МЕТОДЫ СПЛАЙН-ФУНКЦИЙ В ЧИСЛЕННОМ АНАЛИЗЕ
(Вычислительные системы)
1983 год Выпуск 98

УДК 519.62

ЧИСЛЕННОЕ РЕШЕНИЕ СИСТЕМЫ ЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ С ТРЕХДИАГОНАЛЬНОЙ МАТРИЦЕЙ МЕТОДОМ ОРТОГОНАЛЬНЫХ ВСТРЕЧНЫХ ПРОГОНОК

С.И. Фадеев

I. К р а т к а я х а р а к т е р и с т и к а м е т о д а. Предлагаемый метод решения системы линейных алгебраических уравнений с трехдиагональной матрицей является вариантом метода встречных прогонок с использованием последовательности ортогональных преобразований Хаусхолдера. С тем же успехом мы могли бы обратиться к последовательности плоских вращений. Наш выбор связан с возможностью простого обобщения метода на случай, когда система имеет ленточную матрицу общего вида. В результате встречных прогонок формируются подсистемы относительно пар неизвестных. Так как при ортогональных преобразованиях сохраняются значения сингулярных чисел матрицы системы, то обусловленность матрицы каждой из подсистем как отношение наибольшего сингулярного числа к наименьшему по крайней мере не превосходит обусловленности матрицы исходной системы. В условиях машинной реализации алгоритма это утверждение нуждается в информации об оценках погрешности округлений. Благодаря использованию ортогональных преобразований, технология определения оценок оказывается достаточно простой. Трудоемкость метода характеризуется числом арифметических операций порядка N, где N - размерность системы.

По затрагиваемым в статье вопросам имеется обширная литература (см., например, [I]). В частности, сошлемся на работу [2], где приводится алгоритм встречных прогонок на основе исключения неизвестных по методу Гаусса, и работу [3], в которой подробно описан метод вычисления оценок погрешностей, возникающих при преобразованиях Хаусхолдера или, как принято называть, преобразова - ниях отражения.

2. О п р е д е л е н и я. Пусть в N-мерном эвклидовом пространстве даны вектор $a = (a_1, a_2, \ldots, a_N)$, гиперплоскость с нормалью $\nu = (\nu_1, \nu_2, \ldots, \nu_N)$ и пусть $b = (b_1, b_2, \ldots, b_N)$ - отраженный вектор: $b = a - 2\frac{(\nu, a)}{(\nu, \nu)}\nu = Pa$. Здесь P - ортогональная симметрическая матрица отражения размерности N x N с элементами: $P_{i,j} = \delta_{i,j} - 2\frac{\nu_i \nu_j}{(\nu, \nu)}$, где $\delta_{i,j}$ - индекс Кронекера. Обозначим через P[i,i+1] матрицу отражения, если отраженный вектор b таков, что

$$b_i = -\sigma\sqrt{a_i^2 + a_{i+1}^2}, \quad \sigma = \begin{cases} +1, & a_i \geq 0 \\ -1, & a_i < 0 \end{cases}, \quad |b_i| \neq 0, \; b_{i+1} = 0, \qquad (1)$$

а все остальные элементы совпадают с соответствующими элементами вектора a: $b_k = a_k$, $k \neq i$, $k \neq i+1$. Наряду с этим мы будем использовать обозначение P[i+1,i], если

$$b_{i+1} = -\sigma\sqrt{a_i^2 + a_{i+1}^2}, \quad \sigma = \begin{cases} +1, & a_{i+1} \geq 0 \\ -1, & a_{i+1} < 0 \end{cases}, \quad |b_{i+1}| \neq 0, \; b_i = 0, \qquad (2)$$

а остальные элементы векторов a и b совпадают. Матрицы P[i,i+1] и P[i+1,i] отличаются от единичной матрицы лишь элементами с индексами (i,i), (i,i+1), (i+1,i) и (i+1, i+1). Эти элементы образуют подматрицы Q[i,i+1] и Q[i+1,i] соответственно:

$$Q[i,i+1] = \begin{bmatrix} -p_i & -q_i \\ -q_i & p_i \end{bmatrix}, \quad p_i = -\frac{a_i}{b_i}, \quad q_i = -\frac{a_{i+1}}{b_i}, \qquad (3)$$

согласно (I), и

$$Q[i+1,i] = \begin{bmatrix} p_i & -q_i \\ -q_i & -p \end{bmatrix}, \quad p_i = -\frac{a_{i+1}}{b_{i+1}}, \quad q_i = -\frac{a_i}{b_{i+1}}, \qquad (4)$$

согласно (2). Подматрицы Q[i,i+1] и Q[i+1,i] являются матрицами отражения вектора (a_i, a_{i+1}) в пространстве двух измерений, причем (b_i, b_{i+1}) - отраженный вектор:

$$\begin{bmatrix} b_i \\ 0 \end{bmatrix} = Q[i,i+1]\begin{bmatrix} a_i \\ a_{i+1} \end{bmatrix}, \quad \begin{bmatrix} 0 \\ b_{i+1} \end{bmatrix} = Q[i+1,i]\begin{bmatrix} a_i \\ a_{i+1} \end{bmatrix}.$$

3. Оценка погрешности расчета отраженного вектора. Следуя [3], воспользуемся положительными параметрами ε_1 и ε_2 - характеристиками разрядности ЭВМ, использующей представление числа с плавающей запятой - при оценке погрешности вычислений. Здесь ε_1 таково, что $1+\varepsilon_1$ - наименьшее по модулю машинное число, большее единицы; $\varepsilon_2/2$ и $1/\varepsilon_2$ - соответственно нижняя и верхняя границы модулей машинных чисел. Предполагается, что все вычисления производятся с одинарной точностью.

Пусть, например, известна матрица $P[i,i+1]$ и требуется найти вектор $y = (y_1, y_2, \ldots, y_N)$ как отражение вектора $x = (x_2, x_2, \ldots, x_N)$:

$$y = P[i,i+1]x, \quad \|y\| = \|x\| = \sqrt{(x,x)}. \tag{5}$$

В силу структуры $P[i,i+1]$ имеем: $y_k = x_k$, $k \neq i$, $k \neq i+1$,

$$\begin{bmatrix} y_i \\ y_{i+1} \end{bmatrix} = Q[i,i+1]\begin{bmatrix} x_i \\ x_{i+1} \end{bmatrix} = \begin{bmatrix} -p_i x_i - q_i x_{i+1} \\ -q_i x_i + p_i x_{i+1} \end{bmatrix}, \tag{6}$$

где p_i и q_i определены по формулам (3). Из-за ошибок округлений, возникающих а) при вычислении p_i и q_i и б) при вычислении y_i и y_{i+1} по формулам (6), мы вместо y получим вектор $\tilde{y} = (\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_N)$: $\tilde{y} = (P[i,i+1] + \delta P)(x + \delta x) = y+h$, где вектор $h = (h_1, h_2, \ldots, h_N)$ моделирует влияние ошибок округлений, матрица δP и вектор δx имеют нормы порядка ε_1, причем $\tilde{y}_k = y_k$, $h_k = 0$, $k \neq i$, $k \neq i+1$. С точностью до малых более высокого порядка

$$\|h\| \leq \|\delta P\|\|x\| + \|\delta x\|. \tag{7}$$

Рассмотрим оценки погрешностей в случае "а". Имеем: $(p_i)_{\text{выч}} = p_i + \delta p_i + u_i$, $(q_i)_{\text{выч}} = q_i + \delta q_i + v_i$, где

$$|\delta p_i| \leq 4\varepsilon_1 |p_i|, \quad |\delta q_i| \leq 4\varepsilon_1 |q_i|, \quad |u_i| \leq \frac{\varepsilon_2}{2}, \quad |v_i| \leq \frac{\varepsilon_2}{2}.$$

Здесь использовано предположение о том, что, привлекая при необходимости масштабирование элементов a_i и a_{i+1} (значения p_i и q_i от этого не изменяются), можно считать ограниченным снизу модуль b_i: $|b_i| > \sqrt{\frac{3\varepsilon_2}{4\varepsilon_1}}$. Матрица δP отличается от нулевой только элементами с индексами (i,i), $(i+1,i)$, $(i,i+1)$ и $(i+1,i+1)$, кото-

рые образуют подматрицу δQ:

$$\delta Q = \begin{bmatrix} -\delta p_i & -\delta q_i \\ -\delta q_i & \delta p_i \end{bmatrix} + \begin{bmatrix} -u_i & -v_i \\ -v_i & u_i \end{bmatrix}.$$

Поэтому

$$\|\delta P\| \le \|\delta P\|_E = \|\delta Q\|_E = \sqrt{2(\delta p_i+u_i)^2 + 2(\delta q_i+v_i)^2} \le$$

$$\le \sqrt{2(\delta p_i^2+\delta q_i^2)} + \sqrt{2(u_i^2+v_i^2)} \le 4\sqrt{2}\varepsilon_1+\varepsilon_2 .$$

Здесь норма $\|\delta P\|$ оценивается сверху евклидовой нормой $\| \delta P \|_E$, квадрат которой равен сумме квадратов всех элементов матрицы δP.

В случае "б" оценивается норма вектора $\delta x=(\delta x_1,\delta x_2,\dots,\delta x_N)$, причем $\delta x_k = 0$, $k \ne i$, $k \ne i+1$:

$$|\delta x_1| \le 2\varepsilon_1(|p_i||x_i|+|q_i||x_{i+1}|) + \frac{3}{2}\varepsilon_2 \le 2\varepsilon_1\sqrt{x_i^2 + x_{i+1}^2} + \frac{3}{2}\varepsilon_2,$$

$$|\delta x_2| \le 2\varepsilon_1(|q_i||x_i|+|p_i||x_{i+1}|) + \frac{3}{2}\varepsilon_2 \le 2\varepsilon_1\sqrt{x_i^2 + x_{i+1}^2} + \frac{3}{2}\varepsilon_2.$$

Отсюда следует, что $\|\delta x\| = \sqrt{\delta x_1^2+\delta x_2^2} \le \sqrt{8\varepsilon_1^2(x_i^2+x_{i+1}^2)} + \sqrt{\frac{9}{2}\varepsilon_2^2} \le$ $\le 2\sqrt{2}\varepsilon_1\|x\| + \frac{3\sqrt{2}}{2}\varepsilon_2 \le 4\sqrt{2}\,\varepsilon_1\|x\|$, если выполнено неравенство:

$$\|x\| \ge \frac{3\varepsilon_2}{4\varepsilon_1}. \qquad (8)$$

Пренебрегая в (7) отличием нормы $\|x\|$ от нормы $\|\tilde{y}\|$, получим:

$$\|h\| \le (8\sqrt{2}\varepsilon_1 + \varepsilon_2)\|\tilde{y}\| \le 12\varepsilon_1\|\tilde{y}\|, \qquad (9)$$

поскольку заведомо $\varepsilon_2 < 0.6\varepsilon_1$.

