УДК 518.12+519.34

Сеточный вариант нестандартного тригонометрического базиса и его преимущества относительно аналогичного полиномиального базиса

В.В. Смелов

Институт вычислительной математики и математической геофизики Сибирского отделения Российской академии наук, просп. Акад. М.А. Лаврентьева, 6, Новосибирск, 630090
E-mail: vl.smelov@gmail.com

Смелов В.В. Сеточный вариант нестандартного тригонометрического базиса и его преимущества относительно аналогичного полиномиального базиса // Сиб. журн. вычисл. математики / РАН. Сиб. отд-ние. — Новосибирск, 2014. — Т. 17, № 4. — С. 399–409.

В сеточном варианте предложен основанный на тригонометрии функциональный базис, ориентированный на аппроксимацию с высоким порядком точности гладких и кусочно-гладких функций. Дается сравнительный анализ качеств предложенного и полиномиального базисов. Показано преимущество тригонометрического варианта перед полиномиальным.

Ключевые слова: *функциональный базис, эллиптический оператор, энергетическое скалярное произведение, функционал, обобщенное решение, условие сопряжения.*

Smelov V.V. A network version of the non-standard trigonometric basis and its advantages with respect to a similar polynomial basis // Siberian J. Num. Math. / Sib. Branch of Russ. Acad. of Sci. — Novosibirsk, 2014. — Vol. 17, № 4. — P. 399–409.

A trigonometry-based functional basis as a network version is proposed. It is aimed at the approximation with high orders of accuracy of smooth and piecewise-smooth functions. A comparative analysis of the features of the basis proposed and a polynomial one is made. The trigonometric version offers considerable advantages over the polynomial bases.

Key words: *functional basis, elliptic operator, energy scalar product, functional, generalized solution, conjugation condition.*

Введение

Основное содержание статьи представлено специфическим тригонометрическим базисом в сеточном исполнении, ориентированным на решение практических задач математической физики. Далее этот тригонометрический базис сопоставляется с известным (альтернативным) полиномиальным базисом. В заключение статьи проводится сравнение качеств этих двух базисов и устанавливается преимущество тригонометрического.

В хорошо известной монографии [1, гл. 2, п. 2.4.1] на полиномиальной основе предложен вариант аппроксимации гладких функций с высоким порядком точности. Будем ниже именовать этот сеточный метод как “вариант М”. Уточним, что далее под вариантом М будет пониматься только констатация результатов и фактов из соответствующего раздела монографии [1].

Вариант М заключается в построении базисных функций $\{\varphi_i(x)\}_{i=1}^m$, $m = (p+1)/2$, где p — нечетное положительное число. В канонической форме, т. е. на стандартном отрезке $[-1, 1]$, эти функции подчинены следующим требованиям: $\varphi_i(x) = 0$ при $x \notin [-1, 1]$, на каждом из отрезков $[-1, 0]$, $[0, 1]$ функция $\varphi_i(x)$ есть многочлен степени p, причем в точках $x = -1$ и $x = 1$ функция $\varphi_i(x)$ и все ее производные до $(m-1)$-го порядка включительно равны нулю, а в точке $x = 0$ единственной ненулевой производной является $d^{i-1}\varphi_i(x)/dx^{i-1}\,|_{x=0} = 1$, $1 \le i \le m$.

Обратимся, например, к случаю $p = 3$ ($m = 2$). В этом варианте имеем две базисные функции:

$$\varphi_1(x) = \begin{cases} 0 & \text{при} \quad |x| \ge 1, \\ (1-x)^2(1+2x), & 0 \le x \le 1, \\ \varphi_1(-x) & \text{при} \ -1 \le x \le 0, \end{cases} \qquad \varphi_2(x) = \begin{cases} 0 & \text{при} \quad |x| \ge 1, \\ (1-x)^2 x, & 0 \le x \le 1, \\ -\varphi_2(-x) & \text{при} \ -1 \le x \le 0. \end{cases} \tag{1}$$

На пару любых соседних отрезков $[x_{k-1}, x_k]$ и $[x_k, x_{k+1}]$ равномерной сетки эти функции следует отобразить линейно, и любая линейная комбинация их в сеточном исполнении имеет непрерывную производную.

Аналогичная пара функций на основе предлагаемого нами метода представлена ниже спроектированными на сетку формулами (6) и (11). Будем в дальнейшем именовать данный метод как “вариант С”.

Численные эксперименты по аппроксимации функций с базисными функциями (1) и, соответственно, с (6) и (11) продемонстрировали для обоих методов (на равномерной сетке узлов) практически одинаковые величины погрешностей в норме пространства $C[a, b]$. Вариант М как раз и ориентирован только на равномерную сетку. Отметим, что попытка распространить аппроксимационный метод варианта М на неравномерную сетку выявила бы разрыв первой производной базисной функции $\varphi_2(x)$ в узлах между отрезками разной длины. Учитывая, что все базисные функции варианта М при $p \ge 3$ обладают гладкостью не ниже первого порядка, их использование наиболее целесообразно для аппроксимации гладких функций.

