

На правах рукописи



Никитин Дмитрий Александрович

АВТОМАТИЧЕСКАЯ ИНТЕРПОЛЯЦИЯ ЧИСЛОВЫХ ДАННЫХ
ФУНКЦИЯМИ ИЗ ЗАДАННОГО МНОЖЕСТВА С НАИМЕНЬШИМ
КОЛИЧЕСТВОМ ПАРАМЕТРОВ

05.13.17 — Теоретические основы информатики

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата физико-математических наук

Красноярск 2011

Работа выполнена в ФГБОУ ВПО «Сибирский государственный аэрокосмический университет имени академика М.Ф. Решетнева».

Научный руководитель: доктор физико-математических наук,
доцент Сафонов Константин Владимирович

Официальные оппоненты: доктор физико-математических наук,
профессор Винокуров Сергей Фёдорович

доктор физико-математических наук,
профессор Носков Михаил Валерианович

Ведущая организация: Институт проблем управления
им. В.А. Трапезникова РАН, г. Москва.

Защита состоится «28» октября 2011 г. в 16 .00 часов на заседании диссертационного совета Д 212.099.11 при Сибирском федеральном университете по адресу 660074, г. Красноярск, ул. Киренского, 26, ауд. УЛК 115.

С диссертацией можно ознакомиться в библиотеке Сибирского федерального университета.

Автореферат разослан «26» сентября 2011 г.

Ученый секретарь диссертационного совета
кандидат технических наук, доцент



Л. И. Покидышева

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Существует ряд классических методов построения функций, график которых точно проходит через заданный набор точек. Это различные интерполяционные формулы, а также различные дискретные преобразования. Полученную функцию (интерполянт) можно использовать для вычисления значений функции между исходными точками (интерполяция), за их пределами (экстраполяция), а также компактной записи или дальнейшего анализа исходных данных. Для интерполяции и экстраполяции также могут использоваться разностные схемы, хотя они и не дают точной записи интерполянта.

В качестве интерполянтов чаще всего используются многочлены, суммы экспонент, Фурье-суммы, сплайны. Количество слагаемых в указанных функциях определяется количеством исходных точек. Вид интерполянта в конкретной задаче обычно выбирается на основании дополнительной информации о зависимости между исходными значениями. То есть упомянутые методы не предназначены для поиска наиболее подходящей модели данных — каждый из них уже предполагает конкретную модель. Кроме того, классические методы не предназначены для построения интерполянта в виде сумм функций из разных классов (например, сумма экспонент и синусоид).

Таким образом, классические методы построения функций, график которых точно проходит через заданный набор точек, удобно использовать, когда вид интерполянта заранее выбран и остаётся только определить его параметры (коэффициенты). В случае если вид интерполянта заранее неизвестен, либо в заданном множестве функций необходимо определить интерполирующую функцию с наименьшим количеством параметров, то использование данных методов становится затруднительным, так как для каждого отдельного вида функции приходится выполнять отдельную процедуру построения интерполянта.

При обработке больших наборов данных в различных областях науки и техники кроме методов интерполяции не менее часто используются методы аппроксимации, образующие вместе группу методов приближения функций. Это связано с тем, что в значениях измеряемой величины практически всегда присутствует погрешность. Методы аппроксимации могут быть как модификациями методов интерполяции (например, аппроксимирующий сплайн), так и независимыми методами.

В связи с этим разработка метода, позволяющего автоматически определять из множества одномерных известных функциональных зависимостей такую, которая описывает исходные данные с наименьшим количеством параметров, представляет интерес. При этом полезными свойствами такого метода могли бы быть: возможность искать интерполянт в виде суммы функций из разных классов, а также возможность определения

функции, которая хорошо аппроксимирует заданный набор точек, если интерполирующей функции найти не удалось.

Цель исследования: разработка эффективного метода автоматического определения интерполирующей функции с наименьшим количеством параметров из некоторого множества функций, позволяющей настраивать множество видов искомым интерполянтов.

Задачи исследования:

1. Провести анализ современных методов интерполяции.
2. Разработать алгоритм синтеза рекурсивного цифрового фильтра (фильтра с бесконечной импульсной характеристикой, БИХ-фильтра), значения отсчётов импульсной характеристики которого задаются некоторой одномерной функциональной зависимостью.
3. Разработать метод определения вида интерполянта и вычисления её параметров исходя из коэффициентов БИХ-фильтра, получаемого с помощью предложенного алгоритма синтеза.
4. Разработать программную реализацию, использующую предложенный метод автоматической интерполяции числовых данных.

Методы исследования. В ходе исследования использовались основные положения цифровой обработки сигналов, линейной алгебры, функционального анализа, численных методов алгебры и анализа, а также теории алгоритмов.

Научная новизна работы состоит в следующем.

1. Доказана возможность синтеза рекурсивного цифрового фильтра по значениям отсчётов его импульсной характеристики, описываемой любой функцией f из определённого множества Φ (оно названо *множеством функций, известных алгоритму*). Показан алгоритм синтеза таких фильтров.
2. Разработан метод автоматической интерполяции на основе разработанного алгоритма синтеза рекурсивного цифрового фильтра, отличающийся от аналогов возможностью одновременного обнаружения функций нескольких видов, множество которых можно настраивать.
3. Разработаны способы расширения множества Φ ; показана применимость алгоритма синтеза рекурсивного цифрового фильтра по импульсной характеристике, когда f является любой линейной комбинацией функций из множества Φ .

Программная реализация результатов работы. Программа поиска функциональных зависимостей в числовых последовательностях «FunSearch 1.0» прошла экспертизу и зарегистрирована в Отраслевом фонде алгоритмов и программ при Федеральном агентстве по образованию (№ государственной регистрации 50200700388, рег. № ОФАП 7729).

Практическая значимость исследований. Разработанный алгоритм синтеза БИХ-фильтра по началу его импульсной характеристики имеет ряд

возможных приложений. В данной работе наиболее глубоко исследовано его применение для автоматической интерполяции числовых данных. Особенности алгоритма позволили разработать вычислительно эффективный метод для одновременного поиска в данных нескольких классов функций.

