

УДК 519.651

О ПРИМЕНЕНИИ ФУНКЦИЙ ГАУССА В СОЧЕТАНИИ С ТЕОРЕМОЙ КОЛМОГОРОВА ДЛЯ АППРОКСИМАЦИИ ФУНКЦИЙ МНОГИХ ПЕРЕМЕННЫХ

© 2020 г. А. В. Чернов^{1,2}

¹ 603950 Нижний Новгород, пр-т Гагарина, 23, Нижегородский государственный университет
им. Н.И. Лобачевского, Россия

² 603950 Нижний Новгород, ул. Минина, 24, Нижегородский государственный технический университет
им. Р.Е. Алексеева, Россия

e-mail: chavnn@mail.ru

Поступила в редакцию 04.02.2019 г.
Переработанный вариант 11.11.2019 г.
Принята к публикации 14.01.2020 г.

Исследуется специальный класс аппроксимаций непрерывных функций многих переменных на единичном координатном кубе. Основу построения этого класса составляет теорема Колмогорова о представлении функций указанного типа в виде конечной суперпозиции непрерывных функций одного переменного и их аппроксимация линейными комбинациями квадратичных экспонент (функций Гаусса). Эффективность такого представления основана на ранее доказанном автором утверждении о возможности сколь угодно точной аппроксимации на любом фиксированном конечном отрезке материнского вейвлета “мексиканская шляпа” линейной комбинацией двух функций Гаусса. Доказывается всюду плотность изучаемого класса аппроксимаций в классе непрерывных функций многих переменных на координатном кубе. Приводятся результаты численных экспериментов, подтверждающие эффективность аппроксимаций изучаемого класса на примере непрерывных функций двух переменных. Библ. 25. Фиг. 11. Табл. 3.

Ключевые слова: аппроксимация непрерывных функций многих переменных, функции Гаусса, квадратичные экспоненты, теорема Колмогорова.

DOI: 10.31857/S0044466920050075

ВВЕДЕНИЕ

Проблема аппроксимации функций многих переменных имеет важное теоретическое и прикладное значение. Автора эта проблема интересует, главным образом, с точки зрения отыскания эффективных способов понижения размерности аппроксимируемых задач математического программирования, возникающих в результате параметрической дискретизации управляющей функции при численном решении распределенных задач оптимального управления.

При дискретизации задач оптимального управления традиционно используется кусочно-постоянная или кусочно-линейная аппроксимация управляющей функции, что даже в случае подвижной (управляемой) сетки приводит к большой размерности аппроксимирующей задачи математического программирования. В частности, в рамках техники параметризации управления (см. [1]–[6], а также указанную там библиографию) за счет предположения об однозначной разрешимости управляемой системы для каждого допустимого управления функционалы задачи сводятся к функциям конечного числа переменных – параметров аппроксимации управляющей функции, а исходная бесконечномерная задача оптимизации – к конечномерной, которая в этом случае и понимается как аппроксимирующая задача. При этом способы аппроксимации управления могут быть, вообще говоря, различными, но естественно стремиться к использованию наиболее экономных и эффективных способов. Иначе говоря, актуальным является поиск такого класса \mathcal{C} функций, аппроксимирующих неизвестное управление, который, с одной стороны, обеспечивал бы высококачественную аппроксимацию для управляющих функций из достаточно обширного множества при использовании не слишком большого количества параметров аппроксимации, а с другой, позволял бы программировать те или иные полезные свойства

управляющих функций (например, отсутствие быстрых осцилляций, разрушающих соответствующее техническое устройство, устойчивость приближенного решения оптимизационной задачи к искажению входных параметров – помехоустойчивость и т.п.). При использовании аппроксимации управления функциями из класса \mathcal{E} значения искомого оптимального управления в контрольных точках нам, разумеется, не известны (и в этом состоит принципиальное отличие от обычной задачи аппроксимации – интерполяции). В этом случае неизвестные параметры аппроксимации определяются путем минимизации значения целевого функционала, вычисляемого на функциях из класса \mathcal{E} , если они являются допустимыми по смыслу задачи, либо на допустимых суперпозициях, содержащих функции класса \mathcal{E} (например, в рамках метода синус-параметризации, используемого для учета ограничений на значения управления в виде условия принадлежности заданному параллелепипеду в пространстве \mathbb{R}^s). Указанное выше принципиальное отличие следует учитывать при сравнении различных подходов к аппроксимации. Действительно, в рамках обычной задачи аппроксимации (при заданном количестве параметров) на первый план выходят вопросы, связанные с минимизацией расхода времени и (прочих ресурсов) при отыскании неизвестных параметров аппроксимации. В рамках же численного решения задач оптимального управления главное – это повышение точности аппроксимации в отношении произвольной функции (неизвестного управления) при заданном количестве ее параметров, уменьшение количества параметров аппроксимации при сохранении приемлемой точности (по управлению или хотя бы по функционалу) и т.п. При этом точность аппроксимации понимается в абсолютном смысле (на всем множестве $\Pi \subset \mathbb{R}^n$ независимых переменных, либо на достаточно плотной сетке), а не в смысле (лишь) заданного количества узлов, соответствующего количеству параметров.

Указанная выше проблема аппроксимации тесным образом примыкает к проблеме точного представления функций многих переменных. А она, в свою очередь, имеет непосредственное отношение к 13-й проблеме Гильберта: возможно ли точное представление функции многих переменных в виде суперпозиции функций меньшего числа переменных? Для непрерывных функций, заданных на координатном кубе в \mathbb{R}^n , полное решение 13-й проблемы Гильберта было получено А.Н. Колмогоровым: оказалось, что любую такую функцию можно представить с помощью операций сложения, умножения и суперпозиции непрерывных функций одного переменного, см. [7].

Теорема А.Н. Колмогорова. *При любом целом $n \geq 2$ существуют такие определенные на единичном отрезке $E^1 = [0; 1]$ непрерывные действительные функции $\psi^{p,q}(x)$, что каждая определенная на n -мерном единичном кубе E^n непрерывная действительная функция $f(x_1, \dots, x_n)$ представима в виде*

$$f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} \chi_q \left[\sum_{p=1}^n \psi^{p,q}(x_p) \right],$$

где функции $\chi_q(y)$ действительны и непрерывны.

Впоследствии это утверждение неоднократно уточнялось и развивалось, см., например, [8]–[10] (там же см. дальнейшую библиографию).

Между тем необходимо отметить, что внутренние и внешние функции в представлении А.Н. Колмогорова не годятся для практического использования по следующим причинам: 1) алгоритм их построения очень сложен; 2) будучи непрерывными, они ни в одной точке не дифференцируемы; 3) они не допускают построения графика. Однако для практических целей точное представление не требуется, достаточно лишь указать способ их эффективной аппроксимации функциями некоторой простой, унифицированной структуры. Довольно естественно здесь возникает мысль использовать теорему Вейерштрасса об аппроксимации непрерывной функции многочленами (или одно из ее обобщений), см., например, [10]. Однако представление многочленами не всегда удобно ввиду того, что может понадобиться учет высоких степеней (что приводит к неустойчивости аппроксимации) и того, что отдельные слагаемые в таком представлении неравноценны. Поэтому актуальным является поиск иных способов.

Задача аппроксимации функций многих переменных часто возникает из потребностей компьютерной томографии, теории нейронных сетей и теории обучения и поэтому вообще достаточно актуальна. Подходы к аппроксимации функций многих переменных, идейно близкие теореме Колмогорова, условно можно разделить на следующие категории.

1. *Нейросети.* В [11] в качестве непосредственного следствия теоремы [8], уточняющей теореме Колмогорова, было доказано следующее. Пусть $\varphi : E^n \rightarrow \mathbb{R}^m$ — непрерывная вектор-функция, $\varphi(x) = y$. Тогда существуют число $\lambda \in \mathbb{R}$ и непрерывная монотонно возрастающая функция ψ (не зависящие от φ , но зависящие от n) и рациональное число $\varepsilon \in (0; \delta]$ при выбранном произвольно $\delta > 0$ такие, что

$$y_i = \sum_{k=1}^{2n+1} g_i(z_k), \quad i = \overline{1, m}; \quad z_k = \sum_{j=1}^n \lambda^k \psi(x_j + \varepsilon k) + k, \quad k = \overline{1, 2n+1}$$

($y_i, i = \overline{1, m}$, — процессоры верхнего слоя нейросети; $z_k, k = \overline{1, 2n+1}$, — процессоры среднего слоя нейросети; нижний слой — это процессоры $x_i, i = \overline{1, n}$, обработки входных сигналов), где $g_i, i = \overline{1, m}$, — непрерывные функции, зависящие от φ и ε . Как указывалось в [11], нет никакого конструктивного метода построения функций g_i . Поэтому предлагалось разработать адаптивный механизм, который позволял бы по заданным входам x и выходам y подбирать функции g_i путем процесса самоорганизации (обучения). В этом и состоит эвристическая идея метода нейросетей. В [12] был предложен более конструктивный подход. А именно, было доказано, что в интегральной метрике с весом Чебышёва—Эрмита возможно сколь угодно точное приближение функции n переменных достаточно общего вида двухслойной нейронной сетью. Математически, речь идет о приближениях вида

$$\sum_{j=1}^q \beta_j \psi(A_j(x)), \quad q \in \mathbb{N}, \quad \beta_j \in \mathbb{R}; \quad x \in \mathbb{R}^n,$$

где $A_j(x)$ — аффинные функции, $\psi(t)$ — заранее заданная сигмоидная функция. Неубывающая функция $\psi : \mathbb{R} \rightarrow [0; 1]$ называется сигмоидной, если

$$\lim_{t \rightarrow -\infty} \psi(t) = 0, \quad \lim_{t \rightarrow +\infty} \psi(t) = 1.$$

2. *Многочлены.* Как уже было сказано, в [10] предлагалось использовать многочлены для аппроксимации функций представления Колмогорова. Вопросы точного представления многочленов нескольких переменных суперпозициями многочленов одного переменного рассматривались, например, в [13].