Что касается варианта С, то здесь все результаты справедливы на любой сетке. Если предметом аппроксимации является кусочно-гладкая функция $f(x)$, то на промежутках ее гладкости любая линейная комбинация базисных функций (6) и (11) в узлах любой сетки сохраняет непрерывность производной. Если у функции $f(x)$, например, в узле $x = x_k$ известно отношение левосторонней и правосторонней производных, т. е. $f'(x_k - 0) = \lambda\, f'(x_k + 0)$, то любая комбинация базисных функций (6) и (11) обеспечивает точное выполнение указанного соотношения. Подробное обоснование этих и других качеств варианта С будет представлено в следующем пункте, а в пункте 3 будет более подробно выполнен и сравнительный анализ качеств обоих вариантов.

Итак, в данном частном случае с базисом (1) и, соответственно, с базисом (6) и (11) налицо существенное преимущество последнего.

1. Построение базисных функций варианта С

Представим теперь кратко технику получения основанных на тригонометрии базисных функций, более подробное изложение которой содержится в работах [2, 3]. Этот процесс начнем с одного из результатов в теории тригонометрических рядов Фурье.

Легко устанавливается [2, 3], что коэффициенты a_n и b_n тригонометрического ряда Фурье:

$$f(x) = \sum_{n=0}^{\infty} a_n \cos nx + \sum_{n=1}^{\infty} b_n \sin nx, \quad 0 \le x \le 2\pi, \tag{2}$$

убывают по нижеследующему закону:

$$a_n = \sigma_n^{(1)} n^{-N}, \quad b_n = \sigma_n^{(2)} n^{-N}, \quad \sum_{n=1}^{\infty} (\sigma_n^{(i)})^2 < \infty \quad (i = 1, 2), \tag{3}$$

если разлагаемая функция $f(x)$:
а) периодическая,
б) на любом конечном отрезке числовой оси принадлежит пространству Соболева H^N (иначе W_2^N) [4, 5].

По отношению к остатку ряда

$$R_m(x) = \sum_{n=m+1}^{\infty} (a_n \cos nx + b_n \sin nx) \equiv f(x) - \left[\sum_{n=0}^{m} a_n \cos nx + \sum_{n=1}^{m} b_n \sin nx \right]$$

несложно получить следующую оценку [2]:

$$|R_m(x)| \le \delta_m m^{-N+1/2}, \quad \text{где } \delta_m \to 0 \text{ при } m \to \infty. \tag{4}$$

При $N \gg 1$ асимптотически быстрый закон убывания остатка ряда $R_m(x)$ может обеспечивать хорошую аппроксимацию функции $f(x)$ малым числом членов ряда Фурье. Когда же речь идет о разложении в ряд Фурье функции, определенной лишь на конечном отрезке, то (вне зависимости от ее гладкости на этом отрезке) в общем случае можно гарантировать асимптотическое убывание коэффициентов Фурье только как $O(1/n)$. Однако и в этом случае существует возможность разложения функций в тригонометрические ряды (но уже не в ряды Фурье!) при остатке ряда $r_m(x)$ с таким же законом убывания, как и в оценке (4).

Для обоснования последнего утверждения примем, не нарушая общности, за область определения функции $f(x)$ отрезок $0 \le x \le \pi$. Отметим, что на этом отрезке тригонометрические функции линейно независимы, но уже не будут в совокупности взаимно ортогональными (ортогональными будут порознь синусы и косинусы).

Сейчас сформулируем важную теорему, подробное доказательство которой содержится в публикациях [2, 3].

Теорема 1. *Всякая функция $f(x) \in H^N(0, \pi)$, $N \ge 1$, допускает бесконечно много представлений в виде рядов*:

$$f(x) = \sum_{n=0}^{\infty} a_n \cos nx + \sum_{n=1}^{\infty} b_n \sin nx, \quad 0 \le x \le \pi, \tag{5}$$

сходящихся в норме пространства $C^{N-1}[0, \pi]$. Относительно остатка

$$r_m(x) = \sum_{n=m+1}^{\infty} (a_n \cos nx + b_n \sin nx)$$

каждого из рядов (5) *справедливы оценки*:

$$\left| \frac{d^\nu r_m(x)}{dx^\nu} \right| \le \delta_m^{(\nu)} m^{-N+\nu+1/2}, \quad \nu = 0, 1, \ldots, N-1,$$

где $\lim_{m\to\infty} \delta_m^{(\nu)} = 0$.

