На правах рукописи

Куликова Людмила Ивановна

СПЕКТРАЛЬНО-АНАЛИТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ ДАННЫХ И РАСПОЗНАВАНИЯ ОБРАЗОВ

Специальность: 05.13.17 – Теоретические основы информатики

Автореферат
диссертации на соискание ученой степени
кандидата физико-математических наук

Москва-2007

Работа выполнена в Институте математических проблем биологии РАН,
в филиале кафедры ММП факультета ВМиК МГУ им. М.В. Ломоносова.

Научный руководитель:

доктор технических наук,
профессор
Флоренц Федорович Дедус

Официальные оппоненты:

доктор технических наук,
профессор
Сергей Николаевич Шиманов

кандидат физико-математических наук
Юрий Викторович Чехович

Ведущая организация: Институт теоретической и экспериментальной биофизики РАН

Защита диссертации состоится «…..»…. ________________ 2007 г.
в __________ на заседании Диссертационного совета Д002.017.02 Вычислительного центра им. А.А.Дородницына РАН
по адресу: 119991, Москва, ГСП-1, ул. Вавилова, дом 40.

С диссертацией можно ознакомиться в библиотеке ВЦ РАН.

Автореферат разослан 2007 г.

Ученый секретарь
Диссертационного совета Д002.017.02
д.ф.-м.н. В.В. Рязанов

<u>Актуальность темы исследования</u>

В настоящее время наблюдается активное развитие спектральных методов обработки информационных массивов данных. Актуальность таких методов обусловлена, в первую очередь, бурным ростом объемов получаемой информации, необходимостью качественной максимально быстрой и полной ее обработки и решения проблем хранения и передачи этой информации в сжатой форме.

Обобщенный спектрально-аналитический метод (ОСАМ) дает возможность представить исследуемый массив данных в аналитическом виде, а правильный выбор ортонормированного базиса позволяет сократить отрезки ортогональных рядов, описывающих сигнал, то есть максимально сжать массив данных при минимальных потерях [Дедус Ф.Ф. и др., 1999, 2004]. Выбор оптимальных условий описания помогает решать задачу получения наиболее точного аналитического представления сигнала при минимальной глубине аппроксимации. Еще одним фактом, говорящим в пользу обобщенного спектрально-аналитического метода, является то, что предлагаемая методика обладает повышенным эффектом сглаживания высокочастотных помех, наложенных на экспериментальные данные.

Возможность проводить операции математического анализа над сигналами в пространстве коэффициентов разложения, не возвращаясь к первоначальному виду исследуемого массива, является важным свойством ОСАМ, и при определенных условиях обработка сигнала сводится к операциям получения коэффициентов разложения искомых решений через коэффициенты разложения исследуемого сигнала. Потому вывод аналитических соотношений в упомянутом пространстве является актуальным для функций, зависящих как от одной, так и двух переменных. Наличие богатой библиотеки аналитических преобразований в арсенале исследователей значительно упрощает процесс обработки данных.

Круг задач, где можно эффективно применять обобщенный спектрально-аналитический метод, очень широк. Помимо задач, связанных со сжатием данных, хранением в компактном виде, обработки и передачи в сжатом виде, имеются задачи анализа изображений и распознавания образов, параметрической идентификации систем, математической обработки генетических последовательностей, спектральной диагностики и классификации систем, прогнозирования управляющих систем, которые успешно решаются с использованием обобщенного спектрально-аналитического метода. Глубокое и всестороннее изучение обобщенного спектрально-аналитического метода способствует расширению спектра проблем, при решении которых может успешно использоваться этот инструмент.

Цель работы

Целью данной работы является создание алгоритмов обработки информационных массивов путем адаптивного аналитического описания данных и исследование спектральных подходов с применением ортонормированных систем функций для решения задач обработки экспериментальных данных и распознавания образов.

Основные задачи:

- Вывод и исследование аналитических преобразований в пространстве коэффициентов разложения, реализующих операции математического анализа над информационными массивами, с использованием ортонормированных базисов одной и двух переменных
- Реализация метода в решении прикладных задач:
  - анализ и классификация данных магнитной энцефалографии;
  - разработка алгоритма выбора оптимальных условий описания контурных объектов;
  - обработка микробиологического эксперимента.

Методы исследования

В процессе работы использовались методы теории аппроксимации, теории функций, математического и функционального анализа, линейной алгебры, теории вероятностей и математической статистики, теории регулирования, вычислительные методы информатики.

Научная новизна

- OCAM – комбинированный спектрально-аналитический метод, сочетающий в себе преимущества аналитических и численных решений. Он позволяет проводить полную обработку экспериментальных данных в пространстве коэффициентов Фурье. Определение точных аналитических соотношений между исходными коэффициентами разложения и коэффициентами, по которым определяются искомые параметры и характеристики, позволяет избежать многократных преобразований информации и воспроизведения данных в исходном виде. В данной работе пополнена математическая библиотека зависимостей для ряда распространенных операций математического анализа над сигналами для полиномов Лежандра и проведено исследование полученных аналитических соотношений. Проведены исследования по возможности и эффективности выполнения аналитического дифференцирования в условиях, когда на сигнал наложен шум. Показана устойчивость операции дифференцирования в рамках обобщенного спектрально-аналитического метода. Выведены аналитические соотношения в пространстве коэффициентов разложения сигналов, зависящих от двух переменных, для реализации некоторых математических операций над массивами данных. Реализован алгоритм описания данных магнитной энцефалографии с помощью функционального базиса сферических гармоник.

- Разработан алгоритм выбора оптимальных условий аналитического описания контурных объектов, выбора оптимальной системы координат при параметрическом описании контурных изображений, при решении задач анализа изображений и распознавания образов.
- Реализован алгоритм нахождения параметров системы, описывающей рост биомассы и потребление глюкозы.

Все основные результаты работы являются приоритетными.

Практическая и теоретическая ценность

Результаты проведенных исследований могут быть использованы при решении задач обработки экспериментальных данных, полученных в ходе научных исследований в различных областях науки, распознавания образов и анализа изображений.

Разработанные подходы для обработки двумерных функций могут быть реализованы в задачах спектральной диагностики и классификации биомедицинских систем.

Разработан и внедрен в Институт биохимии и физиологии микроорганизмов РАН (ИБФМ РАН) диалоговый пакет программ сжатия и сглаживания данных микробиологических экспериментов и параметрической идентификации модели, описывающей рост биомассы и потребления субстрата.

