Численное решение разностных краевых задач с помощью разностных задач эволюционного типа

В. Е. Шаманский

В настоящей статье изложен метод численного решения систем разностных уравнений, получающихся из уравнений в частных производных эллиптического типа, который оказался довольно эффективным в применении на электронных цифровых вычислительных машинах (ЭЦВМ). При построении метода использованы идеи Хоттелинга и Бодевига [1].

Будем рассматривать метод на примере первой краевой задачи для уравнения

$$\frac{\partial}{\partial x}\left(a\frac{\partial u}{\partial x}\right)+\frac{\partial}{\partial y}\left(b\frac{\partial u}{\partial y}\right)-cu=f, \tag{1}$$

где a, b, c, f — достаточно гладкие функции переменных x, y, причем $a>0$, $b>0$, $c\geqslant 0$ в некоторой конечной области D с границей Γ. Покроем плоскость xoy квадратной сеткой, образованной двумя рядами прямых, параллельных координатным осям. Ломаной линией, состоящей из отрезков этих прямых, аппроксимируем границу Γ. Множество узлов сетки, попавшее в область, ограниченную ломаной, будем называть сеточной областью и обозначать через D_h. Множество узлов, лежащих на ломаной линии, назовем границей этой сеточной области и обозначим через Γ_h.

Под строкой или столбцом сеточной области D_h будем понимать множество узлов этой области, принадлежащее прямой, параллельной соответственно оси ox или oy. Занумеруем строки снизу вверх от 1 до m. Первая строка области D_h состоит из узлов, соседних к нижним граничным узлам. Узлы в строках будем нумеровать двумя индексами $1, k$; $2, k$; . . .; N_k, k, где второй индекс означает номер строки. Пусть h — шаг сетки. Разностное уравнение, аппроксимирующее уравнение (1) в узле сетки с координатами x_i, y_i, возьмем в следующем виде:

$$\begin{gathered}a\left(x_i+\frac{h}{2}, y_j\right)u(x_i+h, y_j)+a\left(x_i-\frac{h}{2}, y_j\right)u(x_i-h, y_j)+\\+b\left(x_i, y_j+\frac{h}{2}\right)u(x_i, y_j+h)+b\left(x_i, y_j-\frac{h}{2}\right)u(x_i, y_j-h)-\\-\left[a\left(x_i+\frac{h}{2}, y_j\right)+a\left(x_i-\frac{h}{2}, y_j\right)+b\left(x_i, y_j+\frac{h}{2}\right)+\right.\\\left.+b\left(x_i, y_j-\frac{h}{2}\right)+h^2c(x_i, y_j)\right]u(x_i, y_j)=h^2f(x_i, y_j).\end{gathered} \tag{2}$$

Обозначим:

$$a_{\min} = \min_{D_h} a\left(x_i \pm \frac{h}{2},\ y_j\right), \quad a_{\max} = \max_{D_h} a\left(x_i \pm \frac{h}{2},\ y_j\right);$$

$$b_{\min} = \min_{D_h} b\left(x_i,\ y_j \pm \frac{h}{2}\right), \quad b_{\max} = \max_{D_h} b\left(x_i,\ y_j \pm \frac{h}{2}\right);$$

$$c_{\min} = \min_{D_h} c\,(x_i,\ y_j), \qquad c_{\max} = \max_{D_h} c\,(x_i,\ y_j);$$

u_{ij}, f_{ij} — значения функций $u(x,\ y)$, $f(x,\ y)$ в узле с индексами i, j. Предположим, что $a_{\min} > 0$, $b_{\min} > 0$, $c_{\min} \geqslant 0$.

Введем векторы:

$$u_k = \begin{pmatrix} u_{1k} \\ u_{2k} \\ \cdot \\ \cdot \\ \cdot \\ u_{N_k,k} \end{pmatrix}, \quad f_k = \begin{pmatrix} h^2 f_{1k} \\ h^2 f_{2k} \\ \cdot \\ \cdot \\ \cdot \\ h^2 f_{N_k,k} \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \cdot \\ \cdot \\ \cdot \\ u_m \end{pmatrix}, \quad f = \begin{pmatrix} f_1 \\ f_2 \\ \cdot \\ \cdot \\ \cdot \\ f_m \end{pmatrix},$$