На защиту выносятся:

1. Разработанный алгоритм синтеза рекурсивного цифрового фильтра по началу импульсной характеристики, члены которой задаются математической функцией, самостоятельно определяет необходимый порядок фильтра.
2. Теоремы существования и единственности БИХ-фильтров, импульсные характеристики которых определяются математическими функциями из указанного множества.
3. Метод автоматической интерполяции, основанный на предложенном алгоритме синтеза БИХ-фильтров, способен отыскивать наиболее простой интерполянт в выбранном множестве функций, а также позволяет настраивать это множество функций.

Апробация работы. Основные положения и результаты диссертации докладывались и обсуждались на Конференции-конкурсе «Технологии Microsoft в теории и практике программирования» (2007 г.), Международной конференции IEEE по управлению и связи SIBCON-2007, международной научно-технической конференции «Радиолокация, навигация, связь» (RLNC-2007), Всероссийской научно-практической конференции с международным участием «Информационные технологии и математическое моделирование» (ИТММ-2009), международной научной конференции «Теоретические и прикладные проблемы математики, механики и информатики» (Караганда, 2010), а также на научно-практических семинарах кафедры безопасности информационных технологий СибГАУ (2006, 2009, 2011 г.).

Публикации. По теме диссертации опубликовано 12 печатных работ. Полный список публикаций представлен в конце автореферата. 3 работы опубликованы в изданиях, рекомендованных ВАК РФ.

Структура и объем работы. Диссертация состоит из введения, трех глав, заключения и трех приложений общим объемом 106 с. Список использованной литературы содержит 83 наименования.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Введение представляет цели и задачи диссертационной работы, раскрывает ее актуальность, научную новизну, методологию исследования, практическую значимость и апробацию полученных результатов.

В первой главе проведен анализ современных методов интерполяции дискретных данных. Обосновывается использование БИХ-фильтра в качестве математического объекта, однозначно связывающего две дискретные последовательности разного размера, и требующего для своего хранения объём памяти, меньший, чем большая из последовательностей.

Анализируются существующие методы синтеза БИХ-фильтров по импульсной характеристике. На основе проведённого анализа делается вывод об актуальности разработки метода автоматического определения интерполирующей функции. В соответствии с целью составляется список необходимых исследований.

Поскольку решено разрабатывать новый метод с использованием некоторых методов цифровой обработки сигналов (ЦОС), в первой главе приводится краткое описание используемых алгоритмов ЦОС и уточняется терминология.

Во второй главе решается задача разработки алгоритма синтеза рекурсивного цифрового фильтра, по заданным начальным отсчётам его импульсной характеристики, с учётом того, что отсчёты импульсной характеристики являются значениями некоторой функции, взятыми с постоянным шагом. Данный алгоритм станет основой для предлагаемого метода автоматической интерполяции

Задана последовательность Y длины L . Необходимо рассчитать рекурсивный цифровой фильтр, начало импульсной характеристики которого совпадает с Y . Рекурсивные ЦФ, описываются выражением

$$y(n) = \sum_{k=0}^N b_k x(n-k) - \sum_{k=1}^M a_k y(n-k) \quad (1)$$

Здесь $y(n)$ и $x(n)$ — выходная и входная последовательности фильтра соответственно, a_k и b_k — коэффициенты фильтра. Наибольшее из чисел M и N считается порядком фильтра.

В таких фильтрах присутствует обратная связь, определяемая коэффициентами a_k , за счёт чего импульсная характеристика фильтра получается бесконечной, при конечном числе его коэффициентов. Поскольку последовательность Y конечна, будем считать её началом импульсной характеристики рассчитываемого фильтра. Таким образом, известен выход фильтра (исходная последовательность Y) и вход — единичный импульс (так как мы приняли, что исходная последовательность это импульсная характеристика). Исходя из этого, необходимо рассчитать цифровой фильтр.

Синтез фильтра заключается в нахождении его коэффициентов a_k и b_k . Используя выражение (1) можно построить систему линейных уравнений, определяющую взаимосвязь коэффициентов фильтра с его импульсной характеристикой. У рекурсивного ЦФ такая система будет бесконечной, т.к. бесконечна сама импульсная характеристика. Вот как будет выглядеть система уравнений в общем виде для первых $L > M + N + 1$ отсчётов импульсной характеристики (используем обозначения $x_n \equiv x(n)$, $y_n \equiv y(n)$):

$$\left\{ \begin{array}{l} b_0 = y_0 \\ b_1 - a_1 y_0 = y_1 \\ b_2 - a_1 y_1 - a_2 y_0 = y_2 \\ \dots \\ b_N - (a_1 y_{N-1} + a_2 y_{N-2} + \dots + a_{M-1} y_0) = y_N \\ -(a_1 y_{N-2} + a_2 y_{N-3} + \dots + a_{M-1} y_1 + a_M y_0) = y_{N+1} \\ \dots \\ -(a_1 y_{L-2} + a_2 y_{L-3} + \dots + a_M y_{L-N-1}) = y_{L-1} \end{array} \right. \quad (2)$$

Здесь неизвестными являются коэффициенты a_k и b_k , а выход фильтра $Y=\{y_k\}$ является его импульсной характеристикой и в нашей постановке задачи известен заранее. Система уравнений (2) состоит из L уравнений, в правой части каждого из которых стоит отсчёт импульсной характеристики. Поскольку в (2) $M+N+1$ неизвестных, то для того, чтобы система могла иметь единственное решение, должно выполняться условие

$$L \geq M + N + 1 \quad (3)$$

Чтобы начать решать систему уравнений (2) не хватает только знания M и N . Пусть $N=M-1$ (таким образом, количество коэффициентов a_k и b_k будет равным), а M выберем любое, удовлетворяющее неравенству (3). Теперь построим систему — назовём её *начальной* — из первых $M+N+1$ уравнений системы (2). Для удобства запишем её в матричном виде:

$$\left(\begin{array}{cccccc} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & \ddots & & \\ & & & & 1 & \\ & & & & & 1 \end{array} \right) \begin{pmatrix} -y_0 \\ -y_1 \\ \vdots \\ -y_{N-1} \\ -y_N \\ \vdots \\ -y_{N+M-1} \end{pmatrix} \begin{pmatrix} -y_0 \\ -y_0 \\ \vdots \\ -y_{N-2} \\ -y_{N-1} \\ \vdots \\ -y_{N+M-2} \end{pmatrix} \dots \begin{pmatrix} -y_0 \\ -y_1 \\ \vdots \\ -y_{M-2} \\ -y_{M-1} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_N \\ a_1 \\ \vdots \\ a_M \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_N \\ y_{N+1} \\ \vdots \\ y_{N+M} \end{pmatrix} \quad (4)$$

Матрицу системы уравнений и соответствующую ей расширенную матрицу назовём P и \overline{P} соответственно. Решив систему (4) найдём все коэффициенты ЦФ. Но прежде, чем её решать, необходимо убедиться, что она совместна и определена. Для этого вычислим ранги матриц P и \overline{P} . Далее возможны три случая:

1. Если $\text{rang}(P) = \text{rang}(\overline{P}) = M+N+1$, то система (4) совместна, обладает полным рангом, и её можно решить каким-либо стандартным методом.