Комплексы программ «Спектральный анализ, классификация и диагностика цифровых массивов» и «Спектральный анализ данных, поиск неточных периодов в системах “SpectralRevisor”» имеют свидетельства об официальной регистрации программы в Российском агентстве по патентам: №2004610405 от 10.02.04 и №2007611639 от 06.05.07 соответственно.

По учебному пособию «Классические ортогональные базисы в задачах аналитического описания и обработки информационных сигналов», куда вошли результаты исследований, идет обучение студентов на факультете вычислительной математики и кибернетики Московского государственного университета им. М.В.Ломоносова и магистрантов ПущГУ. Пособие также представляет интерес для аспирантов, научных сотрудников и инженеров.

Апробация работы

Результаты исследований были доложены на межлабораторном семинаре Института математических проблем биологии РАН, а также представлены на следующих научных конференциях, симпозиумах, школах: Международном коллоквиуме «Новые информационные технологии» (Москва, 1991); VIII, IX, XI, XII Всероссийских конференциях “Математические методы распознавания образов” (Москва, 1997, 1999, 2003, 2005); IV, VI Пущинских школах молодых ученых “Задачи компьютерной биологии”, секция «Математическая и вычислительная биология» (Пущино, 1999, 2002); I Всероссийской конференции «Спектральные методы обработки информации в научных исследованиях» («Спектр-2000») (Пущино, 2000); I Национальной конференции «Информационно-вычислительные технологии в решении фундаментальных научных проблем и прикладных задач химии, биологии, фармацевтики, медицины» (Москва, 2002); V Международном конгрессе по математическому моделированию (Дубна, 2002); конференциях «Информационно-вычислительные технологии в фундаментальных и прикладных физико-математических исследованиях» (Москва, 2005, 2006); I Международной конференции «Математическая биология и биоинформатика» (Пущино, 2006); International School “Evolution, Systems Biology and high Performance Computing Bioinformatics” (Novosibirsk, 2006); XXV Межрегиональной научно-технической конференции «Проблемы эффективности и безопасности функционирования сложных технических и информационных систем» (Серпухов, 2006).

Публикации

По теме диссертации опубликовано 34 научные работы, в том числе 1 учебное пособие, получены 2 свидетельства РОСПАТЕНТ об официальной регистрации программ, 10 статей в рецензируемых журналах, книгах и трудах конференций, тезисов докладов - 21.

Структура и объем диссертационной работы

Диссертация изложена на 112 страницах машинописного текста и состоит из введения, 3 разделов, выводов и списка литературы. Список литературы состоит из 148 наименований. Работа содержит 3 таблицы, 28 рисунков.

Во введении дается обоснование актуальности работы, сформулированы цель и задачи диссертационной работы, определена научная новизна и практическая значимость полученных результатов. Описана общая структура диссертации.

Раздел 1 посвящен обзору методов обработки информационных массивов и классов решаемых задач при обработке данных.

В разделе 2 изложены основы обобщенного спектрально-аналитического метода. Рассмотрены свойства классических ортогональных полиномов, особенности разложения информационной последовательности по системе ортогональных полиномов.

Раздел 3. Результаты и обсуждение

Вывод и исследование аналитических соотношений в пространстве коэффициентов разложения для выполнения операций математического анализа над сигналами

В процессе обработки любой функции, полученной в эксперименте или в результате числовых расчетов, в рамках обобщенного спектрально-аналитического метода вычисляются коэффициенты разложения, которые несут в себе основные характеристики и свойства исходного сигнала. И чем меньше ошибка аппроксимации, тем точнее сохраняется информация об исследуемом сигнале в коэффициентах. Применение адаптивных процедур (оценка формы нескольких первых ортогональных функций с формой исследуемого сигнала для правильного выбора базиса, согласование промежутка ортогональности базисных функций с длительностью сигнала) дает возможность получить оптимальное или квазиоптимальное аналитическое описание функции. В этом случае требуемая точность разложения достигается при минимальной глубине разложения $(N = N_{\min})$.

ОСАМ позволяет проводить полную обработку экспериментальных данных в пространстве коэффициентов Фурье. Эта возможность основана на выводе аналитических зависимостей для получения искомых оценок и характеристик через коэффициенты разложения исходных сигналов. Определение точных аналитических соотношений между коэффициентами разложения и коэффициентами, по которым определяются искомые параметры и характеристики, позволяет избежать многократных преобразований информации и воспроизведения данных в первоначальном виде. Необходимость таких аналитических соотношений очевидна. Потому банк формул постоянно пополняется, существует своя библиотека для каждого базиса.

- ***Аналитические зависимости для функциональных базисов одной переменной.***

Выведены аналитические соотношения в пространстве коэффициентов разложения (табл. 1) и помещены в библиотеку зависимостей для ряда распространенных действий (операции математического анализа) над сигналами для полиномов Лежандра. Проведено исследование полученных аналитических преобразований.

Таблица 1.

Наименование операции	Результирующая формула	Аналитические преобразования в пространстве коэффициентов
Суммирование вычитание сигналов $z(t)=x(t)\pm y(t)$	$z(t)=\sum_{n=0}^{N_1} C_n P_n(t)$	$C_n = A_n \pm B_n;$ $N_1 > N_2,$ $B_n \equiv 0, \quad при \quad n > N_2$
Вычисление производной $z(t)=[x(t)]^{/}$	$z(t)=\sum_{n=0}^{N_1-1} C_n P_n(t)$	$C_n = \frac{2}{T}\sum_{k=n+1}^{N}\sqrt{(2k+1)(2n+1)}A_k$ $k=n+1,...$ *только нечетные при четном n* *только четные при нечетном n*
Вычисление интеграла $z(t)=\int_0^{\tau} x(\tau)d\tau$	$z(t)=\sum_{n=0}^{N_1+1} C_n P_n(t)$	$C_0 = \frac{T}{2}\left[A_0 - \frac{A_1}{\sqrt{3}}\right],$ $C_k = \frac{T}{2\sqrt{2k+1}}\left[\frac{A_{k-1}}{\sqrt{2k-1}} - \frac{A_{k+1}}{\sqrt{2k+3}}\right], \quad k \ge 1$
Умножение сигналов $z(t)=x(t)\times y(t)$	$z(t)=\sum_{n=0}^{N_1+N_2} C_n P_n(t)$	$C_n = \sum_{i=0}^{N_1}\sum_{j=0}^{N_2}\delta_{ijn}A_i B_j, \quad где \quad \delta_{ijn} = \int_0^T P_i P_j P_n dt \quad при$ 1) $n - четн. \to (i+j) - четн.;$ $n - нечетн. \to (i+j) - нечетн.;$ 2) $\lvert i-j \rvert \le n;$ 3) $(i+j) \ge n$