т. е. векторы u_k, f_k составлены из значений соответствующих функций в узлах k-й строки, а векторы u, f — из значений этих же функций в узлах области D_h. Тогда систему разностных уравнений типа (2), записанных для всех узлов k-й строки, можно представить в векторном виде так:

$$A_{k,k-1}u_{k-1} + A_{kk}u_k + A_{k,k+1}u_{k+1} = f_k, \tag{3}$$

где A_{kj} $(j = k-1,\ k,\ k+1)$ — матрицы, имеющие N_k строк и N_j столбцов. Матрицы $A_{k,k-1}$, $A_{k,k+1}$ в каждой строке и каждом столбце имеют не более чем по одному ненулевому элементу. Этими элементами являются значения $b\left(x_i,\ y_j \pm \frac{h}{2}\right)$. Матрица A_{kk} имеет в каждой строке и каждом столбце не более трех отличных от нуля элементов: на главной диагонали элемент

$$-\left[a\left(x_i + \frac{h}{2},\ y_j\right) + a\left(x_i - \frac{h}{2},\ y_j\right) + b\left(x_i,\ y_j + \frac{h}{2}\right) + \right.$$

$$\left. + b\left(x_i,\ y_j - \frac{h}{2}\right) + h^2 c\,(x_i,\ y_j)\right]$$

и на соседних диагоналях — элементы $a\left(x_i \pm \frac{h}{2},\ y_j\right)$, т. е. матрица A_{kk} трехдиагональна. Записывая равенства (3) для каждой строки $(k = 1, 2, \ldots, m)$ придем к системе

$$Au = f \tag{4}$$

с клеточно-трехдиагональной матрицей

$$A = \begin{pmatrix} A_{11} & A_{12} & 0 & 0 & \ldots & 0 & 0 \\ A_{21} & A_{22} & A_{23} & 0 & \ldots & 0 & 0 \\ 0 & A_{32} & A_{33} & A_{34} & \ldots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \ldots & A_{m,m-1} & A_{m,m} \end{pmatrix}.$$

Для матриц A_{ik} выполняются соотношения $A_{ik} = A^*_{ki}$, где звездочкой обозначена транспонированная матрица.

Систему (4) будем решать с помощью решения вспомогательных систем вида

$$Bv = z + g. \tag{5}$$

Если предположить, что матрица B неособая, то вектор z при фиксированном векторе g можно выбрать так, чтобы выполнялось равенство $v = u$. Этот вектор z должен удовлетворять следующей системе линейных уравнений:

$$AB^{-1}z = f - AB^{-1}g. \tag{6}$$

Определим теперь матрицу B и вектор g. Возьмем в качестве матрицы B следующую клеточно-треугольную матрицу:

$$B = \begin{pmatrix} A_{11} & 0 & 0 & 0 & \dots & 0 \\ A_{21} & A_{22} & 0 & 0 & \dots & 0 \\ 0 & A_{32} & A_{33} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & A_{mm} \end{pmatrix}.$$

Разностную задачу, которая приводит к системе вида (5) с клеточно-треугольной матрицей B, будем называть разностной задачей эволюционного типа. Такие системы получаются, например, при решении уравнений в частных производных параболического типа разностным методом. Вектор g в (5), вообще говоря, произволен. Его следует выбирать исходя из удобства программирования вычислительной схемы решения системы (5) для ЭЦВМ. В частности, вектор g можно взять равным f либо положить его равным нулю.

Систему (6) будем решать итерационными методами. Вычислительная схема получится наиболее простой, если к (6) применить метод простой итерации [2]:

$$z_{n+1} = z_n - [AB^{-1}(z_n + g) - f] \quad (n = 0,\ 1,\ 2,\ \dots). \tag{7}$$

Обозначая $B^{-1}(z_n + g) = u_n$, получим вычислительную схему метода:

$$\begin{gathered} Bu_n = z_n + g, \\ z_{n+1} = z_n - (Au_n - f) \\ (n = 0,\ 1,\ 2,\ \dots). \end{gathered} \tag{8}$$

Вектор z_0 здесь произволен. Системы $Bu_n = z_n + g$, которые входят в вычислительную схему (8), при известных z_n, g и определенной выше матрице B легко решаются; их решение сводится к последовательному решению систем более низкого порядка N_k с трехдиагональными матрицами A_{kk}:

$$A_{kk}u_{nk} = -A_{k,k-1}u_{n,k-1} + z_{nk} + g_k \quad (k = 1,\ 2,\ \dots,\ m). \tag{9}$$

В равенствах (9) следует принять $A_{1,0} = 0$, векторы u_{nk}, z_{nk}, g_k определяются аналогично векторам u_k, f_k. Для решения систем с трехдиагональными матрицами на ЭЦВМ целесообразно применять метод прогонки [3, 4].

Поскольку формулы (9) имеют рекуррентный вид, то возникает вопрос об устойчивости счета по этим формулам. Кроме того, необходимо установить условия сходимости итерационного процесса (8).

