КОМПЬЮТЕРНАЯ МАТЕМАТИКА

УДК:519.651.2:681.3

А.А. КАЛЕНЧУК-ПОРХАНОВА, Л.П. ВАКАЛ

НАИЛУЧШАЯ ЧЕБЫШЕВСКАЯ АППРОКСИМАЦИЯ ДЛЯ СЖАТИЯ ЧИСЛЕННОЙ ИНФОРМАЦИИ

Рассматривается проблема сжатия численной информации и её решение с применением наилучшей чебышевской аппроксимации. Даётся обоснование преимуществ разработанных алгоритмов аппроксимации, которые связаны с их оптимизацией по точности и быстродействию. Приводятся некоторые результаты расчётов по сжатию числовых массивов с большими коэффициентами сжатия.

Введение. На современном этапе развития уровень информационного обеспечения стал определяющим фактором развития экономики, науки, техники и общества в целом. Можно утверждать, что от количества и качества полученной информации существенно зависит эффективность деятельности общества и властных структур.

На практике информация, как правило, задаётся в виде массивов числовых данных, которые являются дискретным представлением функциональных зависимостей, характеризующих исследуемые объекты и процессы различной природы. Работа с такими массивами связана с рядом серьёзных трудностей, возникающих, например, при

– использовании в задачах математического моделирования и прогнозирования;

– необходимости экономного хранения больших по объёму массивов или при их скоростной передаче по каналам связи;

– восстановлении значений функциональной зависимости на «неосвещенных» замерами участках.

Для преодоления перечисленных трудностей применяется математическая обработка массивов численной информации с использованием аппарата аппроксимации функций. При этом осуществляется приближенная замена массива дискретного представления функциональной зависимости f некоторым аналитическим выражением (аппроксимантом) F с небольшим числом параметров-коэффициентов и выполняется сжатие численной информации.

Качественно новым подходом при выполнении такой замены является использование интеллектуализированных методов приближения функций способом наилучшей чебышевской (равномерной) аппроксимации, который значительно эффективнее и универсальнее, чем интерполяционный и среднеквадратический способы приближения [1].

Главное преимущество чебышевского способа аппроксимации по сравнению с другими способами приближения – обеспечение точности приближения, полученной на некотором множестве точек интервала приближения, во всех точках этого интервала. При этом чебышевский способ наилучшего равномерного приближения имеет существенное преимущество ещё и потому, что даёт лучшую точность приближения, чем наилучшее среднеквадратическое приближение аппроксимантом того же класса [2].

Указанные преимущества чебышевской аппроксимации позволяют решать с высокой точностью не только задачу нахождения аппроксиманта F и, как следствие, сжатие данных дискретно заданной функциональной зависимости f (прямая задача аппроксимации), но и задачу восстановления значений зависимости f на «неосвещенных» замерами участках (обратная задача аппроксимации). Степень сжатия характеризуется коэффициентом C, который определяется по формуле

$$C = \frac{b(f)}{b(F)}, \tag{1}$$

где $b(f)$, $b(F)$ – количество бит, необходимых для хранения соответственно функции f и аппроксиманта F [3].

Аппарат наилучшей чебышевской аппроксимации. Наиболее эффективным подходом приближения функций способом чебышевской аппроксимации является нахождение аппроксиманта наилучшего приближения.

Для заданной функции $f(x)$ *наилучшим чебышевским приближением* с весом $w(x) \neq 0$ на множестве точек E ($E = [\alpha, \beta]$ или $E_N = \{x_1,\ x_2, \ldots, x_N\} \subset [\alpha, \beta]$) называется такой аппроксимант $F^*(x;\,A)$ из заданного класса функций $\{F(x;\,A)\}$, $A = (a_0, a_1, \ldots, a_n)$ и $n < N$, для которого выполняется условие

$$\rho \equiv \max_{x \in E} \left| w(x)\ \left[f(x) - F^*(x;\ A) \right] \right| = \min_{A} \max_{x \in E} \left| w(x)\ \left[f(x) - F(x;\ A) \right] \right|. \tag{2}$$

Величина ρ называется величиной наилучшего чебышевского приближения (погрешностью аппроксимации).

