

Об эффективности автоматизированного поиска оптимальных приближений с гарантированной точностью в системах с плавающей точкой

А. А. Чусов, Ю. И. Ефимова✉

Дальневосточный федеральный университет, Владивосток, Россия

✉ efimova.iui@dvfu.ru

Аннотация. Представлен метод автоматизированной параллельной минимаксной аппроксимации функций $\mathbb{R} \rightarrow \mathbb{R}$ полиномами $\mathbb{Z}_{b^p} \rightarrow \mathbb{Z}_{b^p}$ из $\mathbb{Z}_{b^p}[x]$ – т. е. определенными над подмножеством рациональных b -ичных чисел конечной точности p . Такие полиномы гарантируют существование верхней границы абсолютной ошибки приближения, однако эти гарантии в литературе формулируются относительно всего множества действительных чисел $\mathbb{R}$, что делает их, строго говоря, применимыми лишь в области абстрактного анализа в противоположность реализуемым численно, в конечной точности, систем нелинейных уравнений с помощью многомерного метода Хаусхолдера. Использование многомерного метода позволяет обобщить существующие реализации поиска минимаксных полиномов на случаи произвольных дифференцируемых на области определения функций.

Цель публикации заключается в обеспечении заданной точности аппроксимации в наихудших случаях, поэтому расчет минимаксных полиномов выполняется с помощью длинной рациональной арифметики; его результаты округляются до чисел с плавающей запятой выбранной точности, и при этом обеспечиваются гарантии относительно численных ошибок не более половины единицы-на-последнем-месте (ULP) [1].

Полученные результаты, включающие в себя реализацию алгоритма Ремеза на основе многомерного метода Хаусхолдера и методов длинной рациональной арифметики, реализованы экспериментально и доступны читателю для верификации.

Ключевые слова: гарантии точности, многомерный метод Хаусхолдера, длинная рациональная арифметика, численные ошибки и гарантии ULP, алгоритм Ремеза, неравномерная аппроксимация функций

Для цитирования: Чусов А. А., Ефимова Ю. И. Об эффективности автоматизированного поиска оптимальных приближений с гарантированной точностью в системах с плавающей точкой // Изв. СПбГЭТУ «ЛЭТИ». 2025. Т. 18, № 8. С. 59–67. doi: 10.32603/2071-8985-2025-18-8-59-67.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Финансирование: Работа выполнена в рамках государственного задания Минобрнауки России в сфере научной деятельности по проекту № FZNS-2023-0008 «Разработка новых адаптивных нейросетевых методов повышения помехоустойчивости и энергоэффективности цифровых систем связи».

Original article

On the Effectiveness of Automated Search for Optimal Approximations with Guaranteed Accuracy in Floating-Point Systems

A. A. Chusov, Yu. I. Efimova✉

Far-Eastern Federal University, Vladivostok, Russia

✉ efimova.iui@dvfu.ru

Abstract. The paper presents a method for automated parallel minimax approximation of functions $\mathbb{R} \rightarrow \mathbb{R}$ by polynomials $\mathbb{Z}_{b^p} \rightarrow \mathbb{Z}_{b^p}$ из $\mathbb{Z}_{b^p}[x]$ – that is, defined over a subset of rational b -numbers of finite precision p . Such polynomials guarantee the existence of an upper bound on the absolute approximation error, but

these guarantees are formulated in the literature with respect to the entire set of real numbers $\mathbb{R}$, which makes them, strictly speaking, applicable only in the field of theoretical analysis, as opposed to the numerical, with finite precision, solutions of systems of nonlinear equations using the multidimensional Householder. The use of the multidimensional method makes it possible to generalize existing implementations of the search for minimax polynomials to cases of arbitrary differentiable functions in the domain of definition.

The purpose of the work is to provide guarantees regarding the accuracy of the approximation in the worst cases, therefore, the calculation of minimax polynomials is performed using long rational arithmetic; the results are rounded to floating-point numbers of the selected precision which provides guarantees regarding numerical errors of no more than half the unit-in-last-place (ULP) [1].

The results obtained by the authors include a numerical implementation of the Remez exchange algorithm based on the multidimensional Householder method in arbitrarily finite precision. The experimental implementation is described in the paper as well as the results of the experimental research. The implementation, and therefore the results, are available for readers to be verified.

Keywords: accuracy guarantees, multidimensional Householder method, long rational arithmetic, numerical errors and ULP guarantees, Remez algorithm, non-uniform function approximation

For citation: Chusov A. A., Efimova Yu. I. On the Effectiveness of Automated Search for Optimal Approximations with Guaranteed Accuracy in Floating-Point Systems // LETI Transactions on Electrical Engineering & Computer Science. 2025. Vol. 18, no. 8. P. 59–67. doi: 10.32603/2071-8985-2025-18-8-59-67.

