РЕШЕНИЕ КРАЕВЫХ ЗАДАЧ ДЛЯ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

С. В. КУЗНЕЦОВ

Статья посвящена разработке варианта ортогональной прогонки, первоначально предложенной С. К. Годуновым в работе [1]. Отличие ортогональной прогонки, изложенной ниже, от описанной в [1] состоит в том, что вместо ортогонализации Грамма — Шмидта использована ортогонализация с помощью преобразований отражения. Будет доказано, что погрешности округления, допущенные во время вычислений по рассматриваемому алгоритму, не могут сильно исказить его результат. Для этого применен метод, предложенный С. К. Годуновым в [2], для ортогональной прогонки в случае систем разностных уравнений.

Метод ортогональной прогонки предназначен для численного решения краевой задачи

$$\frac{du}{dx} = A(x)u(x) + f(x);$$
$$Bu(x_0) = \varphi; \quad Cu(\bar{x}) = \psi, \tag{1}$$

где B, C — прямоугольные матрицы размерности $k \times n$ и $p \times n$ соответственно:

$$B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & & & \\ \vdots & & & \\ b_{k1} & b_{k2} & \dots & b_{kn} \end{pmatrix}; \quad C = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & & & \\ \vdots & & & \\ c_{p1} & c_{p2} & \dots & c_{pn} \end{pmatrix},$$

причем $k + p = n$ и ранги матриц B и C равны соответственно k и p. Элементы матрицы $A(x)$ размерности $n \times n$ и вектор-функция $f^T(x) = (f_1(x), \dots, f_n(x))$ предполагаются непрерывными на отрезке $[x_0, \bar{x}]$.

Автор выражает искреннюю благодарность С. К. Годунову, под чьим руководством выполнена эта работа.

§ 1. ОПИСАНИЕ АЛГОРИТМА

1. Предварительные замечания. Для решения задачи (1) имеется следующий способ решения. Возьмем полную систему линейно независимых векторов $\tilde{u}_1(x_0)$, $\tilde{u}_2(x_0)$, ..., $\tilde{u}_p(x_0)$, удовлетворяющих граничному условию $Bu(x_0) = 0$. Вектор $\tilde{u}_0(x_0)$ удовлетворяет условию $Bu(x_0) = \varphi$. Нетрудно доказать: если решение задачи (1) существует, то оно представимо в виде $u(x) = u_0(x) + \sum_{j=1}^{p} \alpha_j u_j(x)$, где вектор-функции $u_j(x)$, $j = \overline{0, p}$ — решения следующих задач:

$$\frac{du_j(x)}{dx} = A(x)u_j(x), \quad j = \overline{1, p};$$
$$u_j(x_0) = \tilde{u}_j(x_0);$$
$$\frac{du_0(x)}{dx} = A(x)u_0(x) + f(x);$$
$$u_0(x_0) = \tilde{u}_0(x_0).$$

Численным интегрированием определяем значения $u_0(\bar{x})$, $u_j(\bar{x})$, $j = \overline{1, p}$. Коэффициенты α_j находим из системы

$$\sum_{j=1}^{p} \alpha_j Cu_j(\bar{x}) = \psi - Cu_0(\bar{x}).$$

Действуя так, часто сталкиваются с затруднением: если собственные значения $A(x)$ сильно различаются по величине вещественной части, то в процессе увеличения x при интегрировании система векторов $\{u_1(x), \ldots, u_p(x)\}$ будет все более и более «сплющиваться» и когда придем в точку $x=\bar{x}$, это приведет к тому, что невозможно будет аккуратно определить α_j, а если даже α_j и будут найдены, то значение решения определится с потерей значащих цифр. Чтобы избавиться от вышеперечисленных недостатков, С. К. Годунов в [1] предложил метод ортогональной прогонки. Здесь приведен его модифицированный вариант.

2. Алгоритм. 1. Разобьем отрезок $[x_0, \bar{x}]$ на l участков точками $x_0 = x_0 < x_1 < x_2 < \ldots < x_l = \bar{x}$ так, что

$$|x_k - x_{k-1}| \leqslant C \Big/ \max_{x\in[x_0,\bar{x}]} \|A(x)\|. \tag{1.1}$$

2. Определяем $z_j(x_0)$, $j=\overline{1, p}$ — полную ортогональную систему векторов, удовлетворяющих условию $Bz_j(x_0)=0$. 3. Находим вектор $z_0(x_0)$, ортогональный ко всем $z_j(x_0)$, удовлетворяющий условию $Bz_0(x_0)=\varphi$. 4. Путем численного интегрирования находим векторы

$$\begin{gathered} y_j(x_s) = X(x_s, x_{s-1})\, z_j(x_{s-1}), \quad j=1, 2, \ldots, p; \\ y_0(x_s) = X(x_s, x_{s-1})\, z_0(x_{s-1}) + \\ + X(x_s, x_{s-1}) \int_{x_{s-1}}^{x_s} [X(\xi, x_{s-1})]^{-1} f(\xi)\, d\xi, \end{gathered} \tag{1.2}$$

где $X(x, x_s)$ — матрицант, т. е.

$$\frac{dX(x, x_s)}{dx} = A(x)\, X(x, x_s);$$

$$X(x_s, x_s) = I_n.$$

5. Систему векторов $\{y_1(x_s), \ldots, y_p(x_s)\}$ ортогонализируем способом, который изложен ниже, и получаем векторы $z_1(x_s), \ldots, z_p(x_s)$ и матрицу $\tilde{R}_s$ — размерности $p \times p$ такую, что

$$\tilde{Y}_s = \tilde{Z}_s \tilde{R}_s, \tag{1.3}$$

где $\tilde{Y}_s = [y_1(x_s), \ldots, y_p(x_s)]$, $\tilde{Z}_s = [z_1(x_s), \ldots, z_p(x_s)]$, причем $\tilde{R}_s$ — будет верхнетреугольной матрицей и $\tilde{Z}_s^* \tilde{Z}_s = I_n$. Затем вычисляем вектор

$$z_0(x_s) = y_0(x_s) - \sum_{j=1}^{p} r_{j,p+1}^s z_j(x_s), \tag{1.4}$$

где $r_{i,p+1}^s = (z_i(x_s), y_0(x_s))$.
Тогда, если

$$Y_s = [y_1(x_s)\, y_2(x_s) \ldots y_p(x_s) y_0(x_s)] = [\tilde{Y}_s y_0(x_s)];$$

$$Z_s = [z_1(x_s)\, z_2(x_s) \ldots z_p(x_s)\, z_0(x_s)] = [\tilde{Z}_s z_0(x_s)];$$

$$R_s = \left[\begin{array}{ccc|c} & \tilde{R}_s & & \begin{matrix} r_{1,p+1} \\ r_{2,p+1} \\ r_{p,p+1} \end{matrix} \\ \hline 0 & 0 \ \ldots \ 0 & & 1 \end{array}\right],$$

то

$$Y_s = Z_s R_s. \tag{1.5}$$

6. Решаем систему линейных уравнений:

$$\sum_{j=1}^{p} \alpha_j C z_j(\bar{x}) = \psi - C z_0(\bar{x}).$$

7. Находим значения искомой вектор-функции $u(x)$ в промежуточных точках $x_0, x_1, x_2, \ldots, x_{l-1}, x_l$ следующим образом:

$$\text{Пусть } \beta^{(s)} = \begin{bmatrix} \beta_1^{(s)} \\ \beta_2^{(s)} \\ \vdots \\ \beta_p^{(s)} \\ 1 \end{bmatrix}.$$

Положим $\beta_i^{(l)} = \alpha_i$, $i = \overline{1, p}$. Тогда

$$u(x_s) = z_0(x_s) + \sum_{j=1}^{p} \beta_i^{(s)} z_i(x_s),$$

где $\beta_i^{(s)}$ находятся из системы уравнений $R_{s+1}\beta^{(s)} = \beta^{(s+1)}$.

3. Способ ортогонализации. Пусть $u^{(1)}, u^{(2)}, \ldots, u^{(p)}$ — система n-мерных векторов $(p < n)$, которую надо проортогонализировать. Опишем теперь, как происходит ортогонализация. Составим матрицу $U^{(0)}$ размерности $n \times p$, первый столбец которой есть вектор $\bar{u}^{(0,1)} = u^{(1)}$, второй — $\bar{u}^{(0,2)} = u^{(2)}$ и т. д., т. е.

$$U^{(0)} = \left[\bar{u}^{(0,1)} \ldots \bar{u}^{(0,p)}\right] = \begin{bmatrix} u_{11}^{(0)} & u_{12}^{(0)} & \ldots & u_{1p}^{(0)} \\ u_{21}^{(0)} & & & u_{2p}^{(0)} \\ \vdots & & & \vdots \\ u_{n1}^{(0)} & \ldots & & u_{np}^{(0)} \end{bmatrix}.$$

Положим $\mathscr{P}^{(0)} = I_n$ — квадратная $n \times n$ единичная матрица, строки которой будем обозначать $\hat{p}^{(0,1)}, \hat{p}^{(0,2)}, \ldots, \hat{p}^{(0,n)}$. Определим вектор $q^{(1)}$:

$$\sigma = \begin{cases} 1, & \text{если } \bar{u}_1^{(0,1)} = u_{11}^{(0)} \geqslant 0; \\ -1, & \text{если } \bar{u}_1^{(0,1)} = u_{11}^{(0)} < 0; \end{cases}$$

$$q_1^{(1)} = \bar{u}_1^{(0,1)} + \sigma \sqrt{\sum_{j=1}^{n} \left[\bar{u}_j^{(0,1)}\right]^2} = u_{11}^{(0)} + \sigma \sqrt{\sum_{j=1}^{n} \left[u_{j1}^{(0)}\right]^2};$$

$$q_j^{(1)} = \bar{u}_j^{(0,1)} = u_{j1}^{(0)} \quad \text{при} \quad j = 2, 3, \ldots, n.$$

Пронормируем вектор $q^{(1)}$:

$$\tilde{q}_j^{(1)} = q_j^{(1)} / \sqrt{(q^{(1)}, q^{(1)})/2}, \quad j = 1, 2, \ldots, n,$$

где

$$(q^{(1)}, q^{(1)}) = \sum_{j=1}^{n} \left[q_j^{(1)}\right]^2.$$

Находим векторы $\bar{u}^{(1,1)}, \bar{u}^{(1,2)}, \ldots, \bar{u}^{(1,p)}$ по формуле

$$\bar{u}_j^{(1,i)} = \bar{u}_j^{(0,i)} - (\bar{u}^{(0,i)}, \tilde{q}^{(1)})\tilde{q}_j^{(1)}$$

и векторы $\hat{p}^{(1,1)}, \hat{p}^{(1,2)}, \ldots, \hat{p}^{(1,n)}$, $\hat{p}_j^{(1,i)} = \hat{p}_j^{(0,i)} - (\hat{p}^{(0,i)}, \tilde{q}^{(1)})\tilde{q}_j^{(1)}$. Заметим, что $\bar{u}_j^{(1,1)} = 0$ при $j = 2, 3, \ldots, n$. Пусть $U^{(1)}$ — матрица размерности $n \times p$, столбцы которой суть векторы $\bar{u}^{(1,1)}, \bar{u}^{(1,2)}, \ldots, \bar{u}^{(1,p)}$, т. е.

$$U^{(1)} = \left[\bar{u}^{(1,1)}\bar{u}^{(1,2)} \ldots \bar{u}^{(1,p)}\right] = \begin{bmatrix} u_{11}^{(1)} & u_{12}^{(1)} & \ldots & u_{1p}^{(1)} \\ 0 & u_{22}^{(1)} & \ldots & u_{2p}^{(1)} \\ \vdots & \vdots & & \vdots \\ 0 & u_{n2}^{(1)} & \ldots & u_{np}^{(1)} \end{bmatrix},$$

и $\mathscr{P}^{(1)}$ — матрица $n \times n$, строки которой есть векторы $\widehat{p}^{(1,1)}$, $\widehat{p}^{(1,2)}$, ..., $\widehat{p}^{(1,n)}$, т. е.

$$\mathscr{P}^{(1)} = \begin{bmatrix} \widehat{p}^{(1,1)} \\ \widehat{p}^{(1,2)} \\ \vdots \\ \widehat{p}^{(1,n)} \end{bmatrix}.$$

Заметим: если $Q^{(1)}$ — матрица размерности $n \times n$, элементы которой вычисляются по формуле

$$Q_{ij}^{(1)} = \delta_{ij} - 2q_i^{(1)} q_j^{(1)} / (q^{(1)}, q^{(1)}),$$

то $U^{(1)} = Q^{(1)} U^{(0)}$, а $\mathscr{P}^{(1)} = \mathscr{P}^{(0)} Q^{(1)}$. Эти тождества поясняют смысл сделанного: каждый столбец матрицы $U^{(0)}$ и каждая строка матрицы $\mathscr{P}^{(0)}$ отражены относительно гиперплоскости с нормалью $q^{(1)}$. В дальнейшем будем действовать следующим образом:

1. Находим нормаль к гиперплоскости, относительно которой будет происходить отражение:

$$\sigma = \begin{cases} 1, & \text{если } \bar{u}_k^{(k-1,k)} \geqslant 0; \\ -1, & \text{если } \bar{u}_k^{(k-1,k)} < 0; \end{cases}$$

$$q_1^{(k)} = \ldots = q_{k-1}^{(k)} = 0;$$

$$q_k^{(k)} = \bar{u}_k^{(k-1,k)} + \sigma \sqrt{\sum_{j=k}^{n} \left[\bar{u}_j^{(k-1,k)}\right]^2};$$

$$q_j^{(k)} = \bar{u}_j^{(k-1,k)} \quad \text{при} \quad j = k+1,\ k+2,\ \ldots,\ n.$$

2. Нормируем вектор $q^{(k)}$:

$$\tilde{q}_j^{(k)} = q_j^{(k)} / \sqrt{(q^{(k)}, q^{(k)})/2}, \quad j = \overline{1, n}.$$

3. Вычисляем отраженные векторы:

$$\bar{u}_j^{(k,i)} = \bar{u}_j^{(k-1,i)} - (\bar{u}^{(k-1,i)}, \tilde{q}^{(k)})\, \tilde{q}_j^{(k)},$$

$$j = \overline{1, n}; \quad i = \overline{1, p};$$

$$\widehat{p}_j^{(k,i)} = \widehat{p}_j^{(k-1,i)} - (\widehat{p}^{(k-1,i)}, \tilde{q}^{(k)})\, \tilde{q}^{(k)}; \qquad (1.6)$$

$$i = 1,\ 2,\ \ldots,\ n; \quad j = 1,\ 2,\ \ldots,\ n.$$

Заметим, что $\bar{u}_j^{k,k} = 0$ при $j = k+1,\ k+2,\ \ldots,\ n$. Вследствие этого $\bar{u}^{(l,k)} = \bar{u}^{(k,k)}$ при $l > k$. Если $U^{(k)} = [\bar{u}^{(k,1)} \bar{u}^{(k,2)} \ldots \bar{u}^{(k,p)}]$, то $U^{(k)}$ имеет вид

$$U^{(k)} = \begin{bmatrix} u_{11}^{(k)} & u_{12}^{(k)} & \ldots & & u_{1p}^{(k)} \\ & u_{22}^{(k)} & \ddots & & \vdots \\ & & u_{kk}^{(k)} & & \\ & & & u_{k+1,k+1}^{(k)} & u_{k+1,p}^{(k)} \\ & 0 & & u_{k+2,k+1}^{(k)} & \\ & & & \vdots & \vdots \\ & & & u_{n,k+1}^{(k)} & u_{np}^{(k)} \end{bmatrix},$$