- ***Вычисление производных в условиях высокочастотного шума***

Исследована возможность и эффективность вычисления первой, второй и производных большего порядка исследуемых функций с использованием выведенной формулы. Если обрабатываемый сигнал является приближенным, задача вычисления производной может стать некорректной, что приводит к затруднениям при ее вычислении. При аналитическом дифференцировании сигнала, как, в прочем, и при выполнении других операций математического анализа в пространстве коэффициентов разложения, на который наложен высокочастотный шум, необходимо решение проблемы оптимального разложения отрезком ортогонального ряда с целью подавления шума, с тем, чтобы коэффициенты разложения содержали в себе свойства чистого сигнала.

- ***Исследование меры обусловленности оператора дифференцирования для семейств ортогональных полиномов.***

В результате работы показано, что оператор дифференцирования для семейств ортогональных полиномов хорошо обусловлен. Адаптивные процедуры при решении аналитического описания исследуемой функции способствуют улучшению обусловленности. Исследования показали, операция дифференцирования в рамках предлагаемого метода устойчива.

- ***Аналитические соотношения для функционального базиса сферических функций двух переменных***

Существуют задачи, приводящие к необходимости использования для описания данных функциональные базисы двух и более переменных.

Сферические функции или сферические гармоники являются важным классом специальных функций, тесно связанных с классическими ортогональными полиномами. Они возникают при решении широкого круга проблем, например, при решении уравнения Лапласа в сферических координатах. Сферические функции имеют общий вид:

$$Y_{lm}(\theta,\varphi) = \sqrt{\frac{2l+1}{4\pi}\cdot\frac{(l-m)!}{(l+m)!}e^{im\varphi}}\,P_l^m(\cos\theta),$$

здесь $P_l^m(x) = (1-x^2)^{m/2}\dfrac{d^m P_l(x)}{dx^m}$

Функции $p_l(x)$ - функции или полиномы Лежандра, называются еще зональными гармониками. Функции $p_l^m(x)$ называются присоединенными функциями Лежандра. Функции $p_n^m \cos m\varphi$, $p_n^m \sin m\varphi$ называются тессеральными гармониками. При $m = l$ тессеральные гармоники называются секторальными гармониками. Присоединенные функции Лежандра обладают свойством ортогональности. Из теории специальных функций известно, что

$$\int_{-1}^{1} P_{nm}(x)P_{kl}(x)dx = \frac{(n+m)!}{(n-m)!}\frac{\delta_{nk}\delta_{ml}}{2n+1}.$$

То же справедливо и для поверхностных гармоник одинаковой степени, любая пара $p_l^m \cos m\varphi$, $p_l^m \sin m\varphi$, $p_l^s \cos s\varphi$, $p_l^s \sin s\varphi$ ортогональна, интеграл от произведения их по φ обращается в нуль, за исключением $m = s$, когда берется интеграл от квадрата гармоники, беря в обоих множителях либо косинус, либо синус. Стандартные гармоники $Y_{lm}(\theta,\varphi)$ линейно независимы.

Система сферических гармоник является полной в пространстве интегрируемых с квадратом функций $f(\theta,\varphi)$, то есть исследуемая функция может быть разложена в ряд по сферическим гармоникам

$$f(\theta,\varphi) = \sum_{n=0}^{\infty}\sum_{s=0}^{n}(a_{ns}p_n^s \cos s\varphi + b_{ns}p_n^s \sin s\varphi)\cdot$$

Это разложение обладает свойством, аналогичным свойству разложения Фурье или любого другого разложения по ортогональным функциям.

Учитывая свойство ортогональности

$$\int_0^{\pi}\int_0^{2\pi}(p_n)^2\sin\theta d\theta d\varphi=\frac{4\pi}{2n+1},$$

$$\int_0^{\pi}\int_0^{2\pi}\left(p_n^k\right)^2(\cos^2 k\varphi,\sin^2 k\varphi)\sin\theta\ d\mu d\varphi=\frac{2\pi}{2n+1}\cdot\frac{(n+k)!}{(n-k)!},$$

вычисляются коэффициенты разложения a_{nk}, b_{nk}.

Имеется также аналог теоремы Парсеваля

$$\iint[f(\theta,\varphi)]^2 d\mu d\varphi=\sum_{n=0}^{\infty}\sum_{s=0}^{n}a_{ns}^2\iint\left(p_n^s\right)^2\cos^2 s\varphi\ d\mu d\varphi+\sum_{n=0}^{\infty}\sum_{s=0}^{n}b_{ns}^2\iint\left(p_n^s\right)^2\sin^2 s\varphi\ d\mu d\varphi$$

Погрешность аппроксимации функции определяется по формуле

$$\delta=1-\frac{\sum_{n=0}^{N}\sum_{s=0}^{n}a_{ns}^2\iint\left(p_n^s\right)^2\cos^2 s\varphi\ d\mu d\varphi+\sum_{n=0}^{N}\sum_{s=0}^{n}b_{ns}^2\iint\left(p_n^s\right)^2\sin^2 s\varphi\ d\mu d\varphi}{\int_0^{\pi}\int_0^{2\pi}f^2(\theta,\varphi)\sin\theta d\theta d\varphi}.$$

Для присоединенных полиномов Лежандра существуют явные выражения, Однако при больших значениях n приходится оперировать очень большими числами, что приводит к потере точности, переполнениям регистра и прочим неудобствам. Поэтому имеет смысл воспользоваться рекуррентными соотношениями для присоединенных полиномов Лежандра:

$$(l-m)P_l^m=x(2l-1)P_{l-1}^m-(l+m-1)P_{l-2}^m,$$

$$P_m^m=(-1)^m(2m-1)!!(1-x^2)^{m/2},\qquad P_{m+1}^m=x(2m+1)P_m^m.$$

Следуя идеологии ОСАМ, возникает задача выполнения алгебраических операций над сферическими функциями и вывод формул для функций математического анализа в пространстве коэффициентов разложения функций, заданных на сфере, например, для получения производных от функции двух переменных $\frac{\partial f(\theta,\varphi)}{\partial\theta}$, $\frac{\partial f(\theta,\varphi)}{\partial\varphi}$. В ходе исследований выведено аналитическое соотношение для коэффициентов разложения производной функции через коэффициенты разложения самой функции:

$$c_{l0}=\frac{l(l+1)}{2}a_{l1},$$

$$c_{ln}=0.5(l+n+1)(l-n)a_{l\,n+1}-0.5a_{l\,n-1},\ n=\overline{1,l-1}.$$

Аналитические соотношения для других операций математического анализа над функциями двух переменных, в основном, аналогичны выведенным формулам для функций одной переменной. Они вошли в созданную и интегрированную в ОСАМ библиотеку математических операций для функционального базиса двух переменных.