2. Если $\text{rang}(P) \neq \text{rang}(\bar{P})$, то система (4) несовместна и при данных M и N не существует БИХ-фильтра, у которого первые $M+N+1$ отсчётов импульсной характеристики совпадают с заданными. Но он может существовать при большем порядке фильтра. Поэтому, если позволяет длина исходной импульсной характеристики, следует увеличить M и N на единицу, заново построить систему (4) и попробовать решить новую систему.
3. Если $\text{rang}(P) = \text{rang}(\bar{P}) < M+N+1$, то система уравнений недоопределена, то есть имеет бесконечное множество решений. Было бы неправильно выбрать какое-то одно решение из этого множества, так как при меньшем порядке фильтра существует единственное решение системы (4). Поэтому следует уменьшить M и N на единицу, заново построить систему (4) и проверить её на наличие единственного решения.

Таким образом, за счёт второго и третьего пунктов, начав с любых M и N придём к минимально возможным M и N , при которых система уравнений (4) имеет решение. Решив систему уравнений, фактически найдём БИХ-фильтр, у которого $M+N+1$ первых элементов импульсной характеристики совпадают с $M+N+1$ первыми элементами Y . То есть соответствие уже будет установлено, но ещё не для всей длины Y . Поэтому следующим шагом алгоритма является проверка остальных $L-M-N-1$ членов заданной импульсной характеристики на соответствие найденному решению.

Из соображений оптимизации скорости и используемого объёма памяти был выбран следующий способ проверки. После решения начальной системы уравнений известны все коэффициенты фильтра a_k и b_k . Поэтому можно сгенерировать его импульсную характеристику, начиная с 1-го отсчёта и до L -го. Если все отсчёты совпали с исходной последовательностью Y , значит решение верное. Назовём такое решение *фильтром, соответствующим последовательности Y* . Блок-схема алгоритма представлена на рис 1.

Интересной задачей является определение того, в каких случаях в системе уравнений (2) $L-M-N-1$ последних линейных уравнений от $M+N+1$ неизвестных линейно зависимы от первых $M+N+1$ уравнений. Исследования и вычислительные эксперименты показали, что это выполняется, когда члены последовательности Y являются отсчётами некоторой функции $f(x)$, взятыми с постоянным шагом. Установление всех таких функций является важным для предложенного алгоритма. В данной работе доказано, что f может быть

- произвольной полиномиальной функцией;
- произвольной показательной функцией;
- произвольной синусоидой или косинусоидой;

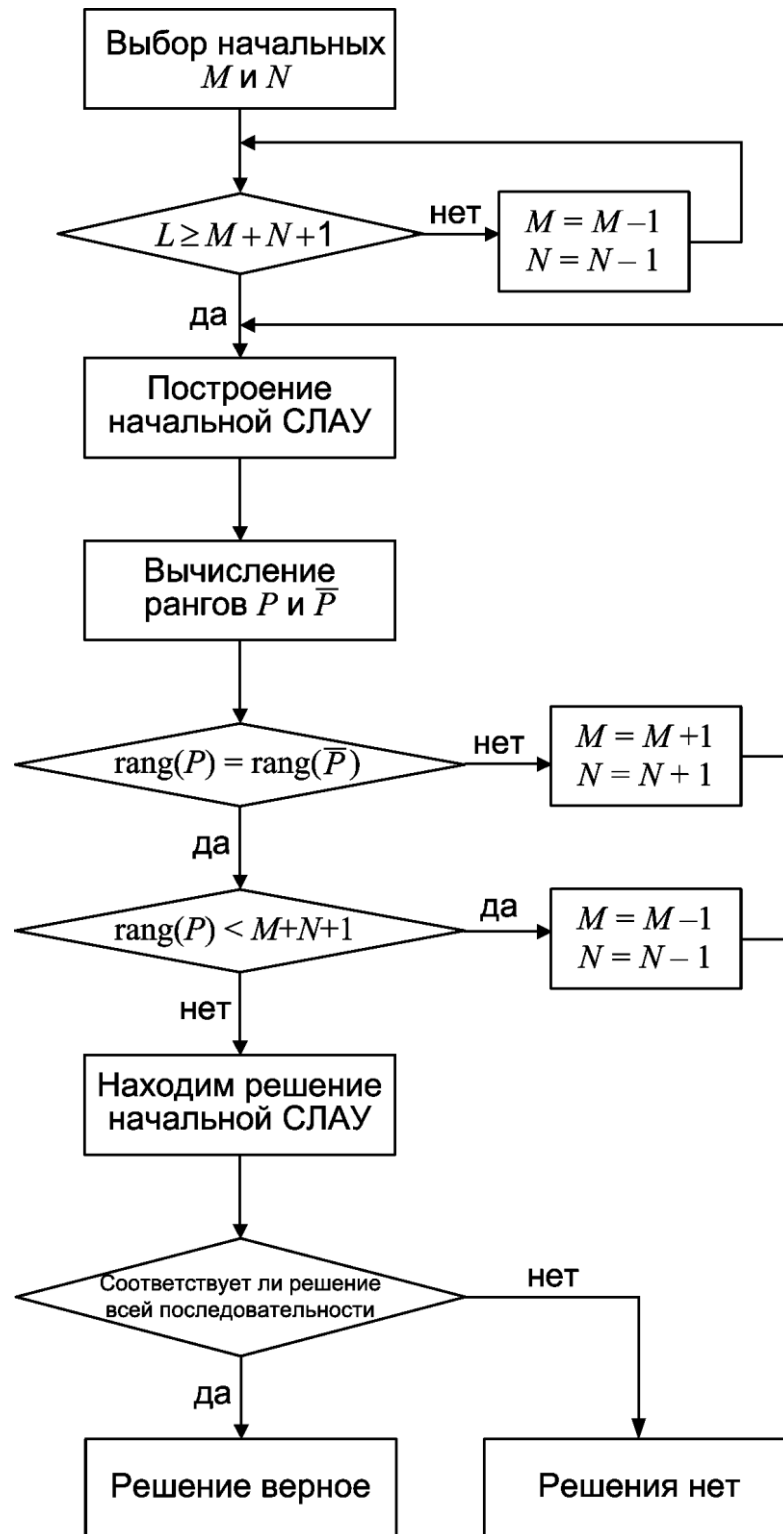


Рис. 1. Блок-схема алгоритма расчёта рекурсивного цифрового фильтра по началу его импульсной характеристики

Формулировки доказанных теорем следующие.