3. *Ридж-функции.* Ридж-функцией называется функция вида $g(a \cdot x)$, где g — функция одной переменной, $a = (a_1, \dots, a_n) \in \mathbb{R}^n$ — заданный вектор, называемый направлением ридж-функции, $x \in \mathbb{R}^n$, $a \cdot x$ — скалярное произведение. Разумеется, далеко не всякую функцию можно представить суммой ридж-функций с различными направлениями. Соответствующие условия представимости можно найти, например, в обзоре [14].

Стоит отметить, что все указанные в нашем кратком обзоре работы (там же см. дальнейшую библиографию), имеют чисто теоретический характер — результатов численных экспериментов в них нет. Отдельно укажем работу [15], где приводятся результаты численных экспериментов по восстановлению графика функции двух переменных с помощью обучаемой нейросети в виде линейного персептрона и упоминается о погрешности аппроксимации в отдельной точке не более 20%.

В [16] исследовались возможности аппроксимации функций одного переменного на конечном отрезке линейными комбинациями квадратичных экспонент (функций Гаусса) с варьируемыми параметрами; было доказано несколько утверждений о дискретно точном характере изучаемых аппроксимаций и приведены результаты численных экспериментов, подтверждающих справедливость этих утверждений и демонстрирующих высокое качество такого способа аппроксимации для гладких функций и достаточно хорошее — для непрерывных и кусочно непрерывных функций. Говоря более конкретно, для одномерного случая в качестве класса \mathcal{E} в работе [16] было предложено использовать класс

$$\begin{aligned} \mathcal{E}_\varepsilon = \{ \Phi = \Phi_v^\varepsilon[\alpha, \beta, \gamma] \in C^\infty[a; b] \mid (\alpha, \beta, \gamma) \in \mathbb{R}^{3v}, v \in \mathbb{N} \}, \\ \Phi_v^\varepsilon[\alpha, \beta, \gamma](x) = \sum_{j=1}^v \alpha_j \varphi_j[\beta_j, \gamma_j](x), \quad \varphi_j[\beta_j, \gamma_j](x) = \exp \left[-\frac{(x - \beta_j)^2}{\varepsilon + \gamma_j^2} \right], \end{aligned} \quad (1)$$

где $\varepsilon > 0$ – некоторое достаточно малое число (нужное для того только, чтобы избежать нуля в знаменателе). В [17] было доказано утверждение о возможности сколь угодно точной аппроксимации на любом конечном отрезке материнского вейвлета “мексиканская шляпа” линейной комбинацией двух квадратичных экспонент. Это утверждение можно рассматривать как строгое теоретическое обоснование эффективности аппроксимации функций одного переменного на любом конечном отрезке линейными комбинациями функций Гаусса, поскольку для вейвлета указанного типа такая эффективность известна. Кроме того, в [17] описана методика применения аппроксимаций управляющих функций линейными комбинациями квадратичных экспонент для эффективной дискретизации сосредоточенных задач оптимального управления; приведены результаты численных экспериментов, подтверждающие эффективность этой методики при численном решении задач оптимального управления.

В данной статье мы исследуем специальный класс аппроксимаций непрерывных функций многих переменных на единичном координатном кубе, основанный на аппроксимации внешних и внутренних функций в представлении, родственном представлению А.Н. Колмогорова, линейными комбинациями квадратичных экспонент. Эффективность такого способа базируется на ранее доказанном автором утверждении о возможности сколь угодно точной аппроксимации на любом фиксированном конечном отрезке материнского вейвлета “мексиканская шляпа” линейной комбинацией двух функций Гаусса. Доказывается всюду плотность изучаемого класса аппроксимаций в классе непрерывных функций многих переменных на координатном кубе. Отметим, что исходной целью автора было построение такого способа аппроксимации функций многих переменных, который при использовании сравнительно малого количества параметров позволял бы получать приближение хорошего качества для функций, устроенных “не слишком плохо” (то есть без большого количества, но в том числе и при наличии умеренного количества резких осцилляций, импульсоподобных всплесков и прочих типов “хаотизации” графика). Приводятся результаты численных экспериментов, подтверждающие эффективность аппроксимаций изучаемого класса в указанном смысле на примере непрерывных функций двух переменных.

Сделаем краткий обзор о применении квадратичных экспонент для аппроксимации функций одного переменного (в частности, непрерывных, гладких, многочленов и т.д.). Интерполяция функциями Гаусса (их еще иногда называют ядрами Гаусса – Gaussian kernels) непрерывных функций рассматривалась, главным образом, на дискретной сетке, как правило, с равномерным шагом на всей числовой оси, либо на конечном множестве отрезка числовой оси, см., например, [18]. Соответствующие алгоритмы аппроксимации иногда называют сеточными (grid). См., например, [19]. Исследовались также и многомерные аналоги, см., например, [20]. В основном, внимание исследователей концентрировалось на оценке погрешности аппроксимаций [21], [22]. Эффективность сеточных алгоритмов аппроксимации функциями Гаусса обосновывалась, например, в [23]. Помимо простой, равномерной сетки, используются также и более сложные конструкции типа многоуровневых сеток, см., например, [24]. Параметрами аппроксимации выступают обычно весовые коэффициенты функций Гаусса в соответствующих линейных комбинациях. Укажем также работу [20], где исследовался вопрос о наилучшем выборе параметров формы.

1. ОСНОВНАЯ ТЕОРЕМА

Далее мы используем модификацию теоремы А.Н. Колмогорова в стиле [8], [9]. А именно, результаты работы [9, разд. “Введение”, “Построение внутренних функций”], опирающейся в свою очередь на [8], можно суммировать в следующем утверждении.

Лемма 1.1. *При произвольно заданных числах – натуральном $n \geq 2$, алгебраическом $\lambda > 0$ степени $\deg \lambda \geq n$, натуральном $\gamma \geq 2(n+1)$ и $\sigma = \frac{1}{\gamma-1}$ существует действительная непрерывная монотонно возрастающая функция одного переменного $\Psi(y)$, $y \geq 0$, такая, что каждая определенная на n -мерном единичном кубе E^n непрерывная действительная функция $f(x_1, \dots, x_n)$ представима в виде*

$$f(x) = \sum_{q=0}^{2n} F_q[G_q(x)], \quad G_q(x) = \sum_{p=1}^n \lambda^p \Psi(x_p + \sigma q), \quad x = (x_1, \dots, x_n) \in E^n,$$

где семейство действительных непрерывных функций одного переменного $\{F_q(y), q = \overline{0, 2n}\}$ определяется по функции f .

Отметим, что способ построения функции Ψ , описанный в [9], сложный и многоэтапный, причем сама функция представляется в виде бесконечного ряда. Поэтому его практическое использование затруднительно. Кроме того, функция Ψ определяется неоднозначно.

Определим класс линейных комбинаций функций Гаусса:

$$\Gamma = \left\{ \Phi = \Phi_v^\varepsilon[\alpha, \beta, \gamma] \in C^\infty[a; b] \mid (\alpha, \beta, \gamma) \in \mathbb{R}^{3v}, v \in \mathbb{N} \right\},$$

где функции $\Phi_v^\varepsilon[\alpha, \beta, \gamma]$ определяются формулой (1) при некотором фиксированном $\varepsilon > 0$, а отрезок $[a; b]$ выбирается достаточно большим для того, чтобы выполнялись условия

$$t + \sigma q \in [a; b], \quad \sum_{p=1}^n \lambda^p \Psi(x_p + \sigma q) \in [a; b] \quad \forall t \in [0; 1], \quad x \in E^n, \quad q = \overline{0, 2n},$$

где числа n, σ, λ и функцию Ψ считаем фиксированными и выбранными в соответствии с тем, как указано в лемме 1.1.

Определим класс аппроксимаций \mathcal{E} как множество

$$\left\{ \sum_{q=0}^{2n} \hat{F}_q[\hat{G}_q(x)], \hat{G}_q(x) = \sum_{p=1}^n \lambda^p \hat{\Psi}(x_p + \sigma q), \hat{\Psi}, \hat{F}_q \in \Gamma, q = \overline{0, 2n} : \hat{G}_q(x) \in [a; b] \quad \forall x \in E^n \right\}.$$

Теорема 1.1. *Класс Γ является всюду плотным в $C[a; b]$.*

Доказательство см. в разд. 2. В качестве непосредственного следствия леммы 1.1 и теоремы 1.1 получаем, что справедлива

Теорема 1.2. *Класс \mathcal{E} является всюду плотным в $C(E^n)$.*

2. ВСПОМОГАТЕЛЬНЫЕ УТВЕРЖДЕНИЯ

Следующее утверждение хорошо известно как теорема Вейерштрасса.