Conflict of interest. The authors declare no conflicts of interest.

Financing: The work was carried out as part of the state assignment of the Ministry of Education and Science of the Russian Federation in the field of scientific activity under project no. FZNS-2023-0008 «Development of New Adaptive Neural Network Methods for Improving the Noise Immunity and Energy Efficiency of Digital Communication Systems».

Введение. Аппроксимация – ключевой инструмент в математике и технике, позволяющий заменять сложные объекты более простыми для решения задач. Нередко аналитического решения не существует, и тогда используются численные методы аппроксимации. Теорема Вейерштрасса показывает, что непрерывную функцию можно с любой точностью приблизить полиномом. Современная теория аппроксимации изучает методы приближения функций с использованием более простых, «базисных» функций, и оценивает погрешности. Выбор метода аппроксимации зависит от условий задачи, требований к точности и доступных ресурсов. Одна из важных задач – реализация приближений в компьютерных библиотеках так, чтобы получаемый результат был максимально точным и соответствовал правилам округления, т. е. с погрешностью не более половины единицы младшего разряда.

Это достигается за счет использования многочлена высокой степени и сужения области, в которой многочлен должен аппроксимировать функцию. Сужение области часто может быть достигнуто за счет использования различных формул сложения или масштабирования для аппроксимируемой функции.

Современные математические библиотеки часто разбивают область определения аппроксимируемой функции на множество малых сегментов и используют полином низкой степени для каждого сегмента. Как только область (обычно интервал на $\mathbb{R}$ или n -мерный куб Лебега на $\mathbb{R}^n$) и степень многочлена выбраны, сам многочлен выбирается таким образом, чтобы минимизировать ошибку наихудшего случая. Т. е. цель состоит в том, чтобы минимизировать на области определения x максимальное значение:

$$\|P(x) - f(x)\|_{\infty},$$

где $P(x)$ – аппроксимирующий полином; $f(x)$ – фактическая функция; обе выполняют отображение $S_1 \rightarrow S_2$, где $S_1 S_2 \subset (\mathbb{R}^n)^2$, $n \in \mathbb{N}$.

Улучшить аппроксимацию точки минимума можно за счет полиномов более высокой степени или сужения интервала аппроксимации. В численных вычислениях ошибки неизбежны из-за ограниченного количества значащих цифр, ошибок округления и других приближений. Инструменты, такие как логарифмические линейки и калькуляторы, также дают приближенные результаты. Компьютерные вычисления имеют ограниченную точность из-за разрядности представления чисел. Точность

можно повысить, используя более точные методы или увеличивая разрядность чисел. Аппроксимация необходима, если число не может быть представлено конечным числом разрядов.

Для реализации вычислений с произвольными гарантиями точности может оказаться недостаточным использование лишь распространенных форматов чисел с плавающей точкой – IEEE-754 и т. п., требуется метод кастомизации точности представления и операций над рациональными значениями, основанный на длинной целочисленной арифметике [2].

Гарантии точности, в свою очередь, оказывают значительный негативный эффект на оперативность и энергоэффективность решения вычислительной проблемы, основанной на вещественной арифметике, причем последняя во многих случаях оказывается в значительной степени ограничивающим фактором, например автономности, мобильности, стоимости и применимости решения.

Предложенные методы могут быть использованы для численного решения оптимизационных задач поиском их минимаксной аппроксимации.

В численном анализе методы Хаусхолдера – это класс алгоритмов поиска корней уравнений функций одной переменной из $\mathbb{R}^n$, которые на всей области определения (для которой осуществляется аппроксимация) должны быть дифференцируемы столько же раз, каков порядок метода Хаусхолдера.

Методы Хаусхолдера представляют собой численные алгоритмы решения нелинейного уравнения $f(x) = 0$. В этом случае функция f должна быть функцией из некоторого евклидова пространства [3]. Эти методы состоят из последовательности итераций. В статье рассмотрены методы двух первых порядков – Ньютона и Хейли.

Метод Ньютона (метод касательных, метод секущих) – алгоритм нахождения корней дифференцируемых функций, который использует итеративную функцию для аппроксимации ее корней. Это – простой и эффективный алгоритм для приближенного нахождения корней действительных функций, т. е. решения уравнений вида $f(x) = 0$. На функцию накладываются требования наличия хотя бы одного корня и дифференцируемости на интервале поиска, мультипликативной инверсии ее производной на этом интервале [4].

Пусть для случая $\mathbb{R} \rightarrow \mathbb{R}$ имеется начальное приближение для корня x_0 (которое, например, подбирается аналитически) и известно, что корень лежит в интервале $[a; b]$. Тогда, используя некоторое априорно известное приближение функции f и производной f' в точке x_0 , например рядом Тейлора, из определения производной следует:

$$f(x) \approx f(x_0) + (x - x_0)f'(x_0).$$

Тогда можно построить рекуррентное соотношение для нахождения корня:

$$x_{i+1} \approx x_i - \frac{f(x_i)}{f'(x_i)},$$

где i – это номер шага алгоритма.