Наиболее известный метод решения задачи (2) – метод последовательных чебышевских интерполяций (п.ч.и.) Ремеза [1], разработанный для случая аппроксимации алгебраическими полиномами $P_n(x) = a_0 + a_1 x + \ldots + a_n x^n$.

Теоретической основой метода Ремеза служит теорема Чебышева, согласно которой полином $P_n(x)$ наилучшего равномерного приближения характеризуется таким необходимым и достаточным условием: на множестве точек E должны найтись, по крайней мере, (n+2) точки $\alpha \le x_0 < x_1 < \ldots < x_{n+1} \le \beta$, в которых функция отклонения $\Delta(x) = w(x)\left[f(x) - P_n(x)\right]$ достигает своего модуль-максимума ρ с чередованием знака

$$\Delta(x_0) = -\Delta(x_1) = \ldots = (-1)^{n+1}\Delta(x_{n+1}) = \rho . \qquad (3)$$

Точки $x_0, x_1, \ldots, x_{n+1}$ называются точками чебышевского альтернанса.

Метод Ремеза является итерационным и основан на последовательных чебышевских интерполяциях, j шагов которых сводятся к построению последовательности (n+2)-точечных наборов $S_j = \left\{x_0^{(j)}, x_1^{(j)}, \ \ldots, \ x_{n+1}^{(j)}\right\} \subset E$, сходящейся к чебышевскому альтернансу.

Скорость сходимости метода п.ч.и. зависит в основном от способа замены наборов S_j. Возможны три варианта замены точек при переходе от S_j к следующему, улучшенному набору S_{j+1}: *оптимальный*, *полуоптимальный* и *допустимый*. На практике в *оптимальном* варианте замены искомый чебышевский альтернанс находится, как правило, за 1–2 итерации, в *полуоптимальном* число итераций может оказаться в несколько раз больше, а в *допустимом* – во много раз больше. В случае оптимального варианта для некоторых классов функций обеспечивается *квадратическая скорость сходимости* [1].

В разработанных в Институте кибернетики алгоритмах п.ч.и. наилучшего чебышевского приближения аппроксимантами разных классов замена наборов S_j осуществляется в соответствии с предложенной в работе [4] процедурой, реализующей *усиленный полуоптимальный вариант* замены, который на практике совпадает с оптимальным и обеспечивает квадратическую скорость сходимости итерационного процесса.

Кроме оптимального варианта замены наборов S_j, указанные алгоритмы имеют ряд возможностей и преимуществ по сравнению с алгоритмами п.ч.и. других авторов, в частности, обеспечивают построение полиномиального приближения с произвольным весом; позволяют находить либо аппроксимант заданной фиксированной степени (вход по степени), либо аппроксимант, обеспечивающий заданную точность приближения (вход по точности); позволяют аппроксимировать как дискретно, так и аналитически заданные функции. При этом дополнительно применяется процедура вычисления значений функции в точках дискретизации.

Для разработанных алгоритмов были получены оценки всех видов погрешностей, в частности, априорные и апостериорные мажорантные детерминированные оценки полной погрешности, причём неулучшаемые для некоторых классов функций [5]. Включение в вычислительные схемы алгоритмов расчетов полной погрешности приближения позволило получить более точную оценку величины наилучшего приближения ρ.

Кроме вышеперечисленных, численная реализация разработанных в Институте кибернетики алгоритмов п.ч.и. имеет также дополнительные преимущества, связанные с оптимизацией алгоритмов по точности и быстродействию [4, 5]. Эти меры оптимизации позволили значительно повысить точность результатов.

Алгоритмы и программные комплексы наилучшей чебышевской аппроксимации эффективно применялись на протяжении многих лет для решения большого числа задач замены дискретного представления функциональных зависимостей разными классами аппроксимантов.