ТЕОРЕМА 1. Для последовательности Y длины $L \geq 2(n+1)$, члены которой являются равноотстоящими отсчётами полиномиальной функции одного переменного $y_i = p_n(i) = \alpha_0 \cdot i^n + \alpha_1 \cdot i^{n-1} + \dots + \alpha_n$, $\alpha_0 \neq 0$, и заданы любые $(n+1)$ подряд идущих члена $y_m, y_{m+1}, \dots, y_{m+n}$, существует единственный

рекурсивный цифровой фильтр порядка $n+1$ ($M=n+1$, $N=n$), импульсная характеристика которого совпадает с Y .

ТЕОРЕМА 2. Для последовательности $Y=y(i)$ длины $L \geq 4$, члены которой являются равноотстоящими отсчётами показательной функции $y_i = f(i) = k \cdot a^{bi} + c$, $a \neq 0$, $a \neq 1$, $b \neq 0$, $k \neq 0$, и заданы первые 4 члена y_0, y_1, y_2, y_3 , существует единственный рекурсивный цифровой фильтр второго порядка ($M=2$, $N=1$), импульсная характеристика которого совпадает с Y .

ТЕОРЕМА 3А. Для последовательности $Y=y(i)$ длины $L \geq 4$, члены которой являются равноотстоящими отсчётами синусоидальной функции $y_i = f(i) = a \cdot \sin(b \cdot i + c)$, $a \neq 0$, $|b| < \pi$, $k \in \mathbb{Z}$ и заданы первые 4 члена y_0, y_1, y_2, y_3 , существует единственный рекурсивный цифровой фильтр второго порядка ($M=2$, $N=1$), импульсная характеристика которого совпадает с Y .

ТЕОРЕМА 3Б. Для последовательности $Y=y(i)$ длины $L \geq 6$, члены которой являются равноотстоящими отсчётами синусоидальной функции $y_i = f(i) = a \cdot \sin(b \cdot i + c) + d$, $a \neq 0$, $|b| < \pi$, $k \in \mathbb{Z}$, $d \neq 0$, и заданы первые 6 членов y_0, y_1, \dots, y_5 , существует единственный рекурсивный цифровой фильтр третьего порядка ($M=3$, $N=2$), импульсная характеристика которого совпадает с Y .

Также, в аналогичном ключе доказано ещё две теоремы о существовании рекурсивного ЦФ для случая, когда отсчёты его импульсной характеристики являются значениями произвольной периодической функцией при условии, что период функции кратен расстоянию между взятыми отсчётами. Данные теоремы представляются как дополнительный результат, так как, в отличие от предыдущих теорем, они не позволяют идентифицировать вид функции, а только позволяют определить периодичность исходной числовой последовательности и длину периода.

ТЕОРЕМА 4А. Для любой периодической последовательности $Y=y(i)$ периода T и длины $L \geq T$ существует рекурсивный цифровой фильтр порядка T ($M=T$, $N=T-1$), импульсная характеристика которого совпадает с Y . Коэффициенты такого фильтра будут равны: $b_0 = y_0, b_1 = y_1, \dots, b_T = y_T, a_1 = 0, a_2 = 0, \dots, a_{T-1} = 0, a_T = -1$, где y_0, \dots, y_T — значение членов последовательности на периоде.

ТЕОРЕМА 4Б. Для любой периодической последовательности $Y=y(i)$ периода $T=2k$ со значениями на периоде вида $(s_0, s_1, \dots, s_{k-1}, -s_0, -s_1, \dots, -s_{k-1})$ и длины $L \geq T$ существует рекурсивный цифровой фильтр порядка k ($M=k$, $N=k-1$), импульсная характеристика которого совпадает с Y . Коэффициенты такого фильтра будут равны: $b_0 = s_0, b_1 = s_1, \dots, b_{k-1} = s_{k-1}, a_1 = 0, a_2 = 0, \dots, a_{T-1} = 0, a_T = 1$.

Предложенный алгоритм позволяет рассчитывать рекурсивный ЦФ наименьшего возможного порядка с заданным началом импульсной

характеристики. Необходимый порядок фильтра зависит от функциональной зависимости f , которой подчиняются члены исходной последовательности Y . Чем сложнее f , тем большего порядка будет ЦФ. Количество коэффициентов фильтров, соответствующие определённым функциональным зависимостям f , а также определяемое ими минимальное необходимое для расчёта количество начальных отсчётов импульсной характеристики, представлены в табл. 1.

Таблица 1. Минимальные M и N , для некоторых функций

Зависимость		M	N	L_{\min}
Линейная	$y_i = k \cdot i + b$	2	1	4
Квадратичная	$y_i = a \cdot i^2 + b \cdot i + c$	3	2	6
Кубическая	$y_i = a \cdot i^3 + b \cdot i^2 + c \cdot i + d$	4	3	8
Полином 4-й степени	$y_i = a \cdot i^4 + b \cdot i^3 + c \cdot i^2 + d \cdot i + e$	5	4	10
Периодическая, T — количество отсчётов, приходящееся на один период	произвольная	T	$T - 1$	$2T$
Показательная	$y_i = k \cdot a^i + c$	2	1	4
Синусоида	$y_i = a \cdot \sin(b \cdot i + c) + d$	3	2	6
Косинусоида	$y_i = a \cdot \cos(b \cdot i + c) + d$	3	2	6

Множество вычислительных экспериментов и результатов, полученных при разработке приложений, описанных в третьей главе, говорят о том, что функция f может быть и любой линейной комбинацией функций из табл. 1. Приведём несколько таких функций, для которых также существует единственный фильтр с указанными минимальными M и N (табл. 2).