В более общем случае $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ аналогичные формы имеют вид

$$f(x) \approx f(x_0) + J_{nm}(x_0)(x - x_0), \quad (1)$$

$$x_{i+1} \approx x_i - J_{nm}^{-1}(x_i)f(x_i),$$

где $J_{nm}(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$ – матрица Якоби, и

$f_i: \mathbb{R} \rightarrow \mathbb{R}$ – i -я компонента вектор-функции f .

Относительно требований к памяти хранение промежуточных значений при итерационном вычислении x_i не требуется, достаточно иметь только две переменные: для текущего i -го значения и для нового $(i + 1)$ -го.

Также требуется реализовать f и ее Якобиан J_{nm} – как вычисляемые функции из $\mathbb{R}^n$.

Метод Ньютона очень важен в вычислительной математике: в большинстве случаев именно его используют для нахождения численных решений уравнений. Метод Ньютона обладает квадратичной сходимостью, т. е. на каждой итерации абсолютное отклонение приближенного значения корня от истинного возводится в квадрат, т. е. число верных знаков удваивается [5].

Метод Хэйли (Halley's method, tangent hyperbolas method, метод касательных гипербол) – это алгоритм нахождения корня, используемый для дважды дифференцируемых функций одной действительной переменной с непрерывной второй производной. Метод состоит из последовательности итераций:

$$x_{i+1} = x_i - \frac{2f(x_i)f'(x_i)}{2f'(x_i)^2 - f(x_i)f''(x_i)}.$$

В общем случае $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ аналогичное значение будет иметь вид

$$x_{i+1} = x_i - \left(2J_{nn}(x_i)^2 - f(x_i)H_{nn}(x_i) \right)^{-1} 2J_{nn}(x_i)f(x_i),$$

где

$$H_{nn}(x) = \begin{bmatrix} \frac{\partial^2 f_1}{\partial x_1 x_1} & \dots & \frac{\partial^2 f_1}{\partial x_1 x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f_1}{\partial x_n x_1} & \dots & \frac{\partial^2 f_1}{\partial x_n x_n} \end{bmatrix} \dots \begin{bmatrix} \frac{\partial^2 f_n}{\partial x_1 x_1} & \dots & \frac{\partial^2 f_n}{\partial x_1 x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f_n}{\partial x_n x_1} & \dots & \frac{\partial^2 f_n}{\partial x_n x_n} \end{bmatrix}$$

– n -мерный тензор 3-го порядка.

По сравнению с методом Ньютона преимущество метода Хэйли заключается в том, что он сходится быстрее (у метода Хэйли кубическая сходимость), однако для этого требуется ресурсоемкое вычисление тензора $H_{nn}(x)$, как видно из (1). Поскольку эти два метода часто взаимозаменяемы (это означает, что для данного корня условия сходимости выполняются для обоих методов), метод Хэйли предпочтителен только в случаях, когда H_{nn} достаточно легко вычислить. В большинстве же многомерных случаев, включая исследованные экспериментально в настоящей статье, метод Хаусхолдера первого порядка оказывается достаточным и предпочтительным.

Постановка задачи. Задача нахождения минимаксных приближений, обеспечивающих наилучшее приближение функции при наихудшем случае, важна во многих областях. Однако поиск таких приближений часто требует значительных вычислительных затрат. Существует потребность разработки методов, обеспечивающих автоматизированный и эффективный поиск минимаксных приближений с гарантированной точностью, при этом минимизируя затраты на вычисления.

Цели исследований:

1. Разработка и анализ новых автоматизированных методов поиска минимаксных приближений.
2. Исследование способов повышения энергоэффективности и скорости работы таких методов.
3. Демонстрация эффективности предложенных методов на примерах.

Численная реализация минимаксного приближения. Алгоритм минимаксной аппроксимации – это метод нахождения приближения мате-

матической функции, которое минимизирует максимальную ошибку [4]. Основным результатом минимаксной аппроксимации служит теорема, принадлежащая Чебышеву: p^* является минимаксным приближением степени n к функции f на $[a, b]$ тогда и только тогда, когда существуют по крайней мере $n+2$ значения $a \leq x_0 < x_1 < \dots < x_{n+1} \leq b$ таких, что

$$p^*(X_i) - f(X_i) = (-1)^i [p^*(X_0) - f(X_0)] = \pm \|f - p^*\|_\infty.$$

Предположим, что $p^*(x) = a_0 + a_1x + a_2x^2$ – минимаксное приближение к $\sin(x)$ на отрезке $\left[0, \frac{\pi}{2}\right]$. Из теоремы Чебышева следует, что существуют по меньшей мере четыре значения x_0, x_3 при которых максимальная ошибка аппроксимации достигается чередующимися знаками. Выпуклость синусоидальной функции подразумевает $x_0 = 0$ и $x_3 = \frac{\pi}{2}$. С помощью минимаксного приближения можно составить систему нелинейных уравнений следующего вида:

$$\begin{cases} a_0 + a_1x_1 + a_2x_1^2 - \sin(x_1) = -\varepsilon, \\ a_1 + 2a_2x_1 - \cos(x_1) = 0; \\ a_0 + a_1x_2 + a_2x_2^2 - \sin(x_2) = -\varepsilon, \\ a_1 + 2a_2x_2 - \cos(x_2) = 0; \\ a_0 + a_1(0) + a_2(0)^2 - \sin(0) = -\varepsilon; \\ a_0 + a_1\frac{\pi}{2} + a_2\left(\frac{\pi}{2}\right)^2 - \sin\left(\frac{\pi}{2}\right) = -\varepsilon. \end{cases} \quad (2)$$

Для решения задачи минимаксного приближения с помощью метода Хаусхолдера был создан следующий алгоритм:

1. Предварительная оценка точек, в которых ожидается максимальное значение ошибки аппроксимации. Абсциссы точек выбираются равными узлам Чебышева $X_{\text{Ч}}$.

2. Предварительная расчет коэффициентов полинома [6]. Для расчета коэффициентов полинома выполняется интерполяция Ньютона, основанная на вычислении разделенных разностей. Компоненты разделенных разностей хранятся в таблице поиска. Значение функции вычисляется иными, менее эффективными методами, приведенными авторами в [7], например с помощью аппроксимации Тейлора, пока относительная ошибка превышает машинную ε [5], [8].

3. Использование алгоритма Ремеза [9] для циклического нахождения оптимального многочлена аппроксимации для заданной функции на области определения, и вычисление этого многочлена методом Хаусхолдера для численного приближения этой функции с использованием найденных коэффициентов многочлена [10].

Сначала фиксируются значения x для нахождения коэффициентов полинома. Их значения вычисляются с помощью многомерного алгоритма Хаусхолдера первого порядка:

$$a = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \\ \varepsilon \end{bmatrix} - \begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n & 1 \\ 1 & x_1 & x_1^2 & \cdots & x_1^n & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n & (-1)^n \\ 1 & x_{n+1} & x_{n+1}^2 & \cdots & x_{n+1}^n & (-1)^{n+1} \end{bmatrix}^{-1} \begin{bmatrix} g_0 \\ g_1 \\ \vdots \\ g_n \\ g_{n+1} \end{bmatrix}, \quad (3)$$

где $g_k(a_0, \dots, a_n, \varepsilon)$ – значение полинома в k -м предположенном узле a_k Чебышева; a – новые значения коэффициентов полинома, которые равны

$$g_k(a_0, \dots, a_n, \varepsilon) = \sum_{i=0}^n a_i x_k^i - (-1)^k, \quad \text{а также}$$

предполагаемой максимальной ошибки полиномиального приближения ε .

Далее с найденными коэффициентами a находим значения x , уточняя расположение в области определения узлов Чебышева на основе ранее предположенных значений $x_0, \dots, x_{n+1}$:

$$x = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{n+1} \end{bmatrix} - \begin{bmatrix} \sum_{i=0}^n a_i x_0^{i-1} - \frac{df}{dx} \Big|_{x=x_0} & 0 & \cdots & 0 & 0 \\ 0 & \sum_{i=0}^n a_i x_1^{i-1} - \frac{df}{dx} \Big|_{x=x_1} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \sum_{i=0}^n a_i x_{n+1}^{i-1} - \frac{df}{dx} \Big|_{x=x_{n+1}} \end{bmatrix}^{-1} \begin{bmatrix} h(x_0) \\ h(x_1) \\ \vdots \\ h(x_{n+1}) \end{bmatrix}, \quad (4)$$

где $h(x_k)$ – это новые значения x , которые равны

$$h(x_k) = \sum_{i=0}^n a_i x_k^i - (-1)^k \varepsilon - f(x) \quad \text{с полученными в}$$

результате (3) коэффициентами $a_0, \dots, a_n$ и ошибкой ε .

4. Оценка того, насколько расходитсся результат шага 3, т. е. a и x , по сравнению с предыдущими итерациями алгоритма Ремеза: если коли-

чество старших разрядов мантииссы каждого числа с плавающей точкой, которые остаются неизменными между итерациями, уменьшается, это воспринимается как признак расходимости реализации [9]. Отметим, что даже если выполняется условие Хаара [2], сходимость алгоритма Ремеза, имеющая место в $\mathbb{R}$, может отсутствовать в конечном множестве чисел с плавающей точкой, аппроксимирующих $\mathbb{R}$.