4. Метод ортогональных встречных прогонок. Рассмотрим систему из N линейных алгебраических уравнений с трехдиагональной матрицей:

$$Cf = F, \qquad (10)$$

или

$$\begin{bmatrix} c_{1,2} & c_{1,3} & & & 0 \\ c_{2,1} & c_{2,2} & c_{2,3} & & \\ & c_{3,1} & c_{3,2} & c_{3,3} & \\ & \cdots & \cdots & \cdots & \\ 0 & c_{N-1,1} & c_{N-1,2} & c_{N-1,3} & \\ & & c_{N,1} & c_{N,2} & \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \cdots \\ f_{N-1} \\ f_N \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ \cdots \\ F_{N-1} \\ F_N \end{bmatrix}.$$

Покажем существование ортогональной матрицы U такой, что система

$$UCf = UF \qquad (11)$$

будет содержать подсистему из двух уравнений относительно неизвестных f_i и f_{i+1} вида:

$$\begin{bmatrix} A^{(i)}_{1,1} & A^{(i)}_{1,2} \\ A^{(i)}_{2,1} & A^{(i)}_{2,2} \end{bmatrix} \begin{bmatrix} f_i \\ f_{i+1} \end{bmatrix} = \begin{bmatrix} G^{(i)} \\ G^{(i)}_2 \end{bmatrix} \qquad (12)$$

Для этого построим последовательность отражений типа $P(k, s)$, $|k-s| = 1$, причем в качестве вектора a на k-м шагу берется k-й вектор-столбец матрицы системы, полученной на (k-1)-м шагу. Если $1 \le i \le N-1$, то последовательность отражений состоит из правого и левого ходов прогонки. При i=1 (т.е. требуется найти f_1 и f_2) достаточно выполнить только левый ход, а при i=N-1 (неизвестные f_{N-1} и f_N) - только правый ход.

На первом шагу правого хода прогонки матрица C и вектор F умножаются на P[1,2]. В силу определения P[1,2] преобразование сводится к вычислению произведения Q[1,2] и подматрицы, состоящей из элементов первой и второй строк системы (I0). Остальные элементы системы (I0) остаются без изменения. Результат представим в виде:

$$Q[1,2]\begin{bmatrix} c_{1,2} & c_{1,3} & 0 & 0\ldots0 & F_1 \\ c_{2,1} & c_{2,2} & c_{2,3} & 0\ldots0 & F_2 \end{bmatrix} = \begin{bmatrix} c^{(1)}_{1,1} & c^{(1)}_{1,2} & c^{(1)}_{1,3} & 0\ldots0 & F^{(1)}_1 \\ 0 & c^{(1)}_{2,2} & c^{(1)}_{2,3} & 0\ldots0 & F^{(1)}_2 \end{bmatrix},$$

где верхний индекс приписан новым значениям необязательно равных нулю элементов первой и второй строк системы:

$$C^{(1)}f = F^{(1)},\ C^{(1)} = P[1,2]C,\ F^{(1)} = P[1,2]F. \qquad (13)$$

На втором шагу матрица $C^{(1)}$ и вектор $F^{(1)}$ умножаются на $P[2,3]$. Преобразование сводится к вычислению произведения $Q[2,3]$ и подматрицы, состоящей из элементов второй и третьей строк системы (I3):

$$Q[2,3]\begin{bmatrix} 0 & c^{(1)}_{2,2} & c^{(1)}_{2,3} & 0 & 0\ldots0 & F^{(1)}_2 \\ 0 & c_{3,1} & c_{3,2} & c_{3,3} & 0\ldots0 & F^{(1)}_3 \end{bmatrix} =$$

$$= \begin{bmatrix} 0 & c^{(2)}_{2,1} & c^{(2)}_{2,2} & c^{(2)}_{2,3} & 0\ldots0 & F^{(2)}_2 \\ 0 & 0 & c^{(2)}_{3,2} & c^{(2)}_{3,3} & 0\ldots0 & F^{(2)}_3 \end{bmatrix},$$

где верхний индекс элементов произведения приписан новым значениям необязательно равных нулю элементов второй и третьей строк системы:

$$C^{(2)}f = F^{(2)},\ C^{(2)} = P[2,3]C^{(1)},\ F^{(2)} = P[2,3]F^{(1)}. \quad (14)$$

Остальные элементы системы (I3) остаются без изменения. После (i-1)-го шага матрица $C^{(i-1)}$ и вектор правых частей $F^{(i-1)}$ имеют вид:

$$C^{(i-1)} = \begin{bmatrix} c^{(1)}_{1,1} c^{(1)}_{1,2} c^{(1)}_{1,3} & & & & \\ c^{(2)}_{2,1} c^{(2)}_{2,2} c^{(2)}_{2,3} & & & & \\ \ddots & \ddots & \ddots & & \\ & c^{(i-1)}_{i-1,1} c^{(i-1)}_{i-1,2} c^{(i-1)}_{i-1,3} & & & \\ & & c^{(i-1)}_{i,2} c^{(i-1)}_{i,3} & & \\ & & c_{i+1,1} c_{i+1,2} c_{i+1,3} & & \\ & & \ddots & \ddots & \ddots \\ & & & c_{N-1,1} c_{N-1,2} c_{N-1,3} & \\ & & & c_{N,1}\ \ c_{N,2} & \end{bmatrix},\ F^{(i-1)} = \begin{bmatrix} F^{(1)}_1 \\ F^{(2)}_2 \\ \ldots \\ F^{(i-1)}_{i-1} \\ F^{(i-1)}_i \\ F_{i+1} \\ \ldots \\ F_{N-1} \\ F_N \end{bmatrix}.$$

Отметим два обстоятельства: i-я строка системы

$$C^{(i-1)}f = F^{(i-1)},\ C^{(i-1)} = P[i-1,i]C^{(i-2)},\ F^{(i-1)} = P[i-1,i]F^{(i-2)}, \quad (15)$$

содержит только неизвестные f_i и f_{i+1}, и элементы строк системы

(I0), начиная с (i+1)-й, остались невозмущенными. Этим завершается правый ход прогонки.

Первый шаг левого хода прогонки связан с умножением матрицы $C^{(i-1)}$ и вектора $F^{(i-1)}$ на $P[N,N-1]$. Преобразование сводится к вычислению произведения $Q[N,N-1]$ и подматрицы, состоящей из элементов N-й и (N-1)-й строк системы (I5):

$$Q[N,N-1]\begin{bmatrix} 0\ldots0\ c_{N-1,1} & c_{N-1,2} & c_{N-1,3} \\ 0\ldots0\ 0 & c_{N,1} & c_{N,2} \end{bmatrix} = \begin{bmatrix} 0\ldots0\ \tilde{c}^{(1)}_{N-1,1} & \tilde{c}^{(1)}_{N-1,2} & 0 \\ 0\ldots0\ \tilde{c}^{(1)}_{N,1} & \tilde{c}^{(1)}_{N,2} & \tilde{c}^{(1)}_{N,3} \end{bmatrix},$$

где волна и верхний индекс приписаны новым значениям необязательно равных нулю элементов N-й и (N-1)-й строк системы:

$$\tilde{c}^{(1)}{}_{f=\tilde{F}}{}^{(1)},\ \tilde{C}^{(1)}=P[N,N-1]C^{(i-1)},\ \tilde{F}^{(1)}=P[N,N-1]F^{(i-1)}. \qquad (16)$$

Остальные элементы системы (I5) остаются без изменения. При этом (N-1)-я строка системы (I6) содержит только неизвестные f_{N-2} и f_{N-1}. На втором шагу матрица $\tilde{C}^{(1)}$ и вектор $\tilde{F}^{(1)}$ умножаются на $P[N-1,N-2]$ и так далее. Завершает левый ход прогонки k-й шаг, $k = N-i-1$. В системе

$$\tilde{c}^{(k)}{}_{f=\tilde{F}}{}^{(k)},\ \tilde{C}^{(k)} = P[i+1,i]\tilde{C}^{(k-1)},\ \tilde{F}^{(k)} = P[i+1,i]F^{(k-1)} \qquad (17)$$

строка с номером i+1 содержит только неизвестные f_i и f_{i+1}. Так как элементы первых i строк систем (I5) и (I7) соответственно совпадают, то матрица $\tilde{C}^{(k)}$ и вектор $\tilde{F}^{(k)}$ имеют вид:

$$\tilde{C}^{(k)} = \begin{bmatrix} c^{(1)}_{1,1} & c^{(1)}_{1,2} & c^{(1)}_{1,3} & & & \\ & c^{(2)}_{2,1} & c^{(2)}_{2,2} & c^{(2)}_{2,3} & & \\ & & \ddots & \ddots & \ddots & \\ & & & c^{(i-1)}_{i-1,1} & c^{(i-1)}_{i-1,2} & c^{(i-1)}_{i-1,3} \\ & & & & c^{(i-1)}_{i,2} & c^{(i-1)}_{i,3} \\ & & & & \tilde{c}^{(k)}_{i+1,1} & \tilde{c}^{(k)}_{i+1,2} \\ & & & & \tilde{c}^{(k)}_{i+2,1} & \tilde{c}^{(k)}_{i+2,2} & \tilde{c}^{(k)}_{i+2,3} \\ & & & & \ddots & \ddots & \ddots \\ & & & & \tilde{c}^{(2)}_{N-1,1} & \tilde{c}^{(2)}_{N-1,2} & \tilde{c}^{(2)}_{N-1,3} \\ & & & & & \tilde{c}^{(1)}_{N,1} & \tilde{c}^{(1)}_{N,2} & \tilde{c}^{(1)}_{N,3} \end{bmatrix},\quad \tilde{F}^{(k)} = \begin{bmatrix} F^{(1)}_1 \\ F^{(1)}_2 \\ \vdots \\ F^{(i-1)}_{i-1} \\ F^{(i-1)}_i \\ \tilde{F}^{(k)}_{i+1} \\ \tilde{F}^{(k)}_{i+2} \\ \vdots \\ \tilde{F}^{(2)}_{N-1} \\ \tilde{F}^{(1)}_N \end{bmatrix}.$$

Очевидно, (I7) и есть искомая система (II), включающая подсистему (I2) относительно f_i и f_{i+1}, где

$$A^{(i)}_{1,1} = C^{(i-1)}_{i,2}, \quad A^{(i)}_{1,2} = C^{(i-1)}_{i,3}, \quad G^{(i)}_1 = F^{(i-1)}_i,$$
$$A^{(i)}_{2,1} = \tilde{C}^{(k)}_{i+1,1}, \quad A^{(i)}_{2,2} = \tilde{C}^{(k)}_{i+1,2}, \quad G^{(i)}_2 = \tilde{F}^{(k)}_{i+1}, \quad k=N-i-1.$$

Ортогональная матрица U является произведением (N-2)-х матриц отражений: $U = P[i+1,i]P[i+2,i+1]\,x\ldots x\,P[N,N-1]\,P[i-1,i]\,P[i-2,i-1]x\ldots xP[1,2]$.