Таблица 2. Минимальные M и N , для некоторых линейных комбинаций функций

Зависимость		M	N	L_{\min}
Сумма линейной и периодической с периодом 2	Например, $y_i = a \cdot i + b + c \cdot (-1)^i$	3	2	6
Сумма квадратичной и периодической с периодом 3		5	4	10
Сумма двух синусоид	$y_i = k_1 \cdot \sin(f_1 \cdot i + \varphi_1) + k_2 \cdot \sin(f_2 \cdot i + \varphi_2) + c$	5	4	10
Сумма показательной, квадратичной и синусоидальной функций	$y_i = k_1 \cdot a^{bi} + k_2 \cdot i^2 + k_3 \cdot i + k_4 \cdot \sin(f \cdot i + \varphi) + c$	6	5	12
Сумма трёх синусоид	$y_i = k_1 \cdot \sin(f_1 \cdot i + \varphi_1) + k_2 \cdot \sin(f_2 \cdot i + \varphi_2) + k_3 \cdot \sin(f_3 \cdot i + \varphi_3) + c$	7	6	14

В третьей главе описывается разработанный метод автоматической интерполяции наборов числовых данных, основой которого стал алгоритм нахождения рекурсивного ЦФ по началу его импульсной характеристики. Предполагается, что искомые функции являются функциями одного переменного f . Интерполянт должен принадлежать множеству функций, известных алгоритму. Опишем сначала, как ищется одна функция, а потом модифицируем метод для отыскания в исходных данных всех участков, описываемых любой из функций, известных алгоритму.

Метод поиска интерполянта состоит из трёх основных шагов:

- 1) Для заданной последовательности чисел $Y=\{y_i\}$ однократно вычисляется соответствующий рекурсивный ЦФ.
- 2) По коэффициентам обратной связи найденного ЦФ, определяется вид функции f , которой подчиняются значения членов Y : $y_i = f(i)$.
- 3) Используя только коэффициенты найденного фильтра, вычисляются значения параметров функции f .

Первый шаг выполняется в соответствии с описанным выше алгоритмом синтеза БИХ-фильтров. Суть второго и третьего шагов метода заключается в следующем. Вид некоторых функциональных зависимостей f можно сразу идентифицировать по значениям коэффициентов обратной связи a_k , а все параметры (константы) функции определяются по коэффициентам b_k . Назовём такие функции «*простыми*». Для других (назовём их «*сложными*») необходимы несложные вычисления, так как в них коэффициенты a_k зависят не только от вида, но и от её параметров.

«*Простыми*» функциями являются полиномы и периодические. В табл. 3 приведены значения коэффициентов обратной связи для различных «*простых*» функций.

Таблица 3. Коэффициенты обратной связи цифровых фильтров, соответствующих некоторым «*простым*» функциональным зависимостям

Зависимость		Коэффициенты a_k
Линейная	$y_i = s_0 \cdot i + s_1$	[-2 1]
Квадратичная	$y_i = s_0 \cdot i^2 + s_1 \cdot i + s_2$	[-3 3 -1]
Кубическая	$y_i = \sum_{k=0}^3 s_k \cdot i^k$	[-4 6 -4 1]
Полином 4-й степени	$y_i = \sum_{k=0}^4 s_k \cdot i^k$	[-5 10 -10 5 -1]
Полином 5-й степени	$y_i = \sum_{k=0}^5 s_k \cdot i^k$	[-6 15 -20 15 -6 1]
Полином 6-й степени	$y_i = \sum_{k=0}^6 s_k \cdot i^k$	[-7 21 -35 35 -21 7 -1]

Таблица 3. Окончание.

Полином 7-й степени	$y_i = \sum_{k=0}^7 s_k \cdot i^k$	[-8 28 -56 70 -56 28 -8 1]
Полином 8-й степени	$y_i = \sum_{k=0}^8 s_k \cdot i^k$	[-9 36 -84 126 -126 84 -36 9 -1]
Периодическая, с периодом T отсчётов	произвольная	$[0 \ 0 \ \dots \ 0 \ 0 \ -1]$ $\underbrace{\hspace{10em}}$ $T-1$ нулей

Например, если исходной последовательности соответствует цифровой фильтр, у которого 3 коэффициента a_k и они равны $-3, 3, -1$, значит, члены последовательности однозначно описываются квадратичной функцией (без погрешности).

Коэффициенты b_k фильтра, у которого значения членов импульсной характеристики задаются «простой» функцией, в свою очередь полностью определяются параметрами функции f . Приведём примеры.

Таблица 4. Зависимость коэффициентов b_k от параметров «простых» функций

Зависимость		Коэффициенты b_k
Линейная	$y_i = k \cdot i + b$	$[k+b \ -b]$
Квадратичная	$y_i = a \cdot i^2 + b \cdot i + c$	$[a+b+c \ a-b-2c \ c]$
Кубическая	$y_i = a \cdot i^3 + b \cdot i^2 + c \cdot i + d$	$[a+b+c+d \ 4a-2c-3d \ a-b+c+3d \ -d]$
Полином 4-й степени	$y_i = a \cdot i^4 + b \cdot i^3 + c \cdot i^2 + d \cdot i + e$	$[a+b+c+d+e \ 11a+3b-c-3d-4e \ 11a-3b-c+3d+6e \ a-b+c-d-4e \ e]$
Периодическая, с периодом T отсчётов	Произвольная	$[y_0 \ y_1 \ \dots \ y_{T-1}]$

Видно, что выразить параметры функции f через коэффициенты b_k найденного фильтра можно, решив несложную систему линейных уравнений. Таким образом, будет найдено точное выражение для искомой функции.

Анализ коэффициентов обратной связи найденного ЦФ с целью определения вида зависимости, присутствующей в исходных данных осложняется тем, что встречаются «сложные» функции, для которых значения коэффициентов a_k не являются постоянными, а зависят кроме вида функции ещё и от её параметров. Среди вышеупомянутых функций такими являются показательная, синусоида и косинусоида. Несмотря на то, что в этом случае коэффициенты фильтра связаны с параметрами зависимости нелинейным преобразованием, найти обратное преобразование и выразить параметры через коэффициенты возможно. Приведём примеры таких функций и зависимость коэффициентов соответствующих им ЦФ от их параметров.