т. е. $u_{ij}^{(k)} = 0$ при $i > j$, $j = 1,\ 2,\ \ldots,\ k$. Если $Q^{(k)}$ — квадратная матрица $n \times n$, элементы которой вычисляются по формулам

$$Q_{ij}^{(k)} = \delta_{ij} - 2q_i^{(k)} q_j^{(k)} / (q^{(k)}, q^{(k)}),$$

то $U^{(k)} = Q^{(k)}U^{(k-1)} = Q^{(k)}Q^{(k-1)} \dots Q^{(1)}U^{(0)}$;

$$\mathscr{P}^{(k)} = \begin{bmatrix} \widehat{p}^{(k,1)} \\ \widehat{p}^{(k,2)} \\ \vdots \\ \widehat{p}^{(k,n)} \end{bmatrix} = Q^{(1)}Q^{(2)} \dots Q^{(k)}. \qquad (1.7)$$

После p преобразований отражения строк матрицы $\mathscr{P}^{(0)}$ и столбцов матрицы $U^{(0)}$, определяемых по формулам (1.6), получим матрицу $U^{(p)}$, имеющую вид

$$U^{(p)} = \begin{bmatrix} u_{11}^{(p)} & u_{12}^{(p)} & \dots & u_{1p}^{(p)} \\ & u_{22}^{(p)} & & \\ & 0 & \ddots & u_{pp}^{(p)} \end{bmatrix}.$$

Вследствие формул (1.7) и того, что $Q^{(j)}Q^{(j)} = I_n$, $j = 1, 2, \dots, p$, получаем $U = U^{(0)} = \mathscr{P}^{(p)}U^{(p)}$. Заметим, что матрицу $U^{(p)}$ можно представить в виде произведения матриц G и $\tilde{R}_1$ размерности $n \times p$ и $p \times p$: $U^{(p)} = G\tilde{R}_1$, где

$$G = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}, \quad \text{т. е. } g_{ij} = \delta_{ij};$$

$$\tilde{R}_1 = \begin{bmatrix} \tilde{r}_{11}^{(1)} & \tilde{r}_{12}^{(1)} & \dots & \tilde{r}_{1p}^{(1)} \\ & \tilde{r}_{22}^{(1)} & & \\ & 0 & \ddots & \tilde{r}_{pp}^{(1)} \end{bmatrix}, \text{ где } \tilde{r}_{ij}^{(1)} = u_{ij}^{(p)},$$

$$i = 1, 2, \dots, p; \quad j = 1, 2, \dots, p.$$

Тогда $U = \mathscr{P}^{(p)}U^{(p)} = \mathscr{P}^{(p)}G\tilde{R}_1 = Z_1\tilde{R}_1$, Z_1 — размерности $n \times p$. Поскольку $\mathscr{P}^{(p)}$ — ортогональная матрица, а столбцы матрицы Z_1 есть первые p столбцов матрицы $\mathscr{P}^{(p)}$ соответственно, получаем $Z_1^*Z_1 = I_p$ — единичную матрицу порядка p.

Пусть $z_1, z_2, \dots, z_p$ — столбцы матрицы Z_1. Из тождества $U = Z_1\tilde{R}_1$ будем иметь

$$z_1 = u^{(1)}/\tilde{r}_{11}^{(1)};$$
$$\dots\dots\dots\dots$$
$$z_j = \left(u^{(j)} - \sum_{k=1}^{j-1} \tilde{r}_{kj}^{(1)} z_j\right) \Big/ \tilde{r}_{jj}^{(1)};$$
$$\dots\dots\dots\dots\dots\dots$$
$$z_p = \left(u^{(p)} - \sum_{k=1}^{p-1} \tilde{r}_{pj}^{(1)} z_j\right) \Big/ \tilde{r}_{pp}^{(1)}.$$

Таким образом, каждый из z_j есть линейная комбинация $u^{(1)}, u^{(2)}, \dots, u^{(p)}$. Векторы $z_1, z_2, \dots, z_p$ и есть векторы, полученные в результате ортогонализации.

§ 2. ОБОСНОВАНИЕ ОРТОГОНАЛЬНОЙ ПРОГОНКИ

1. Понятие хорошей обусловленности. Пусть краевая задача для системы обыкновенных дифференциальных уравнений, сформулированная во введении, имеет решение для любых φ, ψ и $f(x) \in C([x_0, \bar{x}])$.

Тогда решение задачи представляется в виде (см. [4])

$$u(x) = G_1(x)\varphi + G_2(x)\psi + \int_{x_0}^{\bar{x}} G_3(x, \xi) f(\xi)\, d\xi, \tag{2.1}$$

где $G_1(x)$ — матрица размерности $n \times k$, решение задачи

$$\frac{dG_1(x)}{dx} = A(x) G_1(x);$$

$$BG_1(x_0) = I_k, \quad CG_1(\bar{x}) = 0;$$

$G_2(x)$ — матрица размерности $n \times p$, решение задачи

$$\frac{dG_2(x)}{dx} = A(x) G_2(x);$$

$$BG_2(x_0) = 0, \quad CG_2(\bar{x}) = I_p$$

и, наконец, матрица $G_3(x, \xi)$ размерности $n \times n$ — решение задачи

$$\frac{dG_3(x, \xi)}{dx} = A(x) G_3(x, \xi);$$

$$G_3(\xi + 0, \xi) - G_3(\xi - 0, \xi) = I_n.$$

Будем говорить, что задача (1) хорошо обусловлена, если для любых x и ξ из интервала $[x_0, \bar{x}]$ верны оценки

$$\|G_1(x)\| \leqslant K, \quad \|G_2(x)\| \leqslant K, \quad \|G_3(x, \xi)\| \leqslant K.$$

Из представления (2.1) вытекает, что для решения хорошо обусловленной задачи (1) выполнена оценка

$$\max_{x \in [x_0, \bar{x}]} \|u(x)\| \leqslant K\left(\|\varphi\| + \|\psi\| + (\bar{x} - x_0) \max_{x \in [x_0, \bar{x}]} \|f(x)\|\right). \tag{2.2}$$

Покажем, что из (2.2) вытекает: решение задачи (1) устойчиво по отношению к возмущениям.

Теорема. *Пусть $A(x)$ — матрица $n \times n$, $f(x)$ — вектор-функция размерности n, элементы которых кусочно-непрерывные функции, имеющие конечное число разрывов в точках $x_1, x_2, \ldots, x_{l-1}$, причём на участках $(x_0, x_1), (x_1, x_2), \ldots, (x_{l-1}, \bar{x})$ элементы $A(x)$ и $f(x)$ непрерывны. Пусть также выполнены оценки*

$$\|B - \tilde{B}\| \leqslant \varepsilon, \quad \|C - \tilde{C}\| \leqslant \varepsilon, \quad \|\varphi - \tilde{\varphi}\| \leqslant \varepsilon, \quad \|\psi - \tilde{\psi}\| \leqslant \varepsilon;$$

$$\max_{x \in [x_0, \bar{x}]} \|A(x) - \tilde{A}(x)\| \leqslant \varepsilon; \quad \max_{x \in [x_0, \bar{x}]} \|f(x) - \tilde{f}(x)\| \leqslant \varepsilon, \tag{2.3}$$

а $\tilde{u}(x)$ — решение задачи

$$\frac{d\tilde{u}(x)}{dx} = \tilde{A}(x)\tilde{u}(x) + \tilde{f}(x);$$

$$\tilde{C}\tilde{u}(x) = \tilde{\varphi}; \quad B\tilde{u}(\bar{x}) = \tilde{\psi}. \tag{2.4}$$

Тогда при достаточно малом ε будет выполнена оценка

$$\max_{x \in [x_0, \bar{x}]} \|u(x) - \tilde{u}(x)\| \leqslant \mu\varepsilon,$$

где $u(x)$ — решение задачи (1), а константа μ зависит от

$$K, \|\varphi\|, \|\psi\|, \max_{x \in [x_0, \bar{x}]} \|f(x)\|, (\bar{x} - x_0).$$

Прежде чем приступить к доказательству, заметим, что решение задачи (1) с кусочно-непрерывной $f(x)$ также записывается в виде (2.1).

Действительно, пусть $f_j(x)$ — фундаментальная последовательность непрерывных функций, сходящаяся к $f(x)$. Пусть $u_j(x)$ — решение задачи

$$\frac{du_j(x)}{dx} = A(x)u_j(x) + f_j(x);$$

$$Bu_j(x_0) = \varphi; \quad Cu_j(\bar{x}) = \psi,$$

а $f_j(x)$ такая последовательность, что

$$\max_{x\in[x_0,\bar{x}]} \|f_j(x) - f_{j-1}(x)\| \leqslant 1/2^j;$$

$$\max_{x\in[x_0,\bar{x}]} \|f(x) - f_j(x)\| \to 0 \text{ при } j \to \infty.$$

Тогда

$$u_j(x) = G_1(x)\varphi + G_2(x)\psi + \int_{x_0}^{\bar{x}} G_3(x,\xi) f_j(\xi)\, d\xi. \tag{2.5}$$

Отсюда

$$\max_{x\in[x_0,\bar{x}]} \|u_j(x) - u_{j-1}(x)\| \leqslant K(\bar{x} - x_0)/2^j;$$

$$\max_{x\in[x_0,\bar{x}]} \|u_j(x)\| \leqslant K\left(\|\varphi\| + \|\psi\| + (\bar{x} - x_0) \max_{x\in[x_0,\bar{x}]} \|f(x)\| + 1\right)$$

при достаточно большом j. Тогда последовательность $u_j(x)$ сходится к некоторой вектор-функции $u(x)$. Переходя к пределу в уравнении, получаем, что $u(x)$ — решение задачи $du(x)/dx = A(x)u(x) + f(x)$; $Bu(x_0) = \varphi$; $Cu(\bar{x}) = \psi$. Переходя к пределу в тождестве (2.5), имеем

$$u(x) = G_1(x)\varphi + G_2(x)\psi + \int_{x_0}^{\bar{x}} G_3(x,\xi) f(\xi)\, d\xi.$$

Доказательство теоремы. Перепишем задачу (2.4) следующим образом:

$$\frac{d\tilde{u}(x)}{dx} = A(x)\tilde{u}(x) + f(x) + (\tilde{A}(x) - A(x))\tilde{u}(x) + (\tilde{f}(x) - f(x));$$

$$B\tilde{u}(x_0) = \varphi + (\tilde{\varphi} - \varphi) + (B - \tilde{B})\tilde{u}(x_0);$$

$$C\tilde{u}(\bar{x}) = \psi + (\tilde{\psi} - \psi) + (C - \tilde{C})\tilde{u}(\bar{x}).$$

Воспользуемся методом последовательных приближений. Обозначим через $\tilde{u}^{(0)}(x)$ решение задачи

$$\frac{d\tilde{u}^{(0)}(x)}{dx} = A(x)\tilde{u}^{(0)}(x) + f(x) + (\tilde{f}(x) - f(x));$$

$$B\tilde{u}^{(0)}(x_0) = \varphi + (\tilde{\varphi} - \varphi); \quad C\tilde{u}(\bar{x}) = \psi + (\tilde{\psi} - \psi).$$

Пользуясь линейностью системы и неравенством (2.2), получаем

$$\max_{x\in[x_0,\bar{x}]} \|u(x) - \tilde{u}^{(0)}(x)\| = \max_{x\in[x_0,\bar{x}]} \left\| G_1(x)(\tilde{\varphi} - \varphi) + G_2(x)(\tilde{\psi} - \psi) + \right.$$

$$\left. + \int_{x_0}^{\bar{x}} G_3(x,\xi)(\tilde{f}(\xi) - f(\xi))\, d\xi \right\| \leqslant K\varepsilon(2 + (\bar{x} - x_0));$$

$$\max_{x\in[x_0,\bar{x}]} \|\tilde{u}^{(0)}(x)\| \leqslant K\left[\|\varphi\| + \|\psi\| + (\bar{x} - x_0) \max_{x\in[x_0,\bar{x}]} \|f(x)\| + (2 + \bar{x} - x_0)\varepsilon\right].$$

Вектор-функции $u^{(r)}(x)$, $x \in [x_0, \bar{x}]$ определим рекуррентно как решения краевых задач

$$\frac{du^{(r)}(x)}{dx} = A(x)\tilde{u}^{(r)}(x) + f(x) + (f(x) - \tilde{f}(x)) + (\tilde{A}(x) - A(x))\tilde{u}^{(r-1)}(x);$$

$$B\tilde{u}^{(r)}(x_0) = \varphi + (\tilde{\varphi} - \varphi) + (B - \tilde{B})\tilde{u}^{(r-1)}(x_0);$$

$$C\tilde{u}^{(r)}(\bar{x}) = \psi + (\tilde{\psi} - \psi) + (C - \tilde{C})\tilde{u}^{(r-1)}(\bar{x}).$$

Тогда для $u^{(r)}(x) - \tilde{u}^{(r-1)}(x)$ имеют место оценки

$$\max_{x \in [x_0, \bar{x}]} \|\tilde{u}^{(1)}(x) - \tilde{u}^{(0)}(x)\| \leqslant K\varepsilon(2 + \bar{x} - x_0) \max_{x \in [x_0, \bar{x}]} \|\tilde{u}^{(0)}(x)\|;$$

$$\max_{x \in [x_0, \bar{x}]} \|\tilde{u}^{(r)}(x) - \tilde{u}^{(r-1)}(x)\| \leqslant K\varepsilon(2 + \bar{x} - x_0) \max_{x \in [x_0, \bar{x}]} \|\tilde{u}^{(r-1)}(x) - \tilde{u}^{(r-2)}(x)\|.$$

Обозначим $S = \max\{2, \bar{x} - x_0\}$, тогда предыдущие оценки приобретут вид

$$\max_{x \in [x_0, \bar{x}]} \|\tilde{u}^{(r)}(x) - \tilde{u}^{(0)}(x)\| \leqslant 2SK\varepsilon \max_{x \in [x_0, \bar{x}]} \|\tilde{u}^{(0)}(x)\|;$$

$$\max_{x \in [x_0, \bar{x}]} \|\tilde{u}^{(r)}(x) - \tilde{u}^{(r-1)}(x)\| \leqslant 2SK\varepsilon \max_{x \in [x_0, \bar{x}]} \|\tilde{u}^{(r-1)}(x) - \tilde{u}^{(r-2)}(x)\|.$$

Из этих оценок для достаточно малых ε ($\varepsilon < 1/2SK$) вытекают неравенства:

$$\max_{x \in [x_0, \bar{x}]} \|u^{(r)}(x) - \tilde{u}^{(r-1)}(x)\| \leqslant \max_{x \in [x_0, \bar{x}]} \|\tilde{u}^{(0)}(x)\|;$$

$$\max_{x \in [x_0, \bar{x}]} \|\tilde{u}^{(r)}(x)\| = \max_{x \in [x_0, \bar{x}]} \left\| \sum_{i=1}^{r} (\tilde{u}^{(i)}(x) - \tilde{u}^{(i-1)}(x)) + \tilde{u}^{(0)}(x) \right\| \leqslant$$

$$\leqslant (1 - 1/2^r) \max_{x \in [x_0, \bar{x}]} \|\tilde{u}^{(0)}(x)\| / (1 - 1/2).$$

Эти неравенства показывают, что при $r \to \infty$ $\tilde{u}^{(r)}(x)$ сходится к некоторой вектор-функции $\tilde{u}(x)$, которая является решением задачи (2.4). Для $\tilde{u}(x)$ справедливы оценки

$$\max_{x \in [x_0, \bar{x}]} \|\tilde{u}(x)\| \leqslant 2 \max_{x \in [x_0, \bar{x}]} \|\tilde{u}^{(0)}(x)\| \leqslant 2K \left[\|\varphi\| + \|\psi\| + \right.$$

$$\left. + (\bar{x} - x_0) \max_{x \in [x_0, \bar{x}]} \|f(x)\| + \varepsilon(2 + \bar{x} - x_0) \right].$$

Пусть $\varepsilon < \min\{1/2SK, 1/(2 + (\bar{x} - x_0))\}$. Тогда предыдущее неравенство можно переписать следующим образом:

$$\max_{x \in [x_0, \bar{x}]} \|\tilde{u}(x)\| \leqslant 2K \left(\|\varphi\| + \|\psi\| + (\bar{x} - x_0) \max_{x \in [x_0, \bar{x}]} \|f(x)\| + 1 \right).$$

Кроме того,

$$\max_{x \in [x_0, \bar{x}]} \|u(x) - \tilde{u}(x)\| \leqslant \varepsilon \left[2K + 2K^2 (1 + (\bar{x} - x_0)) \left(\|\varphi\| + \|\psi\| + \right. \right.$$

$$\left. \left. + (\bar{x} - x_0) \max_{x \in [x_0, \bar{x}]} \|f(x)\| + 1 \right) \right],$$

что и требовалось доказать.