Реализация экспериментов по измерению эффективности и обсуждение результатов. Все вышеприведенные операции, кроме отображения f и конечной разности, аппроксимирующей значение производной в точке, а также при вычислении узлов Чебышева $X_{\text{Ч}}$ и $f(X_{\text{Ч}})$ для начальной оценки $x_0, \dots, x_{n+1}$ и $a_0, \dots, a_n, \varepsilon$ в (3) и (4), выполняют вычисления в поле рациональных чисел $\mathbb{Q}$. Поэтому, если используемая в реализации числовая система позволяет использовать неограниченно большую, но конечную, точность мантииссы, то в такой системе циклическое вычисление (3) и (4) можно реализовать без потерь в точности, связанной с аппроксимацией до фиксированной точности мантииссы, но с учетом потерь точности при вычислении значений f и df/dx .

Поэтому для реализации экспериментов авторами использована числовая система:

$$Q = \left\{ (-1)^s \left(I + \frac{N}{D} \right) \beta^E \right\} \cup \{-\infty, \infty\} \subset \mathbb{Q} \cup \{-\infty, \infty\},$$

где s – бит знака; беззнаковые целочисленные I, N, D определяют мантииссу числа ($Q \ni x < \infty \Rightarrow D \neq 0$);

целочисленная E со знаком задает экспоненту; β – это основание системы счисления. Очевидно, что если $N=0$, то соответствующий конечный элемент Q будет числом с плавающей точкой (не учитывая дополнительных действий – таких, например, как нормализация двоичных значений IEEE-754, не приводящих к потерям информации). Тогда эвклидовым делением $(ID + N)/D$ можно выполнить округление конечного резуль-

тата описанного алгоритма до числа с плавающей точкой с ошибкой не более 0.5 ULP [1].

Далее приведены результаты измерений оперативности и энергоэффективности реализации поиска минимакса. Хотя измерение оперативности реализуется на современных универсальных процессорах тривиально (приведенные далее значения получены с помощью часов `std::chrono::steady_clock` стандартной библиотеки C++), доступ к датчикам энергопотребления в значительно большей степени ограничен. Например, современные микропроцессоры Intel и AMD предоставляют доступ к этим датчикам посредством регистров MSR (Model-Specific Register), чтение которых осуществляется с помощью инструкции RDMSR. Однако из соображений безопасности выполнение этой инструкции ограничивается только адресным пространством ядра операционных систем Windows и Linux. И хотя последняя, в случае, когда не используется виртуализация, отображает эти регистры на файловую систему по адресу `/dev/cpu/<номер логического ядра>/msr`, для доступа из-под Windows требуется использование драйвера пространства ядра.

Значения энергопотребления считаны посредством обращения к регистрам 0x611 (энергопотребление) и 0x606 (выраженное в джоулях значение единицы энергопотребления) микропроцессора Intel Xeon E5-2695 v2.

Реализация всех экспериментов доступна читателю в GIT-репозитории по веб-адресу <https://gitlab.com/lpsztemp/parallel-remez>.

Были рассмотрены две версии реализации алгоритма посредством написания кода: последовательная и параллельная. Для оценки энергоэффективности и оперативности алгоритма были выбраны критерии оценки: потребление энергии алгоритмом (в джоулях) и время выполнения разных частей алгоритма (в миллисекундах). Для измерения оперативности и энергоэффективности были взяты точки отсчета для последовательной формы: после поиска коэффициентов полинома (2) и после поиска корней полиномов (3), т. е. было измерено время всего цикла, который находит решение Хаусхолдера. Предложенный подход к выбору контрольных точек обоснован их соответствием ключевым этапам вычислительного процесса, оказывающим существенное влияние на общую производительность алгоритма. Данный методологический прием обеспечивает возможность детального анализа динамики работы алгоритма, что позволяет выявить различия в

энергопотреблении и временных характеристиках между последовательной и параллельной реализациями, а также служит основой для валидации их корректности. Полученная детализированная оценка производительности, основанная на методах эмпирического исследования, есть необходимая предпосылка для выявления потенциальных преимуществ параллельной реализации и оптимизации ее параметров.

Результаты измерений были сведены в таблицы и представлены в виде графиков: для значений энергии – табл. 1 и рис. 1, для значений времени – табл. 2 и рис. 2.