Кроме разработанных ранее базовых алгоритмов и программ аппроксимации функций одной и многих переменных полиномами $P_n(x)$ и дробно-рациональными выражениями $R_{mk}(x) = P_m(x)/P_k(x)$ [6, 7], в последнее время были разработаны методы, алгоритмы и программные комплексы построения наилучших чебышевских приближений другими классами аппроксимантов [8]:

экспоненциальными выражениями $a_0\, exp\left(a_1 x + \ldots + a_n x^n\right)$;

логарифмическими выражениями $ln\left(a_0 + a_1 x + \ldots + a_n x^n\right)$;

функциями, являющимися корнем из полинома $\sqrt[l]{a_0 + a_1 x + \ldots + a_n x^n}$.

Для дальнейшего повышения эффективности сжатия больших массивов данных был проведен анализ сравнительных характеристик точности приближения аппроксимантами разных классов. По результатам анализа определён класс обобщённых полиномов $Q_n(x) = z_1\varphi_1(x) + \ldots + z_n\varphi_n(x)$ по системам базисных функций $\{\varphi_1(x), \ldots, \varphi_n(x)\}$ и разработан алгоритм и программный комплекс наилучшего равномерного приближения этими полиномами [7]. В дальнейшем планируется определить перечень наиболее подходящих систем базисных функций и выполнить аппроксимацию полиномами $Q_n(x)$ с использованием различных систем функций для того, чтобы на основе сравнения полученной точности окончательно выбрать такую систему, которая обеспечит наилучшую точность аппроксимации указанными полиномами в конкретном случае.

Для повышения эффективности (точности и быстродействия) всех перечисленных алгоритмов был также реализован подход, основанный на применении сегментной (кусочной) аппроксимации разными классами аппроксимантов [9].

Для работы с программами сжатия и восстановления массивов числовых данных была разработана программная оболочка в системе программирования Delphi. Работа с этими программными средствами позволяет в режиме диалога задавать решение прямой (сжатие) либо обратной (восстановление) задачи, выбирать тот или другой вид аппроксиманта и получать параметры соответствующих результатов, таких, как размер сжатого файла, коэффициент сжатия и др.

Работа в этой программной оболочке не требует от пользователя профессиональных знаний по программированию и не предусматривает необходимости изучения особенностей алгоритмов аппроксимации. Пользователь должен задать только имена файла массива данных и файла результатов сжатия и выбрать желаемое аппроксимирующее выражение из предложенного списка.

Эффективность разработанных алгоритмов сжатия алгебраическими и обобщёнными полиномами, в том числе с использованием сегментной аппроксимации, была подтверждена на примерах сжатия больших массивов числовых данных. Далее приводятся некоторые примеры сжатия (рис. 1).

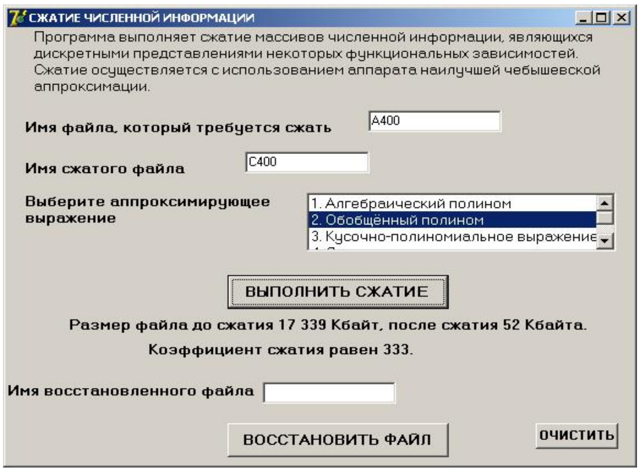

РИС. 1. Вид диалогового окна программной оболочки

Примеры сжатия массивов числовых данных.

Пример 1. Рассмотрим применение разработанного аппарата наилучшего чебышевского приближения для сжатия массива данных измерений солёности воды (всего 1897 замеров), полученных на разных глубинах Чёрного моря, с использованием полиномиальной и кусочно-полиномиальной аппроксимации с разбиением на сегменты.