Таблица 5. Примеры коэффициентов ЦФ, соответствующих «сложным» функциям

Функция		M	N	Коэффициенты
Показательная	$y_i = k \cdot a^{b \cdot i} + c$	2	1	$b_0 = k \cdot a^b + c$ $b_1 = -a^b \cdot (k + c)$ $a_1 = -a^b - 1$ $a_2 = a^b$
Синусоида	$y_i = a \cdot \sin(b \cdot i + c) + d$	3	2	$b_0 = a \cdot \sin(b+c) + d$ $b_1 = a \cdot (\sin(b+c) + \sin(c)) - 2 \cdot d \cdot \cos(b)$ $b_2 = a \cdot \sin(c) + d$ $a_1 = -2 \cdot \cos(b) - 1$ $a_2 = 2 \cdot \cos(b) + 1$ $a_3 = -1$

Исходя из соотношений в табл. 5 определить наличие в данных показательной функциональной зависимости можно, если выполняются следующие условия:

$$M = 2, N = 1, a_1 + a_2 = -1, a_2 > 0, a_2 \neq 1$$

Для синусоидальной функции:

$$M = 3, N = 2, a_1 = -a_2, -1 < a_2 < 3, a_3 = -1$$

Дополнительным преимуществом предложенного метода, наряду с возможностью автоматического определения простейшего интерполянта из трёх указанных классов функций, является возможность включать в это множество функций любую функцию, являющуюся линейной комбинацией уже известных методу функций. Тем не менее, не смотря на то, что среди известного множества функций поиск наилучшего интерполянта выполняется автоматически, также автоматически распространить правила идентификации по коэффициентам ЦФ на все возможные линейные комбинации является сложной задачей. Для этого необходимо сначала определить минимальный порядок системы (4) для произвольной конечной суммы одночленов, синусоид и показательных функций (это несложно). Затем нужно построить такую систему уравнений, доказать её совместность и решить её. В связи с трудоёмкостью последних шагов (из-за операций с матрицами больших размерностей, элементами которых являются функции) данные действия в рамках работы выполнялись только для нескольких классов функций, являющихся линейными комбинациями известных методу функций.

Для суммы «простых» функций коэффициенты a_k соответствующего ЦФ также оказались постоянными и не зависящими от параметров функции. То есть сумма «простых» функций является «простой» функцией и её тоже можно однозначно идентифицировать по коэффициентам обратной связи фильтра. Среди сумм с участием «сложных» функций проверялись

следующие: сумма двух синусоид, сумма трёх синусоид, сумма показательной функции и синуса, сумма квадратичной функции и синусоиды.

Теперь рассмотрим, как на основе этого подхода должен действовать метод отыскания всех участков, интерполируемых известными алгоритму функциями. Программа, реализующая данный метод, должна обладать некоторой базой знаний о функциях, которые она сможет обнаруживать. Такая база знаний должна содержать:

- информацию, необходимую для идентификации определённой функции;
- правила для нахождения значений параметров функции из значений коэффициентов соответствующего фильтра.

В качестве идентифицирующей информации для «простых» функций выступают точные значения коэффициентов обратной связи соответствующего фильтра, а для «сложных» функций — условия, накладываемые на коэффициенты ЦФ.

На первом этапе метода для начала исходной последовательности отсчётов выполняется синтез соответствующего цифрового рекурсивного фильтра. Различным интерполянтам могут соответствовать фильтры разных порядков. Таким образом, для того, чтобы проверить в последовательности наличие всех функций из базы знаний, надо в худшем случае совершить N попыток синтеза БИХ-фильтра, где N — количество разных порядков ЦФ, соответствующих функциям из базы знаний. Алгоритм синтеза, построен так, что непосредственное вычисление БИХ-фильтра каждого порядка не выполняется. Он находит минимальный порядок ЦФ, соответствующего данной последовательности, при котором он является единственным, а уже потом вычисляет его.

После вычисления фильтра некоторого порядка k нужно по его коэффициентам проверить наличие всех функций, которым соответствует фильтр такого порядка. Если для текущего порядка фильтра проверка не дала положительного результата (то есть полученные коэффициенты не соответствуют ни одному правилу из базы знаний), следует проверить фильтры меньших порядков.

Когда найден БИХ-фильтр, коэффициенты которого говорят о том, что элементы последовательности, используемые при его вычислении, описываются функцией известной алгоритму, следует определить, на сколько следующих членов последовательности распространяется найденная функциональная зависимость. Суть такой проверки сводится к вычислению импульсной характеристики найденного фильтра и её поэлементному сравнению с рассматриваемой последовательностью.

После того, как длина участка последовательности, подчиняющегося найденной зависимости, определена, можно сохранить всю полученную информацию об этом участке и повторить проверку для оставшейся части последовательности. В качестве сохраняемой информации могут выступать индекс первого элемента участка, длина участка (количество членов последовательности), идентификатор типа найденной функции, коэффициенты соответствующего фильтра.

Таким образом, после подобного анализа будут известны участки исходной последовательности, интерполируемые полностью определёнными функциями одного переменного.

Далее в 3-й главе рассмотрена задача нахождения аппроксимирующей функции из заданного множества с наименьшим количеством параметров. Пусть в узлах значения исходных данных немного отличаются от точных значений некоторой функции, известной алгоритму. В этом случае коэффициенты обратной связи соответствующего ЦФ уже не будут точно подчиняться правилам идентификации данной функции. Тем не менее, в главе 3 показано, что их значения будут близкими к соответствующим точным значениям (это обусловлено тем, что коэффициенты фильтра являются решением СЛАУ, коэффициенты которой непрерывно зависят только от исходных данных). Поэтому небольшие отклонения исходных данных от идеальных значений будут приводить к малым отклонениям коэффициентов фильтра. Например, если коэффициенты обратной связи вычисленного ЦФ будут равны -3.04 , 3.11 , -0.97 , значит, значения исходных точек очень близки к значениям квадратичной функции. Доказательство данного рассуждения в общем виде для одного вида функции приводится в 3-й главе диссертации. Таким образом, данный метод может быть использован, в том числе, и для аппроксимации числовых данных. Например, по теореме Вейерштрасса, любая непрерывная функция приближается многочленом, однако его степень может быть достаточно высока, и количество параметров – коэффициентов этого многочлена, которые необходимо найти, слишком велико (порядок фильтра для нахождения этого многочлена должен быть выше заданного). При этом может оказаться, что функция с той же точностью приближается, например, экспонентой с меньшим числом параметров, которые и будут найдены указанным методом при заданном порядке фильтра.