Табл. 1. Значения энергий, используемые для нахождения многочленов заданных степеней и с заданной точностью

Tab. 1. Energy values used to find polynomials of specified degrees and with specified accuracy

Степень полинома	Потребление энергии, Дж	
	Поиск коэффициентов полиномов	Поиск корней полиномов
1	3.10463	1.00381
2	6.49492	5.13419
3	17.3419	21.7493
4	38.0201	53.1864
5	93.102	137.128
6	8853.73	267.377
7	359.75	652.989
8	611.58	1102.24
9	12 527	2260.15
10	1867.73	3278.35
11	3337.11	6562.91
12	36 408.8	8384.93
13	7644.48	14 465.7
14	11 285.8	17 429.4
15	15 987.3	35 343.6
16	21 869	44 482.9
17	5616.88	65 195.2
18	42 544.2	19 190.7
19	61 039.3	61 770.5
20	54 853.8	24 511.3

Энергопотребление сильно варьируется в зависимости от степени полинома и этапа вычислений. Нет однозначной тенденции к росту или снижению потребления энергии с увеличением степени полинома. На большинстве степеней полинома этап поиска корней полиномов потребляет больше энергии, чем поиск коэффициентов полиномов.

В целом, время выполнения на всех этапах увеличивается с ростом степени полинома, что и ожидается, поскольку усложнение вычислений

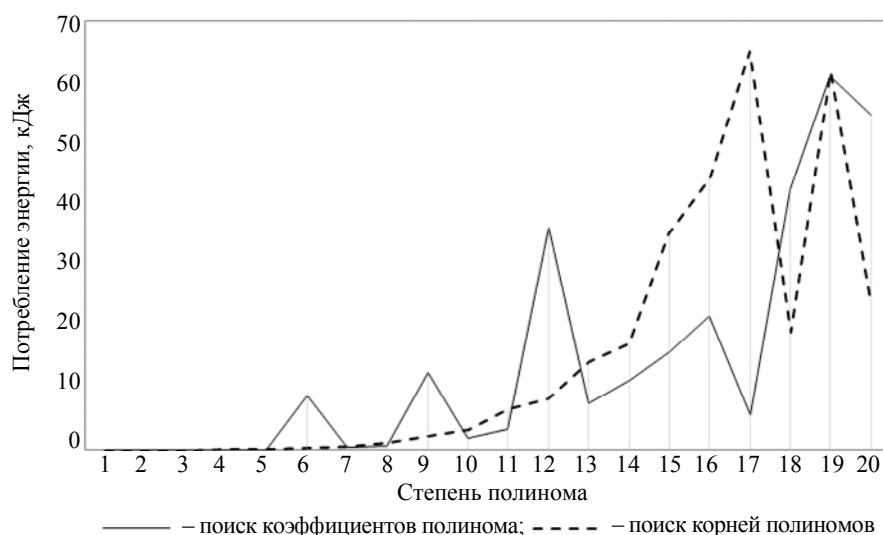


Рис. 1. Значения энергий, используемые для нахождения многочленов заданных степеней и с заданной точностью

Fig. 1. Energy values used to find polynomials of specified degrees and with specified accuracy

приводит к увеличению затрат времени. Видно, что разница в затратах времени между первыми и последними степенями полинома исчисляется порядками величин. Время выполнения на обоих этапах увеличивается с ростом степени полинома. Это соответствует ожидаемому поведению, так как усложнение вычислений приводит к увеличению временных затрат.

Табл. 2. Время, затраченное на поиск многочленов заданных степеней и с заданной точностью
 Tab. 2. The time spent searching for polynomials of specified degrees and with specified accuracy

Степень полинома	Затраченное время, мс	
	Поиск коэффициентов полиномов	Поиск корней полиномов
1	102	34
2	188	184
3	585	793
4	1163	1782
5	2663	3557
6	5056	7067
7	9089	17 343
8	17 001	28 767
9	30 941	59 522
10	47 737	96 448
11	84 265	173 283
12	128 323	251 052
13	198 769	423 983
14	277 733	590 276
15	432 809	926 130
16	570 313	1 232 846
17	843 362	1 885 308
18	1 101 890	2 432 822
19	1 572 285	3 537 045
20	1 991 022	4 466 333

По сравнению с энергией, этап «Поиск корней полиномов» также занимает значительно больше времени, чем этап «Поиск коэффициентов полиномов», что подтверждает необходимость оптимизации именно этого этапа для улучшения общего времени выполнения алгоритма.

Аномалии в энергопотреблении и времени выполнения на некоторых степенях полинома указывают на возможность оптимизации алгоритма. Например, исследование причин высокого потребления энергии и времени на степенях 6 и 12 может привести к значительному улучшению производительности алгоритма в целом.

Выводы и заключение. В данной статье был представлен метод автоматизированной параллельной минимаксной аппроксимации функций полиномами, определенными над множеством рациональных чисел с плавающей запятой. Этот подход направлен на решение проблемы ограниченной применимости традиционных методов минимаксной аппроксимации в реальных численных расчетах, где гарантии точности часто формулируются для всей области действительных чисел, что делает их не всегда применимыми для вычислений с конечной точностью.