2. Некоторые свойства матриц ортогонализации. В этом пункте исходя из хорошей обусловленности задачи (1) получены некоторые оценки для элементов матриц R_s (см. п. 2 § 1), которые нам потребуются в дальнейшем. Сначала покажем, что элементы R_s для любого s

ограничены сверху. Поскольку $y_j(x_s) = X(x_s, x_{s-1}) z_j(x_{s-1})$, $j = 1, 2, \ldots, p$;

$$\|y_j(x_s)\| \leqslant \|X(x_s, x_{s-1})\| \|z_j(x_{s-1})\| = \|X(x_s, x_{s-1})\|,$$

так как $\|z_j(x_{s-1})\| = 1$ для любого j. В силу выбора шага $|x_s - x_{s-1}| \leqslant \leqslant C \Big/ \max\limits_{x \in [x_0, \overline{x}]} \|A(x)\|$, тогда $\|X(x_s, x_{s-1})\| \leqslant e^C$. С другой стороны, в силу (1.5) получаем

$$\|y_j(x_s)\| = \sqrt{\sum_{k=1}^{p} [r_{kj}^{(s)}]^2} \leqslant e^C, \quad j = 1, 2, \ldots, p.$$

Отсюда $|r_{ij}^{(s)}| \leqslant e^C$, $i, j = 1, 2, \ldots, p$. Из (1.4) следует, что

$$|r_{j,p+1}^{(s)}| = |(y_0(x_s), z_j(x_s))| \leqslant \|y_0(x_s)\|.$$

Исходя из тождества (1.2) и того, что $\|X(x_s, x_{s-1})\| \leqslant e^C$ и $\|[X(x_s, x_{s-1})]^{-1}\| \leqslant e^C$, нетрудно оценить $\|y_0(x_s)\|$ сверху:

$$\|y_0(x_s)\| \leqslant \|X(x_s, x_{s-1})\| \|z_0(x_{s-1})\| +$$

$$+ (x_s - x_{s-1}) \|X(x_s, x_{s-1})\| \max_{x \in [x_0, \overline{x}]} \|[X(x, x_{s-1})]^{-1}\| \max_{x \in [x_{s-1}, x_s]} \|f(x)\| \leqslant$$

$$\leqslant e^C \|z_0(x_{s-1})\| + e^{2C} (x_s - x_{s-1}) \max_{x \in [x_{s-1}, x_s]} \|f(x)\|. \tag{2.6}$$

Осталось показать, что $\|z_0(x_{s-1})\|$ ограничена для любого s. Рассмотрим следующую задачу:

$$\frac{d\tilde{u}(x)}{dx} = A(x)\tilde{u}(x) + f(x);$$

$$Bu(x_0) = \varphi; \; Cu(\overline{x}) = 0.$$

Тогда, учитывая (2.2), имеем оценку

$$\max_{x \in [x_0, \overline{x}]} \|\tilde{u}(x)\| \leqslant K \left(\|\varphi\| + (\overline{x} - x_0) \max_{x \in [x_0, \overline{x}]} \|f(x)\| \right).$$

С другой стороны,

$$\tilde{u}(x_{s-1}) = z_0(x_{s-1}) + \sum_{j=1}^{p} \alpha_j z_j(x_{s-1}),$$

где α_j определяются из системы линейных уравнений

$$\sum_{j=1}^{p} \alpha_j C\tilde{z}_j(\overline{x}) = -C\tilde{z}_0(\overline{x}).$$

Здесь $\tilde{z}_j(\overline{x}) = X(\overline{x}, x_{s-1}) z_j(x_{s-1})$, $j = 1, 2, \ldots, p$;

$$\tilde{z}_0(\overline{x}) = X(\overline{x}, x_{s-1}) z_0(x_{s-1}) + X(\overline{x}, x_{s-1}) \int_{x_{s-1}}^{\overline{x}} [X(\xi, x_{s-1})]^{-1} f(\xi)\, d\xi.$$

Отсюда в силу ортогональности $z_1(x_{s-1})$, $z_2(x_{s-1})$, $\ldots$, $z_p(x_{s-1})$, $z_0(x_{s-1})$ получаем

$$\|z_0(x_{s-1})\| \leqslant \|\tilde{u}(x_{s-1})\| \leqslant K \left(\|\varphi\| + (\overline{x} - x_0) \max_{x \in [x_0, \overline{x}]} \|f(x)\| \right). \tag{2.7}$$

Из (2.6) и (2.7) вытекает искомая оценка

$$|r_{i,p+1}^{(s)}| \leqslant \|y_0(x_s)\| \leqslant K e^C \left(\|\varphi\| + (\overline{x} - x_0) \max_{x \in [x_0, \overline{x}]} \|f(x)\| \right) +$$

$$+ e^{2C} (x_s - x_{s-1}) \max_{x \in [x_{s-1}, x_s]} \|f(x)\|. \tag{2.8}$$

Теперь покажем, что все диагональные элементы матриц R_s ограничены снизу. На самом деле достаточно показать ограниченность снизу диагональных элементов матриц $\widetilde{R}_s$, упомянутых в п. 2 § 1, поскольку $r_{p+1,\,p+1}=1$. Пусть $\sigma_1(\widetilde{R}_s)$ наименьшее сингулярное число матрицы $\widetilde{R}_s$, тогда

$$\sigma_1(\widetilde{R}_s)=\min_{\substack{\|x\|=1\\ x\in R^p}}\|\widetilde{R}_s x\|.$$

(см. [3]). Из этого равенства следует, что $|r_{ii}^{(s)}|\geqslant\sigma_1(\widetilde{R}_s)$, $i=1,\,2,\,\ldots$ $\ldots,\,p$. Действительно, набор чисел $\{r_{ii}^{(s)}\}_{i=1}^{p}$ — собственные числа матрицы $\widetilde{R}_s$, так как $\widetilde{R}_s$ — верхнетреугольная матрица. Пусть x_i — собственный вектор, отвечающий $r_{ii}^{(s)}$. Положим $\widetilde{x}_i=x_i/\|x_i\|$, тогда $\|\widetilde{x}_i\|=1$, $i=1,\,2,\,\ldots$ $\ldots,\,p$. Отсюда вытекает

$$\sigma_1(\widetilde{R}_s)=\min_{\substack{\|x\|=1\\ x\in R^p}}\|\widetilde{R}_s x\|\leqslant\|\widetilde{R}_s\widetilde{x}_i\|=|r_{ii}^{(s)}|,\quad i=1,\,2,\,\ldots,\,p.$$

С другой стороны, имеем

$$\|\widetilde{R}_s x\|=\|\widetilde{Z}_s^* Y_s x\|=\|\widetilde{Y}_s x\|=\|X(x_s,x_{s-1})\,\widetilde{Z}_{s-1}x\|=$$
$$=\|X(x_s,x_{s-1})\,x'\|,\ \text{где } x'=\widetilde{Z}_{s-1}x. \tag{2.9}$$

Выше использовано тождество $\widetilde{Y}_s=Z_s\widetilde{R}_s$ и то, что $\widetilde{Z}_s^*\widetilde{Z}_s=I_p$. Заметим, что $\|x'\|=1$, так как $\widetilde{Z}_{s-1}^*\widetilde{Z}_{s-1}=I_p$. Пользуясь тождеством (2.9), легко получить

$$\sigma_1(\widetilde{R}_s)=\min_{\substack{\|x\|=1\\ x\in R^p}}\|\widetilde{R}_s x\|\geqslant\min_{\substack{\|x\|=1\\ x\in R^n}}\|X(x_s,x_{s-1})\,x\|=\sigma_1(X(x_s,x_{s-1})).$$

Ввиду неравенства

$$\|[X(x_s,\,x_{s-1})]^{-1}\|=1/\sigma_1(X(x_s,\,x_{s-1}))\leqslant e^c,$$

которое уже неоднократно использовалось,

$$\sigma_1(\widetilde{R}_s)\geqslant\sigma_1(X(x_s,\,x_{s-1}))\geqslant e^{-c}.$$

Следовательно, исходя из (2.8), имеем $|r_{ii}^{(s)}|\geqslant e^{-c}$, что и необходимо было показать.

3. Вспомогательное утверждение. В этом пункте докажем одно утверждение, которое потребуется в дальнейшем.

Пусть даны:

1. *Ортогональная система векторов* $\{z_1,\,z_2,\,\ldots,\,z_p\}$ *размерности* n, *причем* $p<n$, *т. е.* $(z_i,\,z_j)=0$ *при* $i\neq j$ *и* $\|z_j\|=1$, $j=1,\,2,\,\ldots,\,p$.

2. *Система векторов* $\{\eta_1,\,\eta_2,\,\ldots,\,\eta_p\}$ *размерности* n *такая, что* $\|\eta_i\|\leqslant\varepsilon$, *где* ε *мало. Тогда докажем, что существует матрица* L *размерности* $n\times n$ *такая, что*

$$Mz_j+\eta_j=Lz_j,\quad j=1,\,2,\,\ldots,\,p, \tag{2.10}$$

и $\|L-M\|$ *порядка* ε.

Доказательство этого утверждения будет заключаться в приведении способа нахождения матрицы, обладающей вышеперечисленными свойствами.

Пусть $m_1,\,m_2,\,\ldots,\,m_n$ — строки матрицы M, а $l_1,\,l_2,\,\ldots,\,l_n$ — строки матрицы L. Кроме того, Z — матрица размерности $p\times n$, строки которой суть векторы $z_1,\,z_2,\,\ldots,\,z_p$ соответственно; Y — матрица размерности $p\times n$, строки которой есть векторы $\eta_1,\,\eta_2,\,\ldots,\,\eta_p$ соответственно. Тогда систему линейных уравнений $Mz_j+\eta_j=Lz_j$, $j=1,\,2,\,\ldots,\,p$, в которой

неизвестными являются элементы матрицы L, можно переписать следующим образом:

$$\begin{aligned} Zm_1 &= Zl_1 - Y; \\ Zm_2 &= Zl_2 - Y; \\ &\ldots\ldots \\ Zm_n &= Zl_n - Y. \end{aligned} \tag{2.11}$$

Заметим, что в системе линейных уравнений (2.11) n^2 неизвестных, а количество уравнений pn. Поскольку $p < n$, система линейных уравнений неполна. Добавим к системе (2.11) $n^2 - pn$ уравнений так, чтобы система линейных уравнений оказалась полной.

Пусть $z_{p+1}, z_{p+2}, \ldots, z_n$, $\|z_j\| = 1$, $j = p+1, p+2, \ldots, n$,— такие векторы, что $z_1, z_2, \ldots, z_n$ — ортогональный базис пространства R^n. Тогда рассмотрим систему линейных уравнений

$$\begin{aligned} \tilde{Z}m_1 + \tilde{Y} &= \tilde{Z}l_1; \\ \tilde{Z}m_2 + \tilde{Y} &= \tilde{Z}l_2; \\ &\ldots\ldots \\ \tilde{Z}m_n + \tilde{Y} &= \tilde{Z}l_n, \end{aligned}$$

где $\tilde{Z}$ — матрица $n \times n$, строки которой суть соответственно векторы $z_1, z_2, \ldots, z_n$, а матрица $\tilde{Y}$ размерности $n \times n$, первые p строк которой суть соответственно векторы $\eta_1, \eta_2, \ldots, \eta_p$, остальные строки нулевые. Заметим, что

$$\|\tilde{Y}\| = \max_{\substack{\|x\|=1 \\ x \in R^n}} \|\tilde{Y}x\| \leqslant p \max_{\substack{x \in R^n \\ \|x\|=1}} |(\eta_i, x)| \leqslant p\varepsilon, \quad i = 1, 2, \ldots, p,$$

и $\|\tilde{Z}^{-1}\| = 1$, так как в силу ортогональности $z_1, z_2, \ldots, z_n$ и того, что $\|z_j\| = 1$ для любого $j = 1, 2, \ldots, p$, имеем $\tilde{Z}^*\tilde{Z} = I_n$, поэтому $\|\tilde{Z}^{-1}\| = \|\tilde{Z}^*\| = \|\tilde{Z}\| = 1$. Используя эти оценки, получаем

$$\|L - M\| = \|(M\tilde{Z}^* + \tilde{Y})[\tilde{Z}^*]^{-1} - M\| \leqslant \|\tilde{Y}\| \|(\tilde{Z}^*)^{-1}\| = p\varepsilon.$$

Таким образом, утверждение доказано.