Введем определение устойчивости счета. Счет по формулам (9) на n-й итерации будем называть устойчивым, если для векторов погрешностей $\varepsilon_{n,i}$ и $\varepsilon_{n,i+k}$, допущенных соответственно в векторах $u_{n,i}$ и $u_{n,i+k}$, имеет место неравенство

$$\| \varepsilon_{n,i+k} \| \leqslant \| \varepsilon_{n,i} \|$$

при всех $i = 0,\ 1,\ 2,\ \dots,\ m$ и таких целых $k \geqslant 0$, что $i + k \leqslant m$.

Пользуясь этим определением, легко установить, что для устойчивости счета по формулам (9) достаточно выполнение неравенства

$$\| A_{kk}^{-1} A_{k,k-1} \| < 1.$$

Имеет место следующая теорема.

Теорема 1. *Если выполнено условие*

$$p = \frac{b_{\max}}{b_{\min} + h^2 c_{\min}} \leqslant 1,$$

то счет по формулам (9) *устойчив.*

Доказательство. Оценим нормы матриц A_{kk}^{-1} и $A_{k,k-1}$ при всех $k = 1, 2, \ldots, m$. Для этого возьмем произвольный вектор w_k размерности N_k и оценим квадратичную форму $-(A_{kk}w_k, w_k)$ снизу. Эта квадратичная форма имеет следующий вид:

$$-(A_{kk}w_k, w_k) = \sum_{i=0}^{N_k} \left\{ a\left(x_i + \frac{h}{2}, y_j \right)(w_{i+1,k} - w_{i,k})^2 + \right.$$

$$\left. + \left[b\left(x_i, y_j + \frac{h}{2} \right) + b\left(x_i, y_j - \frac{h}{2} \right) + h^2 c(x_i, y_j) \right] w_{i,k}^2 \right\},$$

где считается, что $w_{0,k} = w_{N_k+1,k} = 0$. Отсюда получаем

$$-(A_{kk}w_k, w_k) \geqslant (2b_{\min} + h^2 c_{\min}) \| w_k \|^2.$$

Из этого неравенства следует, что для любого $k = 1, 2, \ldots, m$ будет

$$\| A_{kk}^{-1} \| \leqslant \frac{1}{2b_{\min} + h^2 c_{\min}}.$$

Замечая, что в каждом столбце и каждой строке матрицы $A_{k,k-1}$ имеется не более чем по одному ненулевому элементу, получаем следующую оценку ее нормы:

$$\| A_{k,k-1} \| \leqslant b_{\max}.$$

Таким образом,

$$\| A_{kk}^{-1} A_{k,k-1} \| \leqslant \| A_{kk}^{-1} \| \cdot \| A_{k,k-1} \| \leqslant \frac{b_{\max}}{2b_{\min} + h^2 c_{\min}},$$

откуда и следует утверждение теоремы.

Следствие. Если в предыдущих построениях заменить строки сеточной области D_h столбцами, то счет по формулам (9) будет устойчив при выполнении неравенства.

$$\frac{a_{\max}}{2a_{\min} + h^2 c_{\min}} \leqslant 1.$$

Теорема 2. *Итерационный процесс* (8) *сходится, если выполнено неравенство*

$$q = p \frac{1 - p^m}{1 - p} < 1$$

с быстротой геометрической прогрессии, которая имеет знаменатель q.

Доказательство. Обозначим через z точное решение системы (6) и введем вектор $y_n = z_n - z$ размерности $\sum_{k=1}^{m} N_k$. Тогда из (6) и (7) получим, что

$$y_n = (E - AB^{-1}) y_{n-1}.$$

Разобьем вектор y_n на векторы y_{nk} ($k = 1, 2, \ldots, m$) размерности N_k. Введем норму вектора y_n следующим образом:

$$\| y_n \|_1 = \max_k \| y_{nk} \|.$$

Здесь индексом единица внизу отмечена введенная норма вектора, двумя вертикальными черточками без индекса обозначена эвклидова норма вектора. Такие же обозначения приняты и для норм матриц, согласованных с этими нормами. Тогда

$$\| E - AB^{-1} \|_1 = \| (B - A) B^{-1} \|_1 \leqslant \| B - A \|_1 \cdot \| B^{-1} \|_1. \tag{10}$$

Из выражений для матриц B и A следует, что

$$\| B - A \|_1 \leqslant \max_{k=1,2,\ldots,m-1} \| A_{k,k+1} \| = \max_{k=2,\ldots,m} \| A_{k,k-1} \| \leqslant b_{\max}. \tag{11}$$