Основное новшество работы заключается в разработке и экспериментальной реализации алгоритма Ремеза на основе многомерного метода Хаусхолдера, который позволяет выполнять минимаксные приближения функций с гарантированной точностью в конечной арифметике. Для повышения производительности и энергоэффек-

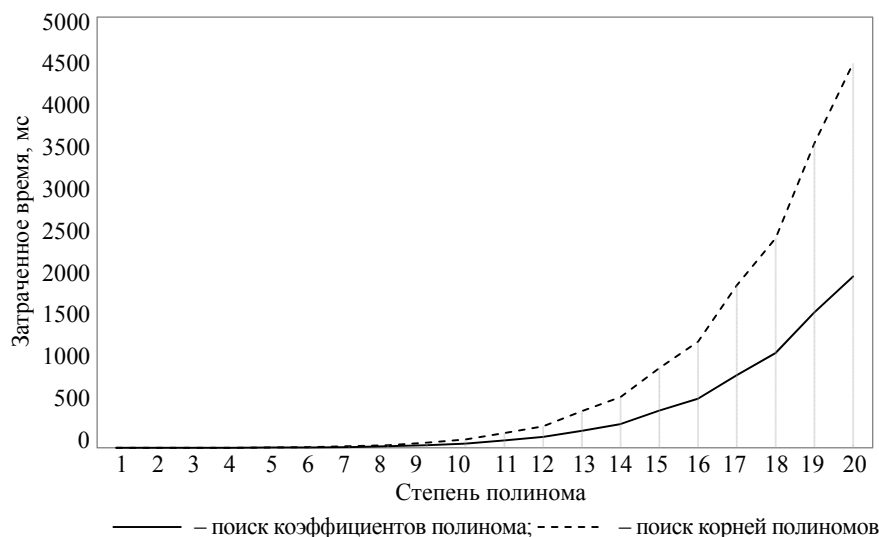


Рис. 2. Время, затраченное на поиск многочленов заданных степеней и с заданной точностью

Fig. 2. The time spent searching for polynomials of specified degrees and with specified accuracy

тивности были использованы параллельные и векторные алгоритмы сложения длинных целых чисел с опережающим переносом, что позволило достичь значительного ускорения вычислений по сравнению с существующими библиотеками.

Экспериментальные результаты показали, что предложенные алгоритмы обеспечивают высокую эффективность, это открывает новые возможности для применения минимаксных приближений в задачах численного анализа, где требуется высокая точность и производительность. В частности, исследование продемонстрировало, что этап поиска корней полиномов потребляет больше энергии и времени по сравнению с поиском ко-

эффициентов полиномов, что подчеркивает необходимость дальнейшей оптимизации этого этапа.

Таким образом, представленный подход решает важную проблему обеспечения гарантий точности в минимаксных аппроксимациях при конечной арифметике и предоставляет основу для дальнейших исследований и разработок в области численного анализа и вычислительной математики. Эти результаты могут быть использованы для улучшения точности и производительности таких различных приложений, требующих высококачественных приближений функций, как моделирование физических процессов, обработка сигналов и другие.

Список литературы

1. Willem M. Minimax theorems. Progress in nonlinear differential equations and their applications (PNLDE Vol. 24). Boston: Birkhäuser, 1996. P. 165–168.
2. Harrison J. A machine-checked theory of floating-point arithmetic // Proc. of the 1999 Intern. Conf. on Theorem Proving in Higher Order Logics (TPHOLs). LNCS. 1999. P. 113–130. doi: 10.1007/3-540-48256-3_9.
3. Hamming R. W. Numerical methods for scientists and engineers. New York: Dover, 1986. 714 p.
4. Chun-Hua G. On Newton's method and Halley's method for the principal root of a matrix // Regina, Linear Algebra and its Appl. 2009. Vol. 432, no. 8. P. 1905–1922.
5. Чусов А. А., Ефимова Ю. И. Анализ точности и формальное обоснование численного решения нелинейных уравнений методами Ньютона и Хэйли при использовании арифметики с плавающей запятой // Изв. СПбГЭТУ «ЛЭТИ». 2022. Т. 15, № 10. С. 35–44. doi: 10.32603/2071-8985-2022-15-10-35-44.

6. Minimax approximation of representation coefficients from generalized samples / T. G. Dvorkind, H. Kirshner, M. Porat, Y. Eldar // IEEE Transactions on Signal Proc. 2007. Vol. 55(9). P. 4430–4443.
7. Чусов А. А., Шкляр А. А., Ефимова Ю. И. Об энергоэффективности и оперативности полиномиального и табличного методов приближения математических функций в системах с плавающей точкой // Вестн. Томского гос. ун-та. Управление, вычислительная техника и информатика. 2024. № 68. С. 112–122. doi: 10.17223/19988605/68/12.
8. Goldberg D. What every computer scientist should know about floating-point arithmetic // ACM Comp. Surveys. 1991. Vol. 23, no. 1. P. 1–44.
9. De Groot E. D., Dr. Sterk A. E., Dr. Trentelman H. L. Finding best minimax approximations with the Remez algorithm: Bachelor's Project Mathematics. University of Groningen, Groningen, 2017. 49 p. URL: <https://fse.student>

theses.ub.rug.nl/15985/1/BSc_Math_2017_deGroot_ED.pdf
(дата обращения: 13.03.2025).