Обратим внимание на то, что правый ход и левый ход прогонки, приводящей к системе (I7), в действительности выполняются независимо. Поэтому систему (I7) можно составить из i первых строк системы (I5) и (N-i) последних строк системы $\tilde{U}Cf = \tilde{U}F$, где

$$\tilde{U} = P[i+1,i]P[i+2,i+1]x\ldots xP[N,N-1].$$

При $i = N-1$ коэффициенты подсистемы (I2) относительно f_{N-1} и f_N имеют вид:

$$A^{(i)}_{1,1} = C^{(i-1)}_{i,2}, \quad A^{(i)}_{1,2} = C^{(i-1)}_{i,3}, \quad G^{(i)}_1 = F^{(i-1)}_i,$$
$$A^{(i)}_{2,1} = C_{N,1}, \quad A^{(i)}_{2,2} = C_{N,2}, \quad G^{(i)}_2 = F_N.$$

В этом случае U является произведением матриц отражений, определяемых правым ходом прогонки: $U=P[N-2,N-1]\,P[N-3,N-2]x\ldots xP[1,2]$.

При $i = 1$ коэффициенты подсистемы относительно f_1 и f_2 имеют вид:

$$A^{(i)}_{1,1} = C_{1,2}, \quad A^{(i)}_{1,2} = C_{1,3}, \quad G^{(i)}_1 = F_1,$$
$$A^{(i)}_{2,1} = \tilde{C}^{(N-2)}_{2,1}, \quad A^{(i)}_{2,2} = \tilde{C}^{(N-2)}_{2,2}, \quad G^{(i)}_2 = \tilde{F}^{(N-2)}_2,$$

а U - произведение матриц отражений, определяемым левым ходом прогонки: $U = P[3,2]P[4,3]x\ldots xP[N,N-1]$.

5. О б у с л о в л е н н о с т ь м а т р и ц ы п о д с и с т е м ы. Пусть $\max \sigma(C)$ и $\min \sigma(C)$ - наибольшее и наименьшее сингулярные числа матрицы C, $0 < \min \sigma(C) \le \max \sigma(C)$, и пусть $\mu(C)$ - обусловленность C: $\mu(C) = \frac{\max \sigma(C)}{\min \sigma(C)}$. Обозначим через $A^{(i)}$ матрицу системы (I2). Для оценки $\mu(A^{(i)})$ нам потребуется теорема отделения Штурма [4]. Приведем ее формулировку.

Рассмотрим последовательность симметрических матриц A_r размерности $r \times r$, где $r = 1,2,\ldots,N$. Пусть $\lambda_j(A_r)$, $j = 1,2,\ldots,r$, обозначает j-е собственное число матрицы A_r, причем $\lambda_1(A_r) \le \lambda_2(A_r) \le \ldots \le \lambda_r(A_r)$. Тогда $\lambda_j(A_{i+1}) \le \lambda_j(A_i) \le \lambda_{j+1}(A_{i+1})$.

Опираясь на теорему отделения, легко показать, что $\mu(A^{(i)}) \le \mu(C)$. Действительно, матрица $\tilde{C}^{(k)}$ системы (I7) имеет следующую структуру:

$$\tilde{C}^{(k)} = \left[\begin{array}{c|c} B_1 & O \\ \hline B_2 & B_3 \end{array}\right],$$

где B_1 - подматрица, включающая $A^{(i)}$ как блок в правом нижнем углу, O - нулевая подматрица. По этой причине матрица $B_1B_1^T$ имеет в правом нижнем углу подматрицу $A^{(i)}[A^{(i)}]^T$. Применив теорему отделения к матрицам $A^{(i)}[A^{(i)}]^T$ и $B_1B_1^T$, получим: $\min \sigma^2(B_1) \le \min \sigma^2(A^{(i)}) \le \max \sigma^2(A^{(i)}) \le \max \sigma^2(B_1)$, т.е. $\mu(A^{(i)}) \le \mu(B_1)$. В свою очередь матрица $\tilde{C}^{(k)}[\tilde{C}^{(k)}]^T$ имеет в левом верхнем углу подматрицу $B_1B_1^T$. Следовательно, $\mu(B_1) \le \mu(\tilde{C}^{(k)})$. Остается вспомнить, что $\mu(\tilde{C}^{(k)}) = \mu(C)$, так как $\tilde{C}^{(k)} = U C$, где U - ортогональная матрица.

6. О ц е н к а п о г р е ш н о с т и р е ш е н и я п о д с и с т е м ы. Во многих отношениях целесообразно дополнить прогонку еще одним шагом. Умножим слева матрицу U на $P[i,i+1]$. Тогда система

$$Sf = V, \quad S = P[i,i+1]UC, \quad V = P[i,i+1]UF, \tag{18}$$

$$\|S\| = \|C\|, \quad \|V\| = \|F\|,$$

будет содержать подсистему относительно f_i и f_{i+1} с верхней треугольной матрицей:

$$L^{(i)}z_i = R^{(i)}, \tag{19}$$

$$L^{(i)} = Q[i,i+1]A^{(i)} = \begin{bmatrix} l_i & m_i \\ 0 & n_i \end{bmatrix}, \quad R^{(i)} = Q[i,i+1]G^{(i)} = \begin{bmatrix} r_i \\ r_{i+1} \end{bmatrix}, \quad z_i = \begin{bmatrix} f_i \\ f_{i+1} \end{bmatrix}.$$

Приведение системы (I0) к (I8) сопровождается накоплением ошибок округлений. В результате вместо (I8) мы получим возмущенную сис-

тему относительно $\tilde{f}$, $\tilde{f} = (\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_N)$:

$$\tilde{S}\tilde{f} = \tilde{V}, \quad \tilde{S} = S + \theta, \quad \tilde{V} = V + \Lambda, \tag{20}$$

где матрица θ и вектор Λ моделируют влияние ошибок округлений вследствие вычислений, связанных с (N-1) ортогональными преобразованиями отражений. Аналогично вместо (19) мы будем иметь систему с матрицей $\tilde{L}^{(i)}$ и правой частью $\tilde{R}^{(i)}$ относительно $\tilde{f}_i$ и $\tilde{f}_{i+1}$:

$$\tilde{L}^{(i)}\tilde{z}_i = \tilde{R}^{(i)}, \tag{21}$$

$$\tilde{L}^{(i)} = \begin{bmatrix} \tilde{l}_i & \tilde{m}_i \\ 0 & \tilde{n}_i \end{bmatrix} = L^{(i)} + B^{(i)}, \quad \tilde{R}^{(i)} = \begin{bmatrix} \tilde{r}_i \\ \tilde{f}_{i+1} \end{bmatrix} = R^{(i)} + H^{(i)}, \quad \tilde{z}_i = \begin{bmatrix} \tilde{f}_i \\ \tilde{f}_{i+1} \end{bmatrix}.$$

Нашей ближайшей задачей является получение оценок для $\|B^{(i)}\|$ и $\|H^{(i)}\|$ - норм возмущений $L^{(i)}$ и $H^{(i)}$.

Пусть C_k, S_k, $\tilde{S}_k$ и θ_k, $k = 1,2,\ldots,N$, - вектор-столбцы матриц $C, S, \tilde{S}$ и θ соответственно: $\tilde{S}_k = S_k + \theta_k$, $\|S_k\| = \|C_k\|$. Согласно [3], оценка нормы вектора Λ с учетом погрешности машинного представления Λ имеет вид: $\|\Lambda\| \le N\alpha e^{N\alpha}\|\tilde{V}\|$, где $\alpha = 12\varepsilon_1$ в соответствие с (9). Так как $\|\tilde{V}\| \le \|V\| + \|\Lambda\|$, то $\|\Lambda\|$ может быть выражена через $\|F\|$: $\|\Lambda\| \le N\alpha e^{N\alpha}\|F\|/(1-N\alpha e^{N\alpha})$, $N\alpha e^{N\alpha} < 1$. Подобная оценка справедлива и для $\|\theta_k\|$. Ее, однако, можно уточнить, если принять во внимание, что не более трех элементов вектора C_k отличны от нуля. Тогда $\|\theta_k\| \le 5\alpha e^{5\alpha}\|\tilde{S}_k\|$ или $\|\theta_k\| \le 5\alpha e^{5\alpha}\|C_k\|/(1 - 5\alpha e^{5\alpha})$, $5\alpha e^{5\alpha} < 1$. В приведенных оценках, как следствие (8), предполагается выполненным условие: $\min(\min_k \|C_k\|, \|F\|) > \frac{3\varepsilon_2}{4\varepsilon_1}$. Мы вправе считать, что матрица θ имеет ту же структуру, что и S, и поэтому содержит подматрицу $B^{(i)}$, так же как элементы вектора $H^{(i)}$ являются элементами Λ. В результате имеем следующие неравенства:

$$\|H^{(i)}\| \le \|\Lambda\| \le N\alpha e^{N\alpha}\|\tilde{V}\|,$$

$$\|B^{(i)}\| \le \|B^{(i)}\|_E \le 5\alpha e^{5\alpha}\sqrt{\|S_2\|^2 + \|S_{i+1}\|^2} \tag{22}$$

или

$$\|H^{(i)}\| \le \frac{N\alpha e^{N\alpha}}{1-N\alpha e^{N\alpha}}\|F\|, \quad \|B^{(i)}\| \le \frac{5\alpha e^{5\alpha}}{1-5\alpha e^{5\alpha}}\sqrt{\|C_i\|^2 + \|C_{i+1}\|^2}.$$