4. Влияние вычислительных погрешностей при прямой прогонке. Пусть $z_1(x_0), z_2(x_0), \ldots, z_p(x_0)$ — полная ортогональная система векторов, удовлетворяющих условию $Bz_j(x_0) = 0$, $j = 1, 2, \ldots, p$, а $z_0(x_0)$ — ортогональный к ним вектор, подчиненный неоднородному уравнению $Bz_0(x_0) = \varphi$. Предположим, что вместо этих векторов найдены близкие к ним

$$\begin{aligned} \bar{z}_1(x_0) &= z_1(x_0) + \zeta_0^{(1)}; \\ &\ldots\ldots\ldots\ldots \\ \bar{z}_p(x_0) &= z_p(x_0) + \zeta_0^{(p)}; \\ \bar{z}_0(x_0) &= z_0(x_0) + \zeta_0^{(0)}, \end{aligned}$$

где $\zeta_0^{(j)}$ погрешности и $\|\zeta_0^{(j)}\| \leqslant \varepsilon_3$. Приведем формулы s шага прямой прогонки с учетом совершаемой погрешности

$$\bar{y}_0(x_s) = X(x_s, x_{s-1})\bar{\bar{z}}_0(x_{s-1}) + X(x_s, x_{s-1}) \int_{x_{s-1}}^{x_s} [X(\xi, x_{s-1})]^{-1} f(\xi)\, d\xi + \eta_s^{(0)}; \tag{2.12}$$

$$\bar{y}_j(x_s) = X(x_s, x_{s-1})\bar{\bar{z}}_j(x_{s-1}) + \eta_s^{(j)},$$

где $\eta_s^{(j)}, \eta_s^{(0)}$ — погрешности интегрирования, причем

$$\|\eta_s^{(j)}\| \leqslant \varepsilon_4; \quad \|\eta_s^{(0)}\| \leqslant \varepsilon_4.$$

Далее, несколько схематизируем вычислительный процесс, предположив, что ортогонализация векторов $\bar{y}_1(x_s), \bar{y}_2(x_s), \ldots, \bar{y}_p(x_s), \bar{y}_0(x_s)$ проводится по точным формулам, указанным в п. 3 § 1, после чего ее результат искажается внесением погрешностей $\theta_{pq}^{(s)}$, $\zeta_s^{(j)}$, где $\|\zeta_s^{(j)}\| \leqslant \varepsilon_3$ и $\|\theta_i^{(s)}\| \leqslant \varepsilon_5$,

$$\Theta^{(s)} = \begin{bmatrix} \theta_{11}^{(s)} & \ldots & \theta_{1p+1}^{(s)} \\ 0 & \ddots & \\ & & \theta_{p+1,p+1}^{(s)} \end{bmatrix}, \quad \theta_i^{(s)} \text{ — столбцы матрицы } \Theta^{(s)}.$$

Причем будем считать $\theta_{p+1,p+1}^{(s)} = 0$, так как $\bar{r}_{p+1,p+1}^{(s)} = 1$ для любого s. Таким образом, имеем

$$\begin{aligned} \bar{\bar{r}}_{pq}^{(s)} &= \bar{r}_{pq}^{(s)} + \theta_{pq}^{(s)}; \\ \bar{\bar{z}}_j(x_s) &= \bar{z}_j(x_s) + \zeta_s^{(j)}. \end{aligned} \tag{2.13}$$

Подставим в формулы (2.12) вместо $\bar{\bar{z}}_j(x_{s-1})$ их выражения из (2.13), получим

$$\bar{y}_0(x_s) = X(x_s, x_{s-1})\,\bar{z}_0(x_{s-1}) + X(x_s, x_{s-1})\,\zeta_s^{(0)} + \eta_s^{(0)} +$$

$$+ X(x_s, x_{s-1}) \int_{x_{s-1}}^{x_s} [X(\xi, x_{s-1})]^{-1} f(\xi)\, d\xi = X(x_s, x_{s-1})\,\bar{z}_0(x_{s-1}) +$$

$$+ \bar{\eta}_s^{(0)} + X(x_s, x_{s-1}) \int_{x_{s-1}}^{x_s} [X(\xi, x_{s-1})]^{-1} f(\xi)\, d\xi;$$

$$\bar{y}_j(x_s) = X(x_s, x_{s-1})\,\bar{z}_j(x_{s-1}) + X(x_s, x_{s-1})\,\zeta_s^{(j)} + \eta_s^{(j)} =$$

$$= X(x_s, x_{s-1})\,\bar{z}_j(x_{s-1}) + \bar{\eta}_s^{(j)}.$$

Тогда

$$\|\bar{\eta}_s^{(j)}\| = \|X(x_s, x_{s-1})\,\zeta_s^{(j)} + \eta_s^{(j)}\| \leqslant e^C \varepsilon_3 + \varepsilon_4 = \varepsilon_6;$$

$$\|\bar{\eta}_s^{(0)}\| = \|X(x_s, x_{s-1})\,\zeta_s^{(0)} + \eta_s^{(0)}\| \leqslant e^C \varepsilon_3 + \varepsilon_4 = \varepsilon_6.$$

Из утверждения предыдущего пункта следует, что существует такая матрица $\tilde{X}(x_s, x_{s-1})$, что

$$X(x_s, x_{s-1})\,\bar{z}_j(x_{s-1}) + \bar{\eta}_s^{(j)} = \tilde{X}(x_s, x_{s-1})\,\bar{z}_j(x_{s-1})$$

и

$$\|\tilde{X}(x_s, x_{s-1}) - X(x_s, x'_{s-1})\| \leqslant p\varepsilon_6.$$

Для остальных $x \in [x_{s-1}, x_s]$ полагаем

$$\tilde{X}(x_s, x_{s-1}) = X(x, x_{s-1}) + \frac{(x - x_{s-1})}{(x_s - x_{s-1})}\left(\tilde{X}(x_s, x_{s-1}) - X(x_s, x_{s-1})\right). \tag{2.14}$$

Продифференцируем это тождество по x и в результате будем иметь

$$\frac{d}{dx}\tilde{X}(x, x_{s-1}) = \frac{d}{dx}X(x, x_{s-1}) + \left(\tilde{X}(x_s, x_{s-1}) - X(x_s, x_{s-1})\right)/(x_s - x_{s-1}) =$$

$$= A(x)\,X(x, x_{s-1}) + \left(\tilde{X}(x_s, x_{s-1}) - X(x_s, x_{s-1})\right)/(x_s - x_{s-1}).$$

Нетрудно проверить, что левая часть этого тождества есть $(A(x) + A_s(x))\,\tilde{X}(x, x_{s-1})$, где

$$A_s(x) = [(\tilde{X}(x_s, x_{s-1}) - X(x_s, x_{s-1}))/(x_s - x_{s-1}) - (\tilde{X}(x_s, x_{s-1}) -$$

$$- X(x_s, x_{s-1}))\,A(x)\,(x - x_{s-1})/(x_s - x_{s-1})][\tilde{X}(x, x_{s-1})]^{-1}. \tag{2.15}$$

Оценим погрешность возмущения матрицы $A(x)$:

$$\max_{x\in[x_{s-1},x_s]}\|A_s(x)\|\leqslant\Big[\big\|\widetilde{X}(x_s,x_{s-1})-X(x_s,x_{s-1})\big\|\Big(1+\max_{x\in[x_s,x_{s-1}]}(\|A(x)\|(x-$$

$$-x_{s-1}))\Big)\times\big\|(\widetilde{X}(x,x_{s-1}))^{-1}\big\|\Big]\Big/(x_s-x_{s-1})\leqslant$$

$$\leqslant(C+1)\,p\varepsilon_6\big\|[\widetilde{X}(x,x_{s-1})]^{-1}\big\|/(x_s-x_{s-1}). \qquad (2.16)$$

Выше использовано неравенство

$$\max_{x\in[x_{s-1},x_s]}[\|A(x)\|(x-x_{s-1})]\leqslant C.$$

Заметим, что

$$\|[X(x,x_{s-1})]^{-1}\|\left\|\frac{(x-x_{s-1})}{(x_s-x_{s-1})}(\widetilde{X}(x_s,x_{s-1})-X(x_s,x_{s-1}))\right\|\leqslant$$

$$\leqslant e^C\|\widetilde{X}(x_s,x_{s-1})-X(x_s,x_{s-1})\|\leqslant p\varepsilon_6 e^C,$$

так как $\|[X(x,x_{s-1})]^{-1}\|\leqslant e^C$ в силу выбора шага, а $(x-x_{s-1})/(x_s-x_{s-1})\leqslant$ $\leqslant 1$ для любого x из интервала $[x_{s-1},x_s]$. Считая ε_6 достаточно малым, по крайней мере $\varepsilon_6\leqslant 1/2pe^C$, получаем

$$\|[X(x,x_{s-1})]^{-1}\|\left\|\frac{(x-x_{s-1})}{(x_s-x_{s-1})}(\widetilde{X}(x_s,x_{s-1})-X(x_s,x_{s-1}))\right\|\leqslant 1/2.$$

Тогда по теореме Банаха об обратных операторах (см. [5], с. 81) $\widetilde{X}(x,x_{s-1})$, определенная равенством (2.14), для любого x из интервала $[x_{s-1},x_s]$ имеет обратную матрицу и из доказательства этой же теоремы будет следовать

$$\|[\widetilde{X}(x,x_{s-1})]^{-1}\|\leqslant 2\|[X(x,x_{s-1})]^{-1}\|\leqslant 2e^C. \qquad (2.17)$$

Таким образом, левая часть выражения (2.15) имеет смысл, а из (2.16) и (2.17) вытекает неравенство

$$\max_{x\in[x_{s-1},x_s]}\|A_s(x)\|\leqslant 2e^C(C+1)\,p\varepsilon_6/(x_s-x_{s-1}).$$

Обратим теперь внимание на первое равенство (2.12). Его можно переписать следующим образом:

$$\bar{y}_0(x_s)=\widetilde{X}(x_s,x_{s-1})\,z_0(x_{s-1})+\bar{\eta}_s^{(0)}+\big(X(x_s,x_{s-1})-$$

$$-\widetilde{X}(x_s,x_{s-1})\big)\bar{z}_0(x_{s-1})+\widetilde{X}(x_s,x_{s-1})\int_{x_{s-1}}^{x_s}[\widetilde{X}(\xi,x_{s-1})]^{-1}f_1^{(s)}(\xi)\,d\xi,$$

где

$$f_1^{(s)}(\xi)=\widetilde{X}(\xi,x_{s-1})\big[\widetilde{X}(x_s,x_{s-1})\big]^{-1}X(x_s,x_{s-1})\,[X(\xi,x_{s-1})]^{-1}f(\xi).$$

Оценим погрешность возмущения правой части:

$$\max_{x\in[x_s,x_{s-1}]}\|f(x)-f_1^{(s)}(x)\|\leqslant\max_{x\in[x_{s-1},x_s]}\|f(x)\|\max_{x\in[x_{s-1},x_s]}\|I_n-$$

$$-\widetilde{X}(x,x_{s-1})\big[\widetilde{X}(x_s,x_{s-1}\big]^{-1}X(x_s,x_{s-1})\,[X(x,x_{s-1})]^{-1}\|\leqslant$$

$$\leqslant 2e^C\max_{x\in[x_{s-1},x_s[}\|[X(x,x_{s-1})]^{-1}\|\|X(x_s,x_{s-1})\|\max_{x\in[x_{s-1},x_s]}\|X(x,x_s)\times$$

$$\times[X(x_s,x_{s-1})]^{-1}-\widetilde{X}(x,x_{s-1})\big[\widetilde{X}(x_s,x_{s-1})\big]^{-1}\|\max_{x\in[x_s,x_{s-1}]}\|f(x)\|\leqslant$$

$$\leqslant 2e^C\max_{x\in[x_{s-1},x_s]}\|f(x)\|\max_{x\in[x_{s-1},x_s]}\|X(x,x_s)[X(x_s,x_{s-1})]^{-1}-$$

$$-(x-x_{s-1})\big(\widetilde{X}(x_s,x_{s-1})-X(x_s,x_{s-1})\big)\big[\widetilde{X}(x_s,x_{s-1})\big]^{-1}/(x_s-x_{s-1})-$$

$$-X(x,x_{s-1})\big[\widetilde{X}(x,x_{s-1})\big]^{-1}\|\leqslant 2(e^{4C}+e^{3C})\,p\varepsilon_6\max_{x\in[x_{s-1},x_s]}\|f(x)\|. \qquad (2.18)$$

Введем обозначение $\varkappa_s = \bar{\eta}_s^{(0)} + (X(x_s, x_{s-1}) - \widetilde{X}(x_s, x_{s-1}))\bar{z}_0(x_{s-1})$. Рассмотрим следующую функцию на интервале $[x_{s-1}, x_s]$:

$$f_2^{(s)}(\xi) = \widetilde{X}(\xi, x_{s-1})\left[\widetilde{X}(x_s, x_{s-1})\right]^{-1}\varkappa_s/(x_s - x_{s-1}).$$

Теперь можно переписать выражение для вычисления $\bar{y}_0(x_s)$ следующим образом:

$$\bar{y}_0(x_s) = \widetilde{X}(x_s, x_{s-1})\bar{z}_0(x_{s-1}) + \widetilde{X}(x_s, x_{s-1})\int_{x_{s-1}}^{x_s}\left[\widetilde{X}(\xi, x_{s-1})\right]^{-1}(f_1^{(s)}(\xi) + f_2^{(s)}(\xi))\,d\xi. \tag{2.19}$$

Исходя из (2.14), очевидна оценка

$$\max_{x\in[x_{s-1},x_s]}\|f_2^{(s)}(x)\| \leqslant \max_{x\in[x_{s-1},x_s]}\|\widetilde{X}(x, x_{s-1})\|\,\|[X(x_s, x_{s-1})]^{-1}\| \times$$
$$\times\|\varkappa_s\|/(x_s - x_{s-1}) \leqslant e^C(e^C + e^C)\|\varkappa_s\|/(x_s - x_{s-1}) \leqslant$$
$$\leqslant 2e^{2C}\varepsilon_6(1 + p\|\bar{z}_0(x_{s-1})\|)/(x_s - x_{s-1}), \tag{2.20}$$

так как $\varepsilon_6 \leqslant 1/2pe^C$ в силу предположения сделанного выше. Пусть кроме этого ограничения на ε_6 выполнено неравенство

$$\varepsilon_6 \leqslant \min_{s=\overline{1,l-1}}\min\left\{1\Big/\left[2SK\left[4e^{4C}\max_{x\in[x_{s-1},x_s]}\|f(x)\| + 2e^{2C}\Big(1 + \right.\right.\right.$$
$$\left.\left.\left. + 2pK\left(\|\varphi\| + \|\psi\| + (\bar{x} - x_0)\max_{x\in[x_0,\bar{x}]}\|f(x)\| + 1\right)\right)\Big/(x_s - x_{s-1})\right]\right], \tag{2.21}$$
$$(x_s - x_{s-1})/[4SKe^Cp(C+1)]\Big\} = S_1.$$

Докажем методом математической индукции, что при выполнении (2.21) и $\varepsilon_6 \leqslant \min\{S_1, 1/2pe^C\}$ для любого s будет выполнено неравенство

$$\|\bar{z}_0(x_s)\| \leqslant 2K\left(\|\varphi\| + \|\psi\| + (\bar{x} - x_0)\max_{x\in[x_0,\bar{x}]}\|f(x)\| + 1\right). \tag{2.22}$$

Для $s = 1$ в силу хорошей обусловленности задачи

$$\frac{d\bar{u}}{dx} = A(x)\bar{u}(x);$$
$$B\bar{u}(x_0) = \varphi; \quad C\bar{u}(\bar{x}) = 0$$

выполнено неравенство $\max_{x\in[x_0,\bar{x}]}\|\bar{u}(x)\| \leqslant K\|\varphi\|$. Поскольку

$$\bar{u}(x_0) = \bar{z}_0(x_0) + \sum_{j=1}^{p}\beta_j\bar{z}_j(x_0)$$

и так как $\bar{z}_0(x_0), \bar{z}_1(x_0), \ldots, \bar{z}_p(x_0)$ ортогональны, то $\|\bar{z}_0(x_0)\| \leqslant \|\bar{u}(x_0)\| \leqslant \max_{x\in[x_0,\bar{x}]}\|\bar{u}(x)\| \leqslant K\|\varphi\|$. Пусть неравенство (2.22) выполнено при $s = k$. Докажем, что оно останется верным при $s = k+1$. Пусть $\widetilde{A}(x) = A(x) + A_s(x)$ и $\widetilde{f}(x) = f_1^{(s)}(x) + f_2^{(s)}(x)$ для каждого x из интервала $[x_{s-1}, x_s]$ для любого $s \leqslant k+1$, и $\widetilde{A}(x) = A(x)$, а $\widetilde{f}(x) = f(x)$ для каждого x из интервала $[x_{k+1}, \bar{x}]$.