10. Hafiz M. A., Bahgat M. Modified of Householder iterative method for solving nonlinear systems // J. Math. Comp. Sci. 2012. Vol. 2, no. 5. P. 1200–1208.

Информация об авторах

Чусов Андрей Александрович – канд. техн. наук, доцент департамента электроники, телекоммуникации и приборостроения Политехнического института, Дальневосточный федеральный университет, г. Владивосток, о. Русский, п. Аякс, 10, Приморский край, 690922, Россия.

E-mail: chusov.aa@dvfu.ru

<https://orcid.org/0000-0002-7931-5368>

Ефимова Юлия Игоревна – магистрант департамента электроники, телекоммуникации и приборостроения Политехнического института, Дальневосточный федеральный университет, г. Владивосток, о. Русский, п. Аякс, 10, Приморский край, 690922, Россия.

E-mail: efimova.iui@dvfu.ru

<https://orcid.org/0000-0002-7694-0694>

References

1. Willem M. Minimax theorems. Progress in nonlinear differential equations and their applications (PNLDE Vol. 24). Boston: Birkhäuser, 1996. P. 165–168.

2. Harrison J. A machine-checked theory of floating-point arithmetic // Proc. of the 1999 Intern. Conf. on Theorem Proving in Higher Order Logics (TPHOLs). LNCS. 1999. P. 113–130. doi: 10.1007/3-540-48256-3_9.

3. Hamming R. W. Numerical methods for scientists and engineers. New York: Dover, 1986. 714 p.

4. Chun-Hua G. On Newton's method and Halley's method for the principal root of a matrix // Regina, Linear Algebra and its Appl. 2009. Vol. 432, no. 8. P. 1905–1922.

5. Chusov A. A., Efimova Ju. I. Analiz tochnosti i formal'noe obosnovanie chislennogo reshenija nelineynyh uravnenij metodami N'jutona i Hjejli pri ispol'zovanii arifmetiki s plavajushhej zapjatoj // Izv. SPbGJeTU «LJeTI». 2022. T. 15, № 10. S. 35–44. doi: 10.32603/2071-8985-2022-15-10-35-44. (In Russ.).

6. Minimax approximation of representation coefficients from generalized samples / T. G. Dvorkind, H. Kir-

shner, M. Porat, Y. Eldar // IEEE Transactions on Signal Proc. 2007. Vol. 55(9). P. 4430–4443.

7. Chusov A. A., Shkljar A. A., Efimova Ju. I. Ob jenergojektivnosti i operativnosti polinomial'nogo i tablichnogo metodov priblizhenija matematicheskikh funkcij v sistemah s plavajushhej tochkoj // Vestn. Tomskogo gos. un-ta. Upravlenie, vychislitel'naja tehnika i informatika. 2024. № 68. S. 112–122. doi: 10.17223/19988605/68/12. (In Russ.).

8. Goldberg D. What every computer scientist should know about floating-point arithmetic // ACM Comp. Surveys. 1991. Vol. 23, no. 1. P. 1–44.

9. E.D. de Groot, Dr. A.E. Sterk, Dr. H.L. Trentelman. Finding best minimax approximations with the Remez algorithm: Bachelor's Project Mathematics. University of Groningen, Groningen, 2017. 49 p. URL: https://fse.studenttheses.ub.rug.nl/15985/1/BSc_Math_2017_deGroot_ED.pdf (data obraschenija: 13.03.2025).

10. Hafiz M. A., Bahgat M. Modified of Householder iterative method for solving nonlinear systems // J. Math. Comp. Sci. 2012. Vol. 2, no. 5. P. 1200–1208.

Information about the authors

Andrey A. Chusov – Cand. Sci. (Eng.), Associate Professor, Department of Electronics, Telecommunications and Instrumentation of the Polytechnic Institute, Far Eastern Federal University, Island Russkiy, Ajax, 10, Primorsky Krai, 690922, Russia.

E-mail: chusov.aa@dvfu.ru

<https://orcid.org/0000-0002-7931-5368>

Yulia I. Efimova – master student, Department of Electronics, Telecommunications and Instrumentation of the Polytechnic Institute, Far Eastern Federal University, Island Russkiy, Ajax, 10, Primorsky Krai, 690922, Russia.

E-mail: efimova.iui@dvfu.ru

<https://orcid.org/0000-0002-7694-0694>

Статья поступила в редакцию 11.02.2025; принята к публикации после рецензирования 14.07.2025; опубликована онлайн 30.10.2025.

Submitted 11.02.2025; accepted 14.07.2025; published online 30.10.2025.