При оценке близости решений систем (I9) и (2I) используется известное неравенство [3]. Отвлечемся на время от ранее принятых обозначений и сопоставим две системы линейных алгебраических уравнений: $Ax = f$ и $(A+B)y = f + g$. При этом

$$\| y - x\| \le \|(A+B)^{-1}\|(\|B\|\,\|x\| + \|g\|) \le$$

$$\le \frac{\|A^{-1}\|}{1-\|B\|\|A^{-1}\|}(\|B\|\|x\|+\|g\|), \quad \|B\|\|A^{-1}\| < 1. \qquad (23)$$

Будем считать, что (I9) является возмущенной системой (2I) с оценками норм возмущений (22). Выпишем основные характеристики матрицы $\tilde{L}^{(i)}$, выражения которых очевидны:

$$\|L\| = \sqrt{\frac{\omega}{2}\left[1+\sqrt{1-\left(\frac{2\Delta}{\omega}\right)^2}\right]}, \quad \|L^{-1}\| = \frac{\|L\|}{\Delta}, \quad \mu = \frac{\|L\|^2}{\Delta}, \qquad (24)$$

$$L \equiv \tilde{L}^{(i)}, \quad \omega = \tilde{l}_i^2 + \tilde{m}_i^2 + \tilde{n}_i^2, \quad \Delta = |\tilde{l}_i\tilde{n}_i|, \quad \mu = \mu(L).$$

Рассмотрим погрешности двух типов. Обозначим через $\tilde{\tilde{f}} = (\tilde{\tilde{f}}_1, \tilde{\tilde{f}}_2, \ldots, \tilde{\tilde{f}}_N)$ вектор, элементы которого определяются как машинное решение системы (2I). В этом случае элементы вектора $\tilde{\tilde{z}}_i + u^{(i)}$, $z_i = (\tilde{f}_i, \tilde{f}_{i+1})$, $u^{(i)} = (u_1^{(i)}, u_2^{(i)})$, являются точным решением системы [3]: $(L + \tilde{B}^{(i)})(\tilde{\tilde{z}}_i + u^{(i)}) = R + \tilde{H}^{(i)}$, $R \equiv \tilde{R}^{(i)}$, причем нормы возмущений оцениваются следующим образом: $\|\tilde{B}^{(i)}\| \le 2\sqrt{2}\varepsilon_1\|L\|$, $\|\tilde{H}^{(i)}\| \le \varepsilon_1\|R\| + \sqrt{2}\varepsilon_2$, $\|u^{(i)}\| \le \varepsilon_2/\sqrt{2}$. Полагая в (23) $x = \tilde{z}_i$, $y = \tilde{\tilde{z}}_i + u^{(i)}$, $\|B\| = 2\sqrt{2}\varepsilon_1\|L\|$, $\|g\| = \varepsilon_1\|R\| + \sqrt{2}\,\varepsilon_2$, получим:

$$\|\tilde{\tilde{z}}_i - z_i\| \le K_i\|\tilde{\tilde{z}}_i\| + \phi_i, \qquad (25)$$

где

$$K_i = \frac{2\sqrt{2}\varepsilon_1\mu}{1-2\sqrt{2}\,\varepsilon_1\mu}, \quad \phi_i = \frac{\|L^{-1}\|}{1-2\sqrt{2}\,\varepsilon_1\mu}(\varepsilon_1\|R\| + \sqrt{2}\varepsilon_2) + \frac{\varepsilon_2}{\sqrt{2}}.$$

Из (25) следует, что $\|\tilde{\tilde{z}}_i\| \le K_i\|\tilde{\tilde{z}}_i\| + \phi_i + \|\tilde{z}_i\|$. Поэтому

$$\|\tilde{\tilde{z}}_i\| \le \frac{\|\tilde{z}_i\| + \phi_i}{1-K_i}, \quad K_i < 1.$$

С учетом последнего неравенства представим (25) в виде:

$$\|\tilde{\tilde{z}}_i - \tilde{z}_i\| \le \rho_1^{(i)} = \frac{K_i\|\tilde{z}_i\| + \phi_i}{1 - K_i}. \qquad (26)$$

Далее, пользуясь (23), сопоставим системы (I9) и (2I). В результате получим оценку нормы погрешности второго рода:

$$\|\tilde{z}_i - z_i\| \le \rho_2^{(i)} = \frac{\|L^{-1}\|}{1-\xi\|L^{-1}\|}\{\xi(\rho_1^{(i)} + \|\tilde{z}_i\|) + \eta\},\ \xi\|L^{-1}\|<1, \quad (27)$$

где, согласно (22),

$$\xi = \frac{5\alpha e^{5\alpha}}{1-5\alpha e^{5\alpha}}\sqrt{\|C_i\|^2 + \|C_{i+1}\|^2},\quad \eta = \frac{N\alpha e^{N\alpha}}{1-N\alpha e^{N\alpha}}\|F\|.$$

Таким образом, значения $\rho_1^{(i)}$ и $\rho_2^{(i)}$ находятся после машинного вычисления $\tilde{f}_i$ и $\tilde{f}_{i+1}$: $\tilde{f}_{i+1} = \tilde{r}_{i+1}/\tilde{n}_i$, $\tilde{f}_i = (\tilde{r}_i - \tilde{m}_i\tilde{f}_{i+1})/\tilde{l}_i$, $\|\tilde{z}_i\| = \sqrt{\tilde{f}_{i+1}^2 + \tilde{f}_i^2}$. Вместе с ними вычисляются оценки абсолютной и относительной погрешностей;

$$\|\tilde{z}_i - z_i\| \le \rho_i = \sqrt{[\rho_1^{(i)}]^2 + [\rho_2^{(i)}]^2},\quad \|\tilde{z}_i - z_i\|/\|\tilde{z}_i\| \le \rho_i/\|\tilde{z}_i\| \quad (28)$$

7. О ц е н к и п о г р е ш н о с т и р е ш е н и я с и с т е м ы. В этом пункте мы получим, используя (28), выражения вычисляемой (и поэтому наиболее эффективной) оценки погрешности решения системы (I0), а затем установим связь оценки относительной погрешности решения системы с обусловленностью матрицы С. Заметим, что определение всех неизвестных f_j по методу встречных прогонок требует организации не менее N_0 "встреч", где

$$N_0 = \begin{cases} N/2, & \text{если } N \text{ четное,} \\ (N+1)/2, & \text{если } N \text{ нечетное.} \end{cases}$$

В случае нечетного N одно из неизвестных вычисляется дважды. Пусть это будет f_{N-1}. Следовательно, $\|\tilde{f}-f\| = \{\sum_{k=1}^{N_0}\|\tilde{z}_{2k-1} - z_{2k-1}\|^2\}^{1/2}$, если N четное, и $\|\tilde{f}-f\| \le \{\sum_{k=1}^{N_0-1}\|\tilde{z}_{2k-1} - z_{2k-1}\|^2 + \|\tilde{z}_{N-1} - z_{N-1}\|^2\}^{1/2}$, если N нечетное. Условимся обозначать знаком Σ сумму N_0 слагаемых как в том, так и в другом случае, объединив их одним неравенством: $\|\tilde{f}-f\| \le \{\Sigma\|\tilde{z}_i - z_i\|^2\}^{1/2}$. Каждое из слагаемых суммы, согласно (28), меньше ρ_i^2. Поэтому

$$\|\tilde{f}-f\| \le \rho = \{\Sigma\,\rho_i^2\}^{1/2},\quad \frac{\|\tilde{f}-f\|}{\|\tilde{f}\|} \le \delta_0 = \frac{\rho}{\|\tilde{f}\|}. \quad (29)$$

Представляет интерес зависимость относительной погрешности $\|\tilde{f}-f\|/\|\tilde{f}\|$ от обусловленности матрицы С или обусловленности матрицы S, $\mu(C) = \mu(S)$, свойственная рассматриваемому методу орто-

гональных встречных прогонок. Для этого нам потребуется равномерная оценка характеристик матрицы $\widetilde{S}$, $\widetilde{S} = S + \theta$, системы (20), чтобы затем оценить характеристики матрицы $L \equiv L^{(i)}$ системы (21). Имеем:

$$\|\widetilde{S}\| \le \|S\| + \|\theta\| = \|C\| + \|\theta\|,$$

$$\|\widetilde{S}^{-1}\| \le \frac{\|S^{-1}\|}{1-\|\theta\|\|S^{-1}\|} = \frac{\|C^{-1}\|}{1-\|\theta\|\|C^{-1}\|},$$

$$\mu(\widetilde{S}) = \|\widetilde{S}\|\|\widetilde{S}^{-1}\| \le \frac{\|C^{-1}\|(\|C\| + \|\theta\|)}{1-\|\theta\|\|C^{-1}\|},$$

где

$$\|\theta\| \le \|\theta\|_E \le K_C\|C\|, \quad K_C = \frac{5N\alpha e^{5\alpha}}{1-5\alpha e^{5\alpha}}. \qquad (30)$$

Согласно п.5, наибольшие и наименьшие сингулярные числа матриц L и $\widetilde{S}$ удовлетворяют неравенствам: $1/\|\widetilde{S}^{-1}\| \le 1/\|L^{-1}\| \le \|L\| \le \|\widetilde{S}\|$. Отсюда следуют оценки характеристик матрицы L:

$$\left.\begin{aligned} &\|L\| \le \|\widetilde{S}\| \le (1+K_C)\|C\|, \\ &\|L^{-1}\| \le \|\widetilde{S}^{-1}\| \le \frac{\|C^{-1}\|}{1-K_C\mu(C)}, \\ &\mu(L) \le \mu(\widetilde{S}) \le \mu_0, \ K_C\mu(C) < 1, \\ &\mu_0 = \frac{1+K_C}{1-K_C\mu(C)}\mu(C). \end{aligned}\right\} \qquad (31)$$

Обратимся вначале к оценке $\|\widetilde{z}_i - \widetilde{z}_i\|$, преобразовав неравенство (25 к виду: $\|\widetilde{z}_i - \widetilde{z}_i\| \le K_\mu\|\widetilde{z}_i\| + \phi$, $K_i \le K_\mu$, $\phi_i \le \phi$, где

$$K_\mu = \frac{2\sqrt{2}\,\varepsilon_1\mu_0}{1-2\sqrt{2}\varepsilon_1\mu_0}, \quad 2\sqrt{2}\,\varepsilon_1\mu_0 < 1. \qquad (32)$$

