

АППРОКСИМАЦИЯ И СГЛАЖИВАНИЕ ФУНКЦИИ НА ОСНОВЕ РЕГУЛЯРИЗАЦИИ ПО ГОДУНОВУ¹⁾

© 2024 г. Э. А. Бибердорф^{1,*}, К. К. Абдишерипов²

¹630090 Новосибирск, пр. Академика Коптюга, 4, Институт математики им. С.Л. Соболева СО РАН, Россия

²630090 Новосибирск, ул. Пирогова, 1, Новосибирский Государственный Университет, Новосибирск, Россия

*e-mail: math@biberdorf.ru

Поступила в редакцию 02.04.2024 г.

Переработанный вариант 02.04.2024 г.

Принята к публикации 02.05.2024 г.

Представлен новый подход к аппроксимации функций, основанный на идеях С.К. Годунова о регуляризации плохо обусловленных систем. Представленный метод позволяет определить значения функций в узлах мелкой сетки, исходя из данных на более крупной сетке, обеспечивая при этом контроль над гладкостью получаемой функции. Обоснованы оценки сходимости и гладкости, а также приведены результаты вычислительных экспериментов, иллюстрирующие эффективность предложенного метода. Библ. 4. Фиг. 6.

Ключевые слова: аппроксимация, сглаживание функций, регуляризация плохо обусловленных СЛАУ.

DOI: 10.31857/S0044466924080017, **EDN:** YBJDNW

ВВЕДЕНИЕ

Благодаря работам по разностным схемам имя Сергея Константиновича Годунова стало известно математикам и физикам всего мира. Несколько менее известны его статьи и книги по вычислительной линейной алгебре, однако именно в них наиболее отчетливо проявляются особенности его научного подхода к решению математических проблем. Обратимся к методу регуляризации по Годунову плохообусловленных систем линейных алгебраических уравнений (СЛАУ). Его возникновение связано с решением СЛАУ, представляющей собой дискретизированное интегральное уравнение Фредгольма первого рода. Известно, что такие СЛАУ, как правило, плохо обусловлены. С.К. Годунов отметил, что не все известные свойства искомого решения, которые присутствуют в исходной физической постановке задачи, нашли отражение в ее математической постановке. В частности, это касалось гладкости решения. Таким образом, получалось, что добавление в СЛАУ уравнений разделенных разностей делало более точным соответствие между физической и математической задачами. Оказалось, что при таком расширении системы значительно уменьшается число обусловленности, а специальный выбор параметра «склейки» двух подсистем приводил к тому, что регуляризация не увеличивает уровень погрешности. Описание этого метода можно найти в монографиях [1–3].

Позднее уже в 2000-х годах в личных беседах С.К. Годунов высказывал мысль, что аналогичный подход мог бы применяться для сглаживания функций, заданных на дискретной сетке. В настоящей работе, развивая эту идею, мы представляем новый метод обработки данных, который сочетает в себе свойства аппроксимации и сглаживания.

Аппроксимация используется во многих прикладных задачах и состоит, как правило, из двух этапов. Первый этап — вычисление коэффициентов аппроксимирующей функции по заданным значениям исходной функции, второй — вычисление значений аппроксимирующей функции в интересующих точках. Основные сложности численной реализации методов аппроксимации связаны с вычислительной трудоемкостью и/или вычислительной неустойчивостью. Особенно очевидны эти проблемы при использовании аппроксимации степенными функциями. Первый этап в этом случае состоит из решения линейной алгебраической системы, которая с ростом степени аппроксимирующего полинома стремительно становится плохо обусловленной.

В настоящей работе предлагается метод аппроксимации, обладающий двумя основными отличительными чертами. Первая из них заключается в том, что метод позволяет вычислять значения аппроксимирующей функции сразу же в процессе решения специальной СЛАУ. Таким образом, два этапа аппроксимации сливаются в

¹⁾ Работа выполнена в рамках государственного задания Института математики им. С.Л. Соболева СО РАН (проект к FWNF-2022-0008), фонд «El-yurt umidi», Узбекистан.

один, что, естественно, приводит к экономии вычислительных затрат. Вторая особенность предлагаемого метода в том, что он основан на приеме регуляризации плохо обусловленных систем. Поэтому можно ожидать, что обусловленность возникающих СЛАУ не слишком велика.

В разд. 1 работы приводится общее описание предлагаемого метода. В разд. 2 и 3 частях дано его обоснование, включая оценки сходимости, гладкости и обусловленности. В завершение приводятся некоторые примеры расчетов.

1. СПОСОБ ПОЛИНОМИАЛЬНОЙ АППРОКСИМАЦИИ

Предлагаемый метод является вариацией метода регуляризации плохо-обусловленных систем, описанного в работе [1]. Суть этого метода состоит в том, что для поиска решения системы, помимо основной информации, используется дополнительная эмпирическая информация об искомой функции, например, в виде оценок производных.

Пусть в точках

$$x_0 < x_1 < \dots < x_{m-1} < x_m, \quad x_j - x_{j-1} = \Delta, \quad j = 1, \dots, m, \quad (1)$$

заданы значения функции $f(x)$. Будем искать значения полинома $g(x)$ степени p в точках ξ_k ($k = 0, 1, \dots, n \cdot m$):

$$x_j = \xi_{jn} < \xi_{jn+1} < \dots < \xi_{(j+1)n} = x_{j+1}, \quad \xi_k - \xi_{k-1} = \delta,$$

аппроксимирующего функцию $f(x)$ (см. фиг. 1).