Опишем некоторые преимущества предлагаемого метода. Во-первых, она позволяет осуществлять потоковую обработку данных. То есть для начала её работы нет необходимости в наличии сразу всех обрабатываемых данных.

Вторым важным свойством метода является возможность изменения множества искомым интерполянтов. Процесс выбора модификации для конкретной решаемой задачи выглядит следующим образом:

- определяется множество функций, приемлемых для описания исходных данных;
- в базу знаний вносятся правила для идентификации каждой такой функции;
- определяется множество порядков фильтров, соответствующих искомым функциям;
- далее в процессе работы перебираются порядки ЦФ только этого множества.

Покажем, что сложность вычислений, выполняемых при работе такой системы, с ростом количества видов искомых функций растет медленнее, чем в классическом случае. Пусть N – количество обнаруживаемых видов функций. Сложность классического случая растёт как $O(N)$, так как для обнаружения каждой функции требуются независимые действия. Пусть M – количество порядков ЦФ, необходимых для обнаружения этих N функций. $M \leq N$ по определению. Соответственно, сложность вычислений при использовании предложенного метода равна $O(M) \leq O(N)$. Даже более точно: сложность равна $O(M) + (N - M) \cdot O(1) \leq O(N)$.

Кроме доказанных видов функций может понадобиться использовать метод для обнаружения интерполянтов другого вида. Поэтому в работе приводится описание способов проверки, работоспособна ли она для нужного вида функций, и обучения конкретной реализации метода для обнаружения таких интерполянтов. Также приводятся листинги программ, выполняемых в системе MATLAB, которые могут в этом помочь.

Если не удалось установить возможность применения метода автоматической интерполяции к конкретному виду функции, можно в ряде случаев использовать разложение в ряд Тейлора интересующей функции f , неизвестной алгоритму, для ее обнаружения с некоторой допустимой ошибкой. Это возможно, если f является аналитической функцией. Такой метод является приближённым и основан на том, что любая аналитическая функция по определению раскладывается в ряд Тейлора, который по сути является полиномом бесконечной степени. А если использовать реализацию метода в реальной цифровой системе, то можно, опираясь на максимальную необходимую точность для решаемой задачи, определить количество значащих членов ряда Тейлора. Отбросив остальные члены ряда, получим полином конечной степени. А для полиномов задача поиска уже решена (теорема 1).

Завершается третья глава описанием преимуществ предложенного метода автоматической интерполяции. Таковыми являются:

- Скорость работы не сильно зависит от их количества обнаруживаемых видов интерполянтов.

- Возможность настройки множества видов обнаруживаемых интерполянтов. Методика настройки описана в работе.
- Минимизация количества параметров функциональных зависимостей. Из всего множества интерполирующих функций находится функция с наименьшим количеством параметров.
- Возможность потоковой работы (обработку данных можно начинать, когда поступили ещё не все данные).

В приложениях содержатся листинги программ, выполняемых в системе MATLAB, на которые имеются ссылки во второй главе диссертации, а также описана программная реализация предложенного метода.

Программа поиска функциональных зависимостей реализована в виде отдельного приложения и осуществляет поиск следующих функциональных зависимостей в одномерных данных: постоянная, линейная, квадратичная, кубическая, полином 4-й степени, показательная, периодические подпоследовательности (периодом от 2 до 5), синусоида, сумма двух синусоид. Для этого вычисляются фильтры 2-го, 3-го, 4-го и 5-го порядков. Выполняется поиск только функций, точно проходящих через имеющиеся точки, то есть задача аппроксимации не решалась. После открытия произвольного файла с данными они отображаются в численном и графическом виде. Предоставляются средства для навигации по данным.

После анализа данных все найденные участки, подчиняющиеся какой-то из перечисленных выше зависимостей, отображаются в виде списков (по одному списку на каждый вид зависимости). Выбранный в списке участок выделяется на численном и графическом отображениях, а на отдельной панели отображается описывающая его функция с точностью до значения свободного члена, длина (количество точек) и индекс начальной точки.

Предусмотрена возможность чтения исходных данных, как байтами, так и словами (двухбайтовыми целыми знаковыми числами). Внешний вид программы изображён на рис. 2.