Лемма 2.1. *Для любого числа $\varepsilon > 0$ и любой функции $f(x_1, \dots, x_n)$, непрерывной на компакте $Q \subset \mathbb{R}^n$, существует многочлен $P(x_1, \dots, x_n)$ такой, что $\max_Q |f(x_1, \dots, x_n) - P(x_1, \dots, x_n)| < \varepsilon$.*

Лемма 2.2. *Для любых $[a; b] \subset \mathbb{R}$ и $\varepsilon > 0$, а также для любых $\beta \in \mathbb{R}$ и $\gamma > 0$ таких, что*

$$\frac{|\beta| + \max(|a|, |b|)}{\gamma} < \sqrt{\min(\varepsilon, 1)}, \quad (2)$$

справедлива оценка $\max_{x \in [a; b]} \left| \exp \left[-\frac{(x - \beta)^2}{\gamma^2} \right] - 1 \right| < \varepsilon$.

Доказательство. Верно разложение в ряд Маклорена

$$e^{-t^2} - 1 = \sum_{n=1}^{\infty} \frac{(-1)^n t^{2n}}{n!} \quad \forall t \in \mathbb{R}.$$

В частности, для каждого фиксированного $t \in [-1; 1]$ справа стоит знакопередающийся ряд с модулем общего члена, строго убывающим до нуля, то есть ряд типа Лейбница. Как известно, модуль суммы такого ряда (так же, как и его остатка) оценивается модулем своего первого члена.

Таким образом, при $t \in [-1; 1]$ имеем $|e^{-t^2} - 1| \leq t^2$. Поэтому чтобы получить нужную нам оценку, достаточно, чтобы $\frac{(x - \beta)^2}{\gamma^2} < \min(\varepsilon, 1)$. Выполнение этого неравенства обеспечивается условием (2).

Лемма доказана.

Лемма 2.3. *Для любых $[a; b] \subset \mathbb{R}$, $a < b$, $C \neq 0$ и $\varepsilon > 0$ найдутся параметры $\alpha, \beta, d \in \mathbb{R}$ и $\gamma > 0$ такие, что*

$$Cx = \alpha \varphi_{\beta, \gamma}(x) + d + \alpha(x), \quad \varphi_{\beta, \gamma}(x) = \exp \left[-\frac{(x - \beta)^2}{\gamma^2} \right], \quad \max_{x \in [a; b]} |\alpha(x)| < \varepsilon.$$

В качестве $\beta \neq 0$ и $\gamma > 0$ можно взять любые числа, удовлетворяющие неравенствам

$$|\beta| > \frac{2|C| \max(|a|, |b|)(b-a)}{\varepsilon}, \quad \gamma > \frac{|\beta| + \max(|a|, |b|)}{\sqrt{\min(\varepsilon/|2|C|(b-a)), 1}}. \quad (3)$$

Доказательство. Предварительно, считая параметры $\beta \in \mathbb{R}$, $\beta \neq 0$, $\gamma > 0$ произвольными, положим $\alpha = \frac{C\gamma^2}{2\beta}$ и оценим разность

$$\begin{aligned} \omega_1(x) &= \left| \alpha \Phi'_{\beta, \gamma}(x) - C \right| = |\alpha| \left| -2 \exp \left[-\frac{(x-\beta)^2}{\gamma^2} \right] \left(\frac{x-\beta}{\gamma^2} \right) - \frac{C}{\alpha} \right| \leq \\ &\leq 2|\alpha| \left| \frac{x}{\gamma^2} \right| + |C| |\Phi_{\beta, \gamma}(x) - 1| = |C| \left(\frac{|x|}{|\beta|} + |\Phi_{\beta, \gamma}(x) - 1| \right). \end{aligned}$$

Обозначим $\varepsilon_1 = \frac{\varepsilon}{b-a}$ и выберем число $\beta \neq 0$ из условия

$$\frac{|C| \max(|a|, |b|)}{|\beta|} < \frac{\varepsilon_1}{2}.$$

После этого выберем число $\gamma > 0$ из условия (2) при замене ε на $\frac{\varepsilon_1}{2|C|}$. Указанный выбор чисел β и γ означает выполнение неравенств (3). Согласно проведенным выкладкам и лемме 2.2, получаем $\omega_1(x) < \varepsilon_1$ для всех $x \in [a; b]$. Рассмотрим интеграл

$$\int_a^x \{ \alpha \Phi_{\beta, \gamma}(t) - Ct \}' dt = \alpha \Phi_{\beta, \gamma}(x) + d - Cx \equiv -\omega(x), \quad x \in [a; b],$$

где $d = Ca - \alpha \Phi_{\beta, \gamma}(a)$. Таким образом, получаем

$$|\omega(x)| \leq \int_a^b \omega_1(t) dt < \varepsilon_1(b-a) = \varepsilon \quad \forall x \in [a; b].$$

Лемма доказана.

Лемма 2.4. Пусть $[a; b] \subset \mathbb{R}$ и $\sigma > 0$ произвольно фиксированы. Тогда любой полином степени $n \geq 0$ можно сколь угодно точно в метрике пространства $C[a; b]$ аппроксимировать некоторой функцией вида $\Phi_v^\sigma[\alpha, \beta, \gamma]$ при $v \leq n+1$.

Доказательство. Рассмотрим на отрезке $[a; b]$, $a < b$, произвольный полином степени n : $\mathcal{P}_n(x) = K_0 + K_1x + \dots + K_nx^n$. Очевидно, нам достаточно доказать, что данный полином можно сколь угодно точно в метрике пространства $C[a; b]$ аппроксимировать некоторой функцией вида $\Phi_v^0[\alpha, \beta, \gamma]$ при $v \leq n+1$, $\gamma_j^2 \geq \sigma$, $j = \overline{1, v}$. Если $n = 0$, достаточно просто сослаться на лемму 2.2. Будем считать, что $n \geq 1$, $C_m = \sqrt[m]{|K_m|}$, $C = 1 + \max_{m=1, n} C_m$, $s_m = \text{sign } K_m$, $m = \overline{1, n}$. Тогда

$$\mathcal{P}_n(x) = K_0 + \sum_{m=1}^n s_m (C_m x)^m.$$

Зафиксируем произвольно число $\varepsilon_1 \in (0; 1)$ и найдем числа $\beta \neq 0$ и $\gamma \geq \sqrt{n\sigma}$ из условий (3) при замене ε на ε_1 . Согласно лемме 2.3, найдутся числа $d_m, \alpha_m \in \mathbb{R}$, $m = \overline{1, n}$, такие, что

$$\mathcal{P}_n(x) = K_0 + \sum_{m=1}^n s_m (\alpha_m \Phi_{\beta, \gamma}(x) + d_m + \omega_m(x))^m, \quad |\omega_m(x)| < \varepsilon_1 \quad \forall x \in [a; b],$$

$m = \overline{1, n}$. Пользуясь формулой бинома Ньютона, получаем

$$(\alpha_m \Phi_{\beta, \gamma}(x) + d_m + \omega_m(x))^m = (\alpha_m \Phi_{\beta, \gamma}(x) + d_m)^m + \delta_m(x),$$

$$\delta_m(x) = \sum_{k=1}^m C_m^k (\alpha_m \Phi_{\beta, \gamma}(x) + d_m)^{m-k} \omega_m^k(x).$$

Заметим, что $|\omega_m(x)| < \varepsilon_1 < 1$, и таким образом,

$$\begin{aligned} |\delta_m(x)| &\leq \varepsilon_1 \sum_{k=1}^m C_m^k |C_m x - \omega_m(x)|^{m-k} \leq \varepsilon_1 \sum_{k=1}^m C_m^k (C|x| + 1)^{m-k} \leq \\ &\leq \varepsilon_1 C^m \sum_{k=1}^m C_m^k (|x| + 1)^{m-k} \leq \varepsilon_1 C^m (|x| + 2)^m \leq \varepsilon_1 C^n (|x| + 2)^n. \end{aligned}$$

Пользуясь теперь произволом в выборе ε_1 , потребуем, чтобы выполнялось условие

$$\varepsilon_1 C^n (\max(|a|, |b|) + 2)^n < \frac{\varepsilon}{2n}.$$

Тогда для величины $\delta(x) = \sum_{m=1}^n s_m \delta_m(x)$ получаем оценку

$$|\delta(x)| < \frac{\varepsilon}{2} \quad \forall x \in [a; b].$$

При этом по построению,

$$\mathcal{P}_n(x) = K_0 + \sum_{m=1}^n s_m (\alpha_m \Phi_{\beta, \gamma}(x) + d_m)^m + \delta(x), \quad x \in [a; b].$$

Еще раз применяя формулу бинома Ньютона, имеем

$$(\alpha_m \Phi_{\beta, \gamma}(x) + d_m)^m = \sum_{k=0}^m C_m^k \alpha_m^{m-k} d_m^k \Phi_{\beta, \gamma}^{m-k}(x).$$