- ***Обработка с помощью обобщенного спектрально-аналитического метода данных магнитной энцефалографии***

В качестве тестирования предлагаемого метода аналитического описания двумерных сигналов рассматривались модельные данные, полученные в расчетах из магнитной энцефалографии (МЭГ). Данное направление исследований – интенсивно развивающаяся область экспериментального изучения высшей нервной деятельности человека, функциональных областей мозга и диагностики различных патологий. Это обусловлено возможностью неинвазивного получения данных о процессах, происходящих как в коре головного мозга, так и в глубоких его отделах. Магнитное поле, в сравнение с электрическим, испытывает значительно меньшие искажения на внутричерепных неоднородностях и покрывающих тканях, что существенно повышает точность локализации источников и снижает требования к знанию структуры внутричерепной среды.

Величина измеряемого магнитного поля, если датчик находится в точке $\mathbf{r}$ и имеет направление нормали $\mathbf{n}$, определяется следующим образом (Sarvas J, 1998):

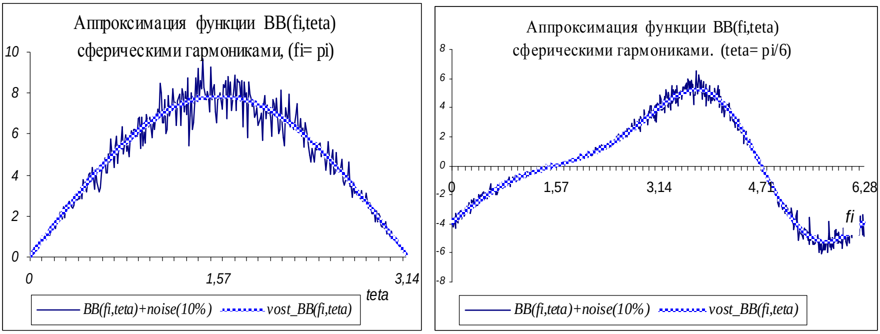
$$B(\mathbf{r}) = \frac{\mu_0}{4\pi F^2}\left(\left(F(\mathbf{Q}\times\mathbf{r}_0) - (\mathbf{Q}\times\mathbf{r}_0, \mathbf{r})\nabla F\right), \mathbf{n}\right),$$

здесь $F = a\left(ar + r^2 - (\mathbf{r}_0, \mathbf{r})\right)$,

$$\nabla F = \left(a^2 r^{-1} + a^{-1}(\mathbf{a}, \mathbf{r}) + 2a + 2r\right)\mathbf{r} - \left(a + 2r + a^{-1}(\mathbf{a}, \mathbf{r})\right)\mathbf{r}_0,$$

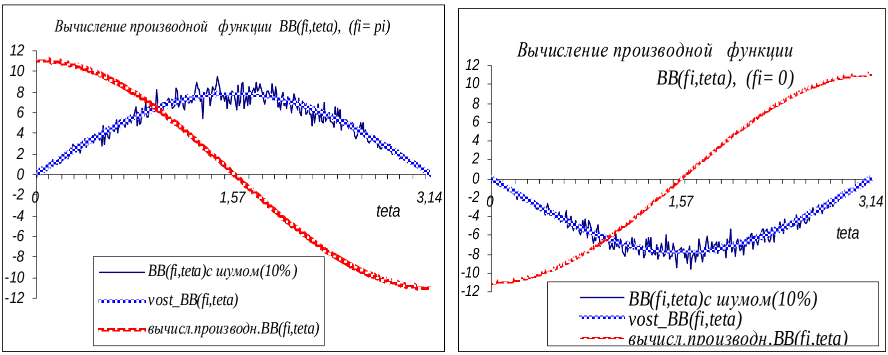
$$\mathbf{a} = \mathbf{r} - \mathbf{r}_0,\ a = |\mathbf{a}|,\ r = |\mathbf{r}|,\ |\mathbf{n}| = 1.$$

Источники магнитной активности головного мозга моделируются точечными токовыми диполями. Каждый такой диполь характеризуется двумя векторами: $\mathbf{r}_0$– радиус-вектор диполя (положение диполя) и $\mathbf{Q}$ – вектор, задающий направление и силу диполя (момент диполя). Для вычисления магнитной индукции на поверхности головы используется модель токового диполя в проводящей сфере. На рис. 1 приводятся результаты аппроксимации пространственного распределения биомагнитного поля сферическими гармониками, на рис. 2. – результат вычисления производной функции магнитного поля.



а) б)

Рис.1. Аппроксимация функции магнитного поля сферическими гармониками. Уровень шума, наложенного на функцию равен 10%. Изображение кривой $BB(\theta,\varphi)$ *при фиксированной переменной: а)* $\varphi=\pi$ *; б)* $\theta=\pi/6$.



а) б)

Рис.2. Вычисление производной функции магнитного поля при фиксированной переменной: а) $\varphi=\pi$ *; б)* $\varphi=0$.

Предлагаемый подход аналитического описания функций двух переменных применяется при решении задач спектральной диагностики и классификации биомедицинских систем. Реализованный алгоритм является составной частью пакета программ «Спектральный анализ, классификация и диагностика цифровых массивов SpectMate», используемый при диагностике таких патологий, как паркинсонизм и болезнь Альцгеймера.

- ***Использование обобщенного спектрально-аналитического метода в задачах анализа изображений и распознавания образов:***

выбор оптимальных условий аналитического описания контурных объектов

Использование обобщенного спектрально-аналитического метода в задачах анализа изображений и распознавания образов имеет свои преимущества. Учитывая, что пространство коэффициентов разложения при правильном выборе системы координат и соответствующего ортогонального базиса отражает все существенные для анализа характеристики изучаемого объекта [Дедус Ф.Ф., Махортых С.А. и др., 1999], задача сводится к выявлению и использованию тех коэффициентов разложения, которые являются наиболее информативными. Коэффициенты разложения выступают в данном случае в качестве признаков объекта. В процессе исследований, выделяются наиболее информативные (оптимальные) наборы признаков для данного объекта и, таким образом, сформировать из исходного пространства коэффициентов признаковое пространство меньшей размерности.