Была выполнена аппроксимация без разбиения на сегменты полиномами степеней 2–15 и аппроксимация с разбиением на 2 сегмента полиномами соответственно степеней 4–11 на первом сегменте и степеней 1–2 на втором сегменте. На рис. 2 показан график функциональной зависимости солёности воды, построенный по данным измерений, и график полинома 5-й степени, аппроксимирующий эту зависимость с погрешностью не более 2.4 %.

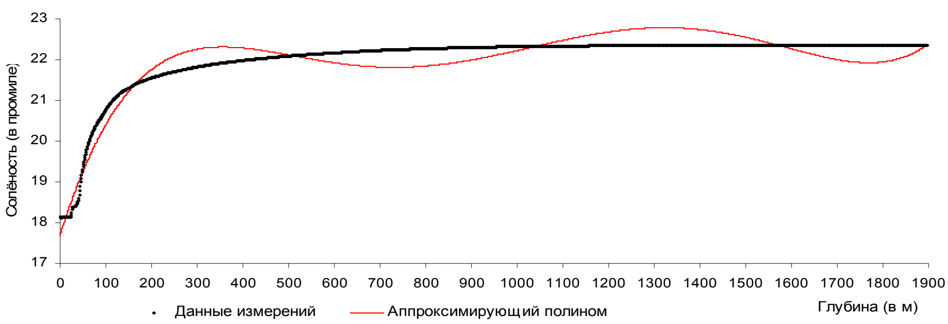

РИС. 2. Зависимость солёности воды от глубины

В сводной таблице приведены значения погрешностей приближений и коэффициентов сжатия для полиномиальной и кусочно-полиномиальной аппроксимации данных измерений солёности воды.

Полиномиальная аппроксимация		Кусочно-полиномиальная аппроксимация			Коэффициент сжатия C
Степень полинома	Погрешность аппроксимации, %	Степень 1-го полинома	Степень 2-го полинома	Погрешность аппроксимации, %	
9	2.1	6	2	0.9	190
10	1.96	7	2	0.77	172
12	1.5	8	2	0.75	146
14	1.2	11	2	0.61	126

Сравнение соответствующих значений позволяет сделать вывод, что при одинаковых коэффициентах сжатия и количестве параметров приближения значительно более эффективной по точности является кусочно-полиномиальная аппроксимация (рис. 3).

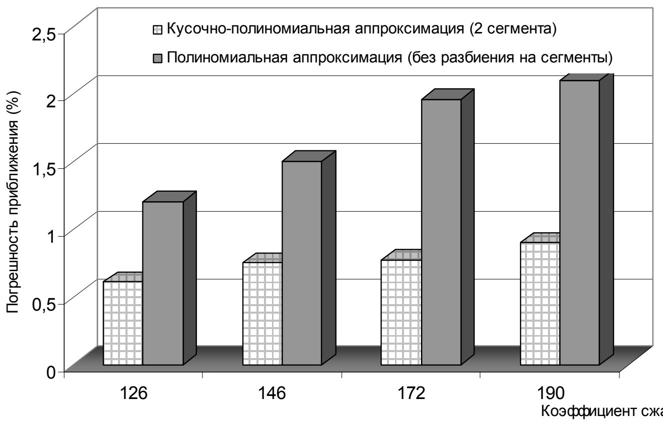

РИС. 3. Соотношение между погрешностями приближения и коэффициентами сжатия

<u>Пример 2.</u> Рассмотрим применение аппроксимации обобщёнными полиномами для сжатия числовой матрицы А размером 1500х1500 (всего 2 250 тыс. действительных чисел). Анализ матрицы показывает, что в размещении элементов в строках наблюдается определённая периодичность. В качестве значений y_j дискретно заданной функциональной зависимости $f_i(x)$ взяты значения элементов i-й строки матрицы А $\left(i=\overline{1,1500}\right)$, а в качестве аргумента x – порядковый номер j значения y_j в строке, т. е. $x_j = j$ и $y_j = f_i(x_j)$ $\left(j=\overline{1,1500}\right)$.