Поскольку выполнено (2.18), то легко получается оценка

$$\max_{x\in[x_{s-1},x_s]}\|\widetilde{A}(x) - A(x)\| \leqslant 2e^C(C+1)pe_6/(x_s - x_{s-1}) \leqslant 1/2SK.$$

для любого интервала $[x_{s-1}, x_s]$. Исходя из (2.18), (2.20) и (2.21), имеем

$$\max_{x\in[x_{s-1},x_s]}\|\tilde f(x)-f(x)\|\leqslant \max_{x\in[x_{s-1},x_s]}\|f_1^{(s)}(x)-f(x)\|+$$

$$+\max_{x\in[x_{s-1},x_s]}\|f_2^{(s)}(x)\|\leqslant 4e^{4C}p\varepsilon_6\max_{x\in[x_{s-1},x_s]}\|f(x)\|+2e^{2C}\varepsilon_6\Big(1+2pK\Big(\|\varphi\|+\|\psi\|+$$

$$+(\bar x-x_0)\max_{x\in[x_0,\bar x]}\|f(x)\|+1\Big)\Big)\Big|(x_s-x_{s-1})\leqslant 1/2SK$$

для любого $s\leqslant k+1$. При $s>k+1$ имеем

$$\max_{x\in[x_{k+1},\bar x]}\|\tilde A(x)-A(x)\|=\max_{x\in[x_{k+1},\bar x]}\|f(x)-\tilde f(x)\|=0.$$

Так как все условия теоремы из п. 1 § 2 выполнены, для решения задачи

$$\frac{d\tilde u(x)}{dx}=\tilde A(x)\tilde u(x)+\tilde f(x);$$

$$B\tilde u(x_0)=\varphi;\quad C\tilde u(\bar x)=\psi$$

справедлива оценка

$$\max_{x\in[x_0,\bar x]}\|\tilde u(x)\|\leqslant 2K\Big(\|\varphi\|+\|\psi\|+(\bar x-x_0)\max_{x\in[x_0,\bar x]}\|f(x)\|+1\Big).$$

Поскольку $\tilde u(x_{k+1})=\bar z_0(x_{k+1})+\sum_{j=1}^{p}\beta_j\bar z_j(x_{k+1})$,

а $\bar z_0(x_{k+1})$, $\bar z_1(x_{k+1})$, ..., $\bar z_p(x_{k+1})$ ортогональны, то нетрудно получить

$$\|\bar z_0(x_{k+1})\|\leqslant 2K\Big(\|\varphi\|+\|\psi\|+(\bar x-x_0)\max_{x\in[x_0,\bar x]}\|f(x)\|+1\Big).$$

Таким образом, требуемое утверждение доказано.

Из всего сказанного видно: $\bar{\bar z}_j(x_s)$ и $\bar{\bar r}_{pq}^{(s)}$ матрицы $\bar{\bar R}_s$ отличаются на величины порядка ε от векторов $\bar z_j(x_s)$ и элементов $\bar r_{pq}^{(s)}$ матрицы $\bar R_s$, а равенства

$$\bar y_j(x_s)=\bar X(x_s, x_{s-1})\bar z_j(x_{s-1}),\quad j=1, 2, \ldots, p;$$

$$y_0(x_s)=\tilde X(x_s, x_{s-1})\bar z_0(x_{s-1})+\tilde X(x_s, x_{s-1})\int_{x_{s-1}}^{x_s}[\tilde X(\xi, x_{s-1})]^{-1}\tilde f(\xi)\,d\xi$$

показывают, что отличие $\bar z_j(x_s)$ и элементов $\bar r_{pq}^{(s)}$ матрицы $\bar R_s$ от $z_j(x_s)$ и элементов $r_{pq}^{(s)}$ матрицы R_s может считаться вызванным тем, что на участке $[x_{s-1}, x_s]$ интегрировали не систему $du/dx=A(x)u(x)+f(x)$, а $du/dx=\tilde A(x)u(x)+\tilde f(x)$, близкую к ней, т. е. величины $\max_{x\in[x_0,\bar x]}\|A(x)-\tilde A(x)\|$ и $\max_{x\in[x_0,\bar x]}\|f(x)-\tilde f(x)\|$ порядка ε.

5. Влияние вычислительных погрешностей при обратной прогонке. Перейдем к описанию обратной прогонки и попытаемся оценить влияние возникающих при этом погрешностей. Обратная прогонка начинается с того, что, представляя $u(\bar x)$ как

$$u(\bar x)=\bar z_0(\bar x)+\sum_{j=1}^{p}\beta_j^{(l)}\bar z_j(\bar x),$$

находим коэффициенты $\beta_j^{(l)}$ из системы $Cu(\bar x)=\psi$. При расчете будем

искать решение в виде

$$\bar{u}(\bar{x}) = \bar{\bar{z}}_0(\bar{x}) + \sum_{j=1}^{p} \bar{\beta}_j^{(l)} \bar{\bar{z}}_j(\bar{x}) =$$

$$= z_0(\bar{x}) + \zeta_l^{(0)} + \sum_{j=1}^{p} \bar{\beta}_j^{(l)} (\bar{z}_j(\bar{x}) + \zeta_l^{(j)}) \qquad (2.23)$$

и для коэффициентов $\bar{\beta}_j^l$ получим

$$\sum_{j=1}^{p} \bar{\beta}_j^{(l)} [C\bar{z}_j(\bar{x}) + C\zeta_l^{(j)}] = \psi - C\bar{z}_0(\bar{x}) - C\zeta_l^{(0)}. \qquad (2.24)$$

Система уравнений

$$\sum_{j=1}^{p} \beta_j^{(l)} C\bar{z}_j(\bar{x}) = \psi - C\bar{z}_0(\bar{x}) = \bar{\psi} \qquad (2.25)$$

имеет решение, для которого выполнена оценка

$$\sqrt{\sum_{j=1}^{p} [\beta_j^{(l)}]^2} \leqslant K \|\bar{\psi}\|, \qquad (2.26)$$

вытекающая из хорошей обусловленности задачи

$$Bu(x_0) = 0; \quad \frac{du(x)}{dx} = \tilde{A}(x) u(x); Cu(\bar{x}) = \bar{\psi}.$$

Покажем, что из оценки (2.26) следует хорошая обусловленность системы (2.25), т. е. покажем, что от изменения на величину порядка ε коэффициентов при $\beta_j^{(l)}$ и правой части решение системы изменится на величину порядка ε. Сделаем это методом последовательных приближений.

Пусть $\mathscr{P} = [C\bar{z}_1(\bar{x}) C\bar{z}_2(\bar{x}) \ldots C\bar{z}_p(\bar{x})]$ — матрица размерности $p \times p$, столбцы которой суть соответственно векторы $C\bar{z}_1(\bar{x})$, $C\bar{z}_2(\bar{x})$, ..., $C\bar{z}_p(\bar{x})$, а $\mathscr{P}_1 = [C\zeta_l^{(1)} C\zeta_l^{(2)} \ldots C\zeta_l^{(p)}]$ — матрица размерности $p \times p$, столбцы которой есть соответственно векторы $C\zeta_l^{(1)}, C\zeta_l^{(2)}, \ldots, C\zeta_l^{(p)}$. Положим $\hat{\psi} = C\zeta_l^{(0)}$. Заметим, что

$$\|\mathscr{P}_1\| = \|\mathscr{P}_1^*\| = \sup_{\substack{x \in R^p \\ \|x\|=1}} \|\mathscr{P}_1^* x\| \leqslant p \sup_{i=\overline{1,p}} \|C\zeta_l^{(i)}\| \leqslant p \|C\| \varepsilon_3;$$

$$\|\hat{\psi}\| \leqslant \|C\| \varepsilon_3. \qquad (2.27)$$

Пусть $\tilde{\mathscr{P}} = \mathscr{P} + \mathscr{P}_1$ — матрица $p \times p$ и $\tilde{\psi} = \bar{\psi} + \hat{\psi}$. Систему

$$\tilde{\mathscr{P}} \bar{\beta} = \tilde{\psi} \qquad (2.28)$$

перепишем следующим образом: $\mathscr{P}\bar{\beta} = \bar{\psi} + (\tilde{\psi} - \bar{\psi}) + (\mathscr{P} - \tilde{\mathscr{P}})\bar{\beta}$. Пусть $\bar{\beta}_0$ — решение системы $\mathscr{P}\bar{\beta}_0 = \bar{\psi} + (\tilde{\psi} - \bar{\psi})$, а $\varepsilon = p\|C\|\varepsilon_3$. Тогда в силу (2.26)

$$\|\beta - \bar{\beta}_0\| \leqslant K\varepsilon; \quad \|\bar{\beta}_0\| \leqslant K(\|\bar{\psi}\| + \varepsilon),$$

где β — решение системы $\mathscr{P}\beta = \bar{\psi}$.

Векторы $\bar{\beta}_r$, $r = 1, 2, \ldots$ определим рекуррентно

$$\mathscr{P}\bar{\beta}_r = \bar{\psi} + (\tilde{\psi} - \bar{\psi}) + (\mathscr{P} - \tilde{\mathscr{P}})\bar{\beta}_{r-1}.$$

Тогда для $\bar{\beta}_r$ имеют место оценки

$$\|\bar{\beta}_1 - \bar{\beta}_0\| \leqslant K\varepsilon \|\bar{\beta}_0\|;$$

$$\|\bar{\beta}_r - \bar{\beta}_{r-1}\| \leqslant K\varepsilon \|\bar{\beta}_{r-1} - \bar{\beta}_{r-2}\|,$$

из которых для достаточно малых ε ($\varepsilon < 1/2K$) получаем

$$\|\bar{\beta}_r - \bar{\beta}_{r-1}\| \leqslant K(\|\bar{\psi}\| + \varepsilon)/2^r;$$

$$\|\bar{\beta}_r\| \leqslant (1 - 1/2^r)(\|\psi\| + \varepsilon)K/(1 - 1/2).$$

Эти неравенства показывают, что при $r \to \infty$ последовательность β_r сходится к некоторому пределу $\bar{\beta}$, который является решением (2.28). Для этого предела справедливы оценки

$$\|\bar{\beta}\| \leqslant 2K(\|\bar{\psi}\| + 1/2K); \quad \|\beta - \bar{\beta}\| \leqslant \varepsilon(K + 2K^2(\|\bar{\psi}\| + 1/2K)). \tag{2.29}$$

Таким образом, ввиду (2.28) и предыдущих рассуждений

$$\|\beta - \bar{\beta}\| \leqslant p\|C\|\varepsilon_3(K + 2K^2(\|\bar{\psi}\| - 1/2K)) \tag{2.30}$$

при $\varepsilon_3 \leqslant 1/2Kp\|C\|$, где

$$\bar{\beta}^{(l)} = \begin{bmatrix} \bar{\beta}_1^{(l)} \\ \bar{\beta}_2^{(l)} \\ \vdots \\ \bar{\beta}_p^{(l)} \end{bmatrix} \text{— решение (2.24), а } \beta^{(l)} = \begin{bmatrix} \beta_1^{(l)} \\ \beta_2^{(l)} \\ \vdots \\ \beta_p^{(l)} \end{bmatrix} \text{— решение} \tag{2.25}$$

Рассмотрим вектор

$$\bar{u}(\bar{x}) = \bar{z}_0(\bar{x}) + \sum_{j=1}^{p} \bar{\beta}_j^{(l)} \bar{z}_j(\bar{x}) + \nu_l \quad (\|\nu_l\| \leqslant \varepsilon_7);$$

$$\bar{u}(\bar{x}) = \bar{z}_0(\bar{x}) + \sum_{j=1}^{p} \bar{\beta}_j^{(l)} \bar{z}_j(\bar{x}) + \sum_{j=1}^{p} (\bar{\beta}_j^{(l)} - \beta_j^{(l)}) \bar{z}_j(\bar{x}) +$$

$$+ \zeta_l^{(0)} + \sum_{j=1}^{p} (\bar{\beta}_j^{(l)} - \beta_j^{(l)}) \zeta_l^{(j)} + \sum_{j=1}^{p} \beta_j^{(l)} \zeta_l^{(j)} + \nu_l =$$

$$= \bar{z}_0(\bar{x}) + \sum_{j=1}^{p} \beta_j^{(l)} \bar{z}_j(\bar{x}) + \xi.$$

Поскольку

$$1. \left\| \sum_{j=1}^{p} (\bar{\beta}_j^{(l)} - \beta_j^{(l)}) \bar{z}_j(\bar{x}) \right\| \leqslant p\|C\|\varepsilon_3(K + 2K^2(\|\bar{\psi}\| + 1/2K)) \leqslant$$

$$\leqslant p\|C\|\varepsilon_3(2K + 2K^2(\|\psi\| + \|C\|\|\bar{z}_0(\bar{x})\|)). \tag{2.31}$$

Так как в силу хорошей обусловленности задачи

$$\frac{du(x)}{dx} = \tilde{A}(x)u(x) + \tilde{f}(x);$$

$$Bu(x_0) = \varphi; \quad Cu(\bar{x}) = 0,$$

имеет место оценка

$$\|\bar{z}_0(\bar{x})\| \leqslant \max_{x \in [x_0, \bar{x}]} \|u(x)\| \leqslant 2K\left(\|\varphi\| + (\bar{x} - x_0) \max_{x \in [x_0, \bar{x}]} \|f(x)\| + 1\right). \tag{2.32}$$

Используя эту оценку в неравенстве (2.31), будем иметь

$$\left\| \sum_{j=1}^{p} (\bar{\beta}_j^{(l)} - \bar{\beta}_j^{(l)}) \bar{z}_j(\bar{x}) \right\| \leqslant p\|C\|\varepsilon_3\left(2K + 4K^3\left(\|\psi\|/2K + \right.\right.$$

$$\left.\left. + \|\psi\| + (\bar{x} - x_0) \max_{x \in [x_0, \bar{x}]} \|f(x)\| + 1\right)\right).$$

$$2. \|\zeta_l^{(0)}\| \leqslant \varepsilon_3 \text{ и } \|\nu_l\| \leqslant \varepsilon_7.$$

$$3. \left\| \sum_{j=1}^{p} \beta_j^{(l)} \zeta_l^{(j)} \right\| \leqslant p\varepsilon_3 K\left(\|\psi\| + \|C\|\|\bar{z}_0(\bar{x})\|\right) \leqslant$$

$$\leqslant K\left(\|\psi\| + \|C\|\left(\|\varphi\| + (\bar{x} - x_0) \max_{x \in [x_0, \bar{x}]} \|f(x)\| + 1\right)\right) p\varepsilon_3,$$

так как верны оценки (2.32) и (2.26).

$$4.\ \left\|\sum_{j=1}^{p}\left(\bar{\beta}_j^{(l)}-\beta_j^{(l)}\right)\zeta_l^{(j)}\right\|\approx\varepsilon_3^2,$$

так как $\|\beta^{(l)}-\beta^{(l)}\|\approx\varepsilon_3$ и $\|\zeta_l^j\|\approx\varepsilon_3$, $j=1, 2, \ldots, p$. Поскольку это член второго порядка малости, им пренебрегаем и получаем

$$\|\xi\|\leqslant\varepsilon_7+p\|C\|\varepsilon_3(2K+4K^3(\|\psi\|/2K+\|\varphi\|+$$
$$+(\bar{x}-x_0)\max_{x\in[x_0,\bar{x}]}\|f(x)\|+1\Big)+p\varepsilon_3K\Big)\|\varphi\|+\|C\|\Big(\|\varphi\|+$$
$$+(\bar{x}-x_0)\max_{x\in[x_0,\bar{x}]}\|f(x)\|+1\Big)\leqslant\max\{\varepsilon_3,\varepsilon_7\}\Big[1+3Kp\|C\|+$$
$$+(2K^2+pK)\|\psi\|+(pK\|C\|+4K^3)\|\varphi\|+(4K^3+pK\|C\|)\times$$
$$\times(\bar{x}-x_0)\max_{x\in[x_0,\bar{x}]}\|f(x)\|\Big]=S_2\max\{\varepsilon_3,\varepsilon_7\}. \qquad (2.33)$$

Отсюда следует, что $\bar{u}(\bar{x})$ получился в результате решения задачи

$$\frac{d\bar{u}(x)}{dx}=\tilde{A}(x)\bar{u}(x)+\tilde{f}(x);$$
$$B\bar{u}(x_0)=\varphi;\quad C\bar{u}(\bar{x})=\psi+\nu,$$

где $\nu=-C\xi$, тогда

$$\|\nu\|\leqslant\|C\|S_2\max\{\varepsilon_3,\ \varepsilon_7\}. \qquad (2.34)$$

Погрешности округления при вычислении правой части в формуле для $\bar{u}(\bar{x})$ обозначены нами через ν_l.