При аппроксимации мы стремимся к выполнению следующих равенств:

$$g(x_j) = f(x_j), \quad j = 1, \dots, m. \quad (2)$$

По аналогии с упомянутым выше методом регуляризации эти соотношения можно рассматривать как основную СЛАУ:

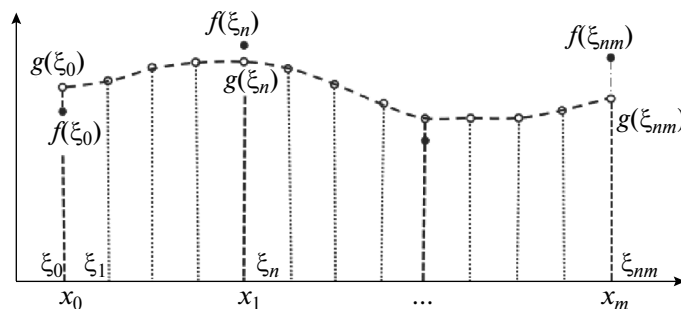
$$Bg = f,$$

где

$$g = \begin{pmatrix} g(\xi_0) \\ g(\xi_1) \\ \vdots \\ g(\xi_{nm}) \end{pmatrix}, \quad f = \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_m) \end{pmatrix}, \quad (3)$$

Матрица B состоит из $m + 1$ строчек. В каждой из них только $(j - 1)n + 1$ -ый элемент (j -номер строки) отличен от нуля и равен 1:

$$B = \begin{matrix} & \begin{matrix} jn+1 \end{matrix} \\ \begin{matrix} j+1 \end{matrix} & \begin{pmatrix} 1 & 0 & \vdots & \dots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \dots & 1 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & \vdots & \dots & 1 \end{pmatrix} \end{matrix}. \quad (4)$$



Фиг. 1. Постановка задачи аппроксимации.

Очевидно, что эта система является недоопределенной и не может быть однозначно решена. Для ее корректной разрешимости необходимо добавить соотношения на производные:

$$g^{(p+1)}(\xi_k) = 0, \quad k = 0, 1, \dots, nm, \quad (5)$$

или разделенные разности порядка $p + 1$:

$$D_{p+1}g(\xi_k) = 0, \quad k = 0, 1, \dots, nm. \quad (6)$$

Для того чтобы система уравнений относительно значений $g(\xi_k)$ была замкнутой, из равенств (6) выбираем

$$D_{p+1}g(\xi_k) = 0, \quad k = \left[\frac{p}{2} \right], \dots, \left[nm - \frac{p+1}{2} \right]. \quad (7)$$

Совокупность равенств (7) образует систему линейных алгебраических уравнений $D_{p+1}g = 0$, где строчки матрицы D_{p+1} содержат коэффициенты разделенных разностей.

Теперь объединим обе системы:

$$Ag + \begin{pmatrix} r \\ \dots \\ \rho \end{pmatrix} = \begin{pmatrix} f \\ \dots \\ 0 \end{pmatrix}, \quad A = \begin{pmatrix} B \\ \dots \\ D_{p+1} \end{pmatrix}. \quad (8)$$

В результате получена переопределенная система, которая может быть несовместной. Поэтому в уравнение вводится невязка. Далее предполагается решать эту систему методом наименьших квадратов, который невязку минимизирует. Заметим, что минимизация невязки ρ , в частности, означает, что аппроксимация обладает также сглаживающим эффектом.

Таким образом, для приближенного вычисления значений аппроксимирующего полинома достаточно решить систему (8).

2. АППРОКСИМАЦИЯ КУСОЧНО-ПОСТОЯННОЙ ФУНКЦИЕЙ

Будем рассматривать случай $p = 0$, когда в качестве дополнительных уравнений берутся разделенные разности первого порядка. Таким образом, матрица D_1 имеет вид

$$D_1 = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & -1 \end{pmatrix}. \quad (9)$$

2.1. Применение метода Лагранжа

Для определения свойств такой аппроксимации можно переформулировать задачу как задачу нахождения условного минимума и воспользоваться методом Лагранжа. С этой целью введем дополнительные обозначения (см. фиг. 2):

$$\begin{aligned} \tilde{f}_j &= g(x_j), \quad j = 1, \dots, m, \\ \delta_{j,k} &= g(\xi_{jn+k}) - g(\xi_{jn+k-1}), \quad j = 1, \dots, m-1, \quad k = 1, \dots, n. \end{aligned} \quad (10)$$

В этих обозначениях метод наименьших квадратов, примененный к системе (8), приводит к вычислению минимума функции

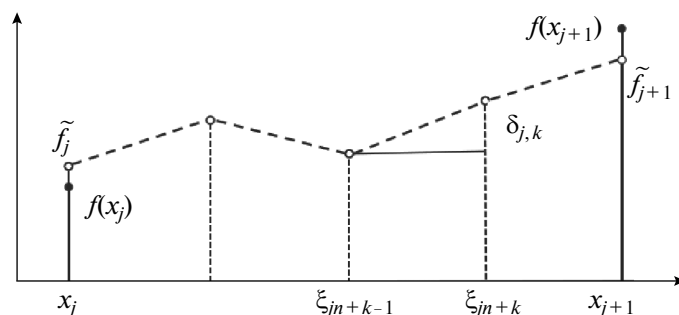
$$F(\tilde{f}_0, \dots, \tilde{f}_m; \delta_{0,1}, \dots, \delta_{m-1,n}) = \sum_{j=0}^m (f_j - \tilde{f}_j)^2 + \sum_{j=0}^{m-1} \sum_{k=1}^n \delta_{j,k}^2$$

при выполнении условий

$$\tilde{f}_{j+1} - \tilde{f}_j = \sum_{k=1}^n \delta_{j,k}, \quad j = 0, \dots, m-1.$$

Для применения метода Лагранжа составляем вспомогательную функцию

$$\Phi(\lambda_0, \dots, \lambda_{m-1}; \delta_{1,1}, \dots, \delta_{l-1,n}) = \sum_{j=0}^m (f_j - \tilde{f}_j)^2 + \sum_{j=0}^{m-1} \sum_{k=1}^n \delta_{j,k}^2 + \sum_{j=0}^{m-1} \lambda_j \left(\tilde{f}_{j+1} - \tilde{f}_j - \sum_{k=1}^n \delta_{j,k} \right).$$


 Фиг. 2. Обозначения для аппроксимации в случае $p = 0$.

Приравниваем к нулю производные функции Φ по всем параметрам:

$$\begin{aligned} \Phi_{\lambda_j} &= \tilde{f}_{j+1} - \tilde{f}_j - \sum_{k=1}^n \delta_{j,k} = 0, \\ \Phi_{\delta_{j,k}} &= 2\delta_{j,k} - \lambda_j = 0, \\ \Phi_{\tilde{f}_0} &= -2(f_0 - \tilde{f}_0) - \lambda_0 = 0, \\ \Phi_{\tilde{f}_1} &= -2(f_1 - \tilde{f}_1) + \lambda_0 - \lambda_1 = 0, \\ \Phi_{\tilde{f}_j} &= -2(f_j - \tilde{f}_j) + \lambda_{j-1} - \lambda_j = 0, \\ \Phi_{\tilde{f}_m} &= -2(f_m - \tilde{f}_m) + \lambda_{m-1} = 0. \end{aligned} \quad (11)$$

Из первых двух равенств получаем, что

$$\delta_{j,k} = \frac{\lambda_j}{2}, \quad (12)$$

$$\tilde{f}_{j+1} - \tilde{f}_j = \sum_{k=1}^n \delta_{j,k} = n\delta_{j,k} = \frac{n}{2}\lambda_j. \quad (13)$$

Замечание. Равенство (12) означает, что в промежутках между точками x_j аппроксимация производится *линейно*. Следовательно, внутри каждого интервала (x_{j-1}, x_j) производная аппроксимирующей функции равна $(\tilde{f}_j - \tilde{f}_{j-1})/\Delta$.

Подставим выражение (13) в равенства для производных Φ_{f_j} :

$$\begin{aligned} \Phi_{\tilde{f}_0} &= -2(f_0 - \tilde{f}_0) - \lambda_0 = -2(f_0 - \tilde{f}_0) - \frac{2}{n}(\tilde{f}_1 - \tilde{f}_0) = 0, \\ \Phi_{\tilde{f}_j} &= -2(f_j - \tilde{f}_j) + \lambda_j - \lambda_{j+1} = -2(f_j - \tilde{f}_j) - \frac{2}{n}(\tilde{f}_{j+1} - \tilde{f}_j) + \frac{2}{n}(\tilde{f}_j - \tilde{f}_{j-1}) = 0, \\ \Phi_{\tilde{f}_m} &= -2(f_m - \tilde{f}_m) + \lambda_{m-1} = -2(f_m - \tilde{f}_m) + \frac{2}{n}(\tilde{f}_m - \tilde{f}_{m-1}) = 0. \end{aligned}$$

В матрично-векторном виде эти равенства принимают вид

$$f = \left(I + \frac{1}{n}T_1\right)\tilde{f}, \quad (14)$$

где

$$T_1 = \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix}.$$

Таким образом,

$$f - \tilde{f} = \frac{1}{n}T_1\tilde{f}.$$

Так как $\|T_1\| \leq 4$, то очевидна оценка

$$\|f - \tilde{f}\| \leq \frac{4}{n}\|\tilde{f}\|.$$

Здесь и далее обозначение $\|\cdot\|$, примененное к вектору, означает евклидову норму, а примененное к матрице — спектральную норму.

Из соотношения

$$\left(I + \frac{1}{n}T_1\right)^{-1}f = \tilde{f}$$

можно вывести и априорную оценку. Так как при $n \geq 5$

$$\left(I + \frac{1}{n}T_1\right)^{-1} = I - \frac{1}{n}T_1 \sum_{j=0}^{\infty} \left(\frac{-1}{n}T_1\right)^j = I + \Omega, \quad \|\Omega\| \leq \frac{4/n}{1 - 4/n},$$

то

$$\|f - \tilde{f}\| \leq \frac{4/n}{1 - 4/n} \|f\|. \quad (15)$$

Тем самым доказана следующая теорема.

Теорема 1. Пусть функция $f(x)$ задана в точках (1), матрицы A, B, D_1 имеют вид (8), (4) и (9) соответственно. Если при этом вектор g является обобщенным нормальным решением системы (8), а вектор $\tilde{f}$ задан равенством (10), то для векторов f и $\tilde{f}$ верна оценка (15), откуда следует, что данная аппроксимация имеет первый порядок точности.

2.2. Гладкость аппроксимации

Пусть матрица R_1 размера $m \times (m+1)$ имеет вид

$$R_1 = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & -1 \end{pmatrix},$$

т.е. она отличается от матрицы D_1 размером. Тогда сглаживающий эффект данной аппроксимации описывается в следующей лемме.

Лемма 1. Для разностей элементов векторов f и $\tilde{f}$ имеет место неравенство

$$\|R_1 f\| \geq \left(1 + \frac{\lambda_{\min}(R_1 R_1^*)}{n}\right) \|R_1 \tilde{f}\|.$$

Для доказательства достаточно умножить обе части уравнения (14) на R_1 и заметить, что $T_1 = R_1^* R_1$, в результате чего получается следующая цепочка:

$$\|R_1 f\| = \|R_1 \tilde{f} + \frac{1}{n} R_1 R_1^* R_1 \tilde{f}\| = \left\| \left(I + \frac{1}{n} R_1 R_1^*\right) R_1 \tilde{f} \right\| \geq \left(1 + \frac{\lambda_{\min}(R_1 R_1^*)}{n}\right) \|R_1 \tilde{f}\|.$$

Таким образом, имеет место убывание разностных производных.

Усилить или ослабить эффект сглаживания можно путем добавления в систему весового параметра η :

$$A = \begin{pmatrix} B \\ \eta D_1 \end{pmatrix}.$$

Очевидно, что чем меньше вес, тем точнее аппроксимирующая функция, и тем меньше ее гладкость, и наоборот.

Повторив описанный в п. 2.1 процесс для этого случая, получим оценки, учитывающие множитель η :

$$\|f - \tilde{f}\| \leq \frac{4\eta^2}{n} \|\tilde{f}\|,$$

$$\|f - \tilde{f}\| \leq \frac{4\eta^2/n}{1 - 4\eta^2/n} \|f\|,$$

а также

$$\|R_1 f\| \geq \left(1 + \frac{4\eta^2}{n} \lambda_{\min}(R_1 R_1^*)\right) \|R_1 \tilde{f}\|.$$

Таким образом, при $p = 0$, несмотря на то что ожидается аппроксимация константой, т.е. полиномом нулевого порядка, в реальности имеет место кусочно-линейная аппроксимация. При этом такая аппроксимация более трудоемка, чем обычная кусочно-линейная интерполяция, когда линейными отрезками соединяются известные значения функции $f(x_j)$. Однако при этом данный подход к аппроксимации позволяет автоматически сгладить исходную функцию.

2.3. Обусловленность системы

Оценку числа обусловленности системы (8) можно получить на основе следующей теоремы (доказательство см. в [4]).

Теорема 2 (об окаймлении). Если $N \times N$ -матрицу A окаймить справа дополнительным столбцом из N элементов, то сингулярные числа $\sigma_1(B), \sigma_2(B), \dots, \sigma_N(B)$ матрицы B размера $N \times (N + 1)$, полученной в результате окаймления, связаны с сингулярными числами $\sigma_1(A), \sigma_2(A), \dots, \sigma_N(A)$ матрицы A условиями перемежаемости

$$\sigma_N(B) \geq \sigma_N(A) \geq \sigma_{N-1}(B) \geq \dots \geq \sigma_1(B) \geq \sigma_1(A).$$

Дополним матрицу D_1 последней строкой матрицы B и обозначим полученную квадратную $(mn + 1) \times (mn + 1)$ матрицу $\hat{D}_1$:

$$\hat{D}_1 = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & -1 \\ 0 & \dots & 0 & 0 & -1 \end{pmatrix}.$$

Так как матрица A , определенная равенством (8), может рассматриваться как окаймление матрицы $\hat{D}_1$, то, применяя теорему 2, получаем неравенства

$$\sigma_{\min}(\hat{D}_1) \leq \sigma_{\min}(A) \leq \sigma_{\max}(\hat{D}_1) \leq \sigma_{\max}(A).$$

Отсюда следует оценка для числа обусловленности матрицы

$$\text{cond}(A) \leq \frac{\sigma_{\max}(A)}{\sigma_{\min}(\hat{D}_1)} = \|A\| \|\hat{D}_1^{-1}\|.$$

При этом имеют место очевидные ограничения на нормы

$$\|A\| \leq 2, \quad \|\hat{D}_1^{-1}\| \leq mn + 1.$$

Тем самым доказана следующая лемма.

Лемма 2. Пусть матрицы B и D_1 размеров $(m + 1) \times (mn + 1)$ и $mn \times (mn + 1)$ соответственно заданы формулами (4) и (9). Тогда число обусловленности матрицы

$$A = \begin{pmatrix} B \\ \dots \\ D_1 \end{pmatrix}$$

не превосходит $2(mn + 1)$.

3. ТОЧНОСТЬ АППРОКСИМАЦИИ КУСОЧНО-ЛИНЕЙНОЙ ФУНКЦИЕЙ

В данном разделе некоторые обозначения будут изменены для удобства выкладок.

Аппроксимация линейной функцией предполагает, что в качестве дополнительных условий используется равенство нулю вторых производных или разделенных разностей. Это означает, что регуляризующая матрица D_2 размера $(mn - 1) \times (mn + 1)$ имеет вид

$$D_2 = \begin{pmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \end{pmatrix}. \quad (16)$$

Введем дополнительные обозначения:

$$d_j = g(\xi_j) - g(\xi_{j-1}), \quad 1 \leq j \leq nm,$$

$$\varepsilon_j = d_{j+1} - d_j = g(\xi_{j+1}) - 2g(\xi_j) + g(\xi_{j-1}), \quad 1 \leq j \leq nm - 1,$$

$$\delta_j = g(x_j) - f_j, \quad 0 \leq j \leq m.$$

Решение системы (8) методом наименьших квадратов означает минимизацию нормы вектора невязки

$$(\delta_0, \dots, \delta_m, \varepsilon_1, \dots, \varepsilon_{nm-1})^T.$$

Чтобы сформулировать дополнительные условия, заметим, что

$$d_{nj+k} = d_{nj+1} + \sum_{i=1}^{k-1} \varepsilon_{nj+i}, \quad g(\xi_{n(j+1)}) - g(\xi_{nj}) = \sum_{i=1}^n d_{nj+i}, \quad 0 \leq j \leq m-1.$$

Подставим первое из этих равенств во второе и получим следующие соотношения:

$$nd_{nj+1} + \sum_{i=1}^{n-1} (n-i)\varepsilon_{nj+i} = f_{j+1} - f_j + \delta_{j+1} - \delta_j, \quad (17)$$

$$nd_{nj} - \sum_{i=1}^{n-1} i\varepsilon_{n(j-1)+i} = f_j - f_{j-1} + \delta_j - \delta_{j-1}. \quad (18)$$

Из равенства (17) вычтем (18), используя то, что $d_{nj+1} - d_{nj} = \varepsilon_{nj}$. Получаем при $j = 1, \dots, m-1$

$$f_{j-1} - 2f_j + f_{j+1} + \delta_{j-1} - 2\delta_j + \delta_{j+1} - \sum_{i=1}^n i\varepsilon_{n(j-1)+i} - \sum_{i=1}^{n-1} (n-i)\varepsilon_{nj+i} = 0. \quad (19)$$

Теперь можно применить метод Лагранжа для вычисления условного минимума. Функция Лагранжа имеет вид

$$\begin{aligned} \Phi = & \sum_{j=0}^m \delta_j^2 + \sum_{i=1}^{nm-1} \varepsilon_i^2 + \\ & + \sum_{j=1}^{m-1} \lambda_j \left(f_{j-1} - 2f_j + f_{j+1} + \delta_{j-1} - 2\delta_j + \delta_{j+1} - \sum_{i=1}^n i\varepsilon_{n(j-1)+i} - \sum_{i=1}^{n-1} (n-i)\varepsilon_{nj+i} \right). \end{aligned}$$

Приравниваем к нулю частные производные функции Лагранжа:

$$\begin{aligned} \Phi_{\varepsilon_j} &= 2\varepsilon_j - \lambda_1 i = 0, \\ \Phi_{\varepsilon_{nj+i}} &= 2\varepsilon_{nj+i} - \lambda_j(n-i) - \lambda_{j+1}i = 0, \quad i = 1, \dots, n, \\ \Phi_{\varepsilon_{n(m-1)+i}} &= 2\varepsilon_{n(m-1)+i} - \lambda_{m-1}(n-i) = 0, \quad j = 1, \dots, m-1; \end{aligned} \quad (20)$$

$$\begin{aligned} \Phi_{\delta_0} &= 2\delta_0 + \lambda_1 = 0, \\ \Phi_{\delta_1} &= 2\delta_1 - 2\lambda_1 + \lambda_2 = 0, \\ \Phi_{\delta_j} &= 2\delta_j + \lambda_{j-1} - 2\lambda_j + \lambda_{j+1} = 0, \quad j = 2, \dots, m-3. \\ \Phi_{\delta_{m-1}} &= 2\delta_{m-1} + \lambda_{m-3} - 2\lambda_{m-2} = 0, \\ \Phi_{\delta_m} &= 2\delta_m + \lambda_{m-2} = 0, \end{aligned} \quad (21)$$

Производные функции Φ по параметрам λ_j составляют систему (19), которую можно записать в виде

$$\Delta + \varphi = E, \quad (22)$$

где при $1 \leq j \leq mn-1$

$$\Delta_j = \delta_{j-1} - 2\delta_j + \delta_{j+1}, \quad E_j = \sum_{i=1}^n i\varepsilon_{n(j-1)+i} + \sum_{i=1}^{n-1} (n-i)\varepsilon_{nj+i}, \quad \varphi_j = f_{j-1} - 2f_j + f_{j+1}.$$

Обращение равенств (20) приводит к выражениям

$$\begin{aligned} \varepsilon_i &= \frac{i\lambda_1}{2}, \quad i = 1, \dots, n, \\ \varepsilon_{nj+i} &= \frac{(n-i)\lambda_j + i\lambda_{j+1}}{2}, \quad i = 1, \dots, n, \quad j = 1, \dots, m-1, \\ \varepsilon_{n(m-1)+i} &= \frac{(n-i)\lambda_{m-1}}{2}, \quad i = 1, \dots, n-1. \end{aligned} \quad (23)$$

Аналогичным образом из (21) выводим выражения

$$\begin{aligned}\delta_0 &= -\frac{\lambda_1}{2}, \\ \delta_1 &= -\frac{-2\lambda_1 + \lambda_2}{2}, \\ \delta_j &= -\frac{\lambda_{j-1} - 2\lambda_j + \lambda_{j+1}}{2}, \quad j = 2, \dots, m-3, \\ \delta_{m-1} &= -\frac{\lambda_{m-2} - 2\lambda_{m-1}}{2}, \\ \delta_m &= -\frac{\lambda_{m-1}}{2},\end{aligned}\tag{24}$$

которые можно представить в виде

$$\delta = -\frac{1}{2}R^*\lambda.\tag{25}$$

Здесь матрица R размера $(m-1) \times (m+1)$ имеет трехдиагональный вид

$$R_2 = \begin{pmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \end{pmatrix}.$$

Отсюда следует, что вектор Δ из (22) можно выразить через λ_j :

$$\Delta = \begin{pmatrix} \delta_0 - 2\delta_1 + \delta_2 \\ \vdots \\ \delta_{m-2} - 2\delta_{m-1} + \delta_m \end{pmatrix} = \begin{pmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \end{pmatrix} \begin{pmatrix} \delta_0 \\ \delta_1 \\ \vdots \\ \delta_m \end{pmatrix} = -\frac{1}{2}R_2R_2^* \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{m-1} \end{pmatrix}.$$

Далее можно показать, что

$$E = \begin{pmatrix} \sum_{i=1}^n i\varepsilon_i + \sum_{i=1}^{n-1} (n-i)\varepsilon_{n+i} \\ \sum_{i=1}^n i\varepsilon_{n+i} + \sum_{i=1}^{n-1} (n-i)\varepsilon_{2n+i} \\ \vdots \\ \sum_{i=1}^n i\varepsilon_{n(m-2)+i} + \sum_{i=1}^{n-1} (n-i)\varepsilon_{n(m-1)+i} \end{pmatrix} = \frac{n}{12}(n^2S + T_2) \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{m-1} \end{pmatrix}.$$

где S и T – симметричные трехдиагональные матрицы

$$S = \begin{pmatrix} 4 & 1 & & & \\ 1 & 4 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & 1 & 4 \end{pmatrix}, \quad T_2 = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & & -1 & 2 \end{pmatrix}.$$

Действительно, из (23) следует

$$\begin{aligned}E &= \frac{1}{2} \begin{pmatrix} \lambda_1 \left(\sum_{i=1}^n i^2 + \sum_{i=1}^{n-1} (n-i)^2 \right) + \lambda_2 \left(\sum_{n=1}^{n-1} i(n-i) \right) \\ \lambda_1 \left(\sum_{n=1}^{n-1} i(n-i) \right) + \lambda_2 \left(\sum_{i=1}^n i^2 + \sum_{i=1}^{n-1} (n-i)^2 \right) + \lambda_3 \left(\sum_{n=1}^{n-1} i(n-i) \right) \\ \vdots \\ \lambda_{m-2} \left(\sum_{n=1}^{n-1} i(n-i) \right) + \lambda_{m-1} \left(\sum_{i=1}^n i^2 + \sum_{i=1}^{n-1} (n-i)^2 \right) \end{pmatrix} = \\ &= \frac{n}{12} \begin{pmatrix} \lambda_1(4n^2 + 2) + \lambda_2(n^2 - 1) \\ \lambda_2(4n^2 + 2) + (\lambda_1 + \lambda_3)(n^2 - 1) \\ \vdots \\ \lambda_{m-1}(4n^2 + 2) + \lambda_{m-2}(n^2 - 1) \end{pmatrix} = \frac{n}{12} \begin{pmatrix} n^2(4\lambda_1 + \lambda_2) + 2\lambda_1 - \lambda_2 \\ n^2(\lambda_1 + 4\lambda_2 + \lambda_3) - \lambda_1 + 2\lambda_1 - \lambda_3 \\ \vdots \\ n^2(4\lambda_{m-1} + \lambda_{m-2}) + 2\lambda_{m-1} - \lambda_{m-2} \end{pmatrix}.\end{aligned}$$

В результате система (22) превращается в СЛАУ

$$\left[\frac{1}{2} R_2 R_2^* + \frac{n}{12} (n^2 S + T_2) \right] \lambda = \varphi,$$

где $\lambda = (\lambda_1, \dots, \lambda_{m-1})^T$. Оценим решение этой системы:

$$\lambda = \frac{12}{n^3} \left[I + \frac{1}{n^2} S^{-1} \left(T_2 + \frac{6}{n} R_2 R_2^* \right) \right]^{-1} S^{-1} \varphi. \quad (26)$$

Простая структура матриц R_2 , S и T_2 позволяет провести следующие оценки:

$$\|T_2\| = 4, \quad \|R_2\| = 4, \quad \|R_2 R_2^*\| \leq 16,$$

$$\lambda_{\min}(S) \geq 4 - 1 - 1 = 2, \quad \|S^{-1}\| = (\lambda_{\min}(S))^{-1} \leq \frac{1}{2}.$$

Введем обозначение

$$\Omega = -\frac{1}{n^2} S^{-1} \left(T_2 + \frac{6}{n} R_2 R_2^* \right),$$

тогда

$$\|\Omega\| \leq \frac{1}{n^2} \|S^{-1}\| \cdot \|T_2 + \frac{6}{n} R_2 R_2^*\| \leq \frac{2}{n^2} \left(1 + \frac{24}{n} \right).$$

Обратим внимание, что при $n \geq 5$

$$\|\Omega\| \leq \frac{1}{2}.$$

Следовательно, при $n \geq 5$ может быть получена также следующая оценка:

$$\| [I + \frac{1}{n^2} S^{-1} (T_2 + \frac{6}{n} R_2 R_2^*)]^{-1} \| = \| (I - \Omega)^{-1} \| \leq 1 + \frac{\|\Omega\|}{1 - \|\Omega\|} \leq 1 + \frac{4}{n^2} \left(1 + \frac{24}{n} \right) \leq 2.$$

Отсюда для (26) и (25) мы получаем

$$\|\lambda\| \leq \frac{12}{n^3} \|\varphi\|, \quad \|\delta\| \leq \frac{24}{n^3} \|\varphi\| \leq \frac{96}{n^3} \|f\|.$$

Таким образом, была доказана теорема о точности описанной аппроксимации.

Теорема 3. Пусть функция $f(x)$ задана в точках (1), матрицы A , B , D_2 имеют вид (8), (4) и (16) соответственно. Если при этом вектор g является обобщенным нормальным решением системы (8), вектор $\tilde{f}$ задан равенством $\tilde{f}_{j+1} = g_{nj+1}$ ($j = 1, \dots, m$), то для векторов f и $\tilde{f}$ верна оценка

$$\|f - \tilde{f}\| \leq \frac{96}{n^3} \|f\|$$

при $n \geq 5$, где $\tilde{f}_j = g(x_j)$.

Таким образом, данная аппроксимация имеет третий порядок по n .

3.1. Гладкость аппроксимации

По определению вектора δ имеет место равенство $\tilde{f} = \delta + f$, где $\tilde{f}_j = g(x_j)$. Умножим обе части этого равенства на R_2 :

$$\|R_2 \tilde{f}\| = \|R_2 \delta + R_2 f\| = \|R_2 f + \frac{1}{2} R_2 R_2^* \Theta R_2 f\| \leq \|I + \frac{1}{2} R_2 R_2^* \Theta\| \|R_2 f\|,$$

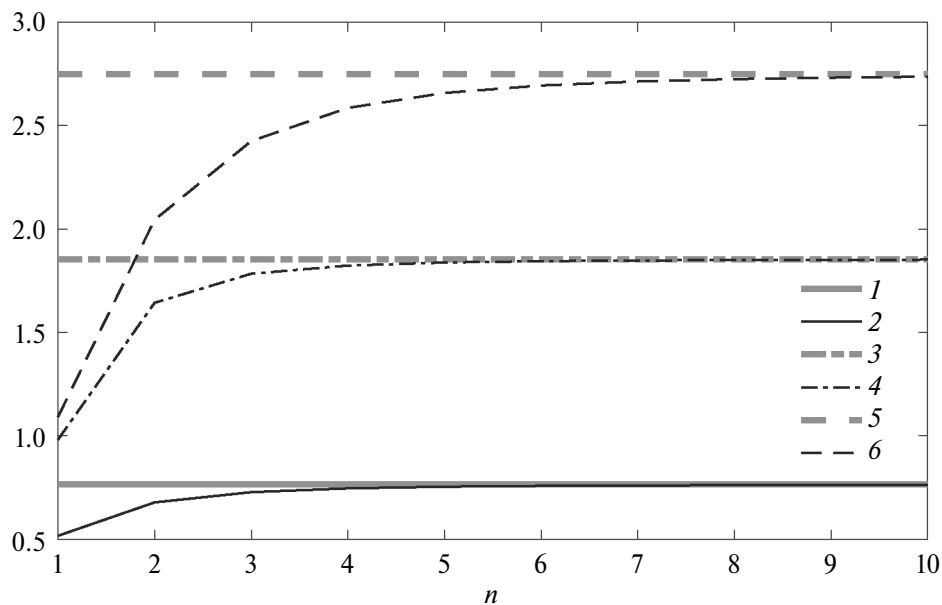
где

$$\Theta = \left(-\frac{1}{2} R_2 R_2^* - \frac{n}{12} (n^2 S + T_2) \right)^{-1}.$$

Используя выкладки предыдущего раздела, получаем оценку

$$\|\Theta\| = \left\| \frac{12}{n^3} (1 - \Omega)^{-1} \right\| \leq \frac{24}{n^3}$$

при $n \geq 5$. Отсюда следует лемма 3.



Фиг. 3. Сравнение $\|R_2 f\|$ и $\|R_2 \tilde{f}\|$, подробное описание см. в тексте.

Лемма 3. Для разностей элементов векторов f и $\tilde{f}$ имеет место неравенство

$$\|R_2 \tilde{f}\| < \left(1 + \frac{16 \cdot 24}{n^3}\right) \|R_2 f\|.$$

Это утверждение значительно более слабое, чем лемма 1, так как оно не показывает уменьшения производной, а значит, сглаживания. Однако проведенные многочисленные численные эксперименты убеждают, что сглаживание все-таки имеет место. На примере фиг. 3 можно видеть сравнение графиков $\|R_2 f\|$ и $\|R_2 \tilde{f}\|$ в зависимости от n при $m = 10$ для функций $\cos(2x)$ (кривые 1, 2 на фиг. 3), $\cos(x) + \sin(3x)$ (кривые 3, 4), $\cos(4x) + \sin(y)$ (кривые 5, 6). Очевидно, что вторые разности самих функций $\|R_2 f\|$ от параметра аппроксимации n не зависят, поэтому им на фиг. 3 соответствуют горизонтальные прямые. Видно, что к ним при растущем n кривые $\|R_2 \tilde{f}\|$ приближаются снизу.

Добавление в матрицу системы параметра η , с помощью которого можно увеличить или уменьшить вес производных,

$$A = \begin{pmatrix} B \\ \eta D_2 \end{pmatrix}$$

изменяет доказанные выше соотношения следующим образом:

$$\|f - \tilde{f}\| \leq \frac{24\eta}{n^3} \left(1 + \frac{2}{n^2} \left(1 + \frac{24\eta}{n}\right)\right) \|f\|.$$

3.2. Обусловленность систем

Для того чтобы оценить обусловленность расширенной матрицы в данном случае, воспользуемся следующими фактами и обозначениями. Обозначим $\tilde{D}_2$ квадратную матрицу размера $(mn+1) \times (mn+1)$, составленную из матрицы D_2 (16) и первой и последней строк матрицы B , и представим ее в блочном

$$\tilde{D}_2 = \begin{pmatrix} 1 & & & & & \\ 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \\ & & & & & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & & & \vdots \\ 0 & \tilde{T} & & 0 \\ \vdots & & & 1 \\ 0 & 0 & \dots & 1 \end{pmatrix}, \quad \|\tilde{D}_2\| \leq 4.$$

При этом имеет место разложение

$$\tilde{T} = - \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & -1 & 0 \\ 0 & \dots & 0 & 1 & -1 \\ 0 & \dots & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & -1 & 0 \\ 0 & \dots & 0 & 1 & -1 \\ 0 & \dots & 0 & 0 & 1 \end{pmatrix}^*.$$

Согласно теореме об окаймлении верна цепочка неравенств

$$\sigma_{\min}(\tilde{T}) \leq \sigma_{\min}(\tilde{D}_2) \leq \sigma_{\min}(A) \leq \sigma_{\max}(\tilde{T}) \leq \sigma_{\max}(\tilde{D}_2).$$

Отсюда следует оценка для числа обусловленности

$$\text{cond}(A) \leq \frac{\sigma_{\max}(A)}{\sigma_{\min}(\tilde{T})} \leq \|A\| \|\tilde{T}^{-1}\| \leq 4(mn+1)^2,$$

которую можно сформулировать в виде леммы.

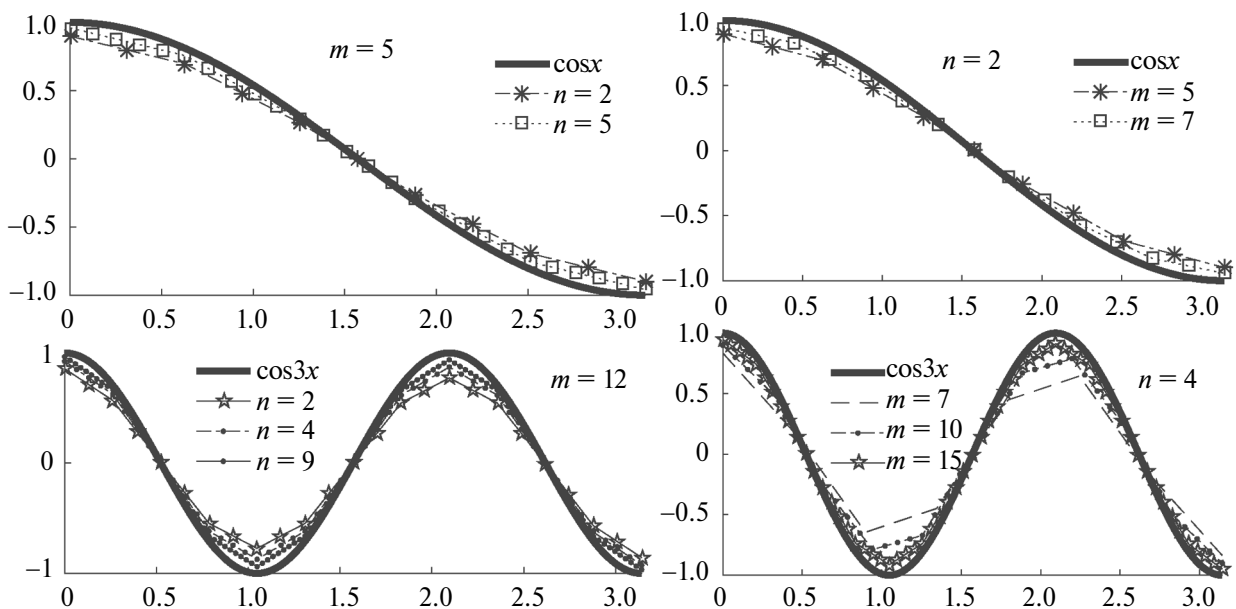
Лемма 4. Пусть матрицы B и D_2 размеров $(m+1) \times (mn+1)$ и $(mn-1) \times (mn+1)$ соответственно заданы формулами (4) и (16). Тогда число обусловленности матрицы

$$A = \begin{pmatrix} B \\ \dots \\ D_2 \end{pmatrix}$$

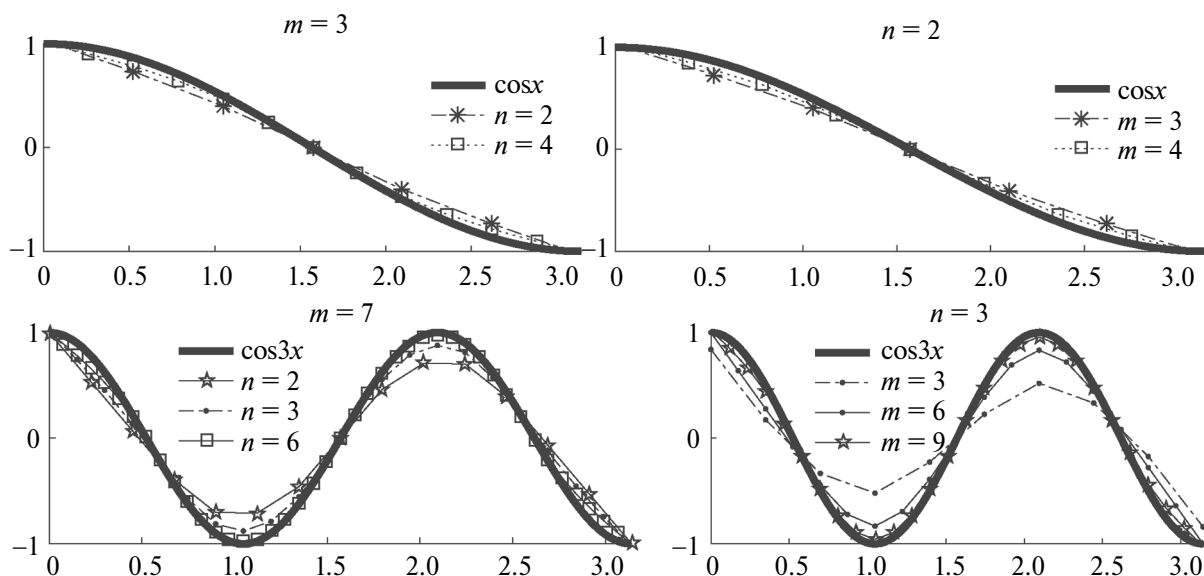
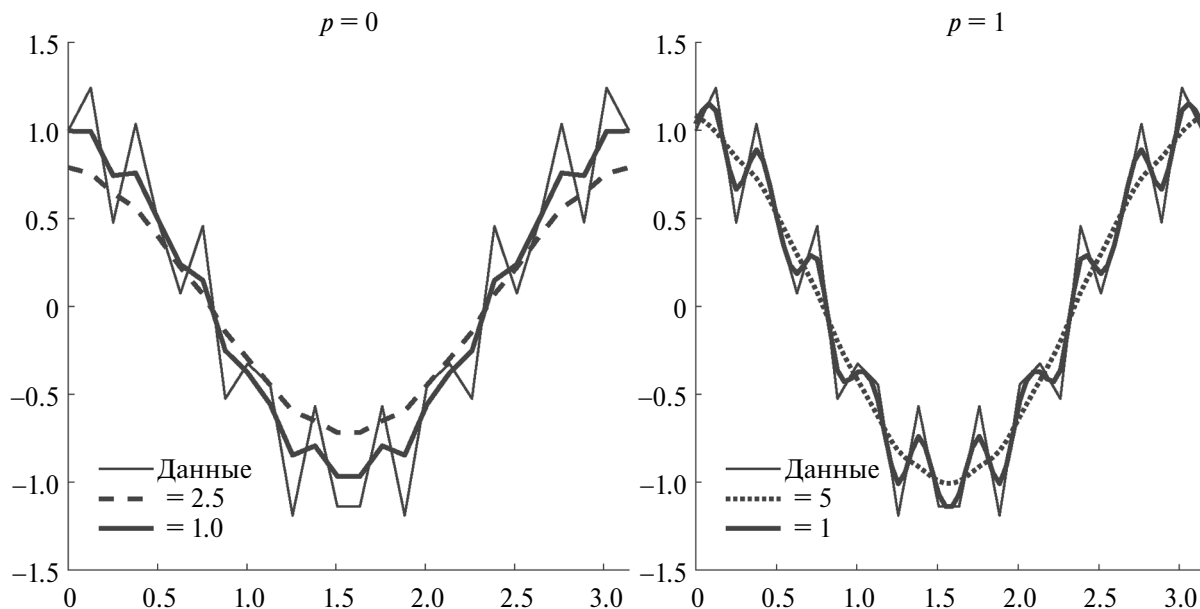
не превосходит $4(mn+1)^2$.

4. РЕЗУЛЬТАТЫ ВЫЧИСЛЕНИЙ

Используем представленный метод для аппроксимации функции $f(x) = \cos kx$ ($k = 1, 3$) на отрезке $[1, 3]$. На фиг. 4, 5 демонстрируется сходимость полученных аппроксимаций в зависимости от параметров m, n , а также от частоты осцилляций функции. На фиг. 6 представлены результаты влияния на сглаживание зашумленных данных весового параметра η . В обоих случаях использованы значения $m = 25, n = 1$.



Фиг. 4. Аппроксимация при $p = 0$.

Фиг. 5. Аппроксимация при $p = 1$.Фиг. 6. Сглаживание при $p = 0$.

ЗАКЛЮЧЕНИЕ

В рамках настоящей работы представлен гибридный метод обработки данных, объединяющий в себе свойства аппроксимации и сглаживания и основанный на оригинальной идее регуляризации плохо обусловленных систем, предложенной С.К. Годуновым. Интересной особенностью метода является то, что значения аппроксимации вычисляются без определения коэффициентов ее разложения по какому-либо базису.

Этот подход может применяться, например, для перехода с крупной сетки на более мелкую при численном решении начально-краевых задач с быстро меняющимися граничными условиями или в задачах, требующих сглаживания. Вероятно, в ходе дальнейших исследований будут выявлены и другие потенциальные возможности, заложенные в представленном методе.

СПИСОК ЛИТЕРАТУРЫ

1. Годунов С.К., Антонов А.Г., Кирилук О.П., Костин В.И. Гарантированная точность решения систем линейных уравнений в евклидовых пространствах. Новосибирск: Наука, 1988. С. 456.

2. Бибердорф Э.А., Попова Н.И. Гарантированная точность современных алгоритмов линейной алгебры. Новосибирск: Изд-во СО РАН, 2006. С. 319.
3. Кабанихин С.И. Обратные и некорректные задачи. Новосибирск: Сиб. науч. изд-во, 2009. С. 458.
4. Годунов С.К. Современные аспекты линейной алгебры. Новосибирск: Науч. книга, 1997. С. 388.

APPROXIMATION AND SMOOTHING OF A FUNCTION BASED ON GODUNOV REGULARIZATION

E. A. Biberdorf^{a,*}, K. K. Abdisheripov^b

^a*Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences, Acad. Koptyug Ave., 4,
Novosibirsk, 630090 Russia*

^b*Novosibirsk State University, Pirogov St., 1, Novosibirsk, 630090 Russia*

*e-mail: math@biberdorf.ru

Received 02 April, 2024

Revised 02 April, 2024

Accepted 02 May, 2024

Abstract. A new approach to function approximation is presented, based on S.K. Godunov's ideas on the regularization of ill-conditioned systems. The proposed method allows for determining function values at nodes of a finer grid from data on a coarser grid while ensuring control over the smoothness of the resulting function. Convergence and smoothness estimates are substantiated, and results from computational experiments illustrate the effectiveness of the proposed method.

Keywords: approximation, function smoothing, regularization of ill-conditioned systems of linear equations.