Определим ϕ. Из выражения ϕ_i и (31) следует, что

$$\phi_i \le \frac{\|C^{-1}\|(\varepsilon_1\|F\| + \sqrt{2}\,\varepsilon_2)}{[1-K_C\mu(C)][1-2\sqrt{2}\varepsilon_1\mu_0]} + \frac{\varepsilon_2}{\sqrt{2}}.$$

Так как $\|F\| \le \|C\|\|f\|$, то

$$\phi_i \le \left\{ \frac{\mu(C)\left(\varepsilon_1 + \frac{\sqrt{2}\,\varepsilon_2}{\|F\|}\right)}{[1-K_C\mu(C)][1-2\sqrt{2}\,\varepsilon_1\mu_0]} + \frac{\varepsilon_2\|C\|}{\sqrt{2}\|F\|} \right\}\|f\|.$$

Потребуем, чтобы норма F была ограничена снизу (для системы (I9) аналогичное ограничение на $\|R^{(i)}\|$ было бы обременительным): $\|F\| \ge \frac{\sqrt{2}\varepsilon_2}{\varepsilon_1}(1 + \frac{1}{2}\|C\|)$. При этом в качестве ϕ можно взять выражение:

$$\phi = \varepsilon_1\left[\frac{2\mu_0}{(1+K_C)(1-2\sqrt{2}\varepsilon_1\mu_0)} + 1\right]\|f\|.$$

Для оценки $\|\tilde{\tilde{f}}-\tilde{f}\|$ просуммируем все $\|\tilde{z}_i-z_i\|^2$: $\|\tilde{\tilde{f}}-\tilde{f}\| \le \sqrt{\Sigma\,\|\tilde{\tilde{z}}_i-\tilde{z}_i\|^2} \le K_\mu\sqrt{\Sigma\|\tilde{z}_i\|^2} + \sqrt{\frac{N+1}{2}}\phi \le \sqrt{2}\,K_\mu\|\tilde{f}\| + \frac{1}{2}\sqrt{N+1}(K_\mu+\sqrt{2}\varepsilon_1)\,\|f\|$. Пусть известно, что $\|\tilde{f}-f\| \le \delta_1'\|f\|$. Тогда $\|\tilde{f}\| \le (1+\delta_1')\|f\|$, и, следовательно,

$$\frac{\|\tilde{\tilde{f}}-\tilde{f}\|}{\|f\|} \le \delta_1'' = \sqrt{2}\,(1+\delta_1')K_\mu + \frac{\sqrt{N+1}}{2}(K_\mu + \sqrt{2}\,\varepsilon_1).$$

Вычисление δ_1' связано с использованием неравенства (27), которое запишем в виде:

$$\|\tilde{z}_i-z_i\| \le \frac{\|L^{-1}\|}{1-\xi\|L^{-1}\|}\,(\xi\|\tilde{z}_i\| + \eta), \quad \xi < K_C\|C\|.$$

С учетом (3I) и неравенства

$$\frac{\eta}{\|C\|} = K_F\frac{\|F\|}{\|C\|} \le K_F\|f\|, \quad K_F = \frac{N\alpha e^{N\alpha}}{1-N\alpha e^{N\alpha}}, \tag{33}$$

получим $\|\tilde{z}_i-z_i\| \le K_f(K_C\|\tilde{z}_i\| + K_F\|f\|)$, где

$$K_f = \frac{\mu_0}{1-K_C(\mu_0-1)}, \quad K_C(\mu_0-1) < 1. \tag{34}$$

Для оценки $\|\tilde{f}-f\|$ просуммируем все $\|\tilde{z}_i-z_i\|^2$:

$$\|\tilde{f}-f\| \le \sqrt{\Sigma\,\|\tilde{z}_i-z_i\|^2} \le K_f\left[K_C\sqrt{\Sigma\,\|\tilde{z}_i\|^2} + \sqrt{\frac{N+1}{2}}K_F\,\|f\|\right] \le$$

$$\le K_f\left[\sqrt{2}\,K_C\|\tilde{f}\| + \sqrt{\frac{N+1}{2}}\,K_F\|f\|\right] \le K_f\left[\sqrt{2}\,K_C(1+\rho_1) + \sqrt{\frac{N+1}{2}}\,K_F\right]\|f\|.$$

Отсюда следует выражение δ_1':

$$\frac{\|\tilde{f}-f\|}{\|f\|} \le \delta_1' = K_f \frac{2K_C + \sqrt{N+1}\,K_F}{\sqrt{2}-2K_C K_F}, \quad \sqrt{2}\,K_C K_f < 1.$$

Теперь мы можем выписать искомую оценку δ_1 относительной погрешности:

$$\frac{\|\tilde{\tilde{f}}-f\|}{\|f\|} \le \delta_1' + \delta_2', \quad \frac{\|\tilde{\tilde{f}}-f\|}{\|\tilde{f}\|} \le \delta_1 = \frac{\delta_1'+\delta_2'}{1-\delta_1'-\delta_2'}, \tag{35}$$

$$\delta_1' = K_f \frac{2K_C + \sqrt{N+1}\,K_F}{\sqrt{2} - 2K_C K_f},$$

$$\delta_1'' = \sqrt{2}\,K_\mu(1+\delta_1') + \frac{\sqrt{N+1}}{2}(K_\mu + \sqrt{2}\,\varepsilon_1),$$

где K_C, K_μ, K_F и K_f определены равенствами (30)-(34).

Формула (35), дающая оценку относительной погрешности решения системы (10), свидетельствует о том, что методу ортогональных встречных прогонок свойственна повышенная чувствительность к возмущениям правых частей. Действительно, если обусловленность матрицы C и размерность системы N "не слишком велики", то числитель и знаменатель дроби $(\delta_1'+\delta_2')/(1-\delta_1'-\delta_2')$ содержат слагаемое порядка $\varepsilon_1 N\sqrt{N}\mu(C)$. Отчасти этот вывод основывается на оценке $\|H^{(1)}\|$ (см. (22)), которая пропорциональна $\|F\|$ и поэтому, как правило, пессимистична. Но вместе с тем вывод отражает и суть явления.

8. О п и с а н и е а л г о р и т м а. Хотя вычисление всех неизвестных, определяемых системой (10), связано с N_0 встречными прогонками, совсем не обязательно всякий раз начинать прогонки заново. Чтобы сформировать N_0 систем (12), достаточно один раз проделать правый ход до $i=N-2$ и левый до $i = 3$, как это предложено в п.4. При этом требуется дополнительный массив W с элементами $W_{i,j}$ размерности $2N_0 \times 3$. Кроме того, текущие значения прогоночных коэффициентов мы будем присваивать коэффициентам α_i, β_i и γ_i. Напомним, что при нечетном N мы условились вычислять дважды f_{N-1}. Поэтому удобнее начинать прогонку с левого хода.

Левому ходу прогонки предшествует заполнение строки с номером $2N_0$ массива W: $W_{2N_0,1} = \beta_0 = C_{N,1}$, $W_{2N_0,2} = \alpha_0 = C_{N,2}$,

$W_{2N_0,3} = \gamma_0 = F_N$. На первом шагу вычисляются p_{N-1} и q_{N-1} - элементы матрицы Q[N,N-1] (см. (2),(4)):

$$p_{N-1} = \frac{\sigma\alpha_0}{\sqrt{\alpha_0^2 + C_{N-1,3}^2}}, \quad q_{N-1} = \frac{\sigma C_{N-1,3}}{\sqrt{\alpha_0^2 + C_{N-1,3}^2}}, \quad \sigma = \begin{cases} +1, & \alpha_0 \geq 0, \\ -1, & \alpha_0 < 0, \end{cases}$$

а затем параметры α_1, β_1 и γ_1 по формулам:

$$\alpha_1 = \tilde{C}_{N-1,2}^{(1)} = p_{N-1}C_{N-1,2} - q_{N-1}\beta_0,$$

$$\beta_1 = \tilde{C}_{N-1,1} = p_{N-1}C_{N-1,1},$$

$$\gamma_1 = \tilde{F}_{N-1}^{(1)} = p_{N-1}F_{N-1} - q_{N-1}\gamma_0.$$

На втором шагу вычисляются p_{N-2} и q_{N-2} - элементы матрицы Q[N-1, N-2] :

$$p_{N-2} = \frac{\sigma\alpha_1}{\sqrt{\alpha_1^2 + C_{N-2,3}^2}}, \quad q_{N-2} = \frac{\sigma C_{N-2,3}}{\sqrt{\alpha_1^2 + C_{N-2,3}^2}}, \quad \sigma = \begin{cases} +1, & \alpha_1 \geq 0, \\ -1, & \alpha_1 < 0, \end{cases}$$

а затем параметры α_2, β_2 и γ_2 по формулам:

$$\alpha_2 = \tilde{C}_{N-2,2}^{(2)} = p_{N-2}C_{N-2,2} - q_{N-2}\beta_1,$$

$$\beta_2 = \tilde{C}_{N-2,1}^{(2)} = p_{N-2}C_{N-2,1},$$

$$\gamma_2 = \tilde{F}_{N-2}^{(2)} = p_{N-2}F_{N-2} - q_{N-2}\gamma_1$$

и так далее: i = 1,2,...,N-2,

$$\alpha_i = \tilde{C}_{N-i,2}^{(i)} = p_{N-i}C_{N-i,2} - q_{N-i}\beta_{i-1},$$

$$\beta_i = \tilde{C}_{N-i,1}^{(i)} = p_{N-i}C_{N-i,1},$$

$$\gamma_i = \tilde{F}_{N-i}^{(i)} = p_{N-i}F_{N-i} - q_{N-i}\gamma_{i-1},$$

где

$$p_{N-i} = \frac{\sigma\alpha_{i-1}}{\sqrt{\alpha_{i-1}^2 + C_{N-i,3}^2}}, \quad q_{N-i} = \frac{\sigma C_{N-i,3}}{\sqrt{\alpha_{i-1}^2 + C_{N-i,3}^2}}, \quad \sigma = \begin{cases} +1, & \alpha_{i-1} \geq 0, \\ -1, & \alpha_{i-1} < 0. \end{cases}$$

При четном i заполняются четные строки массива W: $W_{i,1} = \beta_i$, $W_{i,2} = \alpha_i$, $W_{i,3} = \gamma_i$, $i = 2(N_0 - k)$, $k = 0, 1, \ldots, N_0 - 1$, на этом левый ход прогонки завершается.