В дальнейших вычислениях при проведении обратной прогонки вместо соотношений

$$R_{s+1}\beta^{(s)}=\beta^{(s+1)};$$
$$u(x_s)=z_0(x_s)+\sum_{j=1}^{p}\beta_j^{(s)}z_j(x_s)$$

надо считать, что вычисления проводятся по формулам

$$\bar{R}_{s+1}\bar{\beta}^{(s)}=\bar{\beta}^{(s+1)}+\delta^{(s+1)}; \qquad (2.35)$$
$$\bar{u}(x_s)=\bar{z}_0(x_s)+\sum_{j=1}^{p}\bar{\beta}_j^{(s)}\bar{z}_j(x_s)+\nu_s,$$

где $\|\delta^{(s+1)}\|\leqslant\varepsilon_7$, $\|\nu_s\|\leqslant\varepsilon_7$. Теперь покажем, что $\sqrt{\sum_{j=1}^{p}[\bar{\beta}_j^{(s)}]^2}$ ограничена при достаточно малом ε_7 некоторой постоянной, зависящей лишь от K, $\|\varphi\|$, $\|\psi\|$, $(\bar{x}-x_0)$, $\|f(x)\|$, и что $\|\bar{u}(x_s)-\bar{u}(x_s)\|$ — порядка $\varepsilon=\max_{i=\overline{3,7}}\varepsilon_i$.

Доказательство по индукции. Предположим, что для некоторого s доказаны следующие факты:

1. $R\bar{u}(\bar{x})=\psi+\nu$. На участке $(x_{s+1}, \bar{x})$ вектор-функция $\bar{u}(x)$ удовлетворяет системе $d\bar{u}(x)/dx=\tilde{A}(x)\bar{u}(x)+\tilde{f}(x)+\tilde{f}(x)$.

2. Выполнено неравенство для $n=s+1, s+2, \ldots, l$

$$\sqrt{\sum_{j=1}^{p}[\beta_j^{(n)}]^2}\leqslant 2K\left(\|\varphi\|+\|\psi\|+(\bar{x}-x_0)\max_{x\in[x_0,\bar{x}]}\|f(x)\|+1\right). \qquad (2.36)$$

3. $\|\bar{u}(x_n)-u(x_n)\|$ — порядка ε для $n=s+1, s+2, \ldots, l$, где $\varepsilon=\max_{i=\overline{3,7}}\varepsilon_i$.

Докажем, что 1—3 будут выполнены, если заменить s на $s-1$. Для $s=l$ неравенство (2.36) уже доказано. Рассмотрим первое из соотношений (2.35)

$$\overline{\overline{R}}_{s+1}\overline{\beta}^{(s)}=\overline{\beta}^{(s+1)}+\delta^{(s+1)}.$$

Это тождество можно переписать следующим образом:

$$\left(\overline{\overline{R}}_{s+1}+\widetilde{\Theta}_{s+1}\right)\begin{bmatrix}\overline{\beta}_1^{(s)}\\ \overline{\beta}_2^{(s)}\\ \vdots\\ \overline{\beta}_p^{(s)}\end{bmatrix}=\begin{bmatrix}\overline{\beta}_1^{(s+1)}\\ \overline{\beta}_2^{(s+1)}\\ \vdots\\ \overline{\beta}_p^{(s+1)}\end{bmatrix}-\begin{bmatrix}\overline{r}_{1,p+1}^{(s+1)}\\ \overline{r}_{2,p+1}^{(s+1)}\\ \vdots\\ \overline{r}_{p,p+1}^{(s+1)}\end{bmatrix}-\begin{bmatrix}\theta_{1,p+1}^{(s+1)}\\ \theta_{2,p+1}^{(s+1)}\\ \vdots\\ \theta_{p,p+1}^{(s+1)}\end{bmatrix}+\widetilde{\delta}^{s+1}, \tag{2.37}$$

где $\overline{\overline{R}}_{s+1}$ — главный минор порядка p матрицы $\overline{R}_{s+1}$, $\widetilde{\Theta}_{s+1}$ — главный минор порядка p матрицы Θ_{s+1}, и $\widetilde{\delta}_i^{(s+1)}=\delta_i^{(s+1)}$ для любого $i=1, 2, \ldots, p$. Последнюю $p+1$ компоненту вектора $\delta^{(s+1)}$ можно считать равной нулю, так как $\overline{\beta}_{p+1}^{(s)}=1$ для любого s, по аналогичной причине $\theta_{p+1,p+1}^{(s+1)}=0$. Поскольку

$$\|\widetilde{\theta}_{s+1}^{(i)}\|\leqslant\varepsilon_5,\ \|\widetilde{\Theta}_{s+1}\|\leqslant p\varepsilon_5, \tag{2.38}$$

пренебрегая в разложении

$$\left(I_p+\widetilde{\Theta}_{s+1}\overline{\overline{R}}_{s+1}^{-1}\right)^{-1}=I_p-\widetilde{\Theta}_{s+1}\overline{\overline{R}}_{s+1}^{-1}+\ldots+(-1)^n\left(\widetilde{\Theta}_{s+1}\overline{\overline{R}}_{s+1}^{-1}\right)^n+\ldots$$

членами выше второго порядка малости по ε_5, получаем

$$\left(I_p+\widetilde{\Theta}_{s+1}\overline{\overline{R}}_{s+1}^{-1}\right)^{-1}\approx I_p-\widetilde{\Theta}_{s+1}\overline{\overline{R}}_{s+1}^{-1}. \tag{2.39}$$

Чтобы это было правомерно, необходимо показать, что $\|\overline{\overline{R}}_{s+1}^{-1}\|$ ограничена. Действительно, в силу утверждений, доказанных в п. 2 настоящего параграфа, будет выполнено неравенство

$$1/\sigma_1\left(\overline{\overline{R}}_{s+1}\right)=\|\overline{\overline{R}}_{s+1}^{-1}\|\leqslant 1/\sigma_1\left(\widetilde{X}(x_{s+1}, x_s)\right)=$$
$$=\|[\widetilde{X}(x_{s+1}, x_s)]^{-1}\|\leqslant 2e^C, \tag{2.40}$$

так как верно (2.17).

Пользуясь (2.39), получим тождество

$$\overline{\overline{R}}_{s+1}\begin{bmatrix}\beta_1^{(s)}\\ \vdots\\ \beta_p^{(s)}\end{bmatrix}=\left(I_p-\widetilde{\Theta}_{s+1}\overline{\overline{R}}_{s+1}^{-1}\right)\left[\begin{bmatrix}\overline{\beta}_1^{(s+1)}\\ \overline{\beta}_2^{(s+1)}\\ \vdots\\ \overline{\beta}_p^{(s+1)}\end{bmatrix}-\begin{bmatrix}\overline{r}_{1,p+1}^{(s+1)}\\ \overline{r}_{2,p+1}^{(s+1)}\\ \vdots\\ \overline{r}_{p,p+1}^{(s+1)}\end{bmatrix}-\begin{bmatrix}\theta_{1,p+1}\\ \theta_{2,p+1}\\ \vdots\\ \theta_{p,p+1}\end{bmatrix}+\widetilde{\delta}^{(s+1)}\right], \tag{2.41}$$

которое можно переписать следующим образом: $\overline{\overline{R}}_{s+1}\overline{\beta}^{(s)}=\overline{\beta}^{(s+1)}+\overline{\delta}^{(s+1)}$. Поскольку

$$1.\left\|\widetilde{\Theta}_{s+1}\overline{\overline{R}}_{(s+1)}^{-1}\begin{bmatrix}\overline{\beta}_1^{(s+1)}\\ \vdots\\ \overline{\beta}_p^{(s+1)}\end{bmatrix}\right\|\leqslant$$

$$\leqslant 4pe^C\varepsilon_5K\left(\|\varphi\|+\|\psi\|+(\overline{x}-x_0)\max_{x\in[x_0,\overline{x}]}\|f(x)\|+1\right),$$

так как в силу предположения индукции

$$\|\overline{\beta}^{(s+1)}\|\leqslant 2K\left(\|\varphi\|+\|\psi\|+(\overline{x}-x_0)\max_{x\in[x_0,\overline{x}]}\|f(x)\|+1\right).$$

$$2.\ \|\tilde{\delta}^{(s+1)}\| = \|\delta^{(s+1)}\| \leqslant \varepsilon_7.$$

$$3.\ \left\|\begin{bmatrix}\theta_{1,p+1}\\ \theta_{2,p+1}\\ \vdots\\ \theta_{p,p+1}\end{bmatrix}\right\| = \|\Theta_{p+1}^{(s+1)}\| \leqslant \varepsilon_5.$$

$$4.\ \left\|\tilde{\Theta}_{s+1}\overline{\overline{R}}_{s+1}^{-1}\begin{bmatrix}\overline{r}_{1,p+1}^{(s+1)}\\ \overline{r}_{2,p+1}^{(s+1)}\\ \vdots\\ \overline{r}_{p,p+1}^{(s+1)}\end{bmatrix}\right\| \leqslant 2p^2e^C\varepsilon_5 \max_{i=\overline{1,p}}|\overline{r}_{i,p+1}^{(s+1)}| \leqslant 2p^2e^C\varepsilon_5\|\overline{y}_0(x_{s+1})\| \leqslant$$

$$\leqslant 4p^2e^C\varepsilon_5\left\{K\left(\|\varphi\| + (\overline{x}-x_0)\max_{x\in[x_0,\overline{x}]}\|f(x)\|+1\right)+\right.$$

$$\left.+\,2e^C(x_{s+1}-x_s)\left(\max_{x\in[x_0,\overline{x}]}\|f(x)\|+1\right)\right\}.$$

При получении этой оценки использовано неравенство для $\|\overline{X}(x_s, x_{s-1})\|$ и то, что для решения задачи

$$\frac{d\tilde{u}(x)}{dx} = \tilde{A}(x)\,\tilde{u}(x) + \tilde{f}(x);$$

$$B\tilde{u}(x_0) = \varphi;\ C\tilde{u}(\overline{x}) = \psi$$

верна оценка

$$\max_{x\in[x_0,\overline{x}]}\|\tilde{u}(x)\| \leqslant 2K\left(\|\varphi\| + \|\psi\| + (\overline{x}-x_0)\max_{x\in[x_0,\overline{x}]}\|f(x)\|+1\right).$$

$$5.\ \left\|\tilde{\Theta}_{s+1}\overline{\overline{R}}_{s+1}^{-1}\theta_{p+1}^{(s+1)}\right\| \leqslant 2p\varepsilon_5^2e^C.$$

Как видно, это член второго порядка по ε_5^2 и им можно пренебречь. Аналогично поступаем с членом $\tilde{\Theta}_{s+1}\overline{\overline{R}}_{s+1}^{-1}\tilde{\delta}^{(s+1)}$, так как

$$\left\|\tilde{\Theta}_{s+1}\overline{\overline{R}}_{s+1}^{-1}\tilde{\delta}^{(s+1)}\right\| \leqslant 2p\varepsilon_5\varepsilon_7e^C;$$

получаем следующую оценку:

$$\|\overline{\delta}^{(s+1)}\| \leqslant \max\{\varepsilon_5, \varepsilon_7\}\left[2 + 4pe^C\left\{\left[(p+1)\|\varphi\| + \right.\right.\right.$$

$$+\|\psi\| + (p+1)(\overline{x}-x_0)\max_{x\in[x_0,\overline{x}]}\|f(x)\| + p + 1\Big]K +$$

$$\left.\left.\left.+\,2pe^C(x_{s+1}-x_s)\left(\max_{x\in[x_s,x_{s+1}]}\|f(x)\|+1\right)\right\}\right] = S_3^{(s)}\max\{\varepsilon_3, \varepsilon_7\}.$$

Как видим, $\overline{\delta}^{(s+1)}$ — порядка $\tilde{\varepsilon} = \max\{\varepsilon_5, \varepsilon_7\}$. Пусть

$$\tilde{u}(x_s) = \overline{z}_0(x_s) + \sum_{j=1}^{p}\overline{\beta}_j^{(s)}\overline{z}_j(x_s)$$

и вспомним, что

$$\overline{u}(x_{s+1}) = \overline{\overline{z}}_0(x_{s+1}) + \sum_{j=1}^{p}\overline{\beta}_j^{(s+1)}\overline{\overline{z}}_j(x_{s+1}) + \nu_{s+1} =$$

$$= \overline{z}_0(x_{s+1}) + \sum_{j=1}^{p}\overline{\beta}_j^{(s+1)}\overline{z}_j(x_s) + \nu_{s+1} + \sum_{j=1}^{p}\overline{\beta}_j^{(s+1)}\zeta_j^{(s+1)} + \zeta_0^{(s+1)}.$$

Поэтому

$$\|\bar{v}_{s+1}\| = \left\| v_{s+1} + \sum_{j=1}^{p} \bar{\beta}_j^{(s+1)} \zeta_j^{(s+1)} + \zeta_0^{(s+1)} \right\| \leqslant$$

$$\leqslant \max\{\varepsilon_3, \varepsilon_7\} \left[2 + 2K \left(\|\varphi\| + \|\psi\| + (\bar{x} - x_0) \max_{x \in [x_0, \bar{x}]} \|f(x)\| + 1 \right) \right] =$$

$$= S_4 \max\{\varepsilon_3, \varepsilon_7\},$$

т. е. $\bar{v}_{s+1}$ — порядка $\max\{\varepsilon_3, \varepsilon_7\}$.