Выбор оптимальной системы координат при параметрическом описании контурных изображений имеет целью получить аналитическое описание исследуемого изображения возможно более простым, то есть отрезки ортогональных рядов, описывающих проекции изображения, должны быть по возможности наиболее короткими. Если функция $r(t)$, задающая кривую на плоскости, представлена в системе координат **G** параметрическими уравнениями

$$\mathbf{r}_G(t) = \{x_G(t), y_G(t)$$

в виде цифровых массивов, данная система описывается аналитически набором базисных функций ***B={X*****i*****(t)}**

$$\begin{cases} x_{G,B}(t) = \sum_{i=1}^{N_{G,B}^x} A_i X_i(t),\ y_{G,B}(t) = \sum_{i=1}^{N_{G,B}^y} B_i X_i(t). \end{cases}$$

Тогда, если понимать точность описания в определенном смысле (например, в среднеквадратичном), требуется выбрать такую систему координат **G** и такой ортогональный базис **B**, для которых величины $N_{G,B}^{x,y}$ минимальны при одной и той же точности описания.

Для выбора оптимальных геометрических координат построен алгоритм оценки информативности базиса на примере реализации обобщенного спектрального подхода в системе распознавания букв латинского алфавита. Условие оптимальности геометрического описания формулируется в следующем виде: из всех имеющихся вариантов базисных ортогональных функций ***B={X*****i*****(t)}** и систем координат {**G**} используем такое описание, для которого имеет место максимум функционала

$$\max_{G,B}\left(\sum_{\substack{i,j,l,m \\ i\neq l}} \sum_{\gamma=1}^{N_{G,B}^x+N_{G,B}^y} (A_{ij,\gamma}^{G,B} - A_{lm,\gamma}^{G,B})^2 - \alpha \sum_{\substack{i,j,m \\ j\neq m}} \sum_{\gamma=1}^{N_{G,B}^x+N_{G,B}^y} (A_{ij,\gamma}^{G,B} - A_{im,\gamma}^{G,B})^2 \right).$$

При реализации обучающей процедуры брали различные выборки букв, каждая из которых была представлена несколькими изображениями, отличающимися между собой выбором шрифта, размера, наклона, либо уровнем шума, наложенного на изображение (фрагмент рассматриваемых изображений представлен на рис.3).

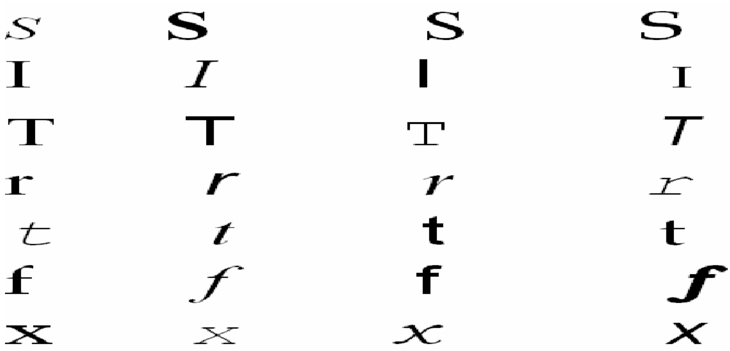


Рис.3.

Рассматриваются параметризации контурных изображений следующего вида:

$$\left\{ x(t)=\sum_{i=0}^{N_1} A_i T_i(t),\ \ y(t)=\sum_{i=0}^{N_1} B_i T_i(t) \right.$$ - в декартовых координатах,

$$\left\{ \rho(t)=\sum_{i=0}^{N_2} C_i T_i(t), \theta(t)=\sum_{i=0}^{N_2} D_i T_i(t) \right.$$ - в полярных координатах,

здесь ρ и θ - радиальная и угловая координаты, суммы представляют собой разложение проекций по ортогональным полиномам. В качестве признаков используются величины $A_1, A_2, ...A_{N_1}, B_1, ..., B_{N_1}, C_0, ..., C_{N_2}, D_0, ...D_{N_2}$ - всего $2(N_1+N_2+1)$ признаков. Предложенная процедура в ходе работы определяла в качестве оптимальной систему координат (полярную или декартову) и проводила распознавание рассматриваемых букв. На рисунке 4 представлен пример решения задачи для случая, когда рассматривались две буквы S и Z, каждая из которых имеет по три различных изображения (см. 1 и 3 квадранты). В данном случае функционал достигает максимума при $\gamma=40,\ \ \gamma=42$, что соответствует 4 и 6 коэффициентам разложения в полярных координатах. Таким образом, процедура в качестве оптимальной предложила полярную систему координат. Точки, соответствующие исследуемым изображениям, расположились на оптимальной признаковой плоскости следующим образом: точки, соответствующие S, - в квадранте 2, Z – в квадранте 4. Следовательно, уже по двум признакам система распознала предложенные буквы.

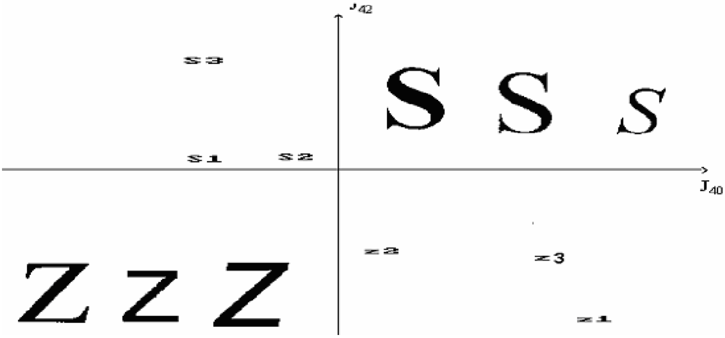


Рис.4.

Разработанный алгоритм позволяет существенно сократить объем вычислений за счет ступенчатой реализации.

- ***Обработка данных микробиологического эксперимента***

Современная микробиология достигла такого уровня, на котором применение математического подхода для точного анализа биологических явлений становится необходимым. Использование математических методов дает возможность не только глубже понять внутреннюю природу микробиологических процессов, но и по-новому взглянуть на некоторые определившиеся проблемы.