Правому ходу прогонки предшествует заполнение первой строки массива W: $W_{1,1} = \alpha_0 = C_{1,2}$, $W_{1,2} = \beta_0 = C_{1,3}$, $W_{1,3} = \gamma_0 = F_1$. В результате имеем систему (I2) (в других обозначениях) относительно f_1 и f_2: $W_{1,1}f_1 + W_{1,2}f_2 = W_{1,3}$; $W_{2,1}f_1 + W_{2,2}f_2 = W_{2,3}$. Здесь элементы $W_{2,1}$, $W_{2,2}$ и $W_{2,3}$ были определены на (N-2)-м шагу левого хода прогонки.

На первом шагу вычисляются p_1 и q_1 - элементы матрицы Q[1,2] (см. (I), (3)):

$$p_1 = \frac{\sigma\alpha_0}{\sqrt{\alpha_0^2 + C_{2,1}^2}}, \quad q_1 = \frac{\sigma C_{2,1}}{\sqrt{\alpha_0^2 + C_{2,1}^2}}, \quad \sigma = \begin{cases} +1, & \alpha_0 \geq 0, \\ -1, & \alpha_0 < 0, \end{cases}$$

а затем параметры α_1, β_1 и γ_1 по формулам:

$$\alpha_1 = C_{2,2}^{(1)} = p_1 C_{2,2} - q_1\beta_0,$$
$$\beta_1 = C_{2,3}^{(1)} = p_1 C_{2,3},$$
$$\gamma_1 = F_2^{(1)} = p_1 F_2 - q_1\gamma_0.$$

На втором шагу вычисляются p_2 и q_2 - элементы матрицы Q[2,3]:

$$p_2 = \frac{\sigma\alpha_1}{\sqrt{\alpha_1^2 + C_{3,1}^2}}, \quad q_2 = \frac{\sigma C_{3,1}}{\sqrt{\alpha_1^2 + C_{3,1}^2}}, \quad \sigma = \begin{cases} +1, & \alpha_1 \geq 0, \\ -1, & \alpha_1 < 0, \end{cases}$$

а затем параметры α_2, β_2 и γ_2 по формулам:

$$\alpha_2 = C_{3,2}^{(2)} = p_2 C_{3,2} - q_2\beta_1,$$
$$\beta_2 = C_{3,3}^{(2)} = p_2 C_{3,3},$$
$$\gamma_2 = F_3^{(2)} = p_2 F_3 - q_2\gamma_1.$$

Кроме того, заполняется третья строка массива W: $W_{3,1} = \alpha_2$, $W_{3,2} = \beta_2$, $W_{3,3} = \gamma_2$. Вместе с этим оказывается сформированной система (I2) относительно f_3 и f_4: $W_{3,1}f_3 + W_{3,2}f_4 = W_{3,3}$ $W_{4,1}f_3 + W_{4,2}f_4 = W_{4,3}$. Здесь элементы $W_{4,1}$, $W_{4,2}$ и $W_{4,3}$ были определены на (N-4)-м шагу левого хода прогонки. Формулы i-го шага имеют вид: $i = 1,2,\ldots,N-2$,

$$\alpha_i = C_{i+1,2}^{(i)} = p_i C_{i+1,2} - q_i^{\beta}{}_{i,i-1},$$

$$\beta_i = C^{(i)}_{i+1,3} = p_i C_{i+1,3},$$

$$\gamma_i = F^{(i)}_{i+1} = p_i F_{i+1} - q_i \gamma_{i-1},$$

где

$$p_i = \frac{\sigma \alpha_{i-1}}{\sqrt{\alpha^2_{i-1} + C^2_{i+1,1}}}, \quad q_i = \frac{\sigma C_{i+1,1}}{\sqrt{\alpha^2_{i-1} + C^2_{i+1,1}}}, \quad \sigma = \begin{cases} +1, & \alpha_{i-1} \geq 0, \\ -1, & \alpha_{i-1} < 0. \end{cases}$$

При четном i заполняются нечетные строки массива W: $W_{i+1,1} = \alpha_i$, $W_{i+1,2} = \beta_i$, $W_{i+1,3} = \gamma_i$, $i = 2K$, $K = 0,1,\ldots, i \leq N-2$, и, таким образом, формируются системы (I2) относительно f_i и f_{i+1}:

$$\begin{aligned} W_{i,1} f_i + W_{i,2} f_{i+1} &= W_{i,3}, \\ W_{i+1,1} f_i + W_{i+1,2} f_{i+1} &= W_{i+1,3}. \end{aligned} \tag{36}$$

Наконец, если N нечетное, то значения $\alpha_{N-2}, \beta_{N-2}$ и γ_{N-2} присваиваются элементам $(2N_0 - 1)$-й строки массива W: $W_{2N_0-1,1} = \alpha_{N-2}$, $W_{2N_0-1,2} = \beta_{N-2}$, $W_{2N_0-1,3} = \gamma_{N-2}$, чтобы образовать систему (I2) относительно f_{N-1} и f_N: $W_{2N_0-1,1} f_{N-1} + W_{2N_0-1,2} f_N = W_{2N_0-1,3}$, $W_{2N_0,1} f_{N-1} + W_{2N_0,2} f_N = W_{2N_0,3}$. Правый ход прогонки, а вместе с ним и формирование N_0 систем (I2), из которых могут быть найдены все неизвестные f_j, завершен.

Далее, каждая из систем (36) приводится к виду (I9) при помощи ортогонального преобразования отражения. Коэффициенты l_i, m_i, r_i и r_{i+1} вычисляются по формулам:

$$l_i = -\sigma \sqrt{W^2_{i,1} + W^2_{i+1,1}},$$

$$m_i = -pW_{i,2} - qW_{i+1,2}, \quad r_i = -pW_{i,3} - qW_{i+1,3},$$

$$n_i = -qW_{i,2} + pW_{i+1,3}, \quad r_{i+1} = pW_{i+1,3} - qW_{i,3},$$

где

$$p = \frac{\sigma W_{i,1}}{\sqrt{W^2_{i,1} + W^2_{i+1,1}}}, \quad q = \frac{\sigma W_{i+1,1}}{\sqrt{W^2_{i,1} + W^2_{i+1}}}, \quad \sigma = \begin{cases} +1, & W_{i,1} \geq 0, \\ -1, & W_{i,1} < 0. \end{cases}$$

Вместе с f_i и f_{i+1}, которые рассчитываются по формулам $f_{i+1} = r_{i+1}/n_i$, $f_i = (r_i - m_i/f_{i+1})/l_i$, метод позволяет найти оценки аб-

солютной и относительной погрешности решения (см.(28)) каждой из систем и тем самым системы (IO) (см.(29)). Обратим также внимание на возможность вычисления оценки относительной погрешности решения в норме $\|f\|_\infty = \max_k |f_k|$.

ЗАМЕЧАНИЕ. Чтобы перейти к следующему шагу прогонки, требуется построить матрицу отражения $Q[i,i+1]$, или $Q[i+1,i]$, и три отраженных вектора. Однако при этом используется только один из двух элементов отраженного вектора, и, следовательно, достаточно учесть только погрешность его вычисления. При оценке нормы возмущения отраженного вектора в ходе прогонки это обстоятельство позволяет в $\sqrt{2}$ раз уменьшить прежнее значение α, определенное неравенством (9). Нетрудно показать, что в оценках погрешности решения системы (IO) можно считать α равным $8\varepsilon_1$.

Обратимся к характеристике трудоемкости метода. На каждом шагу прогонки, всего $2(N-2)$ шагов, выполняются следующие арифметические операции: 3 сложения + 7 умножений + 2 деления + I извлечение квадратного корня. Кроме того, N_0 раз повторяется 5 сложений + II умножений + 3 деления + I извлечение квадратного корня. На практике мы должны также учесть затраты на вычисление коэффициентов системы (IO), которые во многих случаях являются определяющими. Поэтому основное качество трудоемкости метода выражается в линейной зависимости от N числа затрачиваемых арифметических операций.

9. О б о б щ е н и е м е т о д а. Рассмотрим систему линейных алгебраических уравнений с ленточной матрицей:

$$Cf = F, \qquad (37)$$

или

$$\sum_{j=1}^{M_0} C_{i,j} f_j = F_i, \quad i = 1,2,\dots,M_2,$$

$$\sum_{j=1}^{M_1} C_{i,j} f_\nu = F_i, \quad \nu = i+j-M_2-1, \quad i = M_2+1,\dots,N-M_3,$$

$$\sum_{j=1}^{M_0} C_{i,j} f_s = F_i, \quad s = N-M_0+j, \quad i = N-M_3+1,\dots,N.$$

Здесь M_0, M_1, M_2 и M_3 - параметры ленточного массива размерности $N \times M_1$: M_0 - общее число "граничных" уравнений, или порядок разностной краевой задачи, представленной системой (37), $M_1 = M_0+1$ -

"ширина" ленточного массива, M_2 – число "верхних граничных" уравнений, M_3 – число "нижних граничных" уравнений, $M_2 + M_3 = M_0$. Структура матрицы C иллюстрируется рис. I, где $N = 8$, $M_0 = 4$, $M_1 = 5$, $M_2 = 3$, $M_3 = 1$. Крестиками обозначены элементы ленточного массива, часть которых, составляющих граничные строки, заключены в прямоугольники.

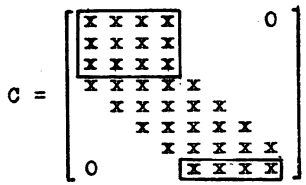

Рис. I

Метод ортогональных встречных прогонок позволяет сформировать N_0 подсистем размерности M_0 каждая, из которых независимо могут быть найдены все элементы вектора f группами по M_0 неизвестных. При этом $N_0 = N/M_0$, если N кратно M_0, и $N_0 = \mathrm{INT}(N/M_0)+1$ – в противном случае. Для запоминания коэффициентов подсистем необходим дополнительный массив W размерности $M_0 N_0 \times M_1$. Как и в случае системы с трехдиагональной матрицей, обусловленность матрицы каждой из подсистем не превышает обусловленности матрицы C. Трудоемкость метода по-прежнему пропорциональна N.

После того как массив W заполнен, требуется выбрать способ решения системы линейных алгебраических уравнений размерности M_0 с заполненной матрицей. В частности, при помощи двусторонних преобразований отражений можно прийти к системе с трехдиагональной матрицей. Этот вариант позволяет наиболее просто оценить погрешности вычислений [3].