Идея доказательства теоремы 1 проста. Скопируем функцию $f(x)$ с отрезка $0 \le x \le \pi$ на отрезок $2\pi \le x \le 3\pi$. На промежуточном отрезке $\pi \le x \le 2\pi$ реализуем такое восполнение недостающего куска функции $f(x)$, чтобы на отрезке $0 \le x \le 3\pi$ построенная функция принадлежала пространству $H^N(0, 3\pi)$. Построение восполняющей функции допускает бесконечное множество вариантов. Это можно сделать, например, посредством полинома достаточно высокой степени. Еще один из вариантов восполнения приведен в книге [2]. Построенная функция может быть теперь продолжена периодически на всю числовую ось с периодом 2π и, в итоге, будут справедливы утверждения (3) и (4) и на всей числовой оси, и на отрезке $0 \le x \le \pi$, что и сформулировано в теореме 1.

Из теоремы 1 следует практический вывод: функции$f(x)$ высокой степени гладкости ($N \gg 1$) на отрезке $0 \le x \le \pi$ могут быть с хорошей точностью равномерно аппроксимированы (вместе с некоторым количеством производных) малым числом слагаемых в приближенном представлении

$$f(x) \approx T(x) \equiv \sum_{n=0}^{m} a_n \cos nx + \sum_{n=1}^{m} b_n \sin nx, \quad 0 \le x \le \pi.$$

Перед предстоящим начальным процессом построения базиса из *непрерывных* гладких и кусочно-гладких функций целесообразно дать предварительные пояснения. Функции $\sin nx$ на концах отрезка $0 \le x \le \pi$ обращаются в нуль и, будучи продолженными нулем вне этого отрезка, будут представлять собой непрерывные на всей числовой оси функции с конечным носителем. По отношению к функциям $\cos nx$ такое непрерывное продолжение невозможно, но оно оказывается возможным при определенных комбинациях этих функций, как это показано ниже.

Перейдем теперь к подробному процессу построения сеточных базисных функций. Пусть на отрезок $[a, b]$ нанесена сетка узлов $x_0 < x_1 < \cdots < x_r$, где $x_0 = a$ и $x_r = b$. Шаг этой сетки обозначим через $h_k = x_k - x_{k-1}$, $k = 1, 2, \ldots, r$. Для всего отрезка $x_0 \le x \le x_r$ кусочно-гладкий базис непрерывных функций с конечными носителями будем комплектовать на основе используемых в равенстве (2) тригонометрических базисных функций. Последние линейно отобразим на каждый из отрезков $[x_{k-1}, x_k]$ и с учетом вышеприведенных пояснений сформируем следующие непрерывные базисные функции с конечными носителями (на каждом из отрезков $[x_{k-1}, x_k]$ равнозначные базисным функциям равенства (2)):

$$\Psi_n^{(k)}(x) = \begin{cases} \sin[\pi n h_k^{-1}(x - x_{k-1})], & x \in [x_{k-1}, x_k], \\ 0, & x \notin [x_{k-1}, x_k], \end{cases} \quad n \ge 1,$$

$$\Phi_n^{(k)}(x) = \begin{cases} \cos[\pi(n+1)h_k^{-1}(x-x_{k-1})] - \cos[\pi(n-1)h_k^{-1}(x-x_{k-1})], & x \in [x'_{k-1}x_k], \\ 0, & x \notin [x'_{k-1}x_k], \end{cases} \quad n \ge 1,$$

$$k = 1, 2, \ldots, r,$$

$$\Lambda_k(x) = \begin{cases} 1 - \cos[\pi h_k^{-1}(x - x_{k-1})], & x \in [x_{k-1}, x_k], \\ 1 + \cos[\pi h_{k+1}^{-1}(x - x_k)], & x \in [x_k, x_{k+1}], \\ 0, & x \notin [x_{k-1}, x_{k+1}]. \end{cases} \quad k = 0, 1, \ldots, r, \tag{6}$$

В последнем равенстве исключаются те составляющие функций $\Lambda_0(x)$ и $\Lambda_r(x)$, которые отвечают фиктивным отрезкам $[x_{-1}, x_0]$ и $[x_r, x_{r+1}]$. Функции $\Psi_n^{(k)}(x)$ и $\Phi_n^{(k)}(x)$ в

зависимости от четности/нечетности нижнего индекса будут нечетными/четными относительно точки $0.5\,(x_{k-1}+x_k)$.

Пусть рассматриваемая непрерывная функция $f(x)$ на каждом отрезке $[x_{k-1},x_k]$ удовлетворяет включению в H^{N_k} $(N_k\ge 1)$. Представим ее в виде

$$f(x)=\sum_{k=1}^{r}\left\{\sum_{n=1}^{\infty}b_n^{(k)}\Psi_n^{(k)}(x)+\sum_{n=1}^{\infty}c_n^{(k)}\Phi_n^{(k)}(x)\right\}+\sum_{k=0}^{r}a_k\Lambda_k(x),\quad x_0\le x\le x_r. \tag{7}$$

В работах [2, 3] доказана следующая теорема.