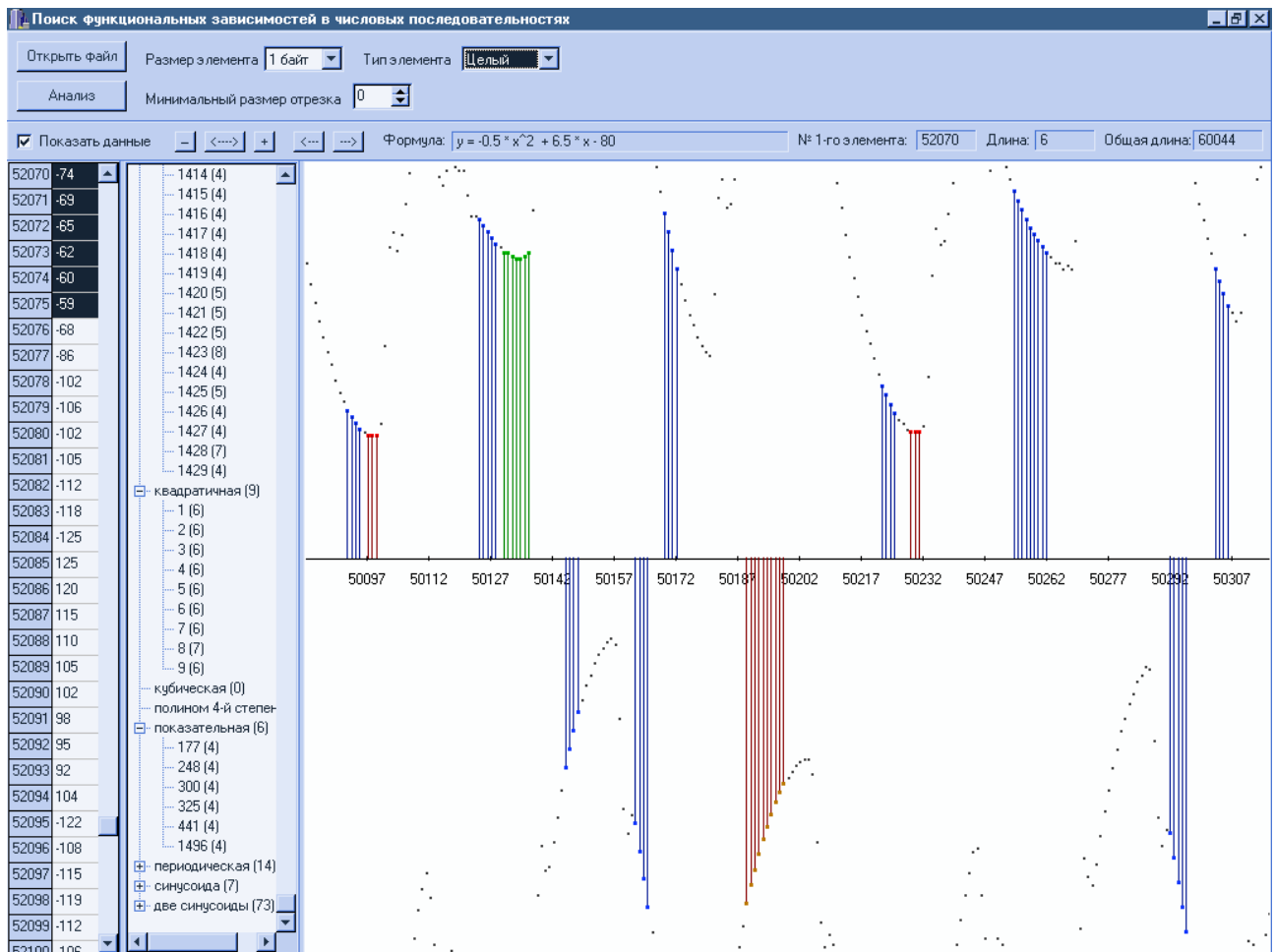


Рис. 2. Внешний вид программы поиска функциональных зависимостей в числовых последовательностях

Основные результаты и выводы.

1. Разработан алгоритм синтеза рекурсивного цифрового фильтра, значения отсчётов импульсной характеристики которого задаются некоторой одномерной функциональной зависимостью. Приведены доказательства корректности данного алгоритма для ряда классов функций. Особенностью алгоритма является определение минимального необходимого порядка фильтра.
2. Разработан метод автоматической интерполяции функциями из некоторого множества, основанный на предложенном алгоритме синтеза рекурсивного цифрового фильтра. Основные преимущества:
 - а) минимизация количества параметров интерполянтов;
 - б) вычислительная эффективность по сравнению с классическими методами при одновременном обнаружении нескольких видов функций;
 - в) возможность настройки множества отыскиваемых функций;
3. Разработаны способы расширения области применения метода: проверка применимости для произвольной функции и замена функции рядом Тейлора.

4. Разработана программная реализация метода: программа поиска функциональных зависимостей в числовых последовательностях.

Основные результаты диссертационной работы представлены в следующих публикациях автора.

1. Никитин Д.А. Об одной численной методике упрощения алгебраических выражений / Д.А. Никитин // Теоретические и прикладные проблемы математики, механики и информатики: Материалы междунар. научной конф. (24-26 июня 2010 г.). – Караганда: КарГУ. – 2010. – С. 150–153.
2. **Никитин Д.А. Приложения алгоритма синтеза рекурсивных цифровых фильтров по импульсной характеристике / Д.А. Никитин // Цифровая обработка сигналов. – 2009. – №4. – С. 8–15.**
3. Никитин Д.А. Теоремы о существовании и порядках цифровых рекурсивных фильтров с импульсными характеристиками определённой формы / Д.А. Никитин // Информационные технологии и математическое моделирование (ИТММ-2009): Материалы VIII Всероссийской научно-практической конференции с международным участием (13-14 ноября 2009 г.). – Томск: Изд-во Том. ун-та. – 2009. – Ч. 2. – С. 144–146.
4. **Никитин Д.А. Синтез рекурсивных цифровых фильтров по импульсной характеристике, определяемой элементарной математической функцией / Д.А. Никитин, Ханов В.Х. // Цифровая обработка сигналов. – 2008. – №3. – С. 10–14.**
5. Nikitin D.A. Compression with Usage of the Functional Dependences Searching / D.A. Nikitin // IEEE International Conference on Control and Communications (SIBCON-2007). Proceedings. – Tomsk: The Tomsk IEEE Chapter & Student. – Russia, Tomsk, April 20-21, 2007. – P. 72–76.
6. Никитин Д.А. Расчёт рекурсивного цифрового фильтра по его импульсной характеристике / Д.А. Никитин // Сборник докладов XIII междунар. науч.-технич. конф. «Радиолокация, навигация, связь» (RLNC-2007) / ООО НПФ «САКВОЕЕ» – Воронеж. – 2007. – Т. 1. – С. 335–341.
7. Никитин Д.А. Сжатие временных рядов с использованием блочной интерполяции / Д.А. Никитин // Информационные технологии моделирования и управления. – 2007. – Вып. 1 (35). – С. 85–89.
8. Никитин Д.А. Сжатие с использованием поиска функциональных зависимостей / Д.А. Никитин // Конференция-конкурс «Технологии Microsoft в теории и практике программирования». – Новосибирск: НГУ. – 2007. – С. 133–135.
9. Никитин Д.А. Программа поиска функциональных зависимостей в числовых последовательностях «FunSearch 1.0» – М.: ВНИИЦ. – 2007. – № 50200700388, рег. № ОФАП 7729.
10. **Никитин Д.А. Алгоритм анализа числовых последовательностей / В.Х. Ханов, Д.А. Никитин // Вестник Сибирского государственного**

аэрокосмического университета имени академика М. Ф. Решетнева. – 2006. – № 6 (13). - С. 11–15.

11. Никитин Д.А. Алгоритм сжатия медиаданных / Д.А. Никитин // X Междунар. науч. конф. с междунар. участием «Решетневские чтения». – Красноярск: СибГАУ. – 2006. – С. 313–314.
12. Никитин Д.А. Программные способы оптимизации алгоритмов цифровой обработки сигналов / Д.А. Никитин // Схемотехника. – 2006. – №2, С. 27–30. – № 3, С. 24–27.