Ограничимся определением последовательности из $N-M_0$ преобразований отражений такой, что система

$$UCf = UF, \tag{38}$$

где U – произведение матриц отражений, будет содержать подсистему с матрицей $A^{(i)}$ и правой частью $R^{(i)}$ относительно M_0 неизвестных:

$$A^{(i)} z_i = R^{(i)}, \quad z_i = (f_i, f_{i+1}, \ldots, f_{i+M_0-1}). \tag{39}$$

Для этого воспользуемся обозначениями, аналогичными P[i,i+1] и P[i+1,i].

Пусть a и b – векторы в N-мерном евклидовом пространстве: $a = (a_1, a_2, \ldots, a_N)$, $b = (b_1, b_2, \ldots, b_N)$. Тогда матрицы отражений P[i,i+m] и P[i+m,i], $m \geq 1$, однозначно определяются из условий:

$b = P[i,i+m]\,a$, если

$$b_i = -\sigma\sqrt{a_i^2+\ldots+a_{i+m}^2}\,, \quad |b_i| \neq 0, \quad \sigma = \begin{cases} +1, & a_i \geq 0, \\ -1, & a_i < 0, \end{cases}$$

$$b_{i+s} = 0, \quad s = 1,2,\ldots,m, \quad b_k = a_k, \quad k \neq i, \quad k \neq i+s;$$

и $b = P[i+m,i]\,a$, если

$$b_{i+m} = -\sigma\sqrt{a_i^2+\quad +a_{i+m}^2}\,, \quad |b_{i+m}| \neq 0, \quad \sigma = \begin{cases} +1, & a_{i+m} \geq 0, \\ -1, & a_{i+m} < 0, \end{cases}$$

$b_{i+s}=0$, $s=0,1,\ldots,m-1$, $b_k=a_k$, $k\neq i+m$, $k\neq i+s$.

В этих обозначениях мы опишем правый и левый ходы прогонки как способ построения матрицы U.

Правый ход состоит из $i-1$ шагов:

$$C^{(j)}f = F^{(j)}, \quad j = 1,2,\ldots,i-1,$$

$$C^{(j)} = P[j,j+M_2]C^{(j-1)}, \quad C^{(0)} \equiv C,$$

$$F^{(j)} = P[j,j+M_2]F^{(j-1)}, \quad F^{(0)} \equiv F.$$

Здесь на каждом шагу в качестве вектора a берется j-й вектор-столбец матрицы $C^{(j-1)}$. Из определения $P[j,j+M]$ следует, что строки с номерами $i, i+1, \ldots, i+M_2-1$ системы

$$C^{(i-1)}f = F^{(i-1)} \tag{40}$$

содержат только элементы вектора z_i. На рис.2 схематично изображен результат правого хода прогонки применительно к системе с матрицей C, представленной на рис.I. Целью прогонки является формирование подсистемы относительно f_3, f_4, f_5 и f_6. Звездочкой отмечены возмущенные элементы матрицы C. В прямоугольник заключены строки системы с матрицей $C^{(2)}$, содержащие f_3, f_4, f_5 и f_6. Кроме того, отмечены элементы C, ставшие нулями.

$$C^{(2)} = \begin{bmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & * & * & * \\ 0 & & * & * & * \\ & & x & x & x \\ & & & x & x \\ & & & & x \end{bmatrix}$$

Рис. 2

Левый ход прогонки состоит из k шагов, $k = N+1 - M_0 - i$:

$$\tilde{C}^{(j)}f = \tilde{F}^{(j)}, \quad j = 1,2,\ldots,k,$$

$$\tilde{C}^{(j)} = P[N+1-j,N+1-M_3-j]\tilde{C}^{(j-1)}, \quad \tilde{C}^{(0)} \equiv C^{(i-1)},$$

$$\tilde{F}^{(j)} = P[N+1-j, N+1-M_3-j]\tilde{F}^{(j-1)}, \quad \tilde{F}^{(0)} \equiv F^{(i-1)}.$$

Здесь на каждом шагу в качестве вектора a берется $(N+1-j)$-й вектор-столбец матрицы $\tilde{C}^{(j-1)}$. Очевидно, строки с номерами $i + M_2$, $i+M_2+1, \ldots, i+M_0-1$ системы

$$\tilde{C}^{(k)}_f = \tilde{F}^{(k)} \qquad (41)$$

содержат только элементы вектора z_i. Приведение к (4I) таково, что строки с номерами $1, 2, \ldots, i+M_2-1$ систем (40) и (4I) соответственно совпадают. Поэтому система (4I) содержит подсистему (39). Прогонка завершена, и мы можем выписать представление матрицы U в виде произведения $N-M_0$ матриц отражений:

$$U = P[M_0+i, M_2+i]x\ldots xP[N, N-M_3]xP[i-1, i-1+M_2]x\ldots xP[1, 1+M_2]. \quad (42)$$

$\tilde{C}^{(2)} =$

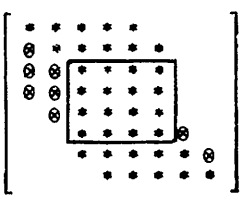

Рис. 3

На рис. 3 схематично представлен результат прогонки в примере, рассматриваемом на рис. I-2. В прямоугольник заключены элементы матрицы подсистемы относительно f_3, f_4, f_5 и f_6.

Заметим, что система (39) может быть сформирована из M_2-х строк системы (40) с номерами $i, i+1, \ldots, i+M_2-1$ и M_3-х строк системы $\tilde{U}Cf = \tilde{U}F$, где $\tilde{U} = P[M_0+i, M_2+i]x\ldots xP[N, N-M_3]$, т.е. правый ход и левый ход прогонки независимы.

I0. Т е с т о в ы е п р и м е р ы. Рассмотрим численное решение 3-х примеров иллюстративного характера с целью продемонстрировать универсальность предлагаемого метода. Каждый из примеров - система из N линейных алгебраических уравнений с трехдиагональной матрицей C, размерность и обусловленность которой не слишком велики. Вычисления были выполнены на системе HP-2000 с параметрами: $\varepsilon_1 = 1{,}1921 \cdot 10^{-7}$, $\varepsilon_2 = 5{,}8775 \cdot 10^{-39}$. Результаты представлены в виде таблиц, выданных на печать ЭВМ, с обозначениями (см. п.п. 6 и 7):

F(I) - значение $\tilde{f}_i$;

ERR[F(I)] - оценка $|\tilde{f}_i - f_i|$, равная ρ_i, $\rho_i = \|\tilde{z}_i - z_i\|$ (см. (28));

COND(L) - значение $\mu(\tilde{L}^{(i)})$, при котором вычислялось $\tilde{f}_i$;

F(I) -T(I) - реальная разность численного и точного решений, т.е. $\tilde{f}_i - f_i$.

Кроме того, вычисляются следующие характеристики решения:

MAX COND(L) - максимальное значение $\mu(\tilde{L}^1)$;

ERR(TØ) - реальное значение $\|\tilde{f}-f\|_\infty / \|\tilde{f}\|_\infty$;

ERR(T2) - реальное значение $\|\tilde{f}-f\| / \|\tilde{f}\|$;

ERR(FØ) - оценка $\|\tilde{f}-f\|_\infty / \|f\|_\infty$, равная $\max \rho_i / \|\tilde{f}\|_\infty$;

ERR(F2) - оценка $\|\tilde{f}-f\| / \|\tilde{f}\|$, равная δ_0 (см. (29)).

При формировании выдачи результатов мы не заботились о том, чтобы печатались только верные знаки.

Приведем еще одну оценку относительной погрешности решения системы (I0), характеризующую другой вариант метода ортогональной прогонки. В результате прямого хода прогонки, состоящего из N-1 шагов правого хода в методе ортогональных встречных прогонок, система (I0) принимает вид: $C^{(N-1)}f = F^{(N-1)}$. Очевидно, согласно (I5), $C^{(N-1)}$ - верхняя треугольная, ленточная матрица с тремя диагоналями [5,6]. Под обратным ходом прогонки понимается последовательное определение элементов вектора f, начиная с f_N. Обусловленность соответствующей $C^{(N-1)}$ возмущенной матрицы, учитывающей погрешности ортогональных преобразований отражений, ограничена сверху величиной μ_*:

$$\mu_* = \frac{1 + 5\sqrt{N}\,\alpha\, e^{5\alpha}}{1-5\sqrt{N}\,\alpha\, e^{5\alpha}\,\mu(C)}\,\mu(C), \quad \alpha = 12\varepsilon_1.$$

Следуя [3], выпишем оценку относительной погрешности:

$$\frac{\|\tilde{f}-f\|}{\|\tilde{f}\|} \le \delta_2 = \frac{\tau_1' + \tau_1''}{1-\tau_1''}, \qquad (43)$$

$$\tau_1' = \frac{5\sqrt{N}\alpha\, e^{5\alpha} + N\alpha\, e^{N\alpha}}{1-5\sqrt{N}\,\alpha\, e^{5\alpha}\,\mu_*}\,\mu_*, \quad \tau_1'' = \varepsilon_1\left(\frac{11\mu_*}{1-9\mu_*} + 1\right).$$

В случае, когда обусловленность матрицы C известна, мы можем сравнить метод ортогональных встречных прогонок и метод ортогональной прогонки при помощи оценок δ_0, δ_1 (см. (29) и (35)) и δ_2. Кроме этих характеристик, мы приведем значения: Δ_1 - реальная относительная погрешность метода ортогональных встречных прогонок и Δ_2 - реальная относительная погрешность метода ортогональной прогонки.

Обратим внимание на соотношение $\mu(C)$ и $\max \mu_i$ – максимальное значение $\mu(\tilde{L}^{(i)})$: $\max \mu_i$ может быть значительно меньше $\mu(C)$. В приведенных примерах $\mu(\tilde{L}^{(i)})$ как функция i имеет различное поведение. В примере I она достигает наибольшего значения к середине интервала по i, а в примере 2 – наименьшего значения, в примере 3 монотонно растет с уменьшением i от N. Представляют интерес данные о "запасе прочности", определяемом различными оценками относительной погрешности в сравнении с реальной относительной погрешностью.

Точное решение тестовых примеров имеет вид: $f_i = \dfrac{1}{2i+1}$, что достигается подбором правых частей рассматриваемых систем по заданной матрице C.

ПРИМЕР I. $N = 20$, $\mu(C) = 178{,}1$,

$$C = \begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ & & 1 & -2 & 1 \\ & & & 1 & -2 \end{bmatrix}.$$

В табл. I численное решение сопоставляется с точным.