Объектом исследований были медленнорастущие микроорганизмы, называемые *к-тактиками* (напр., Rhodococcus minimus 293, Artrobacter globiformis 94А). Исследуемая группа микроорганизмов способна к утилизации многих органических и неорганических соединений, в том числе и токсичных. Поэтому изучение метаболизма данной группы микроорганизмов является чрезвычайно перспективным.

Культура выращивалась в условиях хемостата с постоянной скоростью протока. В ходе экспериментов измерялись четыре величины: содержание биомассы ***X(t),*** содержание белка ***P(t),*** оптическая плотность среды и концентрация глюкозы ***S(t)*** в среде**.** Результаты измерений, представленные в виде графиков и таблиц, явились исходными данными для математической обработки. Кривые были аппроксимированы по полиномам Чебышева и Лежандра. В силу сложности проведения опытов требовалось применять разную степень точности. Для определения оптимальной степени разложения был применен метод Форсайта. В ходе совместной работы был разработан диалоговый пакет программ оптимального сжатия и сглаживания экспериментальных данных.

Исследование изменения концентрации субстрата после одноразового его введения в систему с постоянной скоростью протока послужило модельной задачей для изучения метаболических изменений медленнорастущих микроорганизмов. Математическая теория непрерывного культивирования позволяет определить общие закономерности кинетики процессов биосинтеза в проточных культурах. Основы теории заложены Моно (Monod, 1950), Новиком и Сцилардом (Novick, Szilard, 1950) и получили дальнейшее развитие во множестве других работ (Гительзон, 1965; Иерусалимский, 1966; и др.).

При проточном культивировании динамика концентрации биомассы определяется не только приростом, но и вымыванием части биомассы со скоростью разбавления ***D***.

$$dX / dt = \mu X - DX$$
$$dS / dt = -DS - \mu X / Y \quad (1)$$

Второе уравнение данной системы определяет динамику концентрации субстрата при проточном культивировании, где μ - удельная скорость роста биомассы, ***Y*** - экономический коэффициент, равный количеству субстрата, израсходованному на единицу образовавшейся биомассы.

Для постоянных скорости роста μ и экономического коэффициента Y решения системы очевидны:

$$X = X_0 e^{(\mu - D)t}$$
$$S = (S_0 + X_0 / Y)e^{-Dt} - \frac{X_0}{Y} e^{(\mu - D)t} \quad (2)$$

Если величины X_0, S_0, D и зависимости $x(t)$, $S(t)$ (сглаженные экспериментальные кривые) известны, из выражений (2) легко находятся μ и Y. Для частного случая постоянной в ходе эксперимента концентрации биомассы, модель (2) для глюкозы упрощается и приводится к виду:

$$S(t) = (S_0 + g\ X / D)e^{-Dt} - g\ X / D \quad (3)$$

Дальнейшая работа с уже сглаженными кривыми, по сути, сводилась к параметрической идентификации моделей типа приведенных выше (1-3). А именно, по результатам сглаживания экспериментальных данных были найдены временные зависимости скорости роста биомассы μ***(t)*** экономического ***Y(t)*** и метаболического ***g(t)*** коэффициентов. На рис. 5 и 6 представлены результаты аппроксимации экспериментальных данных, рис. 7 - результат уточнения ненаблюдаемого в ходе эксперимента, но важного параметра: экономического коэффициента.

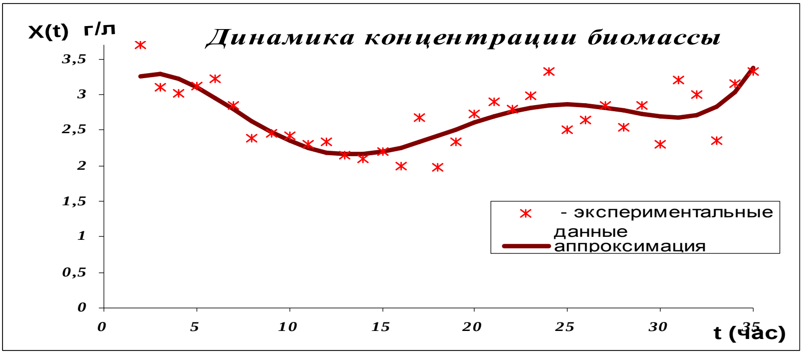


Рис.5.

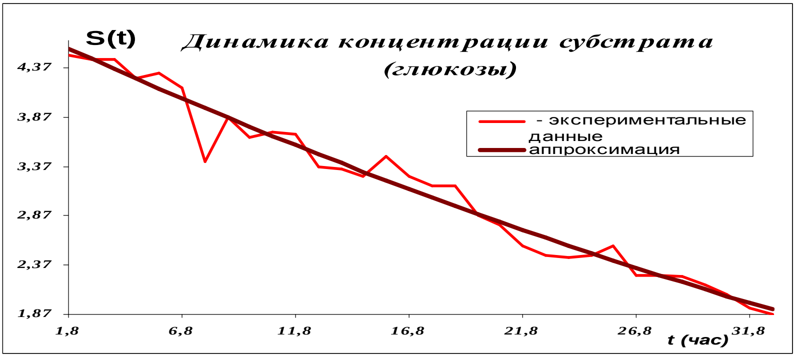


Рис.6.

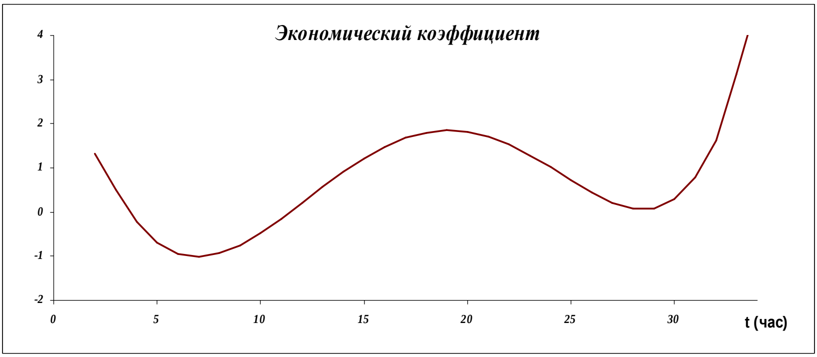


Рис.7.

В ходе решения задачи применялись результаты теоретических выводов: аналитические преобразования в пространстве коэффициентов разложения.