Поскольку матрица B клеточно-треугольная, то можно найти явное выражение для обратной матрицы B^{-1}:

$$B^{-1} = \begin{pmatrix} A_{11}^{-1} & 0 & \ldots & 0 \\ -A_{22}^{-1} A_{21} A_{11}^{-1} & A_{22}^{-1} & \ldots & 0 \\ A_{33}^{-1} A_{32} A_{22}^{-1} A_{21} A_{11}^{-1} & -A_{33}^{-1} A_{32} A_{22}^{-1} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ (-1)^{m-1} A_{mm}^{-1} A_{m,m-1} \ldots A_{11}^{-1} & (-1)^m A_{mm}^{-1} A_{m,m-1} \ldots A_{22}^{-1} & \ldots & A_{mm}^{-1} \end{pmatrix}.$$

Используя ранее найденные оценки норм матриц A_{kk}^{-1} и $A_{k,k-1}$, получим следующую оценку нормы матрицы B^{-1}:

$$\| B^{-1} \|_1 \leqslant \frac{1}{2b_{\min} + h^2 c_{\min}} \max_{k=1,2,\ldots,m} \frac{1 - p^k}{1 - p} = \frac{p}{b_{\max}} \cdot \frac{1 - p^m}{1 - p}. \tag{12}$$

Из (10), (11) и (12) находим

$$\| E - AB^{-1} \|_1 \leqslant p \frac{1 - p^m}{1 - p} = q, \tag{13}$$

т. е.

$$\| z_n - z \|_1 \leqslant q \| z_{n-1} - z \|_1 \leqslant q^n \| z_0 - z \|_1.$$

Отметим, что правая часть оценки (13) зависит только от числа строк, но не зависит от числа узлов в строках. Опытные расчеты на ЭЦВМ показывают, что зависимость быстроты сходимости рассматриваемого итерационного процесса от числа узлов в строках весьма слабая.

Скорость сходимости итерационного процесса (8) можно улучшить, если ввести определенным образом подобранный параметр ω и вместо (8) рассматривать следующий итерационный процесс:

$$\begin{gathered} Bu_n = z_n + g, \\ z_{n+1} = z_n - \omega (Au_n - f) \\ (n = 0, 1, 2, \ldots). \end{gathered} \tag{14}$$

Однако строго доказать в общем случае возможность улучшения сходимости итерационного процесса (14) за счет подбора параметра ω затруднительно. Поэтому мы получим более ограниченный результат для случая, когда вместо уравнения (1) рассматривается уравнение

$$a \frac{\partial^2 u}{\partial x^2} + b \frac{\partial^2 u}{\partial y^2} - cu = f \tag{15}$$

с постоянными коэффициентами a, b, c в прямоугольнике. Тогда справедливо следующее утверждение.

Теорема 3. *При $\omega > 1$, но достаточно близких к единице, выполняется неравенство*

$$\| E - \omega AB^{-1} \|_1 < \| E - AB^{-1} \|_1.$$

Доказательство. Для уравнения (15), рассматриваемого в прямоугольнике, матрицы $A_{k,k-1}$, $A_{k,k+1}$, A_{kk} из (3) принимают вид:

$$A_{k,k-1} = A_{k,k+1} = b\widetilde{E},$$

$$A_{kk} = \begin{pmatrix} -(2a+2b+h^2c) & a & 0 & \ldots & 0 \\ a & -(2a+2b+h^2c) & a & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & & -(2a+2b+h^2c) \end{pmatrix},$$

где $\widetilde{E}$ — единичная матрица, порядок которой равен числу N внутренних узлов в строке. Из приведенных выражений для матриц $A_{k,k-1}$, $A_{k,k+1}$, A_{kk} видно, что они не зависят от индекса k. Обозначим $A_{kk} = \widetilde{A}$. Пусть λ_0 — абсолютное значение наименьшего по модулю собственного числа матрицы $\widetilde{A}$. Из оценки квадратичной формы $(A_{kk}w_k,\ w_k)$, полученной при доказательстве теоремы 1, следует, что $\lambda_0 \geqslant 2b + h^2c$. Подсчитаем теперь элементы матрицы $E - \omega AB^{-1}$. После соответствующих выкладок получим:

$$E - \omega AB^{-1} =$$

$$= \omega \begin{pmatrix} \left(\frac{1}{\omega} - 1\right)\widetilde{E} + b^2\widetilde{A}^{-2} & -b\widetilde{A}^{-1} & 0 & \ldots & 0 \\ -b^3\widetilde{A}^{-3} & \left(\frac{1}{\omega} - 1\right)\widetilde{E} + b^2\widetilde{A}^{-2} & -b\widetilde{A}^{-1} & \ldots & 0 \\ b^4\widetilde{A}^{-4} & -b^3\widetilde{A}^{-3} & \left(\frac{1}{\omega} - 1\right)\widetilde{E} + b^2\widetilde{A}^{-2} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ (-1)^m b^m \widetilde{A}^{-m} & (-1)^{m-1} b^{m-1}\widetilde{A}^{-m+1} & (-1)^{m-2} b^{m-2}\widetilde{A}^{-m+2} & \ldots & -b\widetilde{A}^{-1} \\ 0 & 0 & 0 & \ldots & \left(\frac{1}{\omega} - 1\right)\widetilde{E} \end{pmatrix}.$$

Поскольку

$$b \| \widetilde{A}^{-1} \| = \frac{b}{\lambda_0} \leqslant \frac{b}{2b + h^2c} \leqslant \frac{1}{2},$$

то норма матрицы $E - \omega AB^{-1}$ при $\omega > 1$, но достаточно близких к единице, выразится так:

$$\| E - \omega AB^{-1} \|_1 = \omega \max_{k=3,4,\ldots,m} \left\{ b \| \widetilde{A}^{-1} \| + \left\| \left(\frac{1}{\omega} - 1\right)\widetilde{E} + b^2\widetilde{A}^{-2} \right\| + \right.$$

$$\left. + \sum_{i=3}^{k} b^i \| \widetilde{A}^{-i} \| \right\} = \omega \left[b \| \widetilde{A}^{-1} \| + \left\| \left(\frac{1}{\omega} - 1\right)\widetilde{E} + b^2\widetilde{A}^{-2} \right\| + \right.$$

$$+ \sum_{i=3}^{m} b^i \| \widetilde{A}^{-i} \| \Big] = \omega \left(\frac{1}{\omega} - 1 + \sum_{i=1}^{m} \frac{b^i}{\lambda_0^i} \right) =$$

$$= 1 - \omega \left(1 - \frac{b}{\lambda_0} \cdot \frac{1 - \left(\frac{b}{\lambda_0} \right)^m}{1 - \frac{b}{\lambda_0}} \right).$$

Так как

$$0 < 1 - \frac{b}{\lambda_0} \cdot \frac{1 - \left(\frac{b}{\lambda_0} \right)^m}{1 - \frac{b}{\lambda_0}} < 1,$$

то

$$1 - \omega \left(1 - \frac{b}{\lambda_0} \cdot \frac{1 - \left(\frac{b}{\lambda_0} \right)^m}{1 - \frac{b}{\lambda_0}} \right) < 1 - \left(1 - \frac{b}{\lambda_0} \cdot \frac{1 - \left(\frac{b}{\lambda_0} \right)^m}{1 - \frac{b}{\lambda_0}} \right),$$

что приводит к неравенству

$$\| E - \omega A B^{-1} \|_1 < \| E - A B^{-1} \|_1.$$

Оптимальное значение ω, при котором быстрота сходимости будет наибольшая, можно определить путем опытных расчетов на ЭЦВМ. Для этого, например, можно подсчитать одно и то же небольшое число s итераций процесса (14) при различных $\omega > 1$ и сравнить полученные на последней итерации нормы вектора невязок $r_s = Au_s - f$. То значение ω, при котором получена наименьшая норма вектора невязок, можно принять за оптимальное и продолжать процесс (14) дальше до получения решения.

Рассмотренный нами метод решения разностной краевой задачи (4) был построен на основе применения к системе (6) процесса простой итерации. Для тех случаев, когда этот процесс не сходится или сходится медленно, можно применять другие итерационные процессы, имеющие более широкую область сходимости. Применение метода сопряженных градиентов к решению системы типа (6) рассмотрено в [4] (стр. 188).

ЛИТЕРАТУРА

1. А. С. Хаусхолдер, Основы численного анализа, ИЛ, 1956.
2. Д. К. Фаддеев, В. Н. Фаддеева, Вычислительные методы линейной алгебры, Физматгиз, М., 1960.
3. С. К. Годунов, В. С. Рябенький, Введение в теорию разностных схем, Физматгиз, М., 1962.
4. В. Е. Шаманский, Методы численного решения краевых задач на ЭЦВМ, ч. 1, Изд-во АН УССР, К., 1963.
5. В. С. Рябенький, А. Ф. Филиппов, Об устойчивости разностных уравнений, Гостехиздат, М.—Л., 1956.

Поступила 24.I 1964 г.
Киев