Из показанного на рис. 4 графика функции $f_1(x)$ видно, что функция периодическая. Поэтому для её аппроксимации целесообразно использовать обобщённый полином $Q_2^{(1)}(x) = z_1 + z_2\cos(x-1)$ по системе двух базисных функций 1 и $\cos(x-1)$.

Каждая строка i матрицы А аппроксимируется своим полиномом $Q_2^{(i)}(x)$ $\left(i=\overline{1,1500}\right)$. При этом абсолютная погрешность приближения не превышает половины единицы младшего разряда числа. Например, обобщённый полином, аппроксимирующий функцию $f_1(x)$, имеет вид $Q_2^{(1)}(x) = 500 + 300 \cdot \cos(x-1)$.

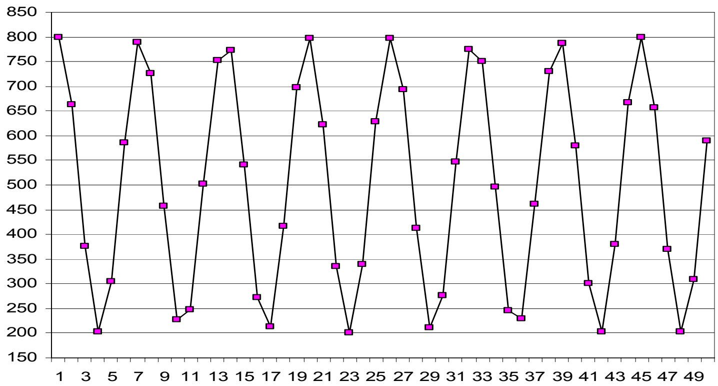

РИС. 4. Фрагмент графика функции $f_1(x)$

В результате аппроксимации матрицу А размером 17339 Кбайт удалось сжать до 52 Кбайт с коэффициентом сжатия 333 (рис. 5).

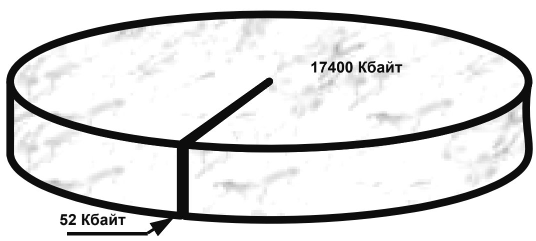

РИС. 5. Размеры матрицы до и после сжатия

Выводы. На практике эффективность разработанных алгоритмов и программ наилучшей чебышевской аппроксимации проверена на большом количестве численных реализаций как в Украине, так и за рубежом. Аппарат аппроксимации на протяжении многих лет использовался в составе прикладного программного обеспечения отечественных ЭВМ и применялся для сжатия массивов данных при решении различных прикладных задач (в том числе, для оборонных целей), при расчётах характеристик сложных систем, например, при расчёте прочностных характеристик летательных аппаратов для НИИ им. Туполева.

В процессе применения аппарата наилучшей чебышевской аппроксимации была не только подтверждена его высокая эффективность, но и проводилось дальнейшее усовершенствование этого аппарата.

В результате численной реализации алгоритмов аппроксимации были разработаны программные комплексы на языках программирования ФОРТРАН, Алгол, Паскаль, а также на С++ для отечественного суперкомпьютера с кластерной архитектурой (СКИТ). В состав прикладного программного обеспечения СКИТ включены две библиотеки программ: библиотека чебышевской аппроксимации функций одной и многих переменных и библиотека для вычисления с повышенной точностью значений элементарных и специальных функций.

Аппарат чебышевской аппроксимации в последнее время применялся для сжатия *больших одномерных массивов-векторов* (с возможным количеством значений до 10 млн. чисел) с целью получения *небольшого числа параметров* аппроксимантов. В результате расчётов были получены большие значения коэффициентов сжатия (в среднем два порядка). Эти работы выполнялись в рамках создания Подсистемы аппроксимации для сжатия больших массивов числовых данных в составе Информационно-аналитической системы «Бюджетный комитет», а также в рамках научно-технического проекта по разработке программно-технических комплексов для решения расчётных задач АНТК «Антонов».