Таким образом, получаем, по сути дела, линейную комбинацию степеней функции $\Phi_{\beta, \gamma}(x)$, максимально возможной из которых является n -я степень. Отсюда ясно, что представление полинома можно переписать в виде

$$\mathcal{P}_n(x) = K_0 + \sum_{j=0}^n f_j \Phi_{\beta, \gamma}^j(x) + \delta(x) = C_0 + \sum_{j=1}^n f_j \Phi_{\beta, \gamma}^j(x) + \delta(x), \quad C_0 = K_0 + f_0.$$

Заметим, что $\Phi_{\beta, \gamma}^j(x) = \Phi_{\beta, \gamma_j}(x)$, $\gamma_j = \frac{\gamma}{\sqrt{j}} \geq \sqrt{\sigma}$. Таким образом,

$$\mathcal{P}_n(x) = C_0 + \sum_{j=1}^n f_j \Phi_{\beta, \gamma_j}(x) + \delta(x).$$

Пользуясь, наконец, леммой 2.2, получаем представление

$$\mathcal{P}_n(x) = C_0 \Phi_{\beta_0, \gamma_0}(x) + \sum_{j=1}^n f_j \Phi_{\beta, \gamma_j}(x) + \omega(x),$$

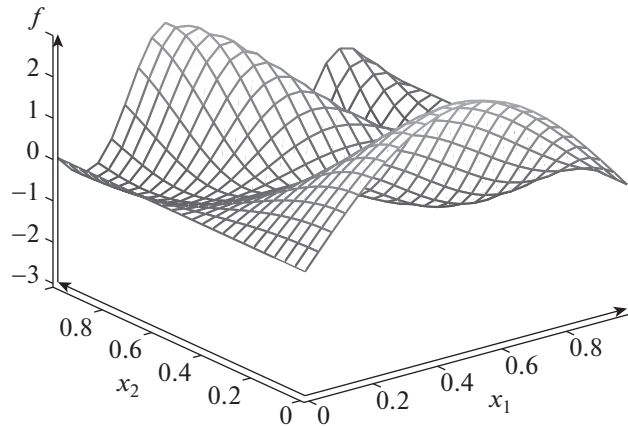
при некоторых $\beta_0 \in \mathbb{R}$, $\gamma_0 \geq \sqrt{\sigma}$, $|\omega(x)| < \varepsilon$ для всех $x \in [a; b]$. Лемма доказана.

Доказательство теоремы 1.1 следует непосредственно из лемм 2.1 и 2.4.

3. РЕЗУЛЬТАТЫ ЧИСЛЕННЫХ ЭКСПЕРИМЕНТОВ

При проведении численных экспериментов везде использовались значения параметров $\lambda = \sqrt[3]{2}$, $\gamma = 2(n+1)$, $\sigma = \frac{1}{\gamma-1}$, $n = 2$. Что касается параметров a и b , в тексте программ их задавать

не требовалось, поскольку функции класса $\Phi_v^\varepsilon[\alpha, \beta, \gamma]$ естественно определены на всей числовой оси, а не только на $[a; b]$. Формально можно считать, что $a < 0$, $b > 0$, причем $|a|, |b|$ достаточно ве-



Фиг. 1. Тест № 1, аппроксимируемая функция f .

лики. Фиксированный (достаточно большой) отрезок $[a; b]$ нужен был только для того, чтобы доказать плотность вложения класса аппроксимаций в класс непрерывных функций.

Тест № 1. Рассмотрим на квадрате E^2 функцию двух переменных, см. фиг. 1, $f(x_1, x_2) = 3 \sin(\pi x_1) \cos(\pi x_2) \cos(\pi^2 x_1 x_2)$. Эта функция выбиралась из следующего соображения: должно быть несколько локальных “всплесков” и “впадин”, чтобы график не имел слишком простой для аппроксимации формы. Выберем разбиение сторон квадрата

$$0 = \xi_1 < \xi_2 < \dots < \xi_\ell = 1, \quad 0 = \eta_1 < \eta_2 < \dots < \eta_m = 1,$$

с равномерным шагом. Будем брать $\ell = m = 25$. Таким образом, получаем сетку из 625 контрольных точек (ξ_i, η_j) , $i = \overline{1, \ell}$, $j = \overline{1, m}$. В этих точках мы будем сравнивать значение аппроксимации и аппроксимируемой функции. При заданном $v \in \mathbb{N}$ определим подклассы

$$\Gamma_v = \{ \Phi = \Phi_v^\varepsilon[\alpha, \beta, \gamma] \in C^\infty[a; b] \mid (\alpha, \beta, \gamma) \in \mathbb{R}^{3v} \},$$

$$\mathcal{E}_v = \left\{ \sum_{q=0}^{2n} \hat{F}_q[\hat{G}_q(x)], \hat{G}_q(x) = \sum_{p=1}^n \lambda^p \hat{\Psi}(x_p + \sigma q), \hat{\Psi}, \hat{F}_q \in \Gamma_v, q = \overline{0, 2n} \right\}.$$

Отметим, что каждая функция класса \mathcal{E}_v определяется набором параметров $v \in \mathbb{R}^\mu$, $\mu = 3v2(n + 1)$, состоящим из троек $(\alpha, \beta, \gamma) \in \mathbb{R}^{3v}$, относящихся к функциям \hat{F}_q , $q = \overline{0, 2n}$, и функции $\hat{\Psi}$ класса Γ_v . Соответствующий класс векторов v будем обозначать $V_v \subset \mathbb{R}^\mu$. В этом смысле будем писать $Q = Q[v]$, $v \in V_v$, для $Q \in \mathcal{E}_v$.

Поставим задачу найти по возможности наилучшую аппроксимацию функции $f(x)$ функцией класса \mathcal{E}_v . Для этого произведем минимизацию квадратичной невязки – суммы квадратов отклонений аппроксимации от аппроксимируемой функции в контрольных точках:

$$\mathcal{H}[v] = \sum_{i=1}^{\ell} \sum_{j=1}^m \{ f(\xi_i, \eta_j) - Q[v](\xi_i, \eta_j) \}^2 \rightarrow \min_{v \in V_v}. \tag{4}$$

Известно, что задача вида (4) является, вообще говоря, некорректной. Поэтому для обеспечения устойчивости процесса численного решения имеет смысл рассматривать ее регуляризованный аналог:

$$\mathcal{H}_\rho[v] = \mathcal{H}[v] + \rho |v|^2 \rightarrow \min_{v \in V_v}, \tag{5}$$

с малым параметром (регуляризации) $\rho > 0$. Мы брали $\rho = 10^{-15}$. Задачу (5) мы решали методами DFP (Дэвидона–Флетчера–Пауэлла) и BFGS (Бройдена–Флетчера–Шенно). Для одномерного поиска использовались методы квадратичной интерполяции с регулировкой шага, Брента (соче-

тание методов золотого сечения и квадратичной интерполяции) и More–Thuente [25]. Другие методы (в том числе квазиньютоновские – BFGS и Бройдена) не давали удовлетворительных результатов. Чтобы убедиться в улучшении качества аппроксимации при увеличении параметра ν , мы рассматривали два случая: $\nu = 5$ (то есть $\mu = 90$) и $\nu = 10$ (то есть $\mu = 180$). Мы брали $\epsilon = 10^{-25}$.

В качестве начального приближения выбирался вектор вида $\kappa\{-1, 2, -3, \dots, (-1)^\mu\mu\}$, где μ – размерность вектора, κ – малая величина (мы брали 10^{-5} и 10^{-4}). Смысл такого выбора в том, чтобы (с учетом равноправия отдельных квадратичных экспонент) повысить вариативность компонент начального приближения (за счет варьирования их знака и абсолютной величины) и при этом ограничить воздействие неблагоприятных вариантов (за счет малости общего множителя κ : поскольку все параметры начального приближения оказываются достаточно малыми, то неудачные значения параметров легко обнуляются в ходе процесса оптимизации, причем обнуление параметров α означает удаление неудачных слагаемых из линейной комбинации функций Гаусса, тогда как удачные значения получают тенденцию к изменению в направлении продвижения к искомому значениям). При использовании методов первого порядка для решения задачи минимизации квадратичной невязки требуется знать формулы частных производных по параметрам

$$w = (w_0, \dots, w_{2n}, w_{2n+1}), \quad w_i = (\alpha^i, \beta^i, \gamma^i) \in \mathbb{R}^{3\nu}, \quad i = \overline{0, 2n+1},$$

функции вида

$$g(x; w) = \sum_{q=0}^{2n} \hat{F}_q(\hat{G}_q(x; w_{2n+1}); w_q), \quad x \in E^n,$$

где

$$\hat{F}_q(t; w_q) = \sum_{j=1}^{\nu} \alpha_j^q \exp \left\{ - \left(\frac{t - \beta_j^q}{\gamma_j^q} \right)^2 \right\}, \quad q = \overline{0, 2n+1},$$

$$\hat{G}_q(x; w_{2n+1}) = \sum_{p=1}^n \lambda^p \hat{F}_{2n+1}(x_p + \sigma q; w_{2n+1}).$$

Отсюда понятно, что

$$\frac{\partial g}{\partial w_{qj}} = \frac{\partial \hat{F}_q}{\partial w_{qj}}, \quad q = \overline{0, 2n}, \quad \frac{\partial g}{\partial w_{2n+1,j}} = \sum_{q=0}^{2n} \frac{\partial \hat{F}_q}{\partial t} \frac{\partial \hat{G}_q}{\partial w_{2n+1,j}}, \quad j = \overline{1, 3\nu}.$$

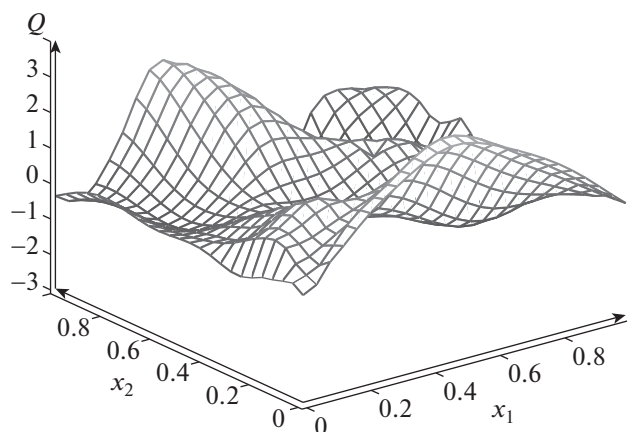
Непосредственным вычислением получаем

$$\frac{\partial \hat{G}_q}{\partial w_{2n+1,j}} = \sum_{p=1}^n \lambda^p \frac{\partial}{\partial w_{2n+1,j}} \hat{F}_{2n+1}(x_p + \sigma q; w_{2n+1}), \quad q = \overline{0, 2n}, \quad j = \overline{1, 3\nu};$$

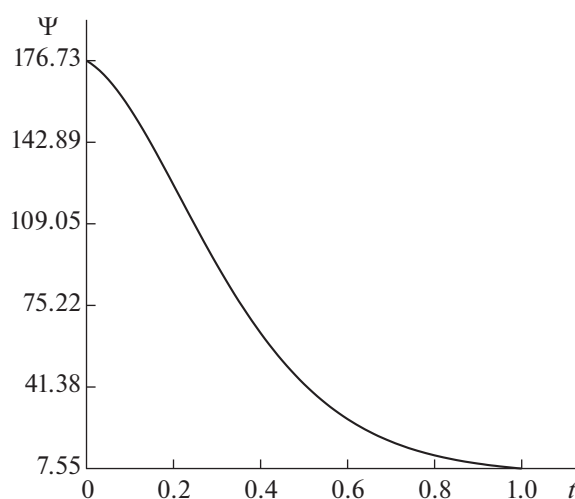
$$\frac{\partial \hat{F}_q}{\partial t} = -2 \sum_{j=1}^{\nu} \alpha_j^q \exp \left\{ - \left(\frac{t - \beta_j^q}{\gamma_j^q} \right)^2 \right\} \frac{t - \beta_j^q}{(\gamma_j^q)^2} = - \sum_{j=1}^{\nu} \frac{\partial \hat{F}_q}{\partial \beta_j^q};$$

$$\frac{\partial \hat{F}_q}{\partial \alpha_j^q} = \exp \left\{ - \left(\frac{t - \beta_j^q}{\gamma_j^q} \right)^2 \right\}, \quad \frac{\partial \hat{F}_q}{\partial \gamma_j^q} = 2 \alpha_j^q \exp \left\{ - \left(\frac{t - \beta_j^q}{\gamma_j^q} \right)^2 \right\} \frac{(t - \beta_j^q)^2}{(\gamma_j^q)^3}.$$

Случай $\nu = 5$. Получены следующие результаты. Значение квадратичной невязки: 25.0890 (отметим, что при минимизации от значения ≈ 29 до указанного продвижение очень медленное в связи с резкими изменениями нормы градиента). В случае “пробуксовки” методов первого порядка использовалось несколько сотен итераций метода Хука–Дживса. Значение нормы градиента квадратичной невязки: 0.041215 (при критерии останова $\|\nabla \mathcal{K}_p\| < 0.05$). Среднее отклонение в контрольных точках: 0.14965. Максимальное отклонение: 0.61687. Учитывая, что аппроксимируемая функция принимает значения от -3 до 3 , соответствующие относительные погрешности: $\delta = 0.024942$, $\delta_m = 0.102812$, см. фиг. 2, 3. Приведем значения параметров по строкам $q = \overline{0, 5}$.



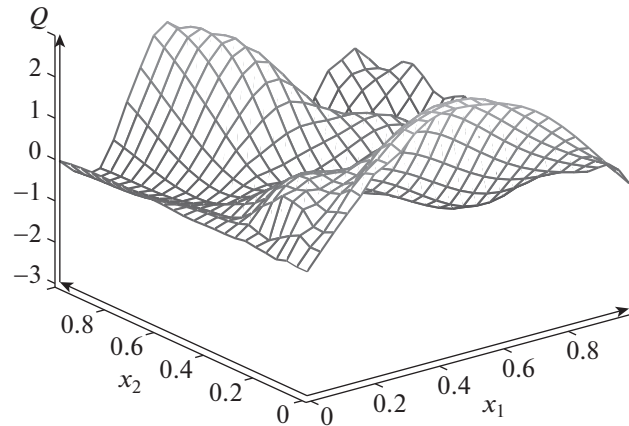
Фиг. 2. Тест № 1, $\nu = 5$, аппроксимация $Q[\nu]$.



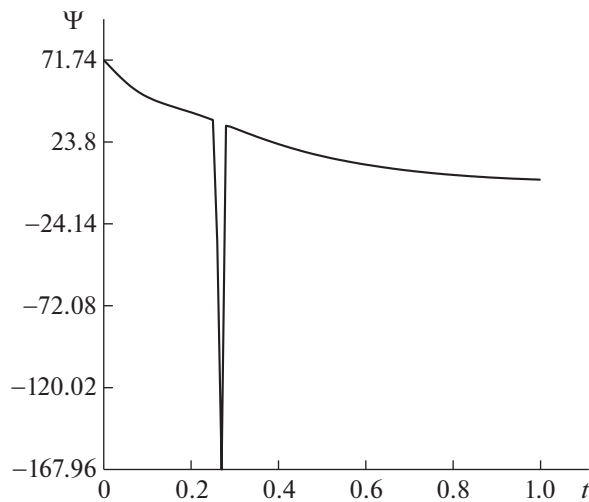
Фиг. 3. Тест № 1, $\nu = 5$, аппроксимация функции Ψ .

$$\alpha : \begin{cases} (-178.9659, 57.1697, 10050.027, 226.2424, 1918.7861), \\ (2570.8992, 1394.4094, 1624.578, -4773.0357, -1944.7895), \\ (-26232.6278, 855.3453, -2.5752, 1.7903, -18.7209), \\ (-660.7367, -86.18, -9650.8862, 6983.1783, 476.1241), \\ (1.7107, -4424.4155, -5380.042, -103.1709, -16.2722), \\ (-5.8677, 33.5005, 94.8855, 36.5603, 25.8181). \end{cases}$$

$$\beta : \begin{cases} (28.8537, 126.1182, 766.2185, 41.4888, -75.1308), \\ (32.2257, 27.8508, 81.1438, 791.3846, 36.2679), \\ (-156.6449, 15.2679, 24.8786, 20.6533, 17.3482), \\ (13.906, 15.1329, 50.4656, -7.3381, 57.5469), \\ (-308.472, 6.5986, 57.2145, 31.0121, 20.0108), \\ (1.0329, 0.061764, -0.039377, -0.074137, -2.0393). \end{cases}$$



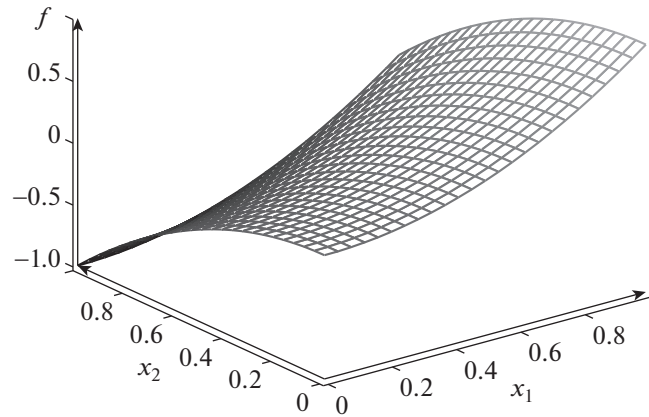
Фиг. 4. Тест № 1, $\nu = 10$, аппроксимация $Q[\nu]$.



Фиг. 5. Тест № 1, $\nu = 10$, аппроксимация функции Ψ .

$$\gamma: \begin{cases} (-22.3296, -102.8432, -1809.3131, 0.068125, 129.2289), \\ (-47.6429, -102.4432, -167.7611, -248.1456, -45.5089), \\ (140.5385, 3.232, 0.11785, -1.7783, 1.8117), \\ (1.4238, 4.6422, 68.6872, 19.1044, 23.0952), \\ (109.0217, -7.3042, -17.7514, 6.5694, -3.6359), \\ (-1.1871, 0.50697, -0.40204, 0.27905, -3.5102). \end{cases}$$

Случай $\nu = 10$. Получены следующие результаты. Значение квадратичной невязки: 11.6649 (отметим, что при минимизации от значения ≈ 15 до указанного продвижение очень медленное в связи с резкими изменениями нормы градиента). Значение нормы градиента квадратичной невязки: 0.049997 (при критерии останова $\|\nabla \mathcal{H}_\rho\| < 0.05$). Среднее отклонение в контрольных точках: 0.096934. Максимальное отклонение: 0.51107. Учитывая, что аппроксимируемая функция принимает значения от -3 до 3 , соответствующие относительные погрешности: $\delta = 0.016156$, $\delta_m = 0.085178$, см. фиг. 4, 5. Значения параметров не приводим в виду громоздкости соответствующего массива данных.



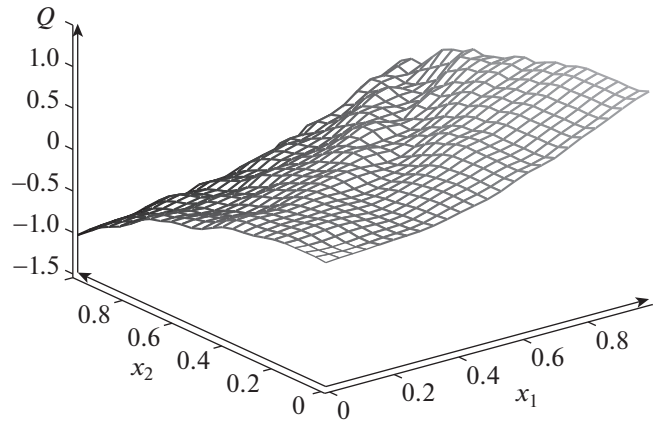
Фиг. 6. Тест № 2, аппроксимируемая функция.

Таким образом, при увеличении количества параметров от $v = 5$ до $v = 10$ точность аппроксимации возросла.

Тест № 2. Рассмотрим на квадрате E^2 непрерывную функцию двух переменных, см. фиг. 6, $f(x_1, x_2) = (x_1)^2 - (x_2)^2$. Остальные параметры выберем такими же, как для теста № 1. Поясним выбор тестовой функции. Когда речь идет об аппроксимации с помощью функций специальной конструкции, полезно убедиться в том, что такая аппроксимация работает и в отношении функций самого простого вида. Кроме того, возможность сколь угодно точной аппроксимации многочлена функциями рассматриваемого класса – это и есть основа доказываемого теоретического результата о всюду плотном вложении. Поэтому здесь мы как бы “убиваем сразу двух зайцев”.

Аппроксимация строилась для случая $v = 5$. При выборе параметра регуляризации $\rho = 10^{-15}$ наилучшее значение квадратичной невязки, которое удается получить, равно $\mathcal{H} \approx 21$. Если же для первых 5000 итераций принять $\rho = 10^{-7}$, а затем уже заменить его на $\rho = 10^{-15}$, то решая задачу (5), получаем следующие результаты. Значение квадратичной невязки: 3.7815. Среднее отклонение в контрольных точках: 0.061349. Максимальное отклонение: 0.23636. Использование “4-ступенчатого” управления параметром регуляризации $10^{-7} \rightarrow 10^{-15} \rightarrow 10^{-20} \rightarrow 0$ позволяет уменьшить значение квадратичной невязки до 3.22, среднее отклонение – до 0.056002, максимальное отклонение – до 0.21441. Кроме того, было опробовано специальное адаптивное управление параметром регуляризации. Суть его состоит в том, что в случае критического замедления сходимости делается пробное уменьшение и увеличение параметра регуляризации и в зависимости от того, какой результат оказывается более удачным, производится коррекция этого параметра в ту или другую сторону. Адаптивное управление параметром регуляризации позволило уменьшить значение квадратичной невязки до 0.99779, среднее отклонение – до 0.031401, максимальное отклонение – до 0.12812. Учитывая, что аппроксимируемая функция принимает значения от -1 до 1 , соответствующие относительные погрешности: $\delta = 0.0157005$, $\delta_m = 0.06406$, см. фиг. 7. Значение нормы градиента квадратичной невязки: 0.0048394 (при критерии останова $\|\nabla K_\rho\| < 0.005$). Приведем значения параметров по строкам $q = \overline{0, 5}$.

$$\alpha : \begin{cases} (128.4279, -18.7144, 1618.5152, 1.3407, 0.60003), \\ (-1800.1866, -1820.761, 4003.9764, 3990.1397, -0.13371), \\ (-652.9142, 8.3762, -3938.1157, 0.26061, 750.6071), \\ (-876.2382, 6069.1399, -1542.2666, -5642.6645, -54.2583), \\ (-89.7812, 5269.2731, 4335.3194, -2.844, -1.0896), \\ (6.4575, 9.6547, 1.2784, -0.39873, -3.0903). \end{cases}$$



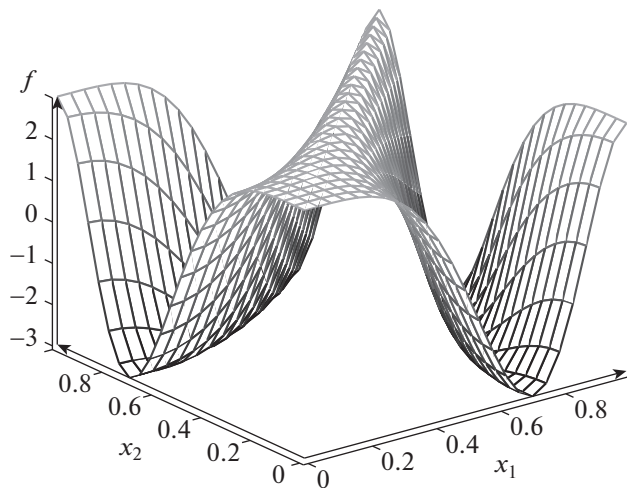
Фиг. 7. Тест № 2, аппроксимация.

$$\beta : \begin{cases} (5.5301, 2.2707, 17.6441, 31.989, 33.2267), \\ (54.4453, -34.8381, -6.0959, -12.7121, 23.609), \\ (-1.3221, -0.27688, -50.8737, 1.5753, -1.5001), \\ (-19.5566, -7.3444, 66.3863, -0.52106, -10.8569), \\ (-0.12464, -19.2713, -1.7185, 0.029663, 8.7288), \\ (-0.059137, 0.56273, -0.10547, 1.2843, 0.95817). \end{cases}$$

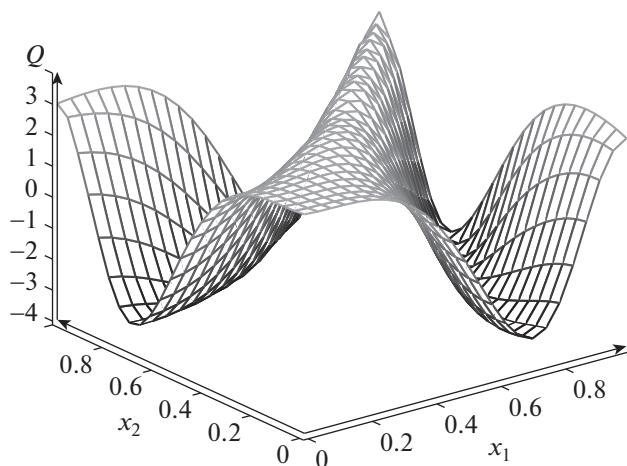
$$\gamma : \begin{cases} (0.080882, 5.4194, 348.7565, 5.9872, -2.463), \\ (7.4272, 18.027, -3.0533, -6.673, -4.8375), \\ (2.9946, 0.99319, 22.5371, 0.029948, 3.0545), \\ (50.7103, -2.9383, -84.1485, -0.26759, -15.2777), \\ (-0.14709, 6.9935, 0.71492, 0.0018687, -10.8164), \\ (0.35551, -0.49688, 0.11507, -0.2726, 0.25905). \end{cases}$$

Тест № 3. Рассмотрим на квадрате E^2 непрерывную функцию двух переменных, см. фиг. 8, $f(x_1, x_2) = 3 \cos(2\pi[(x_1)^2 - (x_2)^2])$. Функция выбиралась из следующего соображения: должно быть несколько “всплесков” и “впадин” ленточного типа (то есть локализованных лишь по одному направлению). Остальные параметры выберем такими же, как для теста № 1. Аппроксимация строилась для случая $\nu = 5$. Если для первых 5000 итераций принять $\rho = 10^{-7}$, затем заменить его на $\rho = 10^{-15}$, а потом при достижении значения $\mathcal{K}_\rho \approx 4$ принять $\rho = 10^{-20}$ и, наконец, при отсутствии убывания нормы градиента, взять $\rho = 0$, то, решая задачу (5), получаем следующие результаты. Значение квадратичной невязки: 3.4369. Значение нормы ее градиента: 0.049951 (при критерии останова $\|\nabla \mathcal{K}_\rho\| < 0.05$). Среднее отклонение в контрольных точках: 0.057943. Максимальное отклонение: 0.29136. Учитывая, что аппроксимируемая функция принимает значения от -3 до 3 , соответствующие относительные погрешности: $\delta = 0.0096572$, $\delta_m = 0.04856$, см. фиг. 9. Приведем значения параметров по строкам $q = 0, 5$.

$$\alpha : \begin{cases} (353.441, -0.33102, -1502.877, -3.0321, -132.2925), \\ (-1.9204, 495.9464, -81.72, -531.1092, 0.36865), \\ (0.20793, 9.3777, 14.9843, -2.3688, 394.0597), \\ (-17.9985, 471.1761, 92.7214, 313.2454, 0.45953), \\ (-67.9944, 216.6915, 80.3268, -136.0645, -1777.0164), \\ (-0.12255, -424803.4305, -34.1087, -11.2601, 49.9432). \end{cases}$$



Фиг. 8. Тест № 3, аппроксимируемая функция.

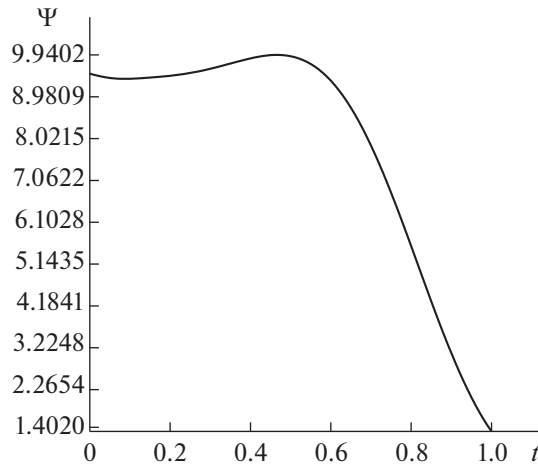


Фиг. 9. Тест № 3, аппроксимация.

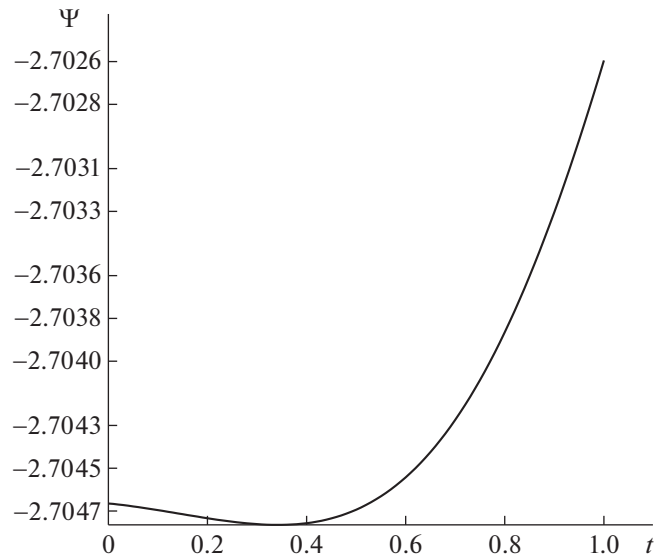
$$\beta : \begin{cases} (92.7012, -0.44829, -9.2283, -9.2311, -9.2351), \\ (-2.217, -9.2265, -9.2178, -9.2265, -2.6155), \\ (-0.39608, -9.2172, 59.3589, 0.065294, 6.0162), \\ (-9.2298, -9.2148, -9.21, 18.8314, -0.37507), \\ (-52.916, -9.2035, -9.2378, -83.662, -170.3084), \\ (1.2878, -14.0943, 13.6156, -3.5857, -19.0742). \end{cases}$$

$$\gamma : \begin{cases} (-325.7388, -0.22068, -0.023797, 0.0013843, -0.0032059), \\ (-0.51626, -0.0034228, -0.0064499, -0.0035007, -0.13142), \\ (0.37298, -5.5662e-005, -28.2643, -1.6457e-012, -252.0657), \\ (0.0077103, -0.084544, 0.0034378, -459.4922, -0.45826), \\ (7.491, 0.0070669, -0.03305, -0.12405, 46.7187), \\ (1.3722, -3.3655, 45.9973, -5.7163, 33.6242). \end{cases}$$

Здесь на то, чтобы достичь заданной малости нормы градиента, с момента выхода квадратичной невязки на значение 3.4369 уходит порядка двадцати тысяч итераций — при том, что значение минимизируемой функции, взятое с точностью до четырех знаков после запятой, за все это вре-



Фиг. 10. Тест № 2, аппроксимация функции Ψ .



Фиг. 11. Тест № 3, аппроксимация функции Ψ .

мя не изменяется никак, оставаясь равным 3.4369. Более того, если не производить переход на $\rho = 0$ по достижении данного значения, то наблюдается убывание нормы градиента до значения порядка 0.4, а затем она снова растет, и т.д. Наблюдается ситуация, когда при достаточной близости значения квадратичной невязки к ее точной нижней грани происходит замедление процесса минимизации, при котором норма градиента может не достигать заданного малого значения за разумное время. Представляется, что при использовании аппроксимаций, основанных на представлении Колмогорова, такая ситуация достаточно характерна. Чтобы понять, в чем тут дело, нужно вспомнить, что функции Колмогорова являются нигде не дифференцируемыми. В результате получается, что мы (под знаком суперпозиции) бесконечно дифференцируемыми функциями пытаемся аппроксимировать функции, нигде не дифференцируемые.

С учетом сказанного, в качестве критерия останова процесса численной оптимизации следует использовать не только условие малости нормы градиента, но и отсутствие существенных улучшений за некоторое количество итераций. Отметим, что такая практика является достаточно распространенной при минимизации сложно устроенных функций. Применение теоремы Колмогорова оправдывается тем, что при использовании малого количества параметров удается получать аппроксимацию хорошего качества для функций, устроенных “не слишком плохо” (а в том, чтобы построить такой способ аппроксимации, как раз и состояла исходная цель автора).

Наконец, сделаем следующее интересное наблюдение, см. фиг. 10, 11. Как указано в формулировке леммы 1.1, функция $\Psi(y)$ должна быть монотонно возрастающей. Однако за счет триви-

Таблица 1. Аппроксимация полиномами

v	\mathcal{K}	$\ \nabla\mathcal{K}\ $	Δ	Δ_{\max}
1	559.3065	0.033963	0.75375	2.5994
2	350.0558	0.041134	0.58236	2.4418
3	77.6166	0.21433	0.27834	1.3729
4	30.3614	3.8951	0.17855	0.68308
5	33.4258	0.91582	0.1826	0.71261
6	27.4957	1.4101	0.1671	0.55097

ального изменения внешних функций $F_q(y)$ (например, домножением аргумента на (-1)) возрастающую функцию $\Psi(y)$ можно заменить на убывающую. Это согласуется с полученными графиками аппроксимации функции Ψ для тестов № 1, 2. Кроме того, можно предположить, что за счет некоторой деформации функций F_q можно для некоторых конкретных функций, подвергаемых аппроксимации, отойти от монотонности функции Ψ в их представлении Колмогорова. И это подтверждается тестом № 3, см. фиг. 11.

Проводились численные эксперименты и с другими тестовыми функциями в виде суперпозиций основных элементарных функций и многочленов второй и третьей степени – в общем и целом картина наблюдалась каждый раз схожая.

Тест № 4. Для функции теста № 1 мы пробовали для аппроксимации одномерных функций в представлении Колмогорова использовать многочлены (заменяя соответствующую подпрограмму в разработанном программном комплексе на языке MATLAB; поэтому количество μ параметров – коэффициентов многочлена степени $(\mu - 1)$ задавалось в виде $\mu = 3v$, где число v передавалось в подпрограмму в качестве входного параметра). Оказалось, что при $v < 5$ такая аппроксимация работает, но такого значения v еще не достаточно для достижения хорошей точности, см. табл. 1. Здесь случай нормы градиента, большей заданного значения 0.05 в критерии останова, означает прерывание вычислений по причине отсутствия улучшений в смысле значения функции квадратичной невязки. При $v = 5$ уже наблюдаются сбои в ходе работы программы (приходится периодически перезапускать ее для продолжения вычислений с другими значениями параметров), связанные с тем, что появляются значения Inf (Infinity – бесконечность) и NaN (Not a Number – неопределенность). При $v = 10$ никаких значений, кроме Inf и NaN, при ненулевой начальной точке получить не удалось; при нулевой начальной точке в качестве аппроксимации получается константа. Причину описанных сбоев нетрудно понять, если провести сравнительную проверку правильности вычисления градиента $\nabla\mathcal{K}$ с помощью конечных разностей (см. следующий тест). Оказывается, что в отличие от аппроксимации функциями Гаусса, при аппроксимации многочленами наблюдается резкий рост погрешности вычисления градиента с ростом параметра v . А это, в свою очередь, происходит вследствие того, что коэффициентами при параметрах многочлена являются степени одномерной переменной t . При $|t| > 1$ и больших показателях степени происходит резкий рост этих коэффициентов, а при $|t| < 1$ их влияние становится исчезающе малым.

Тест № 5. В табл. 2 приведены результаты проверки правильности вычисления градиента $\nabla\mathcal{K}$ функции квадратичной невязки при аппроксимации одномерных функций в представлении Колмогорова многочленами в точке $0.01\mathbf{1}$, где $\mathbf{1}$ – вектор размерности $\mu = 3v$, составленный из единиц. Здесь Δ_{\min} – минимум (по всем частным производным первого порядка) абсолютных погрешностей, Δ_{\max} – максимум абсолютных погрешностей, δ_{\min} – минимум относительных погрешностей, δ_{\max} – максимум относительных погрешностей. При $v = 10$ значения всех погрешностей равны NaN. В табл. 3 приведены результаты проверки правильности вычисления градиента $\nabla\mathcal{K}$ функции квадратичной невязки при аппроксимации одномерных функций в представлении Колмогорова функциями Гаусса в той же точке.

На основе результатов численных экспериментов, описанных выше, можно сделать следующие выводы.

1. Проведенные численные эксперименты подтверждают предположение о том, что предлагаемый метод аппроксимации функций многих переменных на основе приближения одномер-

Таблица 2. Полиномы: проверка правильности вычисления $\nabla \mathcal{K}$

v	Δ_{\min}	Δ_{\max}	δ_{\min}	δ_{\max}
1	$2.28530e-7$	$6.25008e-3$	$3.3848e-7$	$2.6785e-5$
5	$4.89217e+28$	$5.74642e+65$	$2.4827e-10$	1

Таблица 3. Функции Гаусса: проверка правильности вычисления $\nabla \mathcal{K}$

v	Δ_{\min}	Δ_{\max}	δ_{\min}	δ_{\max}
5	$7.33070e-4$	$1.42575e-1$	$6.0807e-6$	0.002413
10	$5.04264e-8$	$1.00985e-1$	$1.0604e-6$	0.0085675

ных функций в их представлении А.Н. Колмогорова (здесь и далее — в стиле леммы 1.1) линейными комбинациями функций Гаусса с варьируемыми параметрами позволяет получать качественные аппроксимации непрерывных функций с “нехаотическим” графиком при сравнительно малом количестве параметров.

2. Устойчивость процесса численного решения задачи (5) наблюдается при использовании методов DFP и BFS в ходе многомерной оптимизации и метода Брента для одномерного поиска, при условии выбора начального приближения в виде вектора $\kappa\{-1, 2, -3, \dots, (-1)^\mu\}$, где μ — количество параметров, κ — малая величина.

3. Погрешность вычисления градиента функционала квадратичной невязки при увеличении количества параметров в рамках предлагаемого подхода не становится критичной, в отличие от подхода, основанного на аппроксимации одномерных функций в представлении А.Н. Колмогорова полиномами. Сравнение с подходом, основанным на аппроксимации одномерных функций полиномами, было необходимо провести, поскольку теорема 1.2 о всюду плотности класса \mathcal{E} в пространстве непрерывных функций доказывается посредством аппроксимации полиномов функциями класса \mathcal{E} . Может возникнуть резонный вопрос: почему бы сразу не использовать полиномы? Помимо уже упомянутых во введении преимуществ использования функций Гаусса, численные эксперименты дают еще один наглядный ответ на этот вопрос.

4. Для повышения точности достигаемой аппроксимации может оказаться полезным постепенное уменьшение параметра регуляризации в ходе численного решения задачи (5) от некоторого малого, но не слишком малого, значения (например, 10^{-7}) до нуля.

5. Функция Ψ в представлении А.Н. Колмогорова, судя по ее аппроксимациям, реализовавшимся в ходе экспериментов, определяется не просто неоднозначно, а проявляет своего рода гибкость, полезную для облегчения численного процесса минимизации квадратичной невязки. Упомянутая гибкость связана с тем, что, как уже было сказано, имеется некоторая компенсируемая неоднозначность в задании функций F_q и Ψ .

6. Для теста № 3 результат получился несколько лучше, чем для теста № 2 (с функцией более простого вида). Вообще говоря, ничего удивительного в этом нет. Например, функцию $y = x$ на отрезке $[0; 1]$ можно сколь угодно точно аппроксимировать рядом Фурье, но для этого потребуются некоторое количество членов разложения. А для аппроксимации (и даже точного представления) функции более сложного вида $y = \cos x$ достаточно всего лишь одного члена разложения $y = \cos x$. Однако, исходя из сравнения результатов численных экспериментов в тестах № 2 и 3, автору было интересно выяснить, насколько в принципе можно улучшить качество аппроксимации при $v = 5$ для теста № 2. Оказалось, что в этом смысле “4-ступенчатое” управление параметром регуляризации $10^{-7} \rightarrow 10^{-15} \rightarrow 10^{-20} \rightarrow 0$ более эффективно, чем “2-ступенчатое”: $10^{-7} \rightarrow 10^{-15}$; адаптивное управление — еще более эффективно.

СПИСОК ЛИТЕРАТУРЫ

1. Teo K.L., Goh C.J., Wong K.H. A unified computational approach to optimal control problems / Pitman Monographs and Surveys in Pure and Applied Mathematics. Vol. 55. Harlow, New York: Longman Scientific & Technical, Wiley, Inc., 1991. ix+329 p.

2. Чернов А.В. О приближенном решении задач оптимального управления со свободным временем // Вестник Нижегородского гос. ун-та им. Н.И. Лобачевского. 2012. № 6(1). С. 107–114.
3. Чернов А.В. О гладких конечномерных аппроксимациях распределенных оптимизационных задач с помощью дискретизации управления // Ж. вычисл. матем. и матем. физ. 2013. Т. 53. № 12. С. 2029–2043.
4. Чернов А.В. О применимости техники параметризации управления к решению распределенных задач оптимизации // Вестник Удмуртского университета. Математика. Механика. Компьютерные науки. 2014. Т. 24. Вып. 1. С. 102–117.
5. Чернов А.В. О гладкости аппроксимированной задачи оптимизации системы Гурса–Дарбу на варьируемой области // Труды Института математики и механики УрО РАН. 2014. Т. 20. № 1. С. 305–321.
6. Чернов А.В. О кусочно постоянной аппроксимации в распределенных задачах оптимизации // Труды ИММ УрО РАН. 2015. Т. 21. № 1. С. 264–279.
7. Колмогоров А.Н. О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения // Докл. АН СССР. 1957. Т. 114. № 5. С. 953–956.
8. Sprecher D.A. On the structure of continuous functions of several variables // Trans. Amer. Math. Soc. 1965. V. 115. P. 340–355.
9. Голубков А.Ю. Построение внешних и внутренних функций представления непрерывных функций многих переменных суперпозицией непрерывных функций одного переменного // Фундамент. и прикл. матем. 2002. Т. 8. Вып. 1. С. 27–38.
10. Бутырский Е.Ю., Кувалдин И.А., Чалкин В.П. Аппроксимация многомерных функций // Научное приборостроение. 2010. Т. 20. № 2. С. 82–92.
11. Hecht-Nielsen R. Kolmogorov's mapping neural network existence theorem // IEEE First Annual Int. Conf. on Neural Networks. San-Diego. 1987. V. 3. P. 11–13.
12. Алексеев Д.В. Приближение функций нескольких переменных нейронными сетями // Фундаментальная и прикл. матем. 2009. Т. 15. Вып. 3. С. 9–21.
13. Горбань А.Н. Обобщенная аппроксимационная теорема и точное представление многочленов от нескольких переменных суперпозициями многочленов от одного переменного // Изв. вузов. Математика. 1998. № 5. С. 6–9.
14. Исмаилов В.Э. Аппроксимация суммами ридж-функций с фиксированными направлениями // Алгебра и анализ. 2016. Т. 28. № 6. С. 20–69.
15. Кульчин Ю.Н., Денисов И.В., Панов А.В., Рыбальченко Н.А. Применение персептронов для нелинейной реконструктивной томографии // Пробл. управления. 2006. Вып. 4. С. 59–63.
16. Чернов А.В. Об использовании квадратичных экспонент с варьируемыми параметрами для аппроксимации функций одного переменного на конечном отрезке // Вестник Удмуртского университета. Математика. Механика. Компьютерные науки. 2017. Т. 27. Вып. 2. С. 267–282.
17. Чернов А.В. О применении квадратичных экспонент для дискретизации задач оптимального управления // Вестник Удмуртского университета. Математика. Механика. Компьютерные науки. 2017. Т. 27. Вып. 4. С. 558–575.
18. Maz'ya V., Schmidt G. Approximate approximations. Mathematical Surveys and Monographs. V. 141. Providence, RI: American Mathematical Society (AMS), 2007. xiv+349 p.
19. Riemenschneider S.D., Sivakumar N. Cardinal interpolation by Gaussian functions: a survey // J. Analysis. 2000. V. 8. P. 157–178.
20. Luh Lin-Tian. The shape parameter in the Gaussian function // Computers and Mathematics with Applications. 2012. V. 63. P. 687–694.
21. Hangelbroek T., Madych W., Narcowich F., Ward J.D. Cardinal interpolation with Gaussian kernels // J. Fourier Anal. Appl. 2012. V. 18. № 1. P. 67–86.
22. Hamm K. Approximation rates for interpolation of Sobolev functions via Gaussians and allied functions // J. Approx. Theory. 2015. V. 189. P. 101–122.
23. Griebel M., Schneider M., Zenger C. A combination technique for the solution of sparse grid problems / Iterative methods in linear algebra. Proc. of the IMACS international symposium, Brussels, Belgium, 2–4 April, 1991. Amsterdam: North-Holland, 1992. P. 263–281.
24. Georgoulis E., Levesley J., Subhan F. Multilevel sparse kernel-based interpolation // SIAM J. Sci. Comput. 2013. V. 35. № 2. P. A815–A831.
25. More J.J., Thuente D.J. Line search algorithms with guaranteed sufficient decrease // ACM Transactions on Mathematical Software. 1994. V. 20. P. 286–307.