Теорема 2. *Существует бесконечно много представлений $f(x)$ в виде рядов* (7) *таких, что на каждом отрезке $[x_{k-1},x_k]$:*
1) *ряд* (7) *сходится в норме пространства $C^{N_k-1}[x_{k-1},x_k]$,*
2) *относительно остатка ряда* (7):

$$R_{n_k}(x)=\sum_{n=n_k+1}^{\infty}[b_n^{(k)}\Psi_n^{(k)}(x)+c_n^{(k)}\Phi_n^{(k)}(x)] \tag{8}$$

справедливы оценки

$$\left|\frac{d^\nu R_{n_k}(x)}{dx^\nu}\right|\le\delta_{n_k}^{(\nu)}\,n_k^{-N_k+\nu+1/2},\quad \nu=0,1,\ldots,N_k-1, \tag{9}$$

где $\lim_{n_k\to\infty}\delta_{n_k}^{(\nu)}=0$.

Замечание 1. Выражение (8) для остатка ряда предполагает, что частичная сумма ряда (7) на разных отрезках гладкости может определяться разным числом слагаемых.

Если функцию $f(x)$ приближенно представить конечным отрезком ряда (7):

$$f(x)\approx T(x)\equiv\sum_{k=1}^{r}\left\{\sum_{n=1}^{n_k}b_n^{(k)}\Psi_n^{(k)}(x)+\sum_{n=1}^{n_k}c_n^{(k)}\Phi_n^{(k)}(x)\right\}+\sum_{k=0}^{r}a_k\Lambda_k(x),\quad a\le x\le b, \tag{10}$$

то результатом будет кусочно-гладкая функция, в общем случае, с разрывами первой производной в узлах сетки. Отметим, что в силу оценок (9) величина этих разрывов в местах непрерывности $f'(x)$ стремится к нулю с увеличением n_k в равенстве (10).

Нарушения гладкости функции $T(x)$ обусловлены только функциями $\Psi_n^{(k)}(x)$, так как производные базисных функций $\Phi_n^{(k)}(x)$ и $\Lambda_k(x)$ во всех узлах сетки равны нулю, т. е. непрерывны. Устранить данные нарушения гладкости $T(x)$ удается тождественными преобразованиями базисных элементов $\Psi_n^{(k)}(x)$. Результаты (вместо $\Psi_n^{(k)}(x)$) представлены ниже

$$\Pi_n^{(k)}(x)=\begin{cases}\dfrac{h_k}{n+1}\sin\left[\pi(n+1)\dfrac{x-x_{k-1}}{h_k}\right]-\dfrac{h_k}{n-1}\sin\left[\pi(n-1)\dfrac{x-x_{k-1}}{h_k}\right], & x\in[x_{k-1},x_k],\\ 0, & x\notin[x_{k-1},x_k],\end{cases}$$

$$k=1,2,\ldots,r,\quad n\ge 2,$$

$$V_k(x) = \begin{cases} \dfrac{h_k}{p_k}\left(\sin\dfrac{2\pi(x-x_{k-1})}{h_k} - 2\sin\dfrac{\pi(x-x_{k-1})}{h_k}\right), & x \in [x_{k-1}, x_k], \\ \dfrac{h_{k+1}}{p_{k+1}}\left(\sin\dfrac{2\pi(x-x_k)}{h_{k+1}} + 2\sin\dfrac{\pi(x-x_k)}{h_{k+1}}\right), & x \in [x_k, x_{k+1}], \\ 0, & x \notin [x_{k-1}, x_{k+1}], \end{cases} \tag{11}$$

$$k = 0, 1, \ldots, r.$$

Относительно параметров p_k и p_{k+1} пояснения будут даны далее.

В равенстве (11) исключаются те составляющие функций $V_0(x)$ и $V_r(x)$, которые отвечают фиктивным отрезкам $[x_{-1}, x_0]$ и $[x_r, x_{r+1}]$. Функции $\Pi_n^{(k)}(x)$ в зависимости от четности/нечетности нижнего индекса будут четными/нечетными относительно точки $0.5\,(x_{k-1} + x_k)$. Нелишне отметить, что справедливо тождество $d\Pi_n^{(k)}(x)/dx \equiv \Phi_n^{(k)}(x)$. Оно может оказаться полезным при вычислениях энергетических скалярных произведений (16), где присутствуют производные.

Объединенную систему функций $U^0 = \Phi_n^{(k)}(x) \cup \Lambda_k(x) \cup \Pi_n^{(k)}(x) \cup V_k(x)$, $U^0 \subset H^1(x_0, x_r)$, примем за окончательный вариант базиса, сохраняющий все качества, представленные теоремой 2. Пространство всех линейных комбинаций функций из U^0 обозначим через $U \subset H^1\,(x_0, x_r)$.