В диссертации также изложены решения еще ряда других прикладных задач, где реализованы предложенные алгоритмы. Данная работа выполнялась с использованием языков программирования C++, FORTRAN, Mathematica.

ВЫВОДЫ

1. Получены аналитические соотношения, реализующие алгебраические операции в пространстве коэффициентов разложения с использованием ортонормированного базиса Лежандра. Показана эффективность использования банка аналитических преобразований. Показана возможность аналитического дифференцирования сигналов в условиях высокочастотного шума.
2. Реализован алгоритм описания данных магнитной энцефалографии с помощью базиса сферических гармоник. Получены аналитические преобразования в пространстве коэффициентов разложения сигналов, зависящих от двух переменных, для реализации некоторых математических операций над массивами данных.
3. Реализован обобщенный спектральный подход в системе распознавания текста. Предложен алгоритм выбора оптимальной системы координат при параметрическом описании контурных изображений в задачах анализа изображений и распознавания образов.
4. Реализован алгоритм оптимального сжатия данных микробиологического эксперимента в рамках обобщенного спектрально-аналитического метода. Проведена параметрическая идентификация математической модели, описывающей рост биомассы и потребление глюкозы.

Список публикаций по теме диссертации

1. Куликова Л.И., Махортых С.А., Устинин М.Н., Чемерис Н.А. Об одном подходе к математической обработке данных по культивированию микроорганизмов и интерпретации полученных результатов. // Материалы Международного коллоквиума «Новые информационные технологии», Москва, октябрь 1991, с.257-259.
2. Дедус Ф.Ф., Куликова Л.И., Махортых С.А., Устинин М.Н., Чемерис Н.А. Обработка данных микробиологического эксперимента и модель биосинтеза. // Материалы Международного симпозиума «Математические теории биологических процессов», Калининград, октябрь 1993, с.86-93.
3. Васильева Т.А., Дедус Ф.Ф., Куликова Л.И. Вывод аналитических зависимостей в пространстве коэффициентов разложения для произведения и возведения в степень исходных сигналов. // VIII Всероссийская конференция “Математические методы распознавания образов”, Москва, 1997, тезисы докладов конференции, с.14-15.
4. Горошникова Т.А., Дедус Ф.Ф., Куликова Л.И. О создании банка формул в пространстве коэффициентов разложения для решения задач распознавания

в широком смысле. // Математические методы распознавания образов. Доклады IX Всероссийской конференции. М., 1999. С.28-30.
5. Горошникова Т.А., Дедус Ф.Ф., Куликова Л.И., Махортых С.А., Панкратов А.Н., Устинин М.Н. Обобщенный спектрально-аналитический метод – универсальная вычислительная технология. // Задачи компьютерной биологии. Тезисы докладов 4-й Пущинской школы молодых ученых, секция «Математическая и вычислительная биология», г. Пущино, 1999, с.4.
6. Горошникова Т.А., Дедус Ф.Ф., Куликова Л.И., Махортых С.А. Спектральные представления решений начальных и краевых задач математической физики // Задачи компьютерной биологии. Тезисы докладов 4-й Пущинской школы молодых ученых, секция «Математическая и вычислительная биология», 19 – 23 апреля 1999 г., г.Пущино, 1999, с. 5.
7. Дедус Ф.Ф., Куликова Л.И., Махортых С.А., Горошникова Т.А. Анализ эффективности и точности полученных соотношений между коэффициентами разложения при типовых операциях с сигналами. В сборнике докладов I Всероссийской конференции «Спектральные методы обработки информации в научных исследованиях» («Спектр-2000»), Пущино, 25-27 октября 2000г, с.124-128.
8. Панкратов А.Н., Куликова Л.И., Дедус Ф.Ф., Махортых С.А. Использование обобщенного спектрально-аналитического метода в задачах обработки экспериментальных данных. // Задачи компьютерной биологии. Тезисы докладов 6ой Пущинской школы молодых ученых, секция «Математическая и вычислительная биология», Пущино, 2002г, с. 187.
9. Куликова Л.И., Дедус Ф.Ф., Махортых С.А. Использование новой информационной технологии обобщенного спектрально-аналитического метода в задачах обработки экспериментальных данных. // Первая Национальная Конференция «Информационно-вычислительные технологии в решении фундаментальных научных проблем и прикладных задач химии, биологии, фармацевтики, медицины». Сборник тезисов докладов, Москва , 2002, с.128-129.
10. Куликова Л.И., Дедус Ф.Ф., Махортых С.А. Обобщенный спектрально–аналитический метод в задачах обработки экспериментальных данных. // В кн. «Горизонты биофизики. От теории к практике (Научные исследования в наукоградах Московской области)». Ответственный редактор чл.–корр. РАН Г.Р.Иваницкий. Пущино, 2003, с.56-62.
11. F.F. Dedus, L.I. Kulikova, S.A. Makhortykh and A.N. Pankratov. Adjustable spectral approaches for biological systems analysis. // V International congress on mathematical modeling, Dubna, september30-October 6, 2002, p. 193.

12. Дедус Ф.Ф., Куликова Л.И., Махортых С.А. Обобщенный спектрально-аналитический метод в применении к задачам обработки экспериментальных данных. // Конференция «От современной фундаментальной биологии к новым наукоемким технологиям». Сборник тезисов, Пущино, 2002, с.65-66.
13. Куликова Л.И., Дедус Ф.Ф., Махортых С.А. Использование новой информационной технологии обобщенного спектрально-аналитического метода в задачах обработки экспериментальных данных. // I Национальная Конференция «Информационно-вычислительные технологии в решении фундаментальных научных проблем и прикладных задач химии, биологии, фармацевтики, медицины». М., 2002, статья в электронной форме http://wwwncgroup.ru/conference/enter/papes.
14. Куликова Л.И., Дедус Ф.Ф., Махортых С.А. Аналитическое описание зашумленных данных и вычисление производных. // II Национальная Конференция «Информационно-вычислительные технологии в решении фундаментальных научных проблем и прикладных задач химии, биологии, фармацевтики, медицины». Сборник материалов, Москва, 2003, с.29.
15. Куликова Л.И. Предварительная обработка данных в процессе их аналитического описания с помощью ортогональных рядов. // XI Всероссийская конференция “Математические методы распознавания образов” (ММРО-11), Москва, 2003, Сборник докладов конференции, с.124-127.
16. Дедус Ф.Ф., Куликова Л.И., Панкратов А.Н., Тетуев Р.К. Классические ортогональные базисы в задачах аналитического описания и обработки информационных сигналов. // Учебное пособие**,** Издательский отдел ВМиК МГУ, М., 2004, 147с.
17. Махортых С.А., Дергузов А.В., Куликова Л.И., Панкратов А.Н. Спектральный анализ, классификация и диагностика цифровых массивов SpectMate. Свидетельство РОСПАТЕНТ об официальной регистрации программы №2004610405 от 10.02.2004.
18. Махортых С.А., Куликова Л.И. Обобщенный спектральный подход в системах распознавания текстов (OCR). // Вестник компьютерных и информационных технологий. - М., № 12, 2005, с. 5-9.
19. Куликова Л.И., Махортых С.А. Спектральное разложение данных магнитной энцефалографии и вывод аналитических соотношений между коэффициентами разложения при типовых операциях с сигналами. // 12-я Всероссийская конференция “Математические методы распознавания образов” (ММРО-12), М., 2005, Сборник докладов конференции, с. 360-362.