Заметим, что $\left\| \sum_{j=1}^{p} \bar{\delta}_j^{(s+1)} \bar{z}_j(x_{s+1}) \right\| = \|\bar{\delta}^{(s+1)}\|$, так как $\bar{z}_1(x_{s+1}), \ldots, z_p(x_{s+1})$ ортогональны, кроме того, $\|\bar{z}_j(x_{s+1})\| = 1$ для любого $j = 1, 2, \ldots, p$. Пусть

$$\xi_{s+1} = \bar{v}_{s+1} - \sum_{j=1}^{p} \bar{\delta}_j^{(s+1)} \bar{z}_j(x_{s+1}).$$

Ясно, что

$$\xi_{s+1} = \tilde{X}(x_{s+1}, x_s) \int_{x_s}^{x_{s+1}} [\tilde{X}(x, x_s)]^{-1} \hat{f}_s(x)\, dx,$$

где $$\hat{f}_s(x) = \tilde{X}(x, x_s) [\tilde{X}(x_{s+1}, x_s)]^{-1} \xi_{s+1}/(x_{s+1} - x_s).$$

Тогда в силу неравенства (2.17) имеем

$$\max_{x \in [x_s, x_{s+1}]} \|\hat{f}_s(x)\| \leqslant 4e^{2C} (\|\bar{v}_{s+1}\| + \|\bar{\delta}^{(s+1)}\|)/(x_{s+1} - x_s) \leqslant$$

$$\leqslant 4e^{2C} (S_3^{(s)} + S_4) \max\{\varepsilon_3, \varepsilon_5, \varepsilon_7\}/(x_{s+1} - x_s). \qquad (2.42)$$

Отсюда $\max_{x \in [x_s, x_{s+1}]} \|\hat{f}_s(x)\|$ — порядка $\max\{\varepsilon_3, \varepsilon_5, \varepsilon_7\}$. Тогда

$$\tilde{u}(x_{s+1}) = \tilde{X}(x_{s+1}, x_s) \tilde{u}(x_s) +$$

$$+ \tilde{X}(x_{s+1}, x_s) \int_{x_s}^{x_{s+1}} [\tilde{X}(x, x_s)]^{-1} (\tilde{f}(x) + \hat{f}(x))\, dx.$$

Определим $\tilde{\beta}^{(s-1)}$, $\tilde{u}(x_{s-1})$, ... при помощи рекуррентных равенств

$$\bar{R}_{n+1} \tilde{\beta}^{(n)} = \tilde{\beta}^{(n+1)};$$

$$\tilde{u}(x_n) = \bar{z}_0(x_n) + \sum_{j=1}^{p} \tilde{\beta}_j^{(n)} \bar{z}_j(x_n).$$

Нетрудно проверить $B\tilde{u}(x_0) = \varphi$. Применяя оценку обусловленности к задаче

$$B\tilde{u}(x_0) = \varphi;$$

$$\frac{d\tilde{u}(x)}{dx} = \tilde{A}(x) \tilde{u}(x) + \tilde{f}(x) \text{ при } x \in [x_0, x_s];$$

$$\frac{d\tilde{u}(x)}{dx} = \tilde{A}(x) \tilde{u}(x) + \tilde{f}(x) \hat{f}_s(x) \text{ при } x \in [x_s, x_{s+1}];$$

$$\frac{d\tilde{u}(x)}{dx} = \tilde{A}(x) \tilde{u}(x) + \tilde{f}(x) + \bar{f}(x) \text{ при } x \in [x_{s+1}, \bar{x}];$$

$$C\tilde{u}(\bar{x}) = \psi + \nu$$

и считая возмущения в правых частях незначительными, получим

$$\|\tilde{u}(x_s)\| \leqslant 2K \left(\|\varphi\| + \|\psi\| + (\bar{x} - x_0) \max_{x \in [x_0, \bar{x}]} \|f(x)\| + 1 \right).$$

Из равенства $\tilde{u}(x_s) = \bar{z}_0(x_s) + \sum_{j=1}^{p} \bar{\beta}_j^{(s)} \bar{z}_j(x_s)$ и ортогональности $\{\bar{z}_j(x_s)\}_{j=1}^{p}$ устанавливаем, что

$$\sqrt{\sum_{j=1}^{p} [\bar{\beta}_j^{(s)}]^2} \leqslant 2K \left(\|\varphi\| + \|\psi\| + (\bar{x} - x_0) \max_{x \in [x_0, \bar{x}]} \|f(x)\| + 1 \right).$$

С помощью этого неравенства будем иметь

$$\bar{\bar{u}}(x_s) = \bar{\bar{z}}_0(x_s) + \sum_{j=1}^{p} \bar{\bar{z}}_j(x_s) \bar{\beta}_j^{(s)} + \nu_s = \bar{z}_0(x_s) + \sum_{j=1}^{p} \bar{\beta}_j^{(s)} \bar{z}_j(x_s) + \bar{\nu}_s,$$

где $\|\bar{\nu}_s\| \leqslant S_4 \max\{\varepsilon_3, \varepsilon_7\}$. Значит, $\|\bar{u}(x_s) - \tilde{u}(x_s)\|$ — порядка $\varepsilon = \max_{i=\overline{3,7}} \varepsilon_i$, отсюда получаем, что $\|\bar{u}(x_s) - u(x_s)\|$ — порядка ε. Применяя наш старый прием, имеем

$$\bar{\nu}_s = \tilde{X}(x_s, x_{s+1}) \int_{x_{s+1}}^{x_s} [\tilde{X}(x, x_{s+1})]^{-1} \hat{\hat{f}}_s(x)\, dx,$$

где

$$\hat{\hat{f}}_s(x) = \tilde{X}(x, x_{s+1}) [\tilde{X}(x_s, x_{s+1})]^{-1} \bar{\nu}_s / (x_s - x_{s+1}).$$

Тогда нетрудно получить оценку

$$\max_{x \in [x_s, x_{s+1}]} \|\hat{\hat{f}}_s(x)\| \leqslant 4e^{2C} S_4 \varepsilon / (x_{s+1} - x_s). \tag{2.43}$$

На участке $[x_s, x_{s+1}]$ определим функцию $\bar{f}(x) = \hat{f}_s(x) + \hat{\hat{f}}_s(x)$. Тогда

$$\bar{u}(x_s) = \tilde{X}(x_s, x_{s+1}) \bar{u}(x_{s+1}) + \tilde{X}(x_s, x_{s+1}) \int_{x_{s+1}}^{x_s} [\tilde{X}(x, x_{s+1})]^{-1} (\tilde{f}(x) + \bar{f}(x))\, dx,$$

где $\tilde{X}(x_s, x_{s+1}) = [\tilde{X}(x_{s+1}, x_s)]^{-1}$ и $\tilde{X}(x, x_{s+1}) = [\tilde{X}(x, x_s)]^{-1}$.

Пусть отныне $\varepsilon = \max_{i=\overline{3,7}} \varepsilon_i$. Оценим $\max_{x \in [x_0, \bar{x}]} \|\bar{f}(x)\|$, используя (2.42) и (2.43):

$$\max_{x \in [x_0, \bar{x}]} \|\bar{f}(x)\| \leqslant \max_{s=\overline{1,l}} \max_{[x_s, x_{s+1}]} \|\bar{f}(x)\| \leqslant$$

$$\leqslant \varepsilon \max_{s=\overline{1,l-1}} [(8e^{2C} S_4 + 4e^{2C} S_3^{(s)}) / (x_{s+1} - x_s)] = \varepsilon S_5.$$

Таким образом, при ограничениях на $\varepsilon \leqslant 1/2pe^C$, $\varepsilon \leqslant S_1$, $\varepsilon \leqslant 1/2Kp\|C\|$; $\varepsilon < 1/2SKS_2\|C\|$, $\varepsilon < S_1/(1 + 2SKS_1S_5)$ в силу теоремы, доказанной в п. 1 § 2, получаем

$$\max_{i=\overline{0,l}} \|u(x_i) - \bar{u}(x_i)\| \leqslant \varepsilon \left[2K + 2K^2 (1 + \bar{x} - x_0) \left(\|\varphi\| + \|\psi\| + (\bar{x} - x_0) \max_{x \in [x_0, \bar{x}]} \|f(x)\| + 1 \right) \right].$$

§ 3. ОЦЕНКИ ПОГРЕШНОСТЕЙ ПРИ ПРОВЕДЕНИИ ОТДЕЛЬНЫХ ЭТАПОВ ПРОГОНКИ

1. Оценка погрешности при ортогонализации. В этом пункте приведем оценки для векторов погрешностей $\zeta_s^{(1)}, \zeta_s^{(2)}, \ldots, \zeta_s^{(p)}$ и элементов $\theta_{pq}^{(s)}$ матрицы $\Theta^{(s)}$, где p и q изменяются от 1 до p (см. п. 4 § 2), кото-

рые возникают при ортогонализации векторов $\bar{y}_1(x_s)$, $\bar{y}_2(x_s)$, ..., $\bar{y}_p(x_s)$ по формулам (1.6). При этом вместо векторов $\bar{z}_1(x_s)$, $\bar{z}_2(x_s)$, ..., $\bar{z}_p(x_s)$ получаются векторы $\bar{\bar{z}}_1(x_s) = \bar{z}_1(x_s) + \zeta_s^{(1)}$, ..., $\bar{\bar{z}}_p(x_s) = \bar{z}_p(x_s) + \zeta_s^{(p)}$, а вместо матрицы $\bar{R}_s$ получается матрица $\bar{\bar{R}}_s$, причем $\bar{\bar{r}}_{pq}^{(s)} = \bar{r}_{pq}^{(s)} + \Theta_{pq}^{(s)}$. Схематично формулы ортогонализации можно переписать следующим образом. Пусть $Y = [\bar{y}_1(x_s), \ldots, \bar{y}_p(x_s)]$ — матрица размерности $n \times p$, столбцы которой суть векторы $\bar{y}_1(x_s)$, $\bar{y}_2(x_s)$, ..., $\bar{y}_{p-1}(x_s)$, $\bar{y}_p(x_s)$, а $\mathscr{P}^{(1)}$, $\mathscr{P}^{(2)}$, ..., $\mathscr{P}^{(p)}$ — матрицы отражения, приводящие матрицу Y к матрице R так, что

$$\begin{aligned} &R^{(0)} = Y; \\ &R^{(i)} = \mathscr{P}^{(i)} Y \quad (i = 1, 2, \ldots, p); \\ &R = R^{(p)}, \end{aligned} \tag{3.1}$$

причем матрица R обладает свойством $r_{ij} = 0$ при $i > j$. А матрица Q размерности $n \times n$ получается так:

$$\begin{aligned} &Q^{(0)} = I_n; \\ &Q^{(j)} = Q^{(j-1)} \mathscr{P}^{(j)} \quad (j = 1, 2, \ldots, p); \\ &Q = Q^{(p)}. \end{aligned} \tag{3.2}$$

Если $q_1, q_2, \ldots, q_p$ — первые p столбцов матрицы Q, то $\bar{z}_i(x_s) = q_i$. Матрица $\bar{\bar{R}}_s$ размерности $p \times p$ определяется соотношениями $\bar{\bar{r}}_{ij}^{(s)} = r_{ij}$, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, p$.

Если учесть погрешности H_i и F_i, допускаемые при вычислении $R^{(i)}$ и $Q^{(j)}$ соответственно, то вместо $R^{(i)}$ и $Q^{(j)}$ следует рассматривать матрицы $\bar{R}^{(i)}$ и $\bar{Q}^{(j)}$, связанные соотношениями

$$\begin{aligned} &\bar{R}^{(0)} = Y; && \bar{Q}^{(0)} = I_n; \\ &\bar{R}^{(i)} = \mathscr{P}^{(i)} \bar{R}^{(i-1)}; && \bar{Q}^{(j)} = \bar{Q}^{(j-1)} \mathscr{P}^{(j)} + F_j; \\ &\bar{R} = \bar{R}^{(p)}; && \bar{Q} = \bar{Q}^{(p)}. \end{aligned} \tag{3.3}$$

Так же, как и в [3], легко доказывается, что при выполнении оценок $\|H_i\|_E \leqslant \delta \|\bar{R}^{(i)}\|_E$ и $\|F_i\|_E \leqslant \alpha \|\bar{Q}^{(i)}\|$ суммарные погрешности при вычислении R и Q оцениваются следующим образом:

$$\begin{aligned} &\|R - \bar{R}\|_E \leqslant p\sqrt{p}\delta e^{p\delta} \|\bar{R}\|_E; \\ &\|Q - \bar{Q}\|_E \leqslant p\sqrt{p}\alpha e^{\alpha p} \|\bar{Q}\|_E, \end{aligned} \tag{3.4}$$

где $\|B\|_E = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{p} b_{ij}^2}$ для матрицы B размерности $n \times p$.

При выполнении ограничений, смысл которых состоит в том, чтобы n было не слишком велико, $\|Y\|_E$ — не слишком велика и не слишком мала, и других ограничений в [3] С. К. Годуновым проведен детальный вывод оценки погрешности при преобразовании квадратной матрицы серией умножений либо слева, либо справа на ортогональные отражения, которая легко переносится на прямоугольную матрицу.

Пусть матрица A подвергается умножению слева на ортогональное отражение $\mathscr{P}$: $B = \mathscr{P}A$, причем каждый столбец $b^{(i)}$ матрицы — произведения B получается из соответствующего столбца $a^{(i)}$ матрицы A в соответствии с алгоритмом, изложенным в п. 3 § 1.

Тогда из [3] имеем неравенство

$$\|\delta B\|_E \leqslant 20\varepsilon_1 \|A\|_E, \tag{3.5}$$

где ε_1 — один из параметров, характеризующий вычислительную машину такой, что наименьшее машинное число (больше единицы) не превосходит $1 + \varepsilon_1$. Второй параметр ε_2 такой, что модуль любого машинного чис-

ла заключен между пределами $\rho_{\min}$ и $\rho_{\max}$ такими, что $\rho_{\min} \leqslant \varepsilon_2/2$, $\rho_{\max} \geqslant \varepsilon_2$. Подробнее об этих числах можно прочесть в [3].

Таким образом, применительно к неравенствам (3.4) можно положить $\alpha = 20\varepsilon_1$, $\delta = 20\varepsilon_1$. В итоге получим неравенства

$$\|\widetilde{\widetilde{R}}_s - \widetilde{\widetilde{R}}_s\|_E \leqslant \|R - \widetilde{R}\|_E \leqslant p\sqrt{p}\delta e^{p\delta}\|Y\| \leqslant 2p\sqrt{p}\delta e^{p\delta}e^C;$$

$$\|\widetilde{\widetilde{Z}}_s - \widetilde{\overline{Z}}_s\| \leqslant \|Q - \widetilde{Q}\|_E \leqslant p\sqrt{p}\delta e^{p\delta}\|Q^{(0)}\| \leqslant p\sqrt{p}\delta e^{p\delta},$$

где $Z_s = [\bar{z}_1(x_s), \bar{z}_2(x_s), \ldots, \bar{z}_p(x_s)]$ — матрица размерности $n \times p$, столбцы которой суть векторы $\bar{z}_1(x_s) = q_1$, $\bar{z}_2(x_s) = q_2$, $\ldots$, $\bar{z}_p(x_s) = q_p$. Поскольку $\widetilde{\widetilde{R}}_s - \widetilde{\widetilde{R}}_s = \widetilde{\Theta}_s$ и $\widetilde{\overline{Z}}_s - Z_s = H = [\zeta_s^{(1)}, \zeta_s^{(2)}, \ldots, \zeta_s^{(p)}]$ — матрица, столбцы которой есть векторы $\zeta_s^{(1)}, \zeta_s^{(2)}, \ldots, \zeta_s^{(p)}$, имеем

$$\|\zeta_s^{(i)}\| \leqslant p\sqrt{p}\delta e^{p\delta}, \|\theta_s^{(i)}\| \leqslant 2p\sqrt{p}\delta e^{p\delta}e^C, \tag{3.6}$$

где $\theta_s^{(i)}$ — столбцы матрицы $\widetilde{\Theta}_s$.