Представим себе, что построенный базис U^0 предполагается использовать для решения задач в многослойных средах с одномерными стационарными уравнениями теплопроводности, диффузии газов или частиц и др. В дифференциальном операторе этих задач

$$A \equiv -\frac{d}{dx}p(x)\frac{d}{dx} + q(x), \quad p(x) \geq p_0 > 0, \quad q(x) \geq 0, \tag{12}$$

коэффициенты $p(x)$ и $q(x)$ будут иметь разрывы на границах слоев с разными физическими свойствами. Если это случилось в узле $x = x_k$ избранной сетки, то решение $\varphi(x)$ в упомянутых задачах имеет реальный смысл только при выполнении следующего условия сопряжения [6]:

$$p_k\varphi'(x_k - 0) = p_{k+1}\varphi'(x_k + 0), \tag{13}$$

где

$$p_k = p(x_k - 0), \qquad p_{k+1} = p(x_k + 0),$$

$$p(x_k \pm 0) = \lim_{\varepsilon\to 0} p(x_k \pm \varepsilon), \qquad \varphi'(x_k \pm 0) = \lim_{\varepsilon\to 0} \varphi'(x_k \pm \varepsilon) \ \ (\varepsilon \geq 0).$$

Условия (13) при решении приведенной ниже краевой задачи будут *естественными* [1, 7]. Это означает, что при любой полной базисной системе эти условия в искомом решении будут автоматически удовлетворены только в пределе, т. е. при $n_k \to \infty$ в нижеприведенном приближении (14). А теперь отметим, что при построении базиса U^0 выявилось благоприятное стечение обстоятельств, когда условия сопряжения (13) оказываются выполненными на любом приближении.

Итак, в базисе U^0 функции $\Phi_n^{(k)}(x)$, $\Pi_n^{(k)}(x)$ и $\Lambda_k(x)$ во всех узлах x_k разрыва коэффициентов оператора A уже удовлетворяют требованиям (13) из-за обращения в нуль производных этих функций в узлах x_k.

Если функция $V_k(x)$ отвечает данному проблемному узлу x_k и если в ней параметры p_k и p_{k+1} отождествить с соответствующими параметрами в равенстве (13), то функция $V_k(x)$ будет удовлетворять условию сопряжения (13). В результате все функции базиса U^0 будут удовлетворять условиям сопряжения равно, как и любая линейная комбинация

$$\varphi(x)=\sum_{k=1}^{r}\left(\sum_{n=1}^{n_k}c_n^{(k)}\Phi_n^{(k)}(x)+\sum_{n=2}^{n_k}b_n^{(k)}\Pi_n^{(k)}(x)\right)+\sum_{k=0}^{r}\left(d_kV_k(x)+a_k\Lambda_k(x)\right). \qquad (14)$$

Везде, вне проблемных узлов, функция $\varphi(x)$ в равенстве (14) будет иметь непрерывную производную, если в равенствах (11) в регулярных узлах x_k принять $p_k = p_{k+1} = 1$. В дальнейшем сформулированные условия относительно параметров p_k и p_{k+1} будут предполагаться выполненными.

Замечание 2. Обратим внимание на случай, когда аппроксимируемая функция $\varphi(x)$ в решаемой задаче в некоторых точках имеет разрыв производной, но отношения между односторонними ее производными типа (13) неизвестны. В этом случае функции $V_k(x)$ в данных точках с любыми фиксированными параметрами p_k и p_{k+1} оказываются некорректными. Выход из этой ситуации очень простой. Пусть $x = x_k$ — вышеописанный проблемный узел. Ту часть функции $V_k(x)$, которая отвечает отрезку $[x_{k-1}, x_k]$, обозначим через $V_k^{(1)}(x)$, а вторую часть — через $V_k^{(2)}(x)$. Естественно при этом принять $p_k = p_{k+1} = 1$. В результате в приближенном представлении функции $\varphi(x)$ в виде (14) вместо слагаемого $d_kV_k(x)$ следует использовать сумму $d_k^{(1)}V_k^{(1)}(x)+d_k^{(2)}V_k^{(2)}(x)$ с искомыми коэффициентами $d_k^{(1)}$ и $d_k^{(2)}$. Все положительные качества нового базиса будут такими же, как и в исходном базисе.

2. Пример решения краевой задачи на основе варианта C

Как уже было объявлено выше, построенный базис U^0 находит естественное применение в вышеупомянутых задачах с оператором (12) при разрывных его коэффициентах. В качестве примера рассмотрим *третью* краевую задачу [4]:

$$A\varphi=f, \qquad \varphi'(x_0)-\eta_0\varphi(x_0)=0, \qquad \varphi'(x_r)+\eta_1\varphi(x_r)=0, \quad \eta_i>0, \ \ i=0,1. \qquad (15)$$

В области определения $D(A)$ оператора A при любых функциях $\varphi(x),\ \psi(x)\in D(A)$ будет справедлив результат

$$(A\varphi,\psi)=\eta_0p(x_0)\varphi(x_0)\psi(x_0)+\eta_1p(x_r)\varphi(x_r)\psi(x_r)+\int_{x_0}^{x_r}\left(p\varphi'\psi'+q\varphi\psi\right)dx, \qquad (16)$$

где (α,β) — символ скалярного произведения функций $\alpha(x)$ и $\beta(x)$ в пространстве $L_2(x_0,x_r)$. Из этого результата следует, что оператор A самосопряжен и положительно определен.