20. Куликова Л.И., Махортых С.А. О спектральном алгоритме распознавания контурных изображений. // Сборник материалов конференции «Информационно-вычислительные технологии в фундаментальных и прикладных физико-математических исследованиях» (ИВТН-2005), М., 2005, с.50.
21. Дергузов А.В., Куликова Л.И., Махортых С.А. Спектральные методы коррекции неполноты данных. // Сборник материалов конференции «Информационно-вычислительные технологии в решении фундаментальных научных проблем и прикладных задач химии, биологии, фармацевтики, медицины» (ИВТН-2005), М., 2005, с. 33. http://www.ivtn.ru/2005/biomedchem/enter/paper.php?p=313
22. Куликова Л.И., Махортых С.А. Выбор оптимальных условий аналитического описания контурных объектов. // Электронный журнал "Исследовано в России", 2005, 144, с. 1501-1511. http://zhurnal.ape.relarn.ru/articles/2005/144.pdf
23. Дедус Ф.Ф., Куликова Л.И., Махортых С.А., Назипова Н.Н., Панкратов А.Н, Тетуев Р.К. Аналитические методы распознавания повторяющихся структур в геномах. // ДОКЛАДЫ АКАДЕМИИ НАУК, 2006, том 411, № 5, с.599-602. F. F. Dedus, L. I. Kulikova, S. A. Makhortykh, N. N. Nazipova, A. N. Pankratov, and R. K. Tetuev Analytical Recognition Methods for Repeated Structures in Genomes// Doklady Mathematics, Vol. 74, № 3, 2006, pp. 926-929.
24. Куликова Л.И., Махортых С.А. Математические операции над двумерными сигналами в базисах сферических гармоник. // Электронный журнал "Исследовано в России", 2006, 60, с.598-608. http://zhurnal.ape.relarn.ru/articles/2006/060.pdf.
25. Дедус Ф.Ф., Куликова Л.И., Тетуев Р.К. Аналитическое дифференцирование сигналов в условиях высокочастотных помех. // Труды XXV Межрегиональной научно-технической конференции «Проблемы эффективности и безопасности функционирования сложных технических и информационных систем», Серпухов, октябрь 2006, 164-167.
26. Панкратов А.Н., Куликова Л.И Способы вычисления производных при равномерном и среднеквадратичном приближении сигналов. // Сборник материалов конференции «Информационно-вычислительные технологии в фундаментальных и прикладных физико-математических исследованиях» (ИВТН-2006), М., 2006, с. 56.
27. Куликова Л.И., Махортых С.А. Спектральная обработка двумерных сигналов и типовые операции в пространстве коэффициентов разложения. // Сборник материалов конференции «Информационно-вычислительные

технологии в фундаментальных и прикладных физико-математических исследованиях» (ИВТН-2006), М., 2006, с. 29.

28. Ruslan Tetuev, Florencz Dedus, Ljudmila Kulikova, Sergey Makhortikh, Anton Pankratov, Nafisa Nazipova “Analytical methods in problems of recognition the structural and functional organization of genetic sequences”, The 2006 BGRS (Bioinformatics of the Genome Regulation and Structure) International Summer School for young scientist “Evolution, Systems Biology and high Performance Computing Bioinformatics”, Novosibirsk, Russia, July 12-15, 2006.
29. Куликова Л.И., Дедус Ф.Ф., Панкратов А.Н, Тетуев Р.К. Обобщенный спектрально-аналитический в задачах исследования динамических систем. // Доклады 1 Международной конференции «Математическая биология и биоинформатика», М.:МАКС Пресс, 2006, с.216-217.
30. Дедус Ф.Ф., Куликова Л.И., Махортых С.А., Назипова Н.Н., Панкратов А.Н., Тетуев Р.К. Аналитические методы в проблемах распознавания структурно-функциональной организации генетических последовательностей. // Доклады 1 Международной конференции «Математическая биология и биоинформатика», М.:МАКС Пресс, 2006, с.173-174.
31. Дедус Ф.Ф., Куликова Л.И., Сулейменова Б.О. Вычисление производных высокого порядка зашумленных сигналов. // Доклады 1 Международной конференции «Математическая биология и биоинформатика», М.:МАКС Пресс, 2006, с. 206.
32. Дедус Ф.Ф., Куликова Л.И., Махортых С.А., Панкратов А.Н, Тетуев Р.К. Спектрально-аналитический метод в задачах математической биологии. // Доклады 1 Международной конференции «Математическая биология и биоинформатика», М.:МАКС Пресс, 2006, с. 175-176.
33. Дедус Ф.Ф., Куликова Л.И., Махортых С.А., Назипова Н.Н., Панкратов А.Н, Тетуев Р.К. Обработка геномных последовательностей в рамках ОСАМ // Вестник МГУ, 2007, №2, с. 12-16.
    F.F. Dedus, L.I. Kulikova, S.A. Makhortykh, N.N. Nazipova, A.N. Pankratov, R.K. Tetuev. Recognition of the Structural-Functional Organization of Genetic Sequences. // Moscow University Computational Mathematics and Cybernetics, 2007, Vol. 31, №2, pp.49-53.
34. Дедус Ф.Ф., Тетуев Р.К., Назипова Н.Н., Куликова Л.И., Махортых С.А., Панкратов А.Н, Ольшевец М.М. Спектральный анализ данных, поиск неточных периодов в системах “SpectralRevisor”. Свидетельство РОСПАТЕНТ об официальной регистрации программы №2007611639 от19.04.2007.