В настоящее время для СКИТ разрабатывается пакет программ аппроксимации функций одной и многих переменных разными способами приближения: интерполяционным, среднеквадратичным и чебышевским [10]. Этот пакет имеет ряд существенных преимуществ по сравнению с известными аналогичными пакетами и специализированными библиотеками, такими, например, как Mathcad, Maple, MATLAB, Mathematica, MATHLIB, NETLIB.

А.О. Каленчук-Порханова, *Л.П. Вакал*

НАЙКРАЩА ЧЕБИШОВСЬКА АПРОКСИМАЦІЯ ДЛЯ СТИСНЕННЯ ЧИСЕЛЬНОЇ НФОРМАЦІЇ

Розглядається проблема стиснення чисельної інформації та її вирішення з використанням найкращої чебишовської апроксимації. Дається обґрунтування переваг розроблених алгоритмів апроксимації, пов'язаних з їх оптимізацією за точністю і швидкодією. Наводяться деякі результати розрахунків по стисненню числових масивів з великими коефіцієнтами стиснення.

A. Kalenchuk-Porkhanova, *L. Vakal*

BEST CHEBYSHEV APPROXIMATIONS FOR NUMERICAL INFORMATION COMPRESSION

The problem of numerical information compression and its solution with the use of best uniform approximation are discussed. The advantages of the algorithms elaborated connected with their accuracy and performance optimization are presented. Some examples of numerical arrays compression with large values of compression coefficients are given.

1. *Ремез Е.Я.* Основы численных методов чебышевского приближения. – Киев: Наук. думка, 1969. – 623 с.
2. *Иванов В.В., Каленчук А.А.* Об эффективности алгоритмов полиномиальных и дробно-рациональных чебышевских приближений // Конструктивная теория функций. – София, 1983. – С. 72–77.
3. *Бердышев В.И.*, *Петрак Л.В.* Аппроксимация функций, сжатие численной информации, приложения. – Екатеринбург: УрО РАН, 1999. – 297 с.
4. *Каленчук-Порханова А.А.* Об одном алгоритме полиномиальной чебышевской аппроксимации // Оптимизация вычислительных методов. – Киев: Ин-т кибернетики АН УССР, 1974. – С. 45–51
5. *Каленчук-Порханова А.А.* Алгоритмы и анализ погрешности наилучшей чебышевской аппроксимации одной переменной // Теория приближения функций: Тр. Междунар. конф. по теории приближения функций, Калуга, 1975. – М., 1977. – С. 213–218.
6. *Каленчук-Порханова А.А.* Аппроксимация функций одной и многих переменных // Численные методы для многопроцессорного вычислительного комплекса ЕС. – М.: Изд-во ВВИА им. Н.Е. Жуковского, 1987. – С. 366–395.
7. *Каленчук-Порханова А.О.*, *Вакал Л.П.* Побудова найкращих рівномірних наближень функцій багатьох змінних // Комп'ютерні засоби, мережі та системи. – 2007. – № 6. – С. 141–148.
8. *Каленчук-Порханова А.О.*, *Вакал Л.П.* Відтворення функціональних залежностей на основі нелінійних наближень деяких видів // Abstracts of International Conf. "Problems of decision making under uncertainties" (May 21–25, 2007). – Chernivtsi, Ukraine, 2007. – P. 135–137.
9. *Вакал Л.П.* Рівномірне кусково-поліноміальне наближення // Комп'ютерні засоби, мережі та системи. – 2006. – № 5. – С. 53–59.
10. *Каленчук-Порханова А.А.*, *Вакал Л.П.* Пакет программ аппроксимации функций // Комп'ютерні засоби, мережі та системи. – 2008. – № 7. – С. 101–107.

Получено 03.12.2008

Об авторах :

Каленчук-Порханова Анжелина Алексеевна,
кандидат физико-математических наук, старший научный сотрудник
Института кибернетики имени В.М. Глушкова НАН Украины,
e-mail: ioanna@public.icyb.kiev.ua

Вакал Лариса Петровна,
научный сотрудник Института кибернетики имени В.М. Глушкова НАН Украины.
e-mail: vakal@ipnet.ua