Для дальнейших целей правую часть результата (16) определим как *энергетическое скалярное произведение* $(\varphi,\psi)_A$ функций $\varphi(x)$ и $\psi(x)$. Возможное наличие разрывов производных в решении $\varphi(x)$ задачи (15) диктует поиск приближенного решения ее, например, в виде функции $\varphi\in H^1(x_0,x_r)$, минимизирующей следующий функционал [7]:

$$J(\varphi)=\eta_0p(x_0)\varphi^2(x_0)+\eta_1p(x_r)\varphi^2(x_r)+\int_{x_0}^{x_r}\left(p\varphi'^2+q\varphi^2\right)dx-2\int_{x_0}^{x_r}f\varphi\,dx. \qquad (17)$$

Расширение области определения функционала $J(\varphi)$ с $D(A)$ до $H^1(x_0,x_r)$ допустимо ввиду того, что граничные условия задачи (15) являются естественными [7] и не обязаны быть удовлетворенными в проблеме минимизации функционала $J(\varphi)$. В решении

этой задачи методом последовательных приближений (например, методом Ритца [7]) эти условия автоматически будут реализованы в пределе.

В скалярном произведении (16) и функционале (17) гладкость функций понижена по сравнению с гладкостью классического решения задачи (15). В связи с этим решение задачи (15) на основе минимизации функционала (17) объявляется ***обобщенным*** решением.

Искомое обобщенное решение $\varphi(x)$ приближенно представим в виде конечной суммы (14). Система уравнений для поиска этого решения получается умножением правой части равенства (14) скалярно (в энергетическом определении) на каждую базисную функцию с приравниваем результата скалярному произведению (уже в пространстве $L_2(x_0, x_r)$) функции $f(x)$ на соответствующую базисную функцию [7].

Относительно прямого решения полученной таким образом системы уравнений следует указать на следующую особенность ее. Неортогональность функций $\sin nx$ и $\cos nx$ на отрезке $0 \le x \le \pi$ влечет ограничение числа сформированных базисных функций U^0 при практическом их использовании (быстро ухудшается обусловленность матрицы Грама системы уравнений). Кроме того, алгоритм решения оказывается весьма непростым.

Как будет показано ниже, предложенный итерационный процесс устраняет все трудности прямого решения получаемой системы уравнений и является устойчивым при любом числе базисных функций.

Приступим теперь к описанию предлагаемого итерационного процесса. Пусть выполнено s итерационных шагов, т. е. найдено приближенное решение в виде (14). На первом "четвертьшаге" $(s+1)$-го итерационного шага представим $\varphi(x)$ в виде:

$$\varphi(x) \approx \sum_{k=1}^{r} \left(\sum_{n=1}^{n_k} c_n^{(k)} \Phi_n^{(k)}(x) + \sum_{n=2}^{n_k} b_n^{(k)} \Pi_n^{(k)}(x) \right) + \sum_{k=0}^{r} (d_k V_k(x) + A_k \Lambda_k(x)),$$

где новые искомые коэффициенты A_k будем определять на основе требования $\min_{A_k} J(\varphi)$, т. е. из следующей системы уравнений:

$$\sum_{j=m-1}^{m+1} A_j(\Lambda_j, \Lambda_m)_A = (f, \Lambda_m) - \sum_{k=m}^{m+1} \left(\sum_{n=1}^{n_k} c_n^{(k)} \left(\Phi_n^{(k)}, \Lambda_m \right)_A + \sum_{n=2}^{n_k} b_n^{(k)} \left(\Pi_n^{(k)}, \Lambda_m \right)_A \right) - \sum_{j=m-1}^{m+1} d_j (V_j, \Lambda_m)_A, \qquad m = 0, 1, \ldots, r.$$

Реализованный процесс минимизации функционала уменьшил (не увеличил) его величину по сравнению с предыдущим его значением. Аналогичная ситуация будет и при всех следующих этапах итерационного процесса. Такая монотонно убывающая последовательность функционалов ограничена снизу, т. е. имеет предел [8]. Это доказывает сходимость итерационного алгоритма.

Матрица приведенной системы является трехдиагональной с диагональным преобладанием, и ее решение легко реализуется методом прогонки [9].