2. Оценка погрешности расчета вектора $\bar{z}_0(x_s)$ при $s \geqslant 1$. Напомним, как вычисляется вектор

$$\bar{z}_0(x_s) = \bar{y}_0(x_s) - \sum_{j=1}^{p} (\bar{y}_0(x_s), \bar{z}_j(x_s))\bar{z}_j(x_s) =$$

$$= \bar{y}_0(x_s) - \sum_{j=1}^{p} (\bar{y}_0(x_s), \bar{z}_j(x_s))\bar{z}_j(x_s) = \bar{y}_0(x_s) - \sum_{j=1}^{p} \bar{r}_{j,p+1}^{(s)}\bar{z}_j(x_s).$$

Оценим точность расчета $\bar{r}_{j,p+1}^{(s)}$. В действительности вместо соотношения $\bar{r}_{j,p+1}^{(s)} = (\bar{y}_0(x_s), \bar{z}_j(x_s))$ вычисления проводятся по формуле $\bar{r}_{j,p+1}^{(s)} = (\widetilde{\bar{y}}_0(x_s), \widetilde{\bar{z}}_j(x_s))$. Воспользуемся оценкой из [3] точности вычисления скалярного умножения при условии, что все расчеты проводятся с помощью регистра с удвоенным числом разрядов

$$|(x, y)_{\text{дейст}} - (x, y)_{\text{маш}}| \leqslant 2\varepsilon_1\|x\|\|y\|.$$

Применим эту оценку следующим образом:

$$|[\bar{r}_{j,p+1}^{(s)}] - [\bar{r}_{j,p+1}^{(s)}]_{\text{маш}}| \leqslant |(\bar{y}_0(x_s), \bar{z}_j(x_s)) -$$
$$- (\bar{y}_0(x_s), \widetilde{\bar{z}}_j(x_s))_{\text{маш}}| \leqslant |(\bar{y}_0(x_s), \zeta_s^{(j)})| + |(\bar{y}_0(x_s), \widetilde{\bar{z}}_j(x_s)) -$$
$$- (\bar{y}_0(x_s), \widetilde{\bar{z}}_j(x_s))_{\text{маш}}| \leqslant 2\varepsilon_1\|\bar{y}_0(x_s)\|(\|\bar{z}_j(x_s)\| + \|\zeta_s^{(j)}\|) + \|\bar{y}_0(x_s)\|\|\zeta_s^{(j)}\| \leqslant$$
$$\leqslant 2\varepsilon_1\|\bar{y}_0(x_s)\|(1 + \delta p\sqrt{p}e^{p\delta}) + \|\bar{y}_0(x_s)\|\delta p\sqrt{p}e^{p\delta} \leqslant$$
$$\leqslant \|y_0(x_s)\|(2\varepsilon_1 + (1 + 2\varepsilon_1)\delta p\sqrt{p}e^{p\delta}), \tag{3.7}$$

где $\delta = 20\varepsilon_1$. Пусть $v = 2\varepsilon_1 + (1 + 2\varepsilon_1)\delta p\sqrt{p}e^{p\delta}$. Тогда

$$|\theta_{i,p+1}^{(s)}| \leqslant \|\bar{y}_0(x_s)\|v \leqslant 2Kv\left(\|\varphi\| + \|\psi\| + (\bar{x} - x_0)\max_{x\in[x_0,\bar{x}]}\|f(x)\| + 1\right)$$

при достаточно малом v.

Воспользуемся также оценкой из [3] точности машинного сложения, вычитания двух векторов размерности N, а также точности умножения вектора на число

$$\|(\alpha + \beta)_{\text{маш}} - (\alpha + \beta)\| \leqslant \varepsilon_1\|\alpha + \beta\| + \varepsilon_2\sqrt{N}/2;$$
$$\|(\alpha - \beta)_{\text{маш}} - (\alpha - \beta)\| \leqslant \varepsilon_1\|\alpha - \beta\| + \varepsilon_2\sqrt{N}/2; \tag{3.8}$$
$$\|(s\alpha)_{\text{маш}} - s\alpha\| \leqslant \varepsilon_1|s|\|\alpha\| + \varepsilon_2\sqrt{N}/2.$$

Тогда, если $s^{(i)} = \bar{r}_{i,p+1}^{(s)}$, имеем, пользуясь (3.7),

$$\|[s^{(i)}\widetilde{\bar{z}}_i(x_{s+1})] - [s_{\text{маш}}^{(i)}\widetilde{\bar{z}}_i(x_{s+1})]_{\text{маш}}\| \leqslant |s^{(i)} - s_{\text{маш}}^{(i)}|\|\widetilde{\bar{z}}_i(x_{s+1})\| +$$
$$+ \varepsilon_1|s_{\text{маш}}^{(i)}|\|\widetilde{\bar{z}}_i(x_{s+1})\| + \varepsilon_2\sqrt{n}/2 \leqslant \|\bar{y}_0(x_s)\|v(1 + p\sqrt{p}\delta e^{p\delta}) +$$
$$+ \varepsilon_1(1 + v)\|\bar{y}_0(x_s)\|(1 + p\sqrt{p}\delta e^{p\delta}) + \varepsilon_2\sqrt{n}/2 = \|\bar{y}_0(x_s)\|v_1 + \varepsilon_2\sqrt{n}/2.$$

Отсюда получаем

$$\left\|\left[s^{(i)}\bar{z}_i(x_s)\right]-\left[s^{(i)}_{\text{маш}}\bar{\bar{z}}_i(x_{s+1})\right]_{\text{маш}}\right\|\leqslant\left\|\bar{y}_0(x_s)\right\|(v+v_1)+\varepsilon_2\sqrt{n}/2.$$

Использование этих неравенств дает

$$\left\|\bar{z}_0(x_s)-\left[\bar{z}_0(x_s)\right]_{\text{маш}}\right\|=\left\|\zeta_s^{(0)}\right\|=$$

$$=\left\|\left[\bar{y}_0(x_s)-\sum_{i=1}^{p}s^{(i)}z_i(x_s)\right]-\left[y_0(x_s)-\sum_{i=1}^{p}s^{(i)}_{\text{маш}}\bar{\bar{z}}_i\right]_{\text{маш}}\right\|\leqslant$$

$$\leqslant\left\|\left[\bar{y}_0(x_s)-\sum_{i=1}^{p}s^{(i)}_{\text{маш}}\bar{\bar{z}}_i(x_s)\right]-\left[\bar{y}_0(x_s)-\sum_{i=1}^{p}s^{(i)}_{\text{маш}}\bar{\bar{z}}_i\right]_{\text{маш}}\right\|+$$

$$+\sqrt{\sum_{i=1}^{p}\left|s^{(i)}_{\text{маш}}-s^{(i)}\right|^2}+\left\|\sum_{i=1}^{p}s^{(i)}_{\text{маш}}\zeta_s^{(i)}\right\|\leqslant$$

$$\leqslant\varepsilon_1\left\|\left[\bar{y}_0(x_s)-\sum_{i=1}^{p}s^{(i)}_{\text{маш}}\bar{\bar{z}}_i(x_s)\right]-\left[\bar{y}_0(x_s)-\sum_{i=1}^{p}s^{(i)}_{\text{маш}}\bar{\bar{z}}_i\right]_{\text{маш}}\right\|+$$

$$+\left\|\bar{y}_0(x_s)\right\|\left(v\sqrt{p}+p^2\delta e^{p\delta}(1+v)\right). \tag{3.9}$$

Из второго соотношения (3.8) легко получить оценку

$$\left\|\left(\alpha-\sum_{i=1}^{p}\beta_i\right)_{\text{маш}}-\left(\alpha-\sum_{i=1}^{p}\beta_i\right)\right\|\leqslant$$

$$\leqslant\varepsilon_1\left[p^2\max_{i=\overline{1,p}}\left\|\beta_i\right\|+\left\|\alpha-\sum_{i=1}^{p}\beta_i\right\|\right]+N\sqrt{N}\,\varepsilon_2/2.$$

Применяя ее для оценивания левой части неравенства (3.9) и пренебрегая членами порядка ε_1^2, будем иметь

$$\left\|\zeta_s^{(0)}\right\|\leqslant\varepsilon_1\left[p^2\left\|\bar{y}_0(x_s)\right\|+\left\|\bar{z}_0(x_s)\right\|\right]+\left\|\bar{y}_0(x_s)\right\|v_3+n\sqrt{n}\,\varepsilon_2/2, \tag{3.10}$$

где

$$v_3=v\sqrt{p}+p^2\delta e^{p\delta}(1+v).$$

Заметим, что оценка (3.10) достаточно грубая и при желании можно вывести более точную, но еще более громоздкую.

3. Оценки погрешности определения начальных приближений. В этом пункте приведены оценки векторов $\zeta_0^{(1)}, \zeta_0^{(2)}, \ldots, \zeta_0^{(p)}, \zeta_0^{(0)}$. Естественно, эти погрешности зависят от способа определения векторов $z_1(x_0), z_2(x_0), \ldots, z_p(x_0), z_0(x_0)$. Сначала изложим вкратце способ их определения, а затем оценим точность.

Предлагается следующий способ нахождения начальных векторов, с которых начинаем прогонку. Пусть $Q^{(1)}, Q^{(2)}, \ldots, Q^{(k)}$ — матрицы отражения, приводящие матрицу B к матрице M такой, что $m_{ij}=0$ при $i<j$, т. е.

$$\begin{aligned}&B^{(0)}=B;\\&B^{(j)}=B^{(j-1)}Q^{(j)};\\&B^{(k)}=M,\end{aligned} \tag{3.11}$$

где

$$M=\begin{bmatrix}m_{11}&&&0\\m_{21}&m_{22}&&\\\vdots&&\ddots&\\m_{k1}&\ldots&&m_{kk}\end{bmatrix}.$$

Поскольку матрицы $Q^{(j)}$ ортогональны и в силу предположения, что B ранга k, выполняется свойство $m_{ii}\neq 0$, где $i=1, 2, \ldots, k$. С учетом этого систему $Bu(x_0)=\varphi$ можно переписать следующим образом:

$$Bu(x_0)=BQ^{(1)}Q^{(2)}\ldots Q^{(k)}Q^{(k)}Q^{(k-1)}\ldots Q^{(1)}u(x_0)=MQ^{*}u(x_0)=\varphi, \tag{3.12}$$

где $Q = Q^{(1)}Q^{(2)} \dots Q^{(k)}$. Для системы уравнений $Mx = 0$ полный ортогональный базис подпространства векторов, удовлетворяющих $Mx = 0$, составляют векторы $e_{p+1}, \dots, e_n$, где e_i — вектор, у которого все компоненты, кроме i-й, равны 0, а i-я компонента равна 1. Тогда система векторов $q_{p+1}, q_{p+2}, \dots, q_n$, где q_i — i-й столбец матрицы Q, составляет полный ортогональный базис подпространства векторов, удовлетворяющих системе уравнений $Bx = 0$. Это следует из того, что Q — ортогональная матрица и в силу (3.12).

Пусть $\bar{M}$ — квадратная $p \times p$ матрица такая, что $\bar{m}_{ij} = m_{ij}$ при $i, j = 1, 2, \dots, k$, т. е. $M = (\bar{M}0)$. Решаем систему линейных уравнений $\bar{M}y = \varphi$, где y — вектор размерности k. Так как $m_{ii} \neq 0$ для любого i, система имеет единственное решение y_0. Пусть x_0 — вектор размерности n, первые k компонент которого равны компонентам вектора y_0, остальные компоненты равны 0. Тогда ясно, что $Mx_0 = \varphi$. Рассмотрим вектор $x_1 = Qx_0$. Поскольку x_1 есть линейная комбинация векторов $q_1, q_2, \dots, q_k$, так как последние p компонент x_0 равны 0, а Q — ортогональная матрица, x_0 ортогонален векторам $q_{k+1}, q_{k+2}, \dots, q_n$. В итоге имеем $z_1(x_0) = q_{k+1}, \dots, z_p(x_0) = q_n$, $z_0(x_0) = x_1$. Рассуждая так же, как и в п. 1 § 3, можно получить оценки: $\|\bar{Q} - Q\| \leqslant k\sqrt{k}\delta e^{k\delta}$, где $\delta = 20\varepsilon_1$, $\|\bar{M} - M\| \leqslant k\sqrt{k}\delta e^{k\delta}\|B\|$. Затем, если воспользоваться оценкой точности решения системы с возмущенной треугольной матрицей, которую можно найти в [6], и использовать оценку точности перемножения двух матриц из [3], то можно оценить точность нахождения вектора x_1.

4. Некоторые замечания относительно погрешностей ν_l и $\delta^{(s)}$: ν_l представляет собой погрешность округления при сложении $p+1$ векторов, которую легко оценить с помощью неравенств для сложения векторов; $\delta^{(s)}$ — погрешность решения системы линейных уравнений с треугольной матрицей на вычислительной машине. Ее тоже можно оценить, тем более что число обусловленности матрицы $\tilde{\tilde{R}}_s$ системы для определения $\tilde{\beta}_s$ ограничено для любого s: $\mu(\tilde{\tilde{R}}_s) \leqslant 4e^{2C}$.

ЛИТЕРАТУРА

1. **Годунов С. К.** О численном решении краевых задач для систем линейных обыкновенных дифференциальных уравнений.— Успехи мат. наук, 1961, т. 16, № 3, с. 171—175.
2. **Годунов С. К.** Метод ортогональной прогонки для решения систем разностных уравнений.— Журн. вычисл. математики и мат. физики, 1962, т. 2, № 6, с. 972—983.
3. **Годунов С. К.** Решение систем линейных уравнений.— Новосибирск: Наука. Сиб. отд-ние, 1980.— 177 с.
4. **Наймарк М. А.** Линейные дифференциальные операторы.— М.: Наука, 1969.— 373 с.
5. **Кутателадзе С. С.** Основы функционального анализа.— Новосибирск: Наука. Сиб. отд-ние, 1983.— 221 с.
6. **Уилкинсон Дж. Х.** Алгебраическая проблема собственных значений.— М.: Наука, 1970.— 564 с.

АЛГОРИТМЫ ИСЧЕРПЫВАНИЯ ТРЕХДИАГОНАЛЬНЫХ СИММЕТРИЧЕСКИХ И ДВУХДИАГОНАЛЬНЫХ МАТРИЦ С ГАРАНТИРОВАННОЙ ОЦЕНКОЙ ТОЧНОСТИ

А. Д. МИТЧЕНКО

ВВЕДЕНИЕ

Работа посвящена выводу нового варианта формул и анализу погрешностей при реализации так называемых алгоритмов ортогонального исчерпывания трехдиагональных симметрических и двухдиагональных