Таблица I

MAX COND(L)= 20.9072

F(I)	ERR(F(I))	COND(L)	F(I)-T(I)
.333333	9.26155E-05	12.4821	-1.78914E-07
.2	8.26155E-05	12.4821	-1.49012E-07
.142857	1.31979E-04	15.9649	-1.49012E-07
.111111	1.31979E-04	15.9649	-1.34110E-07
.090909	1.67385E-04	18.5624	-1.19209E-07
.076923	1.67385E-04	18.5624	-1.04309E-07
6.66665E-02	1.89609E-04	20.1513	-1.19209E-07
5.88234E-02	1.89609E-04	20.1513	-1.26660E-07
5.26314E-02	1.93770E-04	20.9071	-1.71363E-07
4.76189E-02	1.93770E-04	20.9071	-1.71363E-07
4.34782E-02	1.94949E-04	20.9072	-1.04309E-07
3.99999E-0	1.94949E-04	20.9072	-9.63575E-08
3.70369E-02	1.79200E-04	20.1513	-1.26660E-07
3.44827E-02	1.79200E-04	20.1513	-1.04309E-07
.032258	1.48599E-04	18.5624	-6.70552E-08
.030303	1.48599E-04	18.5624	-5.96046E-08
2.85714E-02	1.06320E-04	15.9649	-6.33299E-08
.027027	1.06320E-04	15.9649	-5.21541E-08
.025641	5.22179E-05	12.4821	-3.72529E-08
2.43902E-02	5.22179E-05	12.4821	-1.49012E-08

ERR(T0)= 5.36442E-07 ERR(T2)= 2.35149E-06
ERR(F0)= 5.95311E-04 ERR(F2)= 1.02625E-03

Параметры δ_0, δ_1, δ_2, Δ_1 и Δ_2 имеют следующие значения: $\delta_0 = 1{,}03\cdot10^{-3}$, $\delta_1 = 1{,}17\cdot10^{-2}$, $\delta_2 = 1{,}12\cdot10^{-2}$, $\Delta_1 = 2{,}35\cdot10^{-6}$, $\Delta_2 = 2{,}12\cdot10^{-6}$.

В табл. 2 приведены характеристики численного решения примера при различных N; обусловленность $\mu(C)$ вычислялась по формуле: $\mu(C) = \dfrac{1 + \cos[\pi/(N+1)]}{1 - \cos[\pi/(N+1)]}$.

Таблица 2

N	$\mu(C)\times10^{-2}$	$\delta_1\times10^2$	$\delta_2\times10^2$	$\delta_0\times10^3$	max μ_1	$\Delta_1\times10^6$	$\Delta_2\times10^6$
5	0.139	0.0338	0.0341	0.0605	5.95	0.364	0.156
10	0.484	0.217	0.185	0.196	11.0	0.608	0.876
15	1.03	0.708	0.523	0.498	16.0	1.52	1.57
20	1.78	1.71	1.12	1.03	20.9	2.35	2.12
25	2.73	3.50	2.03	1.87	26.0	3.04	2.57
30	3.89	6.47	3.35	3.09	31.0	3.56	3.15
35	5.25	11.2	5.15	4.77	36.0	3.81	3.17
40	6.81	18.9	7.53	6.98	41.0	3.76	4.24
45	8.57	31.4	10.6	9.80	46.0	3.89	5.80
50	10.5	53.3	14.6	13.2	51.0	3.52	7.01
55	12.7	97.7	19.6	17.5	56.0	3.27	7.73
60	15.1		25.8	22.6	61.0	2.69	8.72
100	41.3			102	101	4.54	17.0
150	92.3			341	151	8.47	24.1
200	164			806	201	11.9	36.3
250	255				251	15.3	50.3
300	368	$\delta_1 > 1$	$\delta_2 > 1$	$\delta_0 > 1$	301	18.4	57.4
350	499				351	26.4	73.5
400	655				401	34.7	90.7

ПРИМЕР 2. N = 21, $\mu(C) = 109{,}2$,

$$C = \begin{bmatrix} -1/10 & 1 & & & \\ 1 & 0 & 1 & & \\ \vdots & \vdots & \vdots & \vdots & \\ & & 1 & 0 & 1 \\ & & & 1 & -1/10 \end{bmatrix}.$$

В табл. 3 численное решение сопоставляется с точным.

Таблица 3

MAX COND(L)= 17.5813

F(I)	ERR(F(I))	COND(L)	F(I)-T(I)
.333333	3.05107E-04	17.5813	-1.78814E-07
.2	3.05107E-04	17.5813	-2.98023E-08
.142857	2.79826E-04	13.0373	-8.94070E-08
.111111	2.79826E-04	13.0373	2.98023E-08
9.09091E-02	2.69702E-04	11.2588	-1.49012E-08
7.69231E-02	2.69702E-04	11.2588	-1.49012E-08
6.66667E-02	2.64724E-04	10.4054	1.49012E-08
5.88235E-02	2.64724E-04	10.4054	7.45058E-09
5.26315E-02	2.61786E-04	10.0433	-5.21541E-08
.047619	2.61786E-04	10.0433	-3.72529E-08
4.34783E-02	2.59895E-04	10.0417	0
.04	2.59895E-04	10.0417	4.47035E-08
3.70371E-02	2.59634E-04	10.3998	7.45059E-09
3.44827E-02	2.59634E-04	10.3998	-2.98023E-08
.032258	2.57797E-04	11.247	-6.70552E-08
.030303	2.57797E-04	11.247	1.86265E-08
2.85716E-02	2.57268E-04	13.0118	1.60187E-07
.027027	2.57268E-04	13.0118	-2.23517E-08
2.56409E-02	2.55978E-04	17.5022	-9.31323E-08
2.43903E-02	2.56978E-04	17.5813	1.49012E-08
2.32559E-02	2.55995E-04	17.5813	1.26660E-07

ERR(T0)= 5.36442E-07 ERR(T2)= 1.49415E-06

ERR(F0)= 9.15320E-04 ERR(F2)= 1.87468E-03

Параметры δ_0, δ_1, δ_2, Δ_1 и Δ_2 имеют следующие значения: $\delta_0 = 1{,}87\cdot10^{-3}$, $\delta_1 = 1{,}10\cdot10^{-2}$, $\delta_2 = 7{,}05\cdot10^{-3}$, $\Delta_1 = 1{,}49\cdot10^{-6}$, $\Delta_2 = 1{,}99\cdot10^{-7}$.

ПРИМЕР 3. $N = 6$, $\mu(C) = 509{,}1$,

$$(C) = \begin{bmatrix} 7/5 & 11/3 & & & & \\ 0 & 7/5 & 11/3 & & & \\ \vdots & \vdots & \ddots & \ddots & \ddots & \\ & & & 0 & 7/5 & 11/3 \\ & & & & 0 & 7/5 \end{bmatrix}.$$

В табл.4 численное решение сопоставляется с точным. Параметры δ_0, δ_1, δ_2, Δ_1 и Δ_2 имеют следующие значения: $\delta_0 = 4.03\cdot10^{-3}$, $\delta_1 = 1{,}48\cdot10^{-2}$, $\delta_2 = 1{,}42\cdot10^{-2}$, $\Delta_1 = 3{,}69\cdot10^{-6}$, $\Delta_2 = 2{,}91\cdot10^{-6}$.

Использование ЭВМ с меньшими значениями ε_1 уменьшило бы "запас прочности" оценок, но не радикальным образом.

Таблица 4

MAX COND(L)= 400.114

F(I)	ERR(F(I))	COND(L)	F(I)-T(I)
.333333	1.781196-03	400.114	-6.556516-07
.2	1.781196-03	400.114	2.983236-07
.142857	2.153596-04	58.3575	-1.192096-07
.111111	2.153596-04	58.3575	5.960466-08
9.090916-02	2.715106-05	8.74506	0
7.692316-02	2.715106-05	8.74506	0

ERR(T0)= 1.966966-06 ERR(T2)= 3.698306-06

ERR(F0)= 5.343596-03 ERR(F2)= 4.032076-03

II. З а к л ю ч е н и е. В сравнении с другими универсальными методами решения системы линейных алгебраических уравнений с трехдиагональной матрицей предлагаемый вариант имеет, на наш взгляд, достоинство, которое выражается в возможностях контроля погрешности в различных нормах. При этом затраты, связанные с контролем, вполне соизмеримы с затратами на само решение. В частности, контролем не отвергается случай, когда, как в примере 3, часть элементов вектора f оказывается вычисленной с приемлемой точностью, но погрешность решения системы может не удовлетворять заданным ограничениям.

Для определения оценки относительной погрешности нам не потребовалось привлекать трудоемкую процедуру вычисления обусловленности матрицы системы. Тем не менее приведенные примеры показывают, что найденная относительная погрешность δ_0 (формула (29)), определяемая из решения системы, эффективнее оценки δ_2 (формула (43)), и, видимо, это следует ожидать вообще (ρ_0 меньше ρ_1 (формула (35)) по построению).

В данной работе не ставилась задача исчерпывающего анализа предлагаемого метода. Мы не можем также сказать в настоящее время, что метод опробован на большом числе практических примеров. Но, как нам кажется, указанных обстоятельств достаточно, чтобы оправдать наш интерес к ортогональным встречным прогонкам.

Автор выражает благодарность С.К.Годунову за внимание и интерес к данной работе, содержание которой во многом отражает попытку автора ответить на его вопросы и замечания, а также участникам семинара, руководимого Ю.С.Завьяловым, и семинара, руководимого С.К.Годуновым, за обсуждение работы.

Литература

I. Вычислительные методы линейной алгебры. Библиографический указатель 1928–1974 г. /Под редакцией В.В.Воеводина.– Новосибирск, 1976. – 418 с.

2. САМАРСКИЙ А.А., НИКОЛАЕВ Е.С.Методы решения сеточных уравнений. М.: Наука, 1978. – 592 с.

3. ГОДУНОВ С.К. Решение систем линейных уравнений. Новосибирск: Наука, 1980. – 177 с.

4. БЕЛЛМАН Р. Введение в теорию матриц.–М.: Наука,1976.–352с.

5. УИЛКИНСОН Дж. РАЙНШ С. Справочник алгоритмов на языке АЛГОЛ. Линейная алгебра. –М.: Машиностроение, 1976. – 389 с.

6. ВОЕВОДИН В.В. Численные методы алгебры (теория и алгоритмы).–М.: Наука, 1966.

Поступила в ред.-изд.отд.
15 июня 1983 года