Далее в обновленном равенстве (14) ($a_k \to A_k$) будем искать новые коэффициенты D_k вместо прежних d_k. Приходим к системам уравнений:

$$\sum_{j=m-1}^{m+1} D_j(V_j, V_m)_A = (f, V_m) - \sum_{k=m}^{m+1} \left(\sum_{n=1}^{n_k} c_n^{(k)} \left(\Phi_n^{(k)}, V_m \right)_A + \sum_{n=2}^{n_k} b_n^{(k)} \left(\Pi_n^{(k)}, V_m \right)_A \right) - \sum_{j=m-1}^{m+1} A_j(\Lambda_j, V_m)_A, \qquad m = 0, 1, \ldots, r. \tag{18}$$

Матрица этой системы и алгоритм ее решения точно такие же, как и в предыдущей системе.

Замечание 3. В рамках случая, описанного в замечании 2, решение системы (18) методом прогонки выполняется изолированно от точки x_0 до первой особой ("проблемной") точки, далее между каждыми двумя особыми точками и, наконец, между последней особой точкой и концом отрезка x_r.

Следующие системы уравнений будут такими:

$$\sum_{n=1}^{n_l} C_n^{(l)}(\Phi_n^{(l)}, \Phi_m^{(l)})_A = (f, \Phi_m^{(l)}) - \sum_{n=2}^{n_l} b_n^{(l)}(\Pi_n^{(l)}, \Phi_m^{(l)})_A - \sum_{k=l-1}^{l} \left[A_k(\Lambda_k, \Phi_m^{(l)})_A + D_k(V_k, \Phi_m^{(l)})_A \right], \tag{19}$$

$$l = 1, 2, \ldots, r, \quad m = 1, 2, \ldots, n_l,$$

$$\sum_{n=2}^{n_l} B_n^{(l)}(\Pi_n^{(l)}, \Pi_m^{(l)})_A = (f, \Pi_m^{(l)}) - \sum_{n=2}^{n_l} C_n^{(l)}(\Phi_n^{(l)}, \Pi_m^{(l)})_A - \sum_{k=l-1}^{l} \left[A_k(\Lambda_k, \Pi_m^{(l)})_A + D_k(V_k, \Pi_m^{(l)})_A \right], \tag{20}$$

$$l = 1, 2, \ldots, r, \quad m = 2, 3, \ldots, n_l.$$

К данным системам уравнений следует дать ряд пояснений и, при одинаковой их структуре, достаточно ограничиться, например системой уравнений (19). В системе (19) при каждом значении l (т. е. "в привязке" к каждому отрезку сетки) эта система представляет собой независимую подсистему из изолированных уравнений. Более того, каждая из этих подсистем расщепляется еще на пару независимых систем по четности/нечетности индексов n и m в левой части равенств (19). У каждой из этих изолированных систем уравнений матрица коэффициентов является трехдиагональной с диагональным преобладанием. В итоге, все системы уравнений итерационного процесса реализуются одним и тем же простым алгоритмом прогонки, и ограничения общего числа базисных функций нет. Уместно, наконец, отметить, что при решении систем (19) и (20) естественным образом применим метод распараллеливания этих систем.

Замечание 4. Увеличение числа базисных функций $\Phi_n^{(k)}(x)$ и $\Pi_n^{(k)}(x)$ на каждом из интервалов сетки позволяет уменьшать число этих интервалов с достижением одинаково желаемого аппроксимационного результата. Погрешность этого результата оценивается в соответствие с неравенством (9) теоремы 2 величиной $\delta_{n_k} n_k^{-N_k+1/2}$, где N_k — порядок гладкости решения задачи (15) на отрезке гладкости $[x_{k-1}, x_k]$, а n_k — числовой параметр в приближенном решении задачи вида (14).[1)]

[1)] Если на отрезке $[x_{k-1}, x_k]$ решение задачи имеет бесконечную гладкость, то за N_k можно принять любое число, и этот "произвол" регулируется коэффициентом $\delta_{n_k} = \delta_{n_k}(N_k)$.

3. Сравнительный анализ вариантов М и С

1) В варианте М предусмотрена только равномерная сетка узлов.

В варианте С реализована аппроксимация гладких и кусочно-гладких функций при произвольной сетке узлов. Если в местах разрыва производной аппроксимируемой функции известны отношения левосторонней и правосторонней производных (например, условия сопряжения в задачах с эллиптическим оператором), то любая комбинация базисных функций строго удовлетворяет этим отношениям.

2) В варианте М при увеличении числа базисных функций (на фиксированной сетке узлов) предыдущий их набор "теряется" и новый их комплект формируется на основе достаточно трудоемкого метода.

В варианте С при увеличении числа базисных функций (на фиксированной сетке узлов) к предыдущему их набору просто добавляются новые базисные функции, аналитические формулы для которых одни и те же и достаточно просты.

