина-Лагерра в диссертационной работе понимаются *обобгценные полиномы Лагерра.* Подобная терминология впервые предложена авторами работы [111] и введена на основании [127].

выводы

В данной работе автором были освещены широкие возможности практического применения ортогональных систем, построенных на классических ортогональных полиномах, и новые методы аналитической обработки сигналов, представленных рядами полиномов.

Во введении рассмотрена общая теория систем классических ортогональных полиномов непрерывного аргумента. Особое внимание уделено ортогональным системам Сонина-Лагерра, на примере которых в дальнейшем изложении были продемонстрированы основные результаты. Так, было показано, что на основании введенных модификаций полиномов и функций Сонина-Лагерра легко реализовать ряд адаптивных процедур в процессе аналитического описания поступающих сигналов.

В первой главе обсуждается возможность аналитической обработки функций на практике посредством изменения спектральных характеристик одномерных сигналов (функций), т.е. коэффициентов разложения спектрального представления. В ранних работах была показана возможность такой обработки, однако способы, предложенные раннее, сводились к длительным вычислениям на практике, зависящим от глубины аппроксимативного разложения сигналов с квадратом.

Во второй главе введен новый метод аналитической обработки сигналов, имеющий линейную зависимость длительности вычислений от глубины представления. Доказана применимость метода в общем случае для всех ортогональных систем и линейных преобразований, для которых существует рекуррентное соотношение вида (2.4.3), что, в частности, рассмотрено на множестве важных примеров. Также рассмотрены некоторые теоретические обобщения метода на случай суперпозиций и обратимости линейных операторов.

В третьей главе показано, как принципы аналитической спектральной обработки сигналов, изложенные ранее, могут служить для реализации быстрых преобразований сигналов на практике. Введен метод спектральных каскада и диффузии, используемый для построения практической схемы быстрых преобразований функций (сигналов). В качестве примера предложена реализация схемы дифференцирования функции, представленной в виде ряда по полиномам Лагерра.

В четвертой главе показано, что изложенные в работе методы и принципы спектральной обработки функций (сигналов) допускают естественное обобщение и в случае реализации соответствующих преобразований сигналов на вычислительных системах с параллельной архитектурой исполнения команд. При этом дополнительного ускорения преобразований функций удается добиться за счет применения принципов векторизации и конвейеризации вычислений, а также благодаря возможности естественного разделения общей схемы более сложных вычислений на самостоятельные простые компоненты (элементарные блоки).

В данное время предложенные принципы работы с ортогональными рядами реализуются в виде законченных программных продуктов, ориентированных как на широкий спектр аналитических задач, так и на некоторые узко специализированные области обработки данных. Результаты работы созданного программного обеспечения полностью согласуются с теоретическими выводами, позволяя нам в дальнейшем судить о практической выгодности использования описанных здесь подходов.

Приведенные в диссертационной работе приложения содержат рекуррентные соотношения особого вида, найденные для систем классических ортогональных полиномов непрерывного и дискретного аргументов. Данные соотношения приведены в табличной форме и требуются для реализации спектральных преобразований на практике. Отметим, что рекуррентные соотношения для полиномов Якоби и Гегенбауэра, соответствующие операторам дифференцирования и интегрирования, получены в рамках выполнения диссертационной работы.

В качестве основных выводов диссертационной работы автор видит

следующее:

1. Создана система правил для аналитических преобразований сигналов, представленных рядами через классические ортогональные полиномы - *алгебра спектральных преобразований.*
2. Задача контурного распознавания визуальных объектов разрешена на основе аналитического вычисления инвариантных признаков контуров.
3. Задача поиска тандемных повторов в геномах разрешена на основе аналитической оценки степени «периодичности» профилей.
4. Применение в задачах обработки данных правил алгебры спектральных преобразований в пространстве коэффициентов разложения способно в значительной степени улучшить и ускорить результаты вычислений.