3) В варианте М носители базисных функций пересекаются, т. е. эти функции не изолированы одна от другой на парах соседних отрезков. Это обстоятельство серьезно осложняет решения практических задач.

В варианте С пересекаются носители только базисных функций (6) и (11). Носители функций $\Phi_n^{(k)}(x)$ и $\Pi_n^{(k)}(x)$ не пересекаются, т. е. эти функции изолированы от отрезка к отрезку. Последний факт существенно облегчает процедуру решения сформированных систем уравнений.

4) В варианте М достаточно просто (например, итерационным методом) решаются задачи только на основе базисных функций, представленных в канонической форме формулами (1). При бóльшем числе базисных функций решение задач становится проблематичным.

В варианте С при любом количестве базисных функций очень простым является итерационный процесс решения задач с использованием только алгоритма прогонки. Здесь системы уравнений представлены только трехдиагональными матрицами.

5)В **варианте М** опущено обоснование результатов — дана только формулировка алгоритмов.

В **варианте С** все результаты строго обоснованы теоремами 1 и 2.

Из приведенного сравнения двух сеточных методов следует очевидное преимущество варианта С.

Дополнение

Предлагаемые вариантом С базисные функции применимы при решениях и других задач, где в уравнениях наличествует эллиптическое слагаемое. Такими уравнениями являются, например зависящее от времени уравнение теплопроводности [6], уравнение диффузии нейтронов [10–12] и др., в которых по пространственной переменной коэффициенты претерпевают разрывы.

Представим в качестве примера задачу с уравнением теплопроводности при однородных граничных условиях произвольного вида. Эту задачу рассмотрим в варианте разностной по времени неявной схемы

$$\frac{\varphi^j(x)-\varphi^{j-1}(x)}{\tau_j}-\frac{d}{dx}\left(p(x)\frac{d\varphi^j(x)}{dx}\right)+q(x)\varphi^j(x)=f^j(x)\,,$$

где $\varphi^0(x)=g(x)$, $\tau_j=t_j-t_{j-1}$, $\varphi^j(x)=\varphi(x,t_j)$. Если переписать это уравнение в виде

$$-\frac{d}{dx}\left(p(x)\frac{d\varphi^j(x)}{dx}\right)+\left(q(x)+\frac{1}{\tau_j}\right)\varphi^j(x)=f^j(x)\ +\frac{1}{\tau_j}\varphi^{j-1}(x),$$

то оно полностью отвечает первому равенству в задаче типа (15). При реализации итерационного алгоритма на временном узле с номером j в качестве начального приближения вида (14) естественно принять функцию $\varphi^{j-1}(x)$, что будет существенно ускорять итерационный процесс на каждом временно́м шаге.

Для решения, например, двумерных задач с эллиптическим оператором методом расщепления [1, 13] операции по каждой переменной можно реализовать с применением базисных функций из варианта С.

Литература

1. **Марчук Г.И.** Методы вычислительной математики. — М.: Наука, 1981.
2. **Смелов В.В.** Задачи Штурма–Лиувилля и разложения функций в быстросходящиеся ряды. — Новосибирск: Изд-во СО РАН, 2000.
3. **Смелов В.В.** О представлении кусочно-гладких функций быстросходящимися тригонометрическими рядами // Сиб. журн. вычисл. математики / РАН. Сиб. отд-ние. — Новосибирск, 1999. — Т. 2, № 4. — С. 385–394.
4. **Михайлов В.П.** Дифференциальные уравнения в частных производных. — М.: Наука, 1983.
5. **Люстерник Л.А., Соболев В.И.** Элементы функционального анализа. — М.: Наука, 1965.
6. **Тихонов А.Н., Самарский А.А.** Уравнения математической физики. — М.: Наука, 1966.
7. **Михлин С.Г.** Вариационные методы в математической физике. — М.: Наука, 1970.
8. **Фихтенгольц Г.М.** Курс дифференциального и интегрального исчисления. Т. 1. — М.-Л.: ГИТТЛ, 1947.
9. **Самарский А.А.** Введение в теорию разностных схем. — М.: Наука, 1971.
10. **Глесстон С., Эдлунд М.** Основы теории ядерных реакторов. — М.: Изд-во иностранной литературы, 1954.
11. **Марчук Г.И., Лебедев В.И.** Численные методы в теории переноса нейтронов. — М.: Атомиздат, 1981.
12. **Смелов В.В.** Лекции по теории переноса нейтронов. — М.: Атомиздат, 1978.
13. **Смелов В.В.** Об обобщенном решении двумерной эллиптической задачи с кусочно-постоянными коэффициентами на основе расщепления дифференциального оператора и использования специфических базисных функций // Сиб. журн. вычисл. математики / РАН. Сиб. отд-ние. — Новосибирск, 2003. — Т. 6, № 1. — С. 59–72.

Поступила в редакцию 16 апреля 2013 г.,
в окончательном варианте 10